Институт проблем передачи информации имени А.А. Харкевича Российская Академия наук



Сборник трудов 42-й междисциплинарной школы-конференции ИППИ РАН

Информационные технологии и системы

Казань, 25-30 сентября 2018

Москва

2018

Междисциплинарная школа-конференция ИППИ РАН "Информационные технологии и системы 2018" (ИТиС 2018) проведена при финансовой поддержке РФФИ в рамках научного проекта № 18-31-10042.

ИТиС 2018: Сборник трудов 42-й междисциплинарной школы-конференции ИППИ РАН "Информационные технологии и системы 2018". — Москва: Институт проблем передачи информации им. А.А. Харкевича РАН, 2018. — 700 с.

Использование синтаксических признаков при построении морфологического теггера на основе GRU*

Мовсесян А.А.¹

Институт проблем передачи информации (ИППИ РАН), derise@iitp.ru

Аннотация В данной работе представлена модель морфологического теггера для русского языка. Векторное представление слов осуществляется на символьном уровне. Морфологические признаки берутся из результата работы морфологического компонента лингвистического процессора ЭТАП. В основе архитектуры лежит двунаправленная GRU. Рассматриваемая модель дает результаты, сравнимые с другими современными морфологическими теггерами для русского языка. При этом в аналогичных работах, использующих морфологические признаки, не рассматриваются синтаксические признаки, являющиеся весьма важными в содержательном плане. В данной работе сделан акцент на рассмотрении синтаксических признаков; в частности, показано улучшение качества работа теггера при их использовании.

Ключевые слова: морфологический теггер, синтаксические признаки, глубокое обучение, ЭТАП

1 Введение

Задача морфологического анализа является одной из ключевых задач при обработке естественного языка. Под морфологическим анализом в данной работе подразумевается получение всех допустимых пар [морфологические характеристики (часть речи, род, число, одушевленность и пр.), лемма (основная форма слова)] для каждой словоформы входного предложения на русском языке. У неомонимичных словоформ такая допустимая пара одна.

Целью работы является улучшение качества морфологического компонента лингвистического процессора ЭТАП [1]. Морфологический анализ в ЭТАПе выполняется, в основном, с использованием морфологического словаря рассматриваемого языка (в контексте данной работы — русского), который для каждой лексемы можно представить в виде биекции (словоформы, наборы морфологических характеристик).

^{*} Работа выполнена в ИППИ РАН за счет гранта Российского научного фонда (проект №16-18-10422)

Несмотря на внушительный объем морфологического словаря для русского языка (порядка 5 млн. словоформ), в силу появления новых слов в языке, а также, например, нецелесообразности хранения в словаре всех имен, фамилий, названий географических объектов и пр., необходима возможность анализа словоформ, не представленных в словаре. Для этого в системе ЭТАП, в частности, используется т. н. гессер [2], который, по имеющейся информации в морфологическом словаре и отдельных словарях лемм, строит гипотезы о морфологических характеристиках для несловарных словоформ.

Поскольку морфологический анализ в ЭТАПе осуществляется пословно и не учитывает контекст, разрешить омонимию на этом шаге невозможно. Ее можно разрешить при синтаксическом анализе, однако с этим связаны несколько проблем:

- Синтаксический анализ является существенно более вычислительно затратной процедурой, чем морфологический анализ; его вычислительная сложность нелинейно зависит от коэффициента омонимии (отношения числа омонимов к числу словоформ в предложении).
- Хотя перед собственно синтаксическим анализом лексическая и грамматическая омонимия частично разрешаются по линейному контексту блоком предсинтаксического анализа, коэффициент омонимии часто остается высоким.
- 3. Число гипотез, порождаемых гессером, тем выше, чем выше требуемая полнота к результату его работы.

Для решения первых двух проблем в ЛП ЭТАП используется статистический морфологический теггер [3], который для каждой словоформы предложения вычисляет распределение вероятностей всех возможных тегов (тегом является набор морфологических характеристик словоформы). Для решения третьей проблемы можно использовать результаты работы гессера в теггере. Однако, за последние несколько лет была показана высокая эффективность методов глубокого обучения в задачах обработки естественного языка (см. раздел 2), поэтому было принято решение внедрить в лингвистический процессор теггер на основе многослойной нейронной сети.

Статья структурирована следующим образом. В разделе 2 описаны методы глубокого обучения, используемые при морфологической разметке. Раздел 3 посвящен предлагаемой модели и некоторым усовершенствованиям в ней. В разделе 4 оценивается качество теггера и его модификаций. В разделе 5 приведены выводы и некоторые идеи для дальнейшей работы.

2 Морфологический анализ и глубокое обучение

Во многих задачах обработки естественного языка, в частности, частеречной разметки, распознавания именованных сущностей и т.п., использование методов глубокого обучения дает существенный прирост качества по сравнению со статистическими моделями. Несмотря на то, что подобные исследования проводятся, в основном, для английского языка (см., например, [4]), часто, одна и та же архитектура нейросети может быть использована для различных естественных языков (применительно к морфологической разметке см., например, [5]).

Данный раздел состоит из двух частей. В первой части описываются некоторые подходы к проектированию признаков, а во второй — возможные архитектуры нейросети.

2.1 Проектирование признаков

Представление слова. Для того, чтобы использовать слово, которое встретилось в предложении, в качестве признака, необходимо получить некоторое (векторное) его представление, понятное компьютеру. Один из простейших методов преобразования слова из словаря размера V в вектор — т. н. one-hot encoding, когда *i*-му слову w_i сопоставляется вектор w_i длины V, причем

$$(\boldsymbol{w}_i)_j = \begin{cases} 1, & \text{если } j = i, \\ 0, & \text{иначе.} \end{cases}$$

Однако такое представление не дает практически никакой информации о самом слове. Кроме того, такой подход неприемлем для большого размера словаря в силу высокой размерности вектора. Одним из возможных улучшений данного подхода является использование некоторой языковой модели для обучения нейросети на большом неразмеченном корпусе таким образом, чтобы внутреннее представление слова было плотным вектором размерности $d \ll V$. Использование известных моделей *CBOW* (нейросеть предсказывает слово по его контексту) и *skip-gram* (нейросеть предсказывает контекст по данному слову), описанных в [6] и [7] (и инструменты на их основе, например, *word2vec*), не только уменьшает размерность векторного представления, но и приводит векторы к такому виду, что близкие по значению векторы близки также в синтаксическом и семантическом смыслах.

В рассмотренных выше методах слово считается неделимым элементом текста. Однако слово состоит из морфем, а морфемы — из букв. При этом число морфем, как и число букв, конечно и невелико, поэтому достаточно представить морфемы (буквы) в виде *one-hot* векторов. Тогда векторным представлением самого слова можно считать сумму векторов, сопоставленных морфемам (буквам) этого слова. При этом, чтобы придать признакам информативность, можно предварительно подать слово как набор символов (морфем) на вход некоторой нейросети и обучить ее с использованием какой-либо языковой модели, или обучать при решении основной задачи (см п. 2.2).

Морфологические признаки В модели морфологии, использующейся в ЛП ЭТАП, 45 морфологических характеристик и 480 уникальных наборов морфологических характеристик. Поскольку все характеристики существенны, а задача состоит в их предсказании, то сами характеристики представляются в виде *one-hot* векторов, а наборы морфологических характеристик являются суммой векторов, соответствующих характеристикам из набора.

Однако можно использовать некоторую дополнительную информацию, чтобы поставить в соответствие наборам морфологических характеристик их вес. Это может быть полезно для омонимичных слов. Например, в работе [8] используется большой размеченный корпус, на котором подсчитываются вероятности каждого набора морфологических характеристик. В итоге, словоформе «стул» будет присвоена вероятность 0.56 для формы именительного падежа и 0.44 — для винительного, поскольку в корпусе словоформы в именительном падеже встречались чаще. В данной работе вероятности вычисляются не по корпусу, а по морфологическому словарю системы ЭТАП.

Синтаксические признаки. В т. н. комбинаторном словаре ЭТАПа каждому имени лексемы сопоставляется различная информация, в т. ч. ее синтаксические и семантические признаки, переводной эквивалент и др. Синтаксические признаки слова характеризуют его способность или неспособность участвовать в тех или иных синтаксических конструкциях.

Важно отметить, что синтаксические признаки, придуманные лингвистами, важны в содержательном плане и могут помочь разрешить омонимию. Например, рассмотрим фразу «родственники и знакомые спортсменов». Признак *АДЪЕКТ*, который приписывается существительным адъективного склонения, может быть полезен при выборе части речи словоформы знакомые.

На сегодняшний день общее число синтаксических признаков составляет 240, которые примерно в 4000 различных комбинациях фигурируют в статьях комбинаторного словаря. Их можно представить в виде, аналогичном представлению морфологических признаков.

2.2 Проектирование архитектуры

Для задачи морфологической разметки можно условно выделить 2 части архитектуры: у первой части входом являются (векторные) представления слов (возможно, предобученные), а выходом — их внутреннее (векторное) представление, а вторая часть по этому представлению (и другим признакам) формирует требуемый выход (распределение вероятностей для морфологических характеристик каждой словоформы).

Что касается первой части, в случае, если слово представлено в виде вектора, то достаточно умножения его на некоторую матрицу, веса которой являются параметрами нейросети. Если слово представлено на уровне символов (или морфем), то есть 2 основных используемых архитектуры: *CNN* (сверточная нейронная сеть) и *LSTM* (long short-term memory [9]). В случае с *LSTM* выходом является последнее состояние. Однако, используются и другие архитектуры, например, в [10] используется полносвязная нейросеть. В работе [5] показано преимущество использования LSTM по сравнению с CNN в задаче морфологической разметки для ряда языков, в т. ч. и русского. Для решения же основной задачи в подавляющем большинстве работ используются архитектуры, в основе которых лежит двунаправленная LSTM.

3 Предлагаемая архитектура нейросети

Предварительно следует отметить, что встречаются ситуации, при которых среди омонимов, полученных от морфологического компонента системы ЭТАП, нет правильного. Кроме того, например, из-за синтаксической омонимии («Автобус обогнал трамвай.», «Кустарник закрывает дом.») иногда принципиально невозможно выбрать правильный набор морфологических характеристик, не рассматривая более широкий контекст, чем одно предложение. В связи с этим, постановка задачи такова: для каждого слова в предложении получить распределение вероятностей для набора морфологических характеристик. В нашем случае таких (различных) наборов 480. Из них выбираются несколько наиболее вероятных и передаются синтаксическому компоненту ЭТАПа. Однако в данной работе, для возможности сравнения результатов, правильным набором характеристик считается набор с наивысшей вероятностью, то есть решается задача классификации.

Все архитектуры реализованы с использованием библиотеки Keras [11], а внутренним модулем выбран Tensorflow [12].

3.1 Базовая архитектура

Представление слова. Несмотря на высокую популярность инструментов типа word2vec, представляется важным сохранить информацию о символах слова в задаче морфологической разметки, особенно для флективных языков (которым является русский), поскольку во флексии содержится значительная информация о морфологических характеристиках словоформы. Таким образом, слово представляется на символьном уровне. В [13] предложено передавать символьное представление слова на вход двунаправленной *LSTM*, чтобы учитывать нелокальные зависимости между символами. Однако наши эксперименты не показали преимуществ двунаправленной архитектуры по сравнению с однонаправленной, что интуитивно объясняется тем, что чтение слов на русском языке осуществляется в одном направлении.

Морфологические признаки. Как было сказано в п. 2.1, наборы морфологических характеристик представляются в виде *one-hot* векторов. Получив такие векторы от всех вариантов разбора словоформы морфологическим компонентом ЭТАПа, они затем суммируются (с весами) и нормализуются. Скрытые слои, выходной слой, регуляризация, опитмизатор. Эта часть архитектуры для базовой модели во многом повторяет архитектуру [8] как победителя соревнования *MorphoRuEval-2017* [14] на открытой дорожке. В основе архитектуры лежат 2 двунаправленные *LSTM*, выход последней передается на полносвязный слой с функцией активации *ReLU*. Функция активации выходного слоя — *softmax*, оптимизатор — *Adam* [15].

Что касается регуляризации, стандартные техники для рекуррентных нейросетей, используемые и в [8] — dropout («прореживание») [16] и batch normalization [17]. Однако в работе [18] показано, что в рекуррентных сетях «номера» нейронов, к которым применяется «прореживание», на каждом временном шаге должны оставаться одними и теми же. Этот подход, называемый variational dropout, используется в данной работе.

3.2 Финальная архитектура

Сама архитектура представлена на рис. 1.



Рис. 1. Итоговая архитектура нейросети

Синтаксические признаки. Основным нововведением является использование синтаксических признаков. Будучи представленными в виде onehot векторов (в случае омонимии вектора суммируются и нормализуются), не учитываются внутренние связи между признаками, поэтому полученные векторы являются входом полносвязного слоя. Аналогичное решение было принято для морфологических признаков.

Gated recurrent unit. Первым и единственным улучшением базовой модели было добавление синтаксических признаков. Несмотря на то, что это повысило качество модели (подробнее в п. 4), дальше качество не улучшалось из-за переобучения. Предложенная в 2014 году альтернатива LSTM - GRU (gated recurrent unit) [19] — имеет более высокое качество результатов при решении различных (но не всех) задач [20], хотя содержит меньшее число компонент. Кроме того, в работе [21] было показано, что *GRU* может быть более устойчива к переобучению, чем *LSTM*. В итоге, в финальной модели *LSTM* (как во входном слое для символьного представления, так и в скрытых слоях) была заменена на *GRU*.

Оптимизатор. В недавно опубликованной работе [22] показывается, что часто применяемые адаптивные оптимизаторы (например, Adam) хуже (а часто — гораздо хуже) справляются с переобучением, чем SGD (стохастический градиентный спуск). Однако, принимая во внимание эмпирический факт, что адаптивные оптимизаторы сходятся быстрее неадаптивных, был использован подход SWATS [23] (Switch from Adam to SGD), при котором на начальных итерациях используется адаптивный оптимизатор (в нашем случае Nadam), а затем — SGD.

4 Эксперименты

Обучение нейросети осуществлялось на синтаксически аннотированном корпусе русского языка (СинТагРус [24]). Корпус был разбит на обучающую и тестовую части в отношении 9 : 1. Информация о корпусе представлена в таб. 1.

Таблица 1. Сведения о корпусе СинТагРус. Хотя знаки препинания, поступающие на вход нейросети, детектировались как слова, здесь они не учитываются.

	Обучающая выборка	Тестовая выборка
Предложений	62041	6893
Слов	891475	99053

На корпусе обучались три нейросети, рассмотренные в п. 3. Как и ожидалось, добавление только лишь синтаксических признаков без изменения других параметров дало улучшение точности. Однако, из-за переобучения, улучшение было небольшим. Качество итоговой модели значительно превзошло остальные. Результаты отражены в таб. 2.

Таблица 2. Сравнение точности различных моделей теггеров на тестовой выборке.

	Точность по словам	Точность по предложениям
Базовая модель	97.39%	70.44%
+синтаксические признаки	97.46%	71.00%
Итоговая модель	99.35%	91.32%

5 Выводы и дальнейшая работа

В данной работе, с использованием методов глубокого обучения, был построен морфологический теггер, дающий высокую точность результатов. Также была показана эффективность использования синтаксических признаков в задаче морфологической разметки.

Несмотря на высокую способность последней модели к генерализации, что видно по тестовой выборке, сейчас мы пытаемся сравнить данную модель с другими морфологическими теггерами с соревнования *MorphoRuEval-*2017 [14]. Это полезно хотя бы потому, что на соревновании используются другие морфологические характеристики, соответственно, модель приходится переобучать, что позволит оценить, как способ разметки влияет на качество.

Одним из недостатков архитектуры является ее громоздкость: нейросеть содержит более 10 миллионов параметров, поэтому скорость обработки предложений теггером низка. Следующим шагом будет предпринята попытка ужать модель, сохранив качество.

Список литературы

- 1. Лингвистическое обеспечение системы ЭТАП-2 / Апресян Ю.Д., Богуславский И.М., Иомдин Л. Л. и др. М:Наука, 1989. 295 с.
- Мовсесян А. А., Сизов В. Г. Разработка морфологического анализатора неизвестных слов на основе морфологических словарей лингвистического процессора ЭТАП. //Информационные технологии и системы 2017 (ИТиС'17). Сборник трудов 41-ой междисциплинарной школы-конференции ИППИ РАН.
- Петроченков В. В., Казенников А. О. Статистический теггер для морфологической разметки русскоязычных текстов //Автоматика и телемеханика. 2013.
 №. 10. С. 154—165.
- Collobert R. et al. Natural language processing (almost) from scratch //Journal of Machine Learning Research. – 2011. – T. 12. – №. Aug. – C. 2493-2537.

- Heigold G., Neumann G., van Genabith J. An extensive empirical evaluation of character-based morphological tagging for 14 languages //Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers. – 2017. – T. 1. – C. 505-513.
- Mikolov T. et al. Efficient estimation of word representations in vector space //arXiv preprint arXiv:1301.3781. – 2013.
- Mikolov T. et al. Distributed representations of words and phrases and their compositionality //Advances in neural information processing systems. – 2013. – C. 3111-3119.
- 8. Anastasiev D. G., Andrianov A. I., Indenbom E. M.: Part-of-speech tagging with rich language description, Computational linguistics and intellectual technologies: Proceedings of the International Conference "Dialogue" (2017)
- Hochreiter S., Schmidhuber J. Long short-term memory //Neural computation. 1997. – T. 9. – №. 8. – C. 1735-1780.
- Anastasiev D. G., Gusev I. O., Indenbom E. M.: Improving part-of-speech tagging via multi-task learning and character-level word representations, Computational linguistics and intellectual technologies: Proceedings of the International Conference "Dialogue" (2018)
- 11. Chollet F. et al. Keras. 2015.
- Abadi M. et al. TensorFlow: A System for Large-Scale Machine Learning //OSDI. – 2016. – T. 16. – C. 265-283.
- Ling W. et al. Finding function in form: Compositional character models for open vocabulary word representation //arXiv preprint arXiv:1508.02096. – 2015.
- Sorokin A. et al. MorphoRuEval-2017: an evaluation track for the automatic morphological analysis methods for Russian. – 2017.
- Kingma D. P., Ba J. Adam: A method for stochastic optimization //arXiv preprint arXiv:1412.6980. – 2014.
- Srivastava N. et al. Dropout: A simple way to prevent neural networks from overfitting //The Journal of Machine Learning Research. – 2014. – T. 15. – №. 1. – C. 1929-1958.
- Ioffe S., Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift //arXiv preprint arXiv:1502.03167. – 2015.
- Gal Y., Ghahramani Z. A theoretically grounded application of dropout in recurrent neural networks //Advances in neural information processing systems. - 2016. - C. 1019-1027.
- Cho K. et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation //arXiv preprint arXiv:1406.1078. – 2014.
- Jozefowicz R., Zaremba W., Sutskever I. An empirical exploration of recurrent network architectures //International Conference on Machine Learning. – 2015. – C. 2342–2350.
- Chung J. et al. Empirical evaluation of gated recurrent neural networks on sequence modeling //arXiv preprint arXiv:1412.3555. – 2014.
- Wilson A. C. et al. The marginal value of adaptive gradient methods in machine learning //Advances in Neural Information Processing Systems. – 2017. – C. 4151-4161.
- Keskar N. S., Socher R. Improving Generalization Performance by Switching from Adam to SGD //arXiv preprint arXiv:1712.07628. – 2017.
- Дяченко П. В. и др. Современное состояние глубоко аннотированного корпуса текстов русского языка (СинТагРус) //Труды Института русского языка им. ВВ Виноградова. – 2015. – №. 6. – С. 272-300.

Анализ воздействия деформаций земной поверхности на процесс подготовки сильных землетрясений

К.Н. Петров, В.Г. Гитис, А.Б. Дерендяев

stranger12@list.ru

Abstract. Исследовано воздействие смещений земной поверхности по данным космической геодезии на процесс подготовки сильных землетрясений. Анализ выполнен с помощью инструментов ГИС ГеоТайм 3. Использовались данные GPS и каталог сильных землетрясений Японии за период 2013-2017 гг. При моделировании по данным GPS прогнозировались землетрясения с глубинами гипоцентров до 60 км и магнитудами М≥6 и М≥5.5. Результаты показывают возможность использования пространственно-временных полей деформаций земной поверхности для прогноза сильных землетрясений.

Keywords: геоинформатика, GPS, прогноз землетрясений.

1 Введение

Землетрясения возникают, когда напряжения в породах земной коры превышает предел прочности. Это вызывает разрушение геологической среды и резкий сдвиг. Такое явление сопровождается выделением огромной энергии, которая причиняет большой социальный и экономический ущерб. Для сокращения ущерба решаются две основные задачи прогноза. Пространственный прогноз состоит в построении карт сейсмического районирования, которые определяют нормы антисейсмической застройки. Пространственно-временной прогноз состоит в оценивании области, временного интервала и магнитуды возможного землетрясения. Обе задачи являются актуальными, но до настоящего времени не имеют достаточно обоснованных решений. Сложность задач обусловлена в основном двумя факторами: недостаточной изученностью моделей сейсмичности, а также неполнотой инструментальных данных как о геологической среде, в которой происходят землетрясения, так и о происходящих в ней процессах подготовки землетрясения.

Данная работа связана с задачей пространственно-временного прогноза землетрясений. Для прогноза привлекаются данные мониторинга сейсмотектонического процесса, информация о строении геологической среды и сведения об исторической сейсмичности. В большинстве работ исследуется возможность прогноза землетрясений только по данным сейсмического мониторинга [1 - 6]. В работе [7] рассмотрены методы и экспериментальные результаты прогноза землетрясений по геофизическим временным рядам. К сожалению, сети геофизического мониторинга имеются только в нескольких странах и открытый доступ к этим наблюдениям ограничен.

В ИППИ РАН разрабатываются методы и система автоматического прогноза землетрясений [8]. Первая версия системы (http://distcomp.ru/geo/prognosis/) работает в реальном времени и использует только данные каталогов землетрясений. Требуется исследовать возможность использования для прогноза землетрясений иных источников информации. Такими наблюдениями, которые в реальном времени публикуются в сети Интернет, являются данные GPS.

Смещения земной поверхности, полученные по измерениям GPS, связаны с геомеханическими процессами, происходящими в верхней части земной коры. В [9] показана связь между деформациями поверхности земли и стационарными характеристиками сейсмического процесса. В [10] исследована статистически значимая взаимосвязь пространственно-временных процессов смещения земной поверхности и землетрясений с магнитудами М≥4. Показано, что сейсмичность усиливается в пространственно-временных областях с экстремальными значениями скорости изменения размеров участков земной поверхности и/или их поворота.

Наша цель состоит в исследовании воздействия полей деформаций земной поверхности по данным GPS на процесс подготовки сильных землетрясений и оценивании возможности использовать данные GPS для прогноза землетрясений. В разделе 2 рассматривается метод анализа. В разделе 3 показаны экспериментальные результаты.

2 Метод анализа

Возможность использования данных GPS для прогноза землетрясений оценивалась с помощью инструментов ГИС ГеоТайм 3. Прогнозировались эпицентры землетрясений с глубинами гипоцентров до 60 км и магнитудами $M \ge 6$ и $M \ge 5.5$. В качестве признаков прогноза взяты пространственно-временные поля дивергенции и ротора скоростей горизонтальных смещений земной поверхности. Предполагалось, что тектонические процессы, влияющие на процесс подготовки землетрясений, отображаются в смещениях земной поверхности. В области подготовки сильных землетрясений значения полей дивергенции и ротора скоростей горизонтальных полей дивергенции и ротора бода-

Для прогноза использовался метод наименьшего объема тревоги [8]. Метод использует модель, в которой должны выполняться два следующих предположения:

- 1. Предвестниками сильных землетрясений являются аномалии, относящиеся к минимальным и максимальным значениям полейпризнаков.
- 2. Значения полей меньшие или равные значениями предвестника землетрясения покомпонентно, также могут быть предвестниками анало-

гичных землетрясений (аналогичное условие относится к значениям полей больших или равных значению предвестника землетрясений).

Рассмотрим алгоритм обучения методом наименьшей тревоги. Пусть даны приведенные к единой сетке поля-признаки F_i, i = 1, ..., I, и выборка эпицентров сильных землетрясений с прогнозируемыми магнитудами M ≥ M*. Согласно условию монотонности для вектора пространства признаков f* опорными точками цилиндров тревоги являются узлы координатной сетки, в которых $f^{(n)}_{i} \ge$ f_i^* по всем i = 1, ..., I. Предвестником события q будем называть вектор $f^{(q)}$, относящийся к такому узлу координатной сетки цилиндра предвестника события q, у которого объем тревоги $v^{(q)} = N^{(q)}/N$ является минимальным среди всех узлов сетки этого цилиндра, где N – все узлы сетки анализируемой области, N^(q) – узлы сетки области тревоги, порождаемой опорными точками вектора f^(q). Очевидно, что при сокращении размерности пространства признаков до J < I опорные точки для события q будут включать в себя все предыдущие опорные точки І-мерного пространства признаков и дополнительные точки, которые могут добавиться из-за уменьшения числа ограничений, задаваемых неравенствами $f_{i}^{(n)} \ge f_{i}^{(q)}$. Таким образом, по определению каждому событию q соответствуют предвестник f^(q) и минимальный для данного пространства признаков объем тревоги v^(q). Для обнаружения событий по значениям прогнозного поля используется пороговый детектор. Значения порога в определяется перебором. При этом значение порога уменьшается, начиная от максимального значения прогнозного поля. По мере уменьшения порога количество обнаруженных событий и объем тревоги увеличиваются. Тем самым определяется как последовательность обнаруживаемых событий q(θ), так и функция увеличения объема тревоги V(θ). Очевидно, что этой последовательности событий соответствует последовательность выбора узлов сетки полей-признаков со значениями предвестников $f^{(q)} = \{f^{(q)}_{i}\}$ в пространстве признаков. Алгоритм наименьшей тревоги позволяет организовать прогнозное так, чтобы при уменьшении значения порога θ узлы сетки со значениями предвестников $\phi^{(q(\theta))}$, а следовательно, обнаруживаемые события q(θ), выбирались в любой последовательности.

Алгоритм состоит из следующих шагов.

1. Сгенерировать учебную выборку $\{f^{(q)}, v^{(q)}\}$, из предвестников землетрясений $f^{(q)}$ и соответствующих им объемов тревоги $v^{(q)}$.

2. Упорядочить предвестники $f^{(q)}$, q = 1, ..., Q, по любому правилу пользователя.

3. Присвоить всем узлам сетки прогнозного поля Ф значение 0.

4. Заменить в узлах сетки прогнозного поля, для которых выполняется условие монотонности $f^{(n)}_{i} \ge f^{(Q)}_{i}$ по всем $i = 1, \ldots, I$, значение 0 на Q; заменить в узлах сетки прогнозного поля, для которых выполняется условие монотонности $f^{(n)}_{i} \ge f^{(Q-1)}_{i}$ по всем $i = 1, \ldots, I$, значение 0 на Q – 1 и далее аналогично заменять в узлах сетки прогнозного поля, для которых выполняется условие монотонности $f^{(n)}_{i} \ge f^{(Q-1)}_{i}$ по всем $i = 1, \ldots, I$, значение 0 на Q – 1 и далее аналогично заменять в узлах сетки прогнозного поля, для которых выполняется условие монотонности $f^{(n)}_{i} \ge f^{(Q-q)}_{i}$ по всем $i = 1, \ldots, I$, значения 0 на Q – 1 – q.

После первого шага алгоритма получаем две выборки:

1. Выборка из предвестников землетрясений $f^{(q)}$ и соответствующих им объемов тревоги $v^{(q)}$.

2. Выборка остальных точек пространства признаков, в которых имеются не выявленные предвестники землетрясений.

Такая задача машинного обучения относится к задаче одноклассового распознавания образов. Задача состоит в том, чтобы прогнозное поле на выборке обнаруживало максимальное число предвестников при объеме тревоги не более заданного.

3 Моделирование

Для исследования выбрана территория Японии в границах 128°-146° в.д., 29°-44° с.ш. Использовались данные GPS-станций геодезической лаборатории Невады [11, ftp://gneiss.nbmg.unr.edu/] и каталог землетрясений, взятый с сайта Международного сейсмологического центра (ISC, www.isc.ac.uk). На территории Японии данные GPS регистрировались 1284 станциями (Рис. 1). Рассматривались только горизонтальные компоненты смещений поверхности земли. Они представлены временными рядами смещений в направлениях - X (восток), и Y (север), с шагом 5 мин от 10 апреля 2013 г. Из сейсмического каталога выбраны землетрясения с магнитудами М≥5.5 и глубинами до 60 км, которые произошли на этом же временном интервале. Каждое землетрясение представлено координатами эпицентра и магнитудой.



Рисунок 1. Карта станций GPS, которые использовались при анализе.

Вычисление пространственно-временных полей деформаций земной поверхности выполнялось в несколько этапов.

На рис. 2 показаны графики 3-х временных рядов GPS данных. Можно видеть, что измерения зашумлены, в них имеются разрывы и значительные выбросы.

На первом этапе ряды были сглажены с экспоненциальным ядром:

$$y_n(t) = \frac{\sum_{i=0}^{H} x_n(t-i)e^{-\frac{i}{N}}}{\sum_{i=0}^{H} e^{-\frac{i}{N}}},$$
(1)

где $y_n(t)$ – искомое сглаженное значение в момент времени $t, x_n(t)$ - значение

временного ряда во времени t, H – размер временного окна для усреднения, $e^{-\dot{N}}$ – экспоненциальная функция, называемая ядром усреднения [12], а N – параметр масштаба. Результат применения этой формулы для сглаживания на временных рядах с рисунка 2, со значением временного окна H = 8640 значений (30 суток) и параметром масштаба n = 4000 (15 суток), показаны на рисунке 3.



Рисунок 2. Примеры временных рядов GPS.



Рисунок 3. Пример сглаженных временных рядов GPS.

Пропущенные значения временных рядов были заполнены с помощью линейной интерполяции. Далее использовались временные ряды с шагом 15 суток.

Следующим этапом было вычисление скоростей смещения поверхности земли по формуле: $z_n(t) = y_n(t) - y_n(t-1)$, где $y_n(t)$ – значение временного ряда во времени t.

После этого временные ряды были интерполированы по пространству в пространственно-временные сеточные поля. Использовалась интерполяция по методу обратных взвешенных расстояний:

$$u_n(t) = \frac{\sum_{k=1}^{K} \frac{z_k(t)}{(r_k)^p}}{\sum_{k=1}^{K} \frac{1}{(r_k)^p}},$$
(2)

где: $u_n(t)$ – значение поля в узле сетки n, K – максимальное количество временных рядов, принимаемых для усреднения, $z_k(t)$ – значение k-го входного ряда k=1,...,K, в момент времени t, r_k – расстояние от точки привязки k-го ряда до точки сетки u_n, p – степень пространственного сглаживания. Если $r_k > R_{max}$ (максимальное расстояние для сглаживания), то k-ый входной ряд не участвует в вычислении $u_n(t)$, если $r_k = 0$, то $u_n = z_k$.

Интерполяция, выполненная по формуле 2, была проведена со значениями радиуса усреднения $R_{max} = 50$ км, степени p = 1 и количеством усредняемых временных рядов K = 5. Результатом интерполяции стали два сеточных поля V_x и V_y .

При отсутствии значений вертикальной компоненты скорости перемещения земной поверхности, параметр дивергенция указывает на относительное сокращение или расширение размера малой горизонтальной площадки в окрестности точки пространственно-временного растра. Эта скалярная величина не зависит от выбранной системы координат и находится по формуле 3.

div
$$V = \frac{\partial V_x}{\partial x} + \frac{\partial V_y}{\partial y}$$
, (3)

Ротор указывает на поворот малой горизонтальной площадки вокруг вертикальной оси:

$$\operatorname{rot} V = \frac{\partial V_{x}}{\partial y} - \frac{\partial V_{y}}{\partial x}, \qquad (4)$$

Для анализа меры воздействия скоростей горизонтальных смещений земной поверхности на процесс подготовки сильных землетрясений по полям дивергенции и ротора прогнозировались землетрясения с магнитудами М≥6 и М≥5.5. Область тестирования ограничена полями вычисленными по данным GPS.

Поля дивергенции и ротора вычислялись в координатной сетке с шагом 0.1° по долготе, 0.1° по широте и 15 суток по времени. Для прогноза использовались два поля-признака. Цилиндр тревоги имеет радиус 20 км и временной интервал 365 суток. Обучение первого временного среза поля тревоги проводилось на интервале 862 суток от 26.04.2013 до 05.09.2015. Прогноз давался с шагом 15 суток, начиная от 06.09.2015 до 01.01.2018. В анализируемом регионе на интер-

вале тестирования оказалось 8 эпицентров землетрясений с магнитудами М≥6 и 18 с магнитудами М≥5.5.

На ринке 4 показан график зависимости (1) долей правильно обнаруженных событий Q^* от всех Q событий $U(\theta) = Q^*/Q$ (уровень обнаружения) и (2) долей числа точек сетки N*, попадающих в область тревоги, от числа всех точек сетки N анализируемой пространственно-временной области V (θ) =N*/N (объем тревоги). На рисунке 4 показаны зависимости вероятности успешного прогноза землетрясений U(θ) от объема тревоги V(θ) для землетрясений с магнитудами М≥6. Из рисунка можно видеть, что при использовании одного поля-признака наилучшие прогнозные зависимости U(V) для объема тревоги V≤0.2 дают положительные значения поля дивергенции (тёмно-красная кривая) и положительные значения поля ротора (фиолетовая кривая). Для объема тревоги V≤0.2 по положительным значения поля дивергенции из 8 событий предсказано 5, а по положительным значениям поля ротора предсказано 6. Это означает, что прогноз с объемом тревоги V≤0.2 по полю дивергенции в 2.5, а по полю ротора в 3 раза лучше случайного. При использовании двух полей (поля положительных значений дивергенции и поля отрицательных значений ротора) прогноз улучшается (оранжевая кривая). Для объема тревоги V≤0.15 предсказано 6. Это означает, что прогноз с объемом тревоги V≤0.15 по двум полям дивергенции и ротора в 4 раза лучше случайного.



Рисунок 4. Графики зависимости вероятности успешного прогноза землетрясений U(θ) от объема тревоги V(θ) для землетрясений с магнитудами M≥6.

На рис. 5 показаны зависимости вероятности успешного прогноза землетрясений по полям положительных значений дивергенции и отрицательных значений ротора от объема тревоги для землетрясений с магнитудами М≥6 и М≥5.5. Можно видеть, что прогноз землетрясений с магнитудами М≥5.5 менее успешен. Это объясняется тем, что при уменьшении пороговой магнитуды прогнозируемых землетрясений аномальные процессы при их подготовке становятся менее выраженными.



Рисунок 5. График зависимости процента спрогнозированных (ось ординат) землетрясений от объёма тревоги (ось абсцисс), с магнитудой ≥ 6 (зелёная кривая) и магнитудой ≥ 5,5 (бордовая кривая).

На рисунках 6 и 7 показаны 8 прогнозируемых эпицентров землетрясений с магнитудами М≥6 и 18 с магнитудами М≥5.5. Закраска эпицентров соответствует значению поля тревоги, при котором давался прогноз. Наиболее темный тон показывает, что данный эпицентр обнаружен полем с объемом тревоги менее 0.05. Интенсивность тона закраски уменьшается в соответствии с величинами объема тревоги: 0.05, 0.1, 0,15, 0.2. Белая закраска показывает, что землетрясение не предсказано при объеме тревоги менее 0.2.



Рисунок 6. Прогнозируемые эпицентры землетрясений с магнитудой М≥6. Закраска эпицентров соответствует значению поля тревоги, при котором давался прогноз.



Рисунок 7. Прогнозируемые эпицентры землетрясений с магнитудой М≥6. Закраска эпицентров соответствует значению поля тревоги, при котором давался прогноз.

Выводы

Показано, что близкие к экстремальным значения скорости сокращения или расширения размеров участков земной поверхности (сильных изменений дивергенции), и/или близкие к экстремальным значения скорости поворота участков земной поверхности (сильных изменений ротора) свидетельствуют о процессе подготовки сильных землетрясений.

Результат требует дополнительной проверки на других сейсмически активных регионах.

Источники литературы

- Завьялов А. Д. Среднесрочный прогноз землетрясений: основы, методика, реализация. М.: Наука, 2006.
- Кособоков В. Г. Прогноз землетрясений, основы, реализация, перспективы. Серия" Вычислительная сейсмология". Вып. 36. I. – 2005.
- 3. Шебалин Н. В. Сильные землетрясения. М. : Изд-во Акад. горн. наук, 1997.
- 4. Lighthill M. J. A critical review of VAN: earthquake prediction from seismic electrical signals. World scientific, 1996.
- 5. King C. Y. Gas geochemistry applied to earthquake prediction: An overview //Journal of Geophysical Research: Solid Earth. 1986. T. 91. №. B12. C. 12269-12281.
- Гитис В. Г., Дерендяев А. Б. Обучение прогнозу землетрясений методом наименьшей тревоги //Информационные процессы. – 2017. – Т. 17. – № 2. – С. 123-136.
- Гитис, Ермаков, 2004. Основы пространственно-временного прогнозирования в геоинформатике // М., ФИЗМАТЛИТ, 2004. 256 С.
- Gitis V.G., Derendyaev A.B. Web-Based GIS Platform for Automatic Prediction of Earthquakes // Computational Science and Its Applications – ICCSA 2018, Springer, 2018
- 9. Мухамедиев Ш.А., Зубович А.В., Кузиков С.И. Выделение блоков земной коры по данным GPS_измерений // Докл. РАН. 2006. Т. 408. № 4. С. 539–542.
- Соболев Г. А., Закржевская Н. А., Акатова К. Н., Гитис В. Г., Дерендяев А. Б., Брагин В. Д., Сычева Н., Кузиков С. И. Динамика взаимодействия полей сейсмичности и деформаций земной поверхности (Бишкекский геодинамический полигон) //Физика Земли. – 2010. – №. 10. – С. 15-37.
- ISPRS T. C., ISPRS III T. C. The international archives of the photogrammetry, remote sensing and spatial information sciences. – 2000.
- Härdle W., Stoker T. M. Investigating smooth multiple regression by the method of average derivatives //Journal of the American statistical Association. – 1989. – T. 84. – №. 408. – C. 986-995.

Ответ на общий вопрос: семантика слова НЕТ

Светлана Тимошенко^[0000-0002-6700-1644]

IITP RAS, Bolshoy Karetny per. 19, build.1, 127051 Moscow, Russia timoshenko@iitp.ru

Аннотация В работе рассматриваются некоторые типы общего вопроса, то есть такого, который допускает в качестве ответа Дa или Hem, и особенности поведения ответа Hem. Рассмотренный материал позволяет установить, что семантика ответа Hem заметно отличается от семантики ответа Дa. Дa является простым подтверждением логической формулы вопроса, а Hem - это самостоятельное негативное утверждение. Если в вопросе есть утверждение с отрицанием, Hemс ним совпадает. Если такого утверждения нет, оно его образует.

Keywords: общий вопрос · отрицание · семантика естественного языка

1 Описание проблемы

Рассмотрим две пары "вопрос - ответ".

- (1) (a) *Вася приходил? Нет.*
 - (b) Вася не приходил? Нет.

Ответ *Hem* в обоих случаях означает 'нет, не приходил'. Но как это получается? Рассмотрим (1а). *Hem* = 'Вася не приходил'. Простейший анализ этого случая: предположить, что структура вопроса такого типа (общего вопроса) - это оператор вопроса, примененный к утверждению 'Вася приходил', и что *Hem* - это логическое отрицание утверждения из вопроса - $\neg P$. Однако если отрицание *не* в вопросе (1b) тоже соответствует логическому отрицанию, ответ *Hem* ведет себя парадоксально: он значит "нет, не приходил", хотя должен значить "нет, приходил", поскольку $\neg(\neg P)$, подчиняясь правилу снятия двойного отрицания, должно быть эквивалентно утверждению *P*.

2 Возможные пути решения

Есть три способа разрешить это затруднение: 1) констатировать, что *Hem* в ответах на вопросы с отрицанием - это другое *Hem*, отличное от того *Hem*, которое служит ответом на общие вопросы без отрицательной частицы и включает в себя логическое \neg ; 2) констатировать, что частица *не* в рассматриваемом типе вопросов не выражает логического \neg и не входит в состав

2 С. Тимошенко

утверждения, находящегося в сфере действия оператора вопроса, а "делает" что-то иное; 3) предложить более сложное описание семантики *Hem*, чем логическое отрицание утверждения, находящегося в сфере действия оператора вопроса.

2.1 Решение 1: два разных Нет

Решение 1 кажется самым неоправданным, поскольку аналитические процедуры, принятые в лексикографии и позволяющие установить, сколько словлексем соответствует одному написанию, к сложебным словам плохо применимы. Кроме того, сам факт, что предполагаемые разные *Hem* находятся в отношении дистрибуции, контекст которой легко описать, убеждает нас в том, что ответ *Hem* - это одна языковая единица.

2.2 Решение 2: два разных не

Имеется аргумент в пользу решения 2 - существование в русском языке вопросительной конструкции с *не*, выражающей вежливую просьбу:

- (2) (а) Лишнего пригласительного не найдется?
 - (b) Caxap не передашь?

Аналогичный пример (*Молока не продашь ли, хозяюшка?*) приводится в "Русской грамматике" под редакцией Н.Ю. Шведовой [1]. В этой грамматике различаются первичная (собственно запрос информации) и вторичные функции вопросительных предложений. Пример с отрицанием снабжен пояснением "содержит просьбу" (стр. 387) и иллюстрирует одну из вторичных функций вопроса. Стратегии ответов на такие вопросы в грамматике не описываются.

На ограничения, которые эта конструкция накладывает на ответ, обратила внимание Т.В. Булыгина [2]. По её мнению, на вопросы (2) возможен только отрицательный ответ, и тогда это не столько ответ, сколько отказ в выполнении просьбы, иногда довольно грубый (*He nepedam*). Булыгина утверждает также, что утвердительный ответ, не сопровождающийся выполнением действия, любой носитель русского языка сочтет не адекватным ситуации [2] (стр. 267). Автор, очевидно, исходит из собственной языковой интуиции возможности экспериментальной проверки здесь весьма ограничены.

Тем не менее, если проанализировать в устном подкорпусе Национального корпуса русского языка употребление вопросов с самым актуальным для этой конструкции глаголом *найтись* (запрос: indic,fut,sg bques), то окажется, что из 44 найденных примеров 42 будут являться просьбами, и 29 из них будут содержать частицу *не*, так что по крайней мере можно говорить о тяготении просьб с этим глаголом к конструкции "вопрос с отрицанием". Относительно других глаголов на материале устного подкорпуса установить ничего определенного не удается, так как даже частотные глаголы в заданной форме и перед знаком вопроса встречаются в нем в среднем по десятку

3

раз, и почти все примеры являются несомненными вопросами с вопросительным словом. Однако принципиальная возможность классифицировать вопросы с любыми глаголами как просьбы остается. В этом случае употребление *не* логически объяснимо: нужно иметь дополнительное средство, позволяющее отличать просьбы от настоящих вопросов, и *не* иногда берет на себя эту функцию.

2.3 Структура общего вопроса

Естественно считать, что когда человек задает общий вопрос, то он заранее имеет некоторое предположение, которое и считает более вероятным. Но всегда ли это так? И возможен ли на русском языке вариант чистого логического запроса - непредвзятого запроса $P \lor \neg P$?

Если спрашивающий формирует общий вопрос на основе некоторого предположения, то разница в форме вопроса может отражать разницу между предположениями спрашивающего. Если это верно, то (1a) и (1b) - два разных вопроса, потому что в основе их лежат два разных предположения.

Подобного взгляда на общий вопрос придерживается Е.В. Падучева. [3] Она вводит понятие исходного предположения (ИП) - дизъюнкции всех мыслимых альтернатив ответа. Однако исходным предположением общего вопроса она считает не дизъюнкцию $P \lor \neg P$, а только тот из дизъюнктивных членов, который говорящий считает более вероятным. Анализируя вопросы

- (3) (a) *Вы хотите чаю?*
 - (b) Вы не хотите чаю?
 - (с) Вы хотите чаю или нет?

Падучева утверждает, что "в смысл вопроса (3а) входит ИП 'Вы хотите чаю', в смысл (3b) - ИП 'Вы не хотите чаю', а вопрос (3c) не содержит никакого ИП." (стр. 239)

С изложенным анализом трудно согласиться. Во-первых, возникает странное противоречие. Понятие исходного предположения как дизъюнкции всех альтернатив ответа применяется к вопросам с вопросительным словом. А у общего вопроса, который по форме ближе всего к дизъюнкции, вдруг нет никакого исходного предположения. Во-вторых, приведенные примеры, как отмечает сам автор, могут быть не столько вопросами, сколько вежливыми предложениями - в зависимости от интонационного оформления. Почему анализ структуры общего вопроса и проверка гипотезы об определенном соотношении форм вопроса и исходных предположений иллюстрируется не самыми типичными примерами?

2.4 Исходное предположение и общий вопрос с частицей разве

Чтобы прояснить вопрос о предположениях спрашивающего в случае вопросов (1a) и (1b), рассмотрим еще одну разновидность общих вопросов:

(4) (а) Разве Вася приходил?

4 С. Тимошенко

(b) Разве Вася не приходил?

Если эксплицитно выразить смысл, который имеет вопрос с частицей *разве*, (4a) будет выглядеть следующим образом:

- 1. Есть основания полагать, что Вася приходил;
- 2. говорящий считает, что Вася не приходил;
- 3. говорящий хочет, чтобы собеседник подтвердил, что Вася не приходил.

А в (4b) все три утверждения будут отличаться на отрицание:

- 1. Есть основания полагать, что Вася не приходил;
- 2. говорящий считает, что Вася приходил;
- 3. говорящий хочет, чтобы собеседник подтвердил, что Вася приходил.

Проверить обязательность утверждений (1-3) для значения вопроса с *разее* можно с помощью мысленного эксперимента. Мысленный эксперимент - это способ задать необходимый экстралингвистический контекст. Предполагается, что формулировка условия в случае мысленного эксперимента является полной. Добавление каких-либо деталей к контексту меняет условия эксперимента, и следовательно, порождает новый эксперимент, отличный от начального.

Эксперимент 1. Допустим, я прихожу в некоторое место. Существует вероятность, что Вася приходил туда раньше, чем я. Я об этой вероятности знаю и хочу узнать, имел ли место приход Васи.

В такой ситуации мне естественно задать вопросы (1a) и (1b), и невозможно - (4a) и (4b), поскольку у меня нет убеждения 2, ни положительного, ни отрицательного. А тот факт, что (1a) и (1b) здесь равно уместны, опровергает анализ подобных общевопросительных пар как различающихся на исходное предположение. Семантических различий между примерами (1a) и (1b) пока не выявлено. Видоизменим наш мысленный эксперимент.

Эксперимент 2. Допустим, я прихожу в некоторое место и вижу на столе знакомую кружку с остатками чая. Я знаю, что из нее здесь обычно пьет Вася, поэтому предполагаю, что Вася приходил, и хочу узнать, действительно ли это так.

Уместно спросить (1а). Невозможно спросить не только (1b), но и (4a), и (4b). (4a) и (4b) исключены потому, что, по условиям эксперимента, отсутствует противоречие между моим предположением и внешней информацией. Если добавить к условиям эксперимента уверенность спрашивающего в том, что Вася не приходил, то наблюдаемая кружка создаст предпосылку для противоречия, и вопрос (4a) (*Pasee Bacs npuxodus?*) станет возможным. (4b) останется по-прежнему невозможным. (1b) в такой ситуации (*Bacs не npuxodus?*) также будет не самым естественным выбором, что снова опровергает гипотезу о наличии у этого вопроса исходного предположения 'Вася не приходил'.

Теперь можно проанализировать функционирование частицы *не* в вопросах типа (4). В (4а) имеется предположение 'Вася не приходил', следовательно, отрицание является частью значения *разее*. В (4b) говорящий предполагет

5

обратное, а именно 'Вася приходил', и это утверждение возникает в результате того, что в сферу действия частицы *разве*, содержащей оператор \neg , помещается отрицательное утверждение с частицей *не*: $\neg \neg P \equiv P$ Если эта интерпретация верна, мы сможем объяснить и более сложные примеры:

- (5) (а) Разве ты знаешь, что Вася приходил?
 - (b) Разве ты не знаешь, что Вася приходил?

(5а) - это, действительно, вопрос. На него можно дать ответ Да ('Да, знаю') или *Hem* ('Нет, не знаю').

А вот (5b) - это не вполне вопрос. И Дa, и *Hem* в качестве ответной реакции выглядят здесь странно. Ср. один из крайне немногочисленных примеров в НКРЯ [4]:

(6) Разве ты не знаешь, что мы, православные, десять лет проливали свою кровь, чтобы свергнуть турецкое иго? - Нет, не знал. [Николай Кокухин. Невидимые старцы, или Афон сокровенный // «Наука и религия», 2010]

Утрату вопросительной функции в случае (5b) можно объяснить взаимодействием *не* со скрытыми утверждениями, содержащимися в вопросе:

- Вася приходил (вклад семантики фактивного глагола *знать*);
- адресату вопроса известно, что Вася приходил (не + разве = $\neg \neg p = p$).

То есть спрашивающий считает, что Вася приходил и что адресату это известно. Иными словами, такая форма вопроса предполагает, что спрашивающий и адресат обладают одинаковой информацией. Следовательно, спрашивать бессмысленно, и конструкция приобретает иную модальность: говорящий выражает удивление, потому что считает, что адресат должен знать о приходе Васи.

2.5 Общий вопрос с не и вопрос вида Р или нет?

Теперь можно вернуться к результатам мысленного эксперимента 2, к тому факту, что при наличии предположения, что Вася приходил, уместно задать вопрос (1a) и неуместно - вопрос (1b). Немного поменяв условия, мы далее установили, что при наличии обратного убеждения ('Вася не приходил') вопрос (1b) все равно не становится более естественным для данной ситуации. Можно предположить, что он возможен только в ситуации полной неопределенности и, таким образом, является формой выражения запроса вида $P \vee \neg P$ на естественном языке. Если это верно, то что означают вопросы типа (3c) (*P или nem?*)?

На вопрос *Вася приходил или нет?* можно ответить *Нет* и ни при каких условиях нельзя ответить $\mathcal{A}a$. Формально этот вопрос содержит дизъюнкцию, а значит, две какие-то пропозиции, и ответ должен быть выбором одной из них. Причем дизъюнкция вопроса - не просто свободная дизъюнкция пропозиций, различающихся на отрицание, так как невозможно:

26

6 С. Тимошенко

(7) (a) **Hem unu Bacs npuxodun?*

- (b) *Вася не приходил или да?
 - (с) *Вася не приходил или нет?

Значит, мы имеем конструкцию, в которой присутствует сложный оператор вопроса 'Я хочу, чтобы ты сказал мне про Р, истинно оно или ложно', и на поверхностном уровне *Hem* в вопросе означает 'Р ложно'. А 'Нет' в ответе - это указание на эту вторую пропозицию. Чтобы указать на другую пропозицию, нужно её как-то обозначить, например, глаголом - *приходил*. На этом примере мы видим ассиметрию *Да* и *Hem*: *Да*, в отличие от *Hem*, является подтверждением всей логической формулы вопроса, и возможно в ответе на вопрос с дизъюнкцией только тогда, когда члены дизъюнкции не исключают друг друга:

(8) Хочешь кофе или чаю? - Да.

(8) прагматически странный, но не невозможный пример, и ответ $\mathcal{A}a$ означает 'Хочу кофе или чай', то есть отвечающему все равно, что пить. Кстати, следствием такой семантики $\mathcal{A}a$ является синонимичность ответов $\mathcal{A}a$ и Hem на вопрос (1b) - оба ответа означают 'Вася не приходил'. Hem устроено сложнее - можно сказать, что это негативное утверждение. Если в вопросе есть утверждение с отрицанием, Hem выступает как его заместитель, своего рода отрицательное местоимение, замещающее собой целое предложение. Если такого утверждения нет, а есть нейтральный вопрос, Hem такое утверждение образует.

3 Выводы

- 1. Структуру общего вопроса не всегда правильно представлять как ?Р
- 2. Запросу $P \vee \neg P$ точнее всего соответствует общий вопрос с ne при глагольной вершине P.
- 3. Ответы Да и $H\!em$ не симметричны.
- 4. *Нет* это негативное утверждение. Если в вопросе есть утверждение с отрицанием, *Нет* с ним совпадает. Если такого утверждения нет, оно его порождает.

Работа выполнена при поддержке Российского научного фонда (грант № 16-18-10422)

Список литературы

- 1. Русская грамматика. Под ред. Н.Ю. Шведовой, Т. II, Издательство Наука, Москва (1980)
- Булыгина, Т.В., Шмелев, А.Д. : Языковая концептуализация мира (на материале русской грамматики), Школа "Языки русской культуры Москва (1997)
- Падучева, Е.В. : Высказывание и его соотнесенность с действительностью: Референциальные аспекты семантики местоимений. 6-е изд. Издательство ЛКИ, Москва (2010)
- 4. Национальный корпус русского языка, режим доступа: http://ruscorpora.ru

Сравнение ядерного алгоритма k-средних с методом Уорда в задачах кластеризации вершин графов

Николай Козырский^{1,2,4}, Павел Чеботарёв^{2,5}, Глеб Балицкий^{1,2,3}

¹ Сколковский институт науки и технологий, Москва, Россия

² Московский физико-технический институт, Москва, Россия

³ Институт проблем передачи информации РАН, Москва, Россия

⁴ Институт радиотехники и электроники имени В. А. Котельникова РАН,

Москва, Россия

⁵ Институт проблем управления имени В.А. Трапезникова РАН, Москва, Россия nikolay.kozyrskiy@phystech.edu, pavel4e@gmail.com,

gleb.balitskiy@phystech.edu

Аннотация Основные цели данной работы – исследовать и сравнить качество кластеризации алгоритмом k-средних и методом Уорда на случайных графах, сгенерированных на основе стохастической блочной модели, и нескольких реальных датасетах. Чтобы повысить качество кластеризации, для каждого алгоритма строились спрямляющие пространства при помощи ряда нелинейных ядер и их логарифмированных версий. Метод-победитель определялся голосованием по правилу Коупленда. Аналогично сравнивались результаты кластеризации на реальных датасетах и исследовалась устойчивость методов при добавлении шума в структуры реальных графов.

Ключевые слова: Случайные графы, стохастическая блочная модель (СБМ), кластеризация на графах, ядро, алгоритм *k*-средних (*k*-means), метод Уорда, голосование Коупленда.

1 Введение

Любые неоднородные данные содержат некоторую базовую структуру, являющуюся отражением их гетерогенной природы. Процесс поиска этой структуры в терминах группировки элементов данных называется кластеризацией, а сами выявленные группы называются кластерами [1].

Графы – структуры, образованные множеством вершин и множеством ребер, обозначающих связи между парами вершин – удобны при формализации многих задач из самых различных предметных областей, будь то поиск групп пользователей социальной сети с общими интересами или оптимизация сети автодорог. Кластеризация на графах – это задача группирования вершин графа в кластеры с учетом структуры ребер графа таким образом, чтобы в каждом кластере было сравнительно много ребер, а между кластерами ребер было относительно мало. В реальности базовые структуры, порождающие графы, весьма разнообразны; различны их природа, характеристики, размеры. По этой причине, по-видимому, не существует простого алгоритма кластеризации, который был бы оптимальным во всех случаях. Однако для семейств графовых моделей с близкими свойствами задача о поиске оптимального алгоритма кластеризации вполне осмысленна. К настоящему моменту разработан целый ряд методов кластеризации на графах. Например, Spectral clustering [3], методы иерархической кластеризации, SCAN, Walktrap и другие [2,6]. Некоторые методы объединяет необходимость/возможность использования той или иной меры близости между вершинами графа. Меры, обладающие определенными свойствами, называют ядрами.

Цели данной работы: 1) сравнение эффективности ряда мер близости вершин графов в задачах кластеризации при использовании их совместно с популярными алгоритмами k-средних (k-means) и Уорда; 2) сравнение эффективности самих алгоритмов k-средних и Уорда, применяемых в комбинации с рассмотренными мерами близости. Исследование в работе ведется на искусственных графах, сгенерированных с помощью стохастической блочной модели [4], и на ряде известных реальных наборах данных (датасетах).

Среди работ близкой тематики отметим статью [7], где метод Louvain сравнивался с ядерным алгоритмом k-средних и применялся критерий модулярности. При этом использовались меры близости Sigmoid Commute Time, Logarithmic Forest и др. Эти две меры близости также участвовали в соревновании мер в диссертации [10], целью которой было найти лучшую меру для кластеризации методом Уорда на определенных датасетах. В работе [8] исследовались свойства таких ядер, как Heat Kernel, Normalized Heat Kernel, Regularized Laplacian Kernel, Absorption Kernel и других. Метрика, соответствующая мере близости Forest accessibility, изучалась в [9]. В работах [5,11] была введена мера Communicability и описаны ее особенности. В [15] было изучено ядро Walk.

2 Определения и обозначения

Граф – это любая пара множеств G = (V, E), где V, множество вершин, – подмножество любого счетного множества, $E \subseteq \{(v_i, v_j) : v_i, v_j \in V\}$ – множество ребер. Матрица смежности A графа G = (V, E) с конечным числом пронумерованных числами от 1 до n вершин – это квадратная матрица размера $n \times n$, в которой элемент a_{ij} , $i, j = \overline{1, n}$, равен числу ребер $(v_i, v_j) \in E$ в графе G. С ее помощью удобно представлять граф с конечным числом вершин в виде матрицы. В данной работе не рассматриваются мультиграфы, и элементы матриц смежности являются индикаторами наличия ребра между вершинами i и j:

$$a_{ij} = \begin{cases} 1, & \exists (v_i, v_j) \in E, \\ 0, & \text{в противном случае.} \end{cases}$$

Матрица смежности неориентированного графа симметрична. Матрицей Лапласа графа G = (V, E) называется квадратная матрица L = D - A, где $D = \text{diag}(A \cdot \mathbf{1}), \mathbf{1}$ – единичный вектор $n \times 1$. В работе используем связные графы; петли разрешены.

Граф, сгенерированный блочной стохастической моделью, будем обозначать $G = (v_n, p_{\rm in}, p_{\rm out})$, где v_n – мощность множества вершин графа, $p_{\rm in}$ и $p_{\rm out}$ – вероятности наличия ребра между двумя вершинами, принадлежащими одному и тому же и разным классам соответственно. Таким образом, данная модель позволяет регулировать плотность внутри классов и межкластерную разреженность.

3 Ядра

Положительно полуопределенная матрица $K = (K_{xy})$ называется ядром графа G. Функция $k(x, y) = K_{xy}$, генерирующая элементы ядра, интерпретируется как мера близости вершин G. Для определенного класса мер близости k(x, y) на конечном множестве A функция $d : A \times A \to \mathbb{R}$, определяемая равенством $d(x, y) = \frac{1}{2}(k(x, x) + k(y, y)) - k(x, y), x, y \in A$, является функцией расстояния. Подробнее это обсуждается в [13].

3.1 Communicability, Katz Kernel, Forest, Heat и Sigmoid Commute Time

Через t будем обозначать параметр, выделяющий ядро из параметрического семейства ядер. Область значений этого параметра определяется для каждого семейства по-своему; для удобства сравнения и отображения на графиках будем преобразовывать ее к отрезку [0; 1].

- Communicability Kernel (Com) экспоненциальное диффузионное ядро определяемое равенством $\mathbf{K}^{\text{Com}}(t) = \exp(tA) = \sum_{i=0}^{\infty} \frac{t^k}{k!} A^k, t > 0.$
- Katz Kernel ("Plain" Walk) диффузионное ядро Каца (иногда называемое ядром Неймана): $\mathbf{K}^{\text{Katz}}(t) = (I - tA)^{-1}, \ 0 < t < (\rho)^{-1},$ где ρ – спектральный радиус матрицы смежности A.
- Matrix-Forest (Forest) Kernel регуляризованное ядро Лапласа: $\mathbf{K}^{\text{For}}(t) = (I + tL)^{-1}, t > 0.$
- Heat Kernel (Heat) экспоненциальное диффузионное ядро Лапласа: $\mathbf{K}^{\text{Heat}}(t) = \exp(-tL) = \sum_{i=0}^{\infty} \frac{(-t)^k}{k!} L^k, \ t > 0.$ — Commute Time. $d^{\text{CT}}(v_i, v_j)$ – средняя длина кольцевого маршрута (v_i –
- Commute Time. $d^{\text{CT}}(v_i, v_j)$ средняя длина кольцевого маршрута $(v_i v_j v_i)$. Соответствующее ядро $K^{\text{CT}} = L^+$ называется Commute Time. Применяя сигмоиду поэлементно к K^{CT} , получаем сигмоидальное ядро Sigmoid Commute Time Kernel (SCT) \mathbf{K}^{SCT} : $[\mathbf{K}^{\text{SCT}}]_{ij} = \frac{1}{1 + \exp\left(\frac{-tk_{ij}}{\sigma}\right)}$,

 $i, j = \overline{1, N}, t > 0, \sigma$ – стандартное отклонение элементов ядра.

3.2 Логарифмические меры

Целесообразность использования логарифмических мер обсуждается в [14]. По аналогии с [15,16] прологарифмируем поэлементно меры, описанные в предыдущем подразделе. Полученные меры обозначим

 $\mathbf{K}^{\mathrm{LogComm}}, \ \mathbf{K}^{\mathrm{LogWalk}}, \ \mathbf{K}^{\mathrm{LogFor}}, \ \mathbf{K}^{\mathrm{LogHeat}}, \ \mathbf{K}^{\mathrm{LogSCT}}$

4 Алгоритмы кластеризации

4.1 Ядерный алгоритм k-средних

Стандартный алгоритм *k*-средних может быть модифицирован использованием ядерной функции. Используя подходящее нелинейное отображение из оригинального входного пространства в некоторое другое пространство, можно обнаружить кластеры, линейно неразличимые в исходном пространстве.

Обозначим кластеры как π_j и множество кластеров как $\{\pi_j\}_{j=1}^k$. При использовании нелинейной функции ϕ целевая функция ядерного алгоритма *k*-средних будет выглядеть следующим образом [17]:

$$D(\{\pi_j\}_{j=1}^k) = \sum_{j=1}^k \sum_{\mathbf{v} \in \pi_j} |\phi(\mathbf{v}) - \mathbf{m}_j|^2, \text{rge } \mathbf{m}_j = \frac{\sum_{\mathbf{v} \in \pi_j} \phi(\mathbf{v})}{|\pi_j|}.$$

Отметим, что \mathbf{m}_j является "лучшим" представителем кластера π_j : $\mathbf{m}_j = \operatorname{argmin}_{\mathbf{z}} \sum_{\mathbf{v} \in \pi_j} |\phi(\mathbf{v}) - \mathbf{z}|^2$. Евклидово растояние от $\phi(\mathbf{v})$ до центра \mathbf{m}_j задается

выражением:
$$\left(\phi(\mathbf{v}) - \frac{\sum\limits_{\tilde{\mathbf{v}} \in \pi_j} \phi(\tilde{\mathbf{v}})}{|\pi_j|}\right)^2 = \phi(\mathbf{v})\phi(\mathbf{v}) - \frac{2\sum\limits_{\tilde{\mathbf{v}} \in \pi_j} \phi(\tilde{\mathbf{v}})\phi(\mathbf{v})}{|\pi_j|} + \frac{\sum\limits_{\tilde{\mathbf{v}}, \mathbf{v} \in \pi_j} \phi(\tilde{\mathbf{v}})\phi(\mathbf{v})}{|\pi_j|^2}$$

Скалярные произведения $\phi(\tilde{\mathbf{v}})\phi(\mathbf{v})$ вычисляются при помощи ядерной функции $k(\tilde{\mathbf{v}}, \mathbf{v})$, и их значения записываются в матрицу K.

4.2 Метод Уорда

Опишем схему работы иерархических агломеративных методов. Сначала каждый объект считается отдельным кластером. Для кластеров с одним элементом функция расстояния определяется как: R(x, y) = d(x, y). Затем начинается процесс слияня кластеров. На каждой итерации вместо пары самых близких кластеров U и V образуется новый кластер $W = U \cup V$. Расстояние от нового кластера W до любого другого кластера S вычисляется через известные на тот момент расстояния R(U, V), R(U, S) и R(V, S): $R(U \cup V, S) = \alpha_U R(U, S) + \alpha_V R(V, S) + \beta R(U, V) + \gamma |R(U, S) - R(V, S)$, где $\alpha_U, \alpha_V, \beta, \gamma$ – числовые параметры (формула Ланса-Уильямса).

Методом Уорда называется алгломеративная кластеризация, использующая расстояние Уорда: $R^{\text{Ward}}(W,S) = \frac{|S||W|}{|S|+|W|} d^2 \left(\sum_{\omega \in W} \frac{\omega}{|W|}, \sum_{s \in S} \frac{s}{|S|} \right); \alpha_U = \frac{|S|+|U|}{|S|+|W|}, \quad \alpha_V = \frac{|S|+|V|}{|S|+|W|}, \quad \beta = \frac{-|S|}{|S|+|W|}, \quad \gamma = 0.$

Методу Уорда можно передать параметр *k* – количество ожидаемых кластеров. Он является критерием остановки построения дендрограммы.

5 Результаты

Для оценки качества кластеризации в работе использовался Adjusted Rand Index (ARI). Значения ARI принадлежат отрезку [-1;1], где 0 соответствует случайному гаданию, 1 – полному совпадению, -1 – результату, максимально удаленному от данных.

5.1 Сгенерированные графы

В данном подразделе будут представлены результаты исследований, которые проводились на сгенерированных семействах случайных графов. Рассматривались модели с двумя классами: G(100, 0.25, 0.05), G(200, 0.2, 0.05) и с тремя классами: G(100, 0.3, 0.1). Для каждой модели генерировалось 100 графов. На каждом графе отрабатывали оба алгоритма со всеми ядрами и большим набором гиперпараметров. Затем проводилось усреднение по сгенерированным графам для каждого значения гиперпараметра.

Для получения общих выводов о работе двух методов проведем соревнование метрик методом Коупленда [19], как это сделано в [20]. Будем использовать следующую схему.

- Генерируем граф. На нем проводим кластеризацию двумя методами с каждым из ядер. Для каждого ядра выбираем значение параметра, соответствующее самому лучшему качеству.
- В каждом туре голосования попарно сравнивается качество методов, определяемых их ядрами. В каждой паре победитель получает +1, проигравший –1. Если два метода дают качество 1.0, то оба получают +1.
- В каждом туре метод характеризуется суммой полученных оценок.
- Проводится несколько туров, результаты их складываются.

Графы генерируем с помощью модели $G(N, p_{in}, p_{out}, m)$, где m – число классов.

	k-means	Ward
$\sum_{\text{PlainWalk}}$	-121	143
\sum_{Walk}	160	-90
\sum_{Comm}	128	-64
\sum_{LogComm}	137	-75
\sum_{Forest}	-46	80
$\sum_{\text{LogForest}}$	141	45
\sum_{Heat}	121	83
\sum_{LogHeat}	137	71
\sum_{SCT}	-48	80
\sum_{LogSCT}	-63	99
Σ	566	372

Таблица 1: Результаты голосования методом Коупленда

Данные таблицы 1 указывают на то, что метод Уорда выигрывает реже, чем метод k-средних. В то же время для k-средних можно выделить явного фаворита – Walk: эта мера получила наибольшую сумму очков. Большинство логарифмических мер показало лучшее качество, чем соответствующие "простые" меры. Это прогнозируемый результат, поскольку для графов, сгенерированных выбранной моделью, естественно свойство разрезной аддитивности, которым обладают логарифмические меры. Суммарно по всем метрикам с отрывом выигрывает k-means.

Приведем оптимальные параметры семейств метрик и соответствующие им ARI для каждого метода. Отметим красным цветом наименьшие значения ARI по столбцам, зеленым – наивысшие. Если в столбце присутствует более одного значения ARI, равного 1.0, то лидеры не выделяются.

	G(100	0, 0.25, 0.05, 2)	G(100	0, 0.3, 0.1, 3)	G(200	(0.3, 0.05, 2)
	t^{opt}	ARI	t^{opt}	ARI	t^{opt}	ARI
k-means PlainWalk	0.675	1.0	0.8	0.5123	0.525	1.0
Ward PlainWalk	0.725	0.9856	0.675	0.8049	-48	48
k-means Walk	0.7	0.9967	0.275	0.9383	0.025	1.0
Ward Walk	0.025	0.9912	0.225	0.8260	44	-40
k-means Comm	0.1	1.0	0.175	0.9553	0.025	1.0
Ward Comm	0.175	0.9904	0.175	0.8400	37	-31
k-means LogComm	0.025	0.9952	0.025	0.9321	0.025	1.0
Ward LogComm	0.15	0.9920	0.05	0.8103	42	-40
k-means Forest	0.65	1.0	0.875	0.3219	0.325	0.98
Ward Forest	0.775	0.9975	0.95	0.6601	-28	28
k-means LogForest	0.025	1.0	0.575	0.9903	0.025	1.0
Ward LogForest	0.05	1.0	0.225	0.9436	42	2
k-means Heat	0.2	1.0	0.375	0.8863	0.075	1.0
Ward Heat	0.3	1.0	0.475	0.8643	31	13
k-means LogHeat	0.05	1.0	0.25	0.9874	0.025	1.0
Ward LogHeat	0.05	1.0	0.275	0.9675	32	32
k-means SCT	0.025	1.0	0.1	0.4007	0.025	1.0
Ward SCT	0.925	0.9984	0.05	0.7310	-31	31
k-means LogSCT	0.025	1.0	0.025	0.3712	0.025	1.0
Ward LogSCT	0.15	0.9991	0.35	0.7217	-32	32

Таблица 2: Лучшие метрики и оптимальные параметры

Генерировались также графы с переменным числом узлов в классах. В каждом графе N = 100 вершин распределялись по двум неравномощным классам. Наилучшие результаты представлены на рис. 1, где по оси абсцисс отложен размер первого класса N_1 , по оси ординат – ARI.



Рис. 1. Модель графа с неравномощными кластерами

Здесь наилучшие результаты показала мера Communicability и ее логарифмическая версия: для обоих алгоритмов качество очень высокое. При вариации от 1 до 9 числа вершин в классе 1 качество быстро растет от нулевого до высокого. Таким образом, данная мера эффективна в случае гетерогенных сетей, где размеры классов могут сильно отличаться, не становясь исчезающе малыми.

Остальные "простые" меры дают хорошее качество только при низком числе вершин в одном из классов либо, напротив, при разделении вершин на два класса практически поровну. При этом логарифмические версии тех же мер хорошо справляются с кластеризацией на случайных графах при значительных вариациях числа вершин в классах.

5.2 Реальные графы

В этом разделе сравним качество алгоритмов кластеризации на реальных графах. Используются известные датасеты: Zachary's Karate club, Newsgroup graph (9 датасетов), а также Football и Political books. Самый маленький среди них – Zachary's Karate club, в котором 34 вершины размечены на 2 класса, а самый большой – Newsgroup graph – содержит 997 вершин и 5 классов.

Исследуем метрики и алгоритмы на устойчивость. Проведем следующий эксперимент: будем добавлять новые связи между вершинами и удалять имеющиеся случайным образом, но так, чтобы число измененных ребер не превышало 10% от их начального количества. Сгенерируем 50 графов на основе существующих датасетов по указанной схеме и усредним по ним значения при оптимальных параметрах.

Существенно лучшее качество показывает k-means. В чистых датасетах он уступил методу Уорда только на датасете Zachary. LogComm в k-means показывает очень высокие результаты на чистых датасетах и при этом выдает один из лучших результатов на зашумленных датасетах. Мера LogHeat
	foo	tball	polt	oooks	Zachary		news	2cl 1
	$t^{\rm opt}$	ARI	t^{opt}	ARI	t^{opt}	ARI	t^{opt}	ARI
k-means PlainWalk	0.95	0.9061	0.075	0.2760	0.375	0.9921	0.025	0.1348
Ward PlainWalk	0.575	0.9063	0.025	0.6898	0.475	1.0	0.925	0.0706
k-means Walk	0.95	0.9034	0.275	0.6114	0.025	0.6686	0.95	0.0231
Ward Walk	0.925	0.9063	0.5	0.6609	0.525	1.0	0.2	0.5996
k-means Comm	0.425	0.9113	0.825	0.3156	0.25	1.0	0.025	0.0945
Ward Comm	0.575	0.9093	0.025	0.6795	0.2	1.0	0.5	0.0482
k-means LogComm	0.575	0.9015	0.275	0.6141	0.025	0.6687	0.75	0.0356
Ward LogComm	0.55	0.9034	0.225	0.6342	0.225	0.8823	0.1	0.4543
k-means Forest	0.95	0.8926	0.95	0.7040	0.825	0.8823	0.025	0.0001
Ward Forest	0.075	0.8966	0.875	0.6745	0.8	1.0	0.025	0.010
k-means LogForest	0.85	0.8966	0.9	0.6818	0.15	0.8823	0.775	0.9214
Ward LogForest	0.025	0.8893	0.95	0.6893	0.2	1.0	0.525	0.8095
k-means Heat	0.3	0.8966	0.45	0.7267	0.625	0.8823	0.025	0.0
Ward Heat	0.075	0.8966	0.875	0.7084	0.625	1.0	0.025	0.0
k-means LogHeat	0.15	0.8869	0.8	0.6689	0.35	0.8823	0.475	0.9214
Ward LogHeat	0.025	0.8893	0.875	0.7084	0.3	1.0	0.475	0.8460
k-means SCT	0.675	0.8935	0.95	0.7267	0.025	0.8823	0.725	0.001
Ward SCT	0.025	0.8966	0.925	0.6955	0.025	1.0	0.025	0.0
k-means LogSCT	0.65	0.8947	0.95	0.7038	0.025	0.8823	0.875	0.002
Ward LogSCT	0.025	0.8966	0.025	0.6934	0.025	1.0	0.025	0.0

Таблица 3: Лучшие метрики при оптимальных параметрах на реальных датасетах

оказалась наилучшей для кластеризации датасета news_2cl_1 – в условиях шума она дала наилучшее качество в сравнении с остальными ядрами в обоих алгоритмах.

6 Заключение

По результатам проведенных в работе экспериментов можно сделать вывод, что для рассмотренных классов графов при использовании ядер с оптимальными параметрами метод k-средних работает в среднем лучше, чем метод Уорда. Особенно ярко это проявляется в экспериментах на реальных датасетах и в результатах голосования по Коупленду.

Отметим следующий результат для ядра Forest. При использовании его в обоих алгоритмах качество кластеризации обычно оказывалось одним из худших. В то же время логарифмированный Forest часто приносил один из лучших результатов для обоих алгоритмов.

В случае случайных графов наиболее устойчивой к изменению числа вершин в классах для обоих алгоритмов оказалась, среди нелогарифмированных мер, Communicability, причем при ее использовании качество двух методов отличается непринципиально.

В присутствии шума метод *k*-средних с ядром LogHeat показал достаточную устойчивость на всех датасетах и наилучший результат на двух семействах зашумленных датасетов.

Благодарности

Исследование выполнено при финансовой поддержке РФФИ в рамках научных проектов No 18-07-01409 A, 18-37-00322.

Список литературы

- J.M. Kleinberg, E. Tardos.: Approximation algorithms for classification problems with pairwise relationships: Metric labeling and Markov random fields. Journal of the ACM. no. 49, pp. 14–23, 2002.
- S. Fortunato.: Community detection in graphs. Physics Reports. no. 486(3), pp. 75–174, 2010.
- U. Luxburg.: A tutorial on spectral clustering. Statistics and Computing. no. 17(4), pp. 395–416, 2007.
- P. W. Holland, K. B. Laskey, S. Leinhardt.: Stochastic blockmodels: First steps. Social Networks. no. 5(2), pp. 109–137, 1983.
- F. Fouss, L. Yen, A. Pirotte, M. Saerens: An experimental investigation of graph kernels on a collaborative recommendation task. Sixth International Conference on Data Mining (ICDM'06). pp. 863–868, 2006, IEEE.
- И. Силин, М. Панов.: Алгоритм адаптивных весов для оценки параметров и выделения сообществ в случайных графа. ITAS. pp. 199–213, 2016.
- F. Sommer, F. Fouss, M. Saerens.: Comparison of Graph Node Distances on Clustering Tasks. ICANN. pp. 192–201, 2016.
- K. Avrachenkov, P. Chebotarev, D. Rubanov.: Kernels on Graphs as Proximity Measures. Algorithms and Models for the Web Graph. pp. 27–41, 2017.
- P. Chebotarev, E. Shamis.: The Forest Metrics for Graph Vertices. Electronic Notes in Discrete Mathematics. pp. 98–107, 2002.
- C. Augustin.: Comparison of some community detection methods for social network analysis. Master's thesis, Louvain School of Management, Universite catholique de Louvain, Louvain, Belgium. pp. 88, 2015.
- 11. E. Estrada.: The communicability distance in graphs. Linear Algebra and its Applications. pp. 4317–4328, 2012.
- P. Chebotarev.: The Walk Distances in Graphs. Discrete Applied Mathematics. pp. 1484–1500, 2012.
- P. Chebotarev, E. Shamis.: On a Duality between Metrics and -Proximities. Automation and Remote Control. pp. 608–612, 1998.
- P. Chebotarev.: The graph bottleneck identity. Advances in Applied Mathematics. v. 47, num. 3, pp. 403–413, 2011.
- P. Chebotarev.: The Walk Distances in Graphs. Discrete Applied Mathematics. pp. 1484–1500, 2012.
- P. Chebotarev.: A Class of graph-geodetic distances generalizing the shortest-path and the resistance distances. Discrete Applied Mathematics. v. 159, pp. 295–302, 2011.
- F. Fouss, M. Saerens, M. Shimbo.: Algorithms and Models for Network Data and Link Analysis. Cambridge University Press. Cambridge, United Kingdom; New York. 2016.
- I. Kivimaki, M. Shimbo, M. Saerens: Developments in the theory of randomized shortest paths with a comparison of graph node distances. Physica A: Statistical Mechanics and its Applications. pp. 600–616, 2014.
- A. Copeland: A Reasonable Social Welfare Function. Department of Mathematics, University of Michigan. 1951
- I. Kivimaki, M. Shimbo, M. Saerens: Developments in the theory of randomized shortest paths with a comparison of graph node distances. Physica A: Statistical Mechanics and its Applications. v. 393, pp. 600–616, 2014.

Гетероскедастичные гауссовские процессы и их применение для байесовской оптимизации

М. Колодяжный¹

А. Зайцев²

 1 МФТИ 2 Сколтех, ИППИ (РАН)

Аннотация Байсовская оптимизация — подход для оптимизации "тяжелых"функций. Идея состоит в замене исходной целевой функции вычисленной регрессионной моделью, построенной по выборке. Для построения модели обычно используют регрессию на основе гауссовских процессов. Она работает плохо, если данные гетероскедастичны: шум зависит от координаты и это сказывается на точности оптимизации. В данной статье предлагается использовать новые алгоритмы регрессии на основе гетероскедастичных гауссовких процессов, для того чтобы улучшить качество оптимизации.

Keywords: Байесовская оптимизация · Гетероскедастичный шум · Вариационный вывод.

1 Вступление

Байсовская оптимизация является очень эффективным подходом для глобальной оптимизации так называемых black-box функций, градиенты которых не известны, а вычисление значений самих функций является очень дорогостоящей операцией. Она широко применяется в разных сферах: от автоматического подбора гиперпараметров для моделей машинного обучения [7] до балансирования роботов [4]. Идея состоит в замене исходной целевой функции вычисленной регрессионной моделью, построенной по выборке. Для построения модели обычно используют регрессию на основе гаусовских процессов. Обычно значения целевой функции зашумлены, что затрудняет процесс оптимизации и влечет за собой ошибки и неточности при поиске глобального минимума. В самом общем случае цель глобальной оптимизации может быть поставлена следующим образом: найти такой **х**^{*}, чтобы

$$\mathbf{x}^* = \max_{\mathbf{x} \subseteq \mathcal{X}} f(\mathbf{x}), \ \mathcal{X} \subseteq R^d$$

При этом мы наблюдаем:

$$y(\mathbf{x}) = f(\mathbf{x}) + \varepsilon(\mathbf{x}).$$

Обычно считается, что шум имеет вид:

$$\varepsilon \sim \mathcal{N}(0, \sigma^2),$$

где дисперсия не зависит от \mathbf{x} . Это не всегда верно. Довольно часто дисперсия имеет не постоянное значение, а зависит от \mathbf{x} :

$$\sigma^2 = r(\mathbf{x}).$$

Шум с распределением: $\varepsilon \sim \mathcal{N}(0, r(\mathbf{x}))$, называется гетероскедастичным. Наличие гетероскедастичности в данных влечет за собой потерю в точности при нахождении глобального минимума.

В данной статье предложен метод, позволяющий бороться с проявлением гетероскедастичного шума при байесовской оптимизации. Вводится модель шума, аналогичная [5]. Качество работы предложенного алгоритма рассматривается на ряде реальных и модельных задач.

2 Теоретическая часть

2.1 Регрессия на основе гауссовских процессов

Регрессия на основе гауссовских процессов позволяет строить нелинейные аппроксимационные модели и обеспечивает аналитическое выражение для прогноза и оценки его неопределенности аппроксимации в заданной точке **x**.

Гауссовский процесс $f(\cdot)$ — случайный процесс, такой, что любой конечный набор значений гауссовского процесса в различных точках пространства имеет многомерное нормальное распределение.

Таким образом, случайный процесс $f(\cdot)$ определяется средним $m(\mathbf{x}) = \mathbb{E}[f(\mathbf{x})]$ и ковариационной функцией $k(\mathbf{x}, \mathbf{x}') = \mathbb{E}[(f(\mathbf{x}) - m(\mathbf{x}))(f(\mathbf{x}') - m(\mathbf{x}'))].$

Обычно при моделировании мы имеем доступ только к зашумленным значениям функции:

$$y(\mathbf{x}) = f(\mathbf{x}) + \varepsilon(\mathbf{x}),$$

где шум $\varepsilon(\mathbf{x})$ моделируется гауссовским белым шумом с дисперсией $\tilde{\sigma}^2$. В таком случае наблюдения $y(\mathbf{x})$ будут гауссовским процессом с нулевым средним и ковариационной функцией $cov(y(\mathbf{x}), y(\mathbf{x}')) = k(\mathbf{x}, \mathbf{x}') + \tilde{\sigma}^2$.

2.2 Регрессия на основе гетероскедастичных гауссовских процессов

Обозначим символом $\mathcal{GP}(\mu_0, k_f(\mathbf{x}, \mathbf{x}'))$ гауссовский процесс со средним μ_0 и ковариационной функцией $k_f(\mathbf{x}, \mathbf{x}')$, где θ_f - гиперпараметры ковариационной функции.

Тогда определим нашу гетероскедастичную модель следующим образом: $f(x) \sim \mathcal{GP}(0, k_f(\mathbf{x}, \mathbf{x}'))$

 $\varepsilon \sim \mathcal{N}(0, r(\mathbf{x})),$ где $r(\mathbf{x}) = e^{g(\mathbf{x})}, \qquad g(x) \sim \mathcal{GP}(\mu_0, k_g(\mathbf{x}, \mathbf{x}'))$ $\theta = \{\theta_f, \theta_q, \mu_0\}$ - гиперпараметры.

То есть будем считать, что не только целевая функция представляет из себя случайный гауссовский процесс, но и логарифм дисперсии шума тоже. Использование экспоненты нужно для того, чтобы избежать отрицательных значений для дисперсии.



Рис. 1: Слева - построенная регрессионная модель, справа - распределения в точке x = -14(точки, с самой большой дисперсией): зеленым цветом показано апостериорное распределение в этой точке, голубым - нормальное распределение с параметрами μ_* и σ_* - значениями, предсказанными для данной точки в результате работы алгоритма.

Так как для такой модели, мы не можем получить правдоподобие в явном виде, то для оценки параметров регрессии будем максимизировать его вариационную нижнюю оценку:

 $F(q(\boldsymbol{f}), q(\boldsymbol{g})) = \log p(\boldsymbol{y}) - KL(q(\boldsymbol{f})q(\boldsymbol{g}) \parallel p(\boldsymbol{f}, \boldsymbol{g} \mid \boldsymbol{y})).$

Так мы используем дополнительное предположение о распределении дисперсии, то это позволяет более точно строить аппроксимационную модель. [5]

После оценки параметров можно получать предсказания для значения среднего и дисперсии функций в новой точке.

2.3 Байесовская оптимизация

Типичный алгоритм байесовской оптимизации представлен ниже.

\mathbf{A}	lgorithm	1	Байесовская	оптимизация
--------------	----------	---	-------------	-------------

 for n = 1, 2, ... do
Фиксируем модель данных, используя регрессию на основе гауссовких процессов
Выбираем новую точку x_{n+1} с максимумом acquisition function α: x_{n+1} = max α(x)
Вычисляем значение целевой функции в этой точке y_{n+1}
Добавляем точку в нашу подвыборку D_{n+1} = {D_n, (x_{n+1}, y_{n+1})}
Делаем апдейт модели
еnd for
Итерируем до тех пор пока не будет исчерпан лимит на количество оценок

Функция выгоды (Acquisition Function) определяет в какой следующей точке следует сделать вычисление тяжелой функции. Обычно используют Expected Improvement(EI), у которой есть явная формула:

если обозначить $\Delta(\mathbf{x}) = y_{best} - \mu_*(\mathbf{x})$, то

 $\alpha_{EI} = \int_{y} max(0, y_{best} - y_{*})p(y_{*}|\mathbf{x})dy_{*} = max(0, \Delta(\mathbf{x})) - \sigma_{*}(\mathbf{x})\varphi(\frac{\Delta(\mathbf{x})}{\sigma_{*}(\mathbf{x})}) + |\Delta(\mathbf{x})| \Phi(-\frac{|\Delta(\mathbf{x})|}{\sigma_{*}(\mathbf{x})})$ (1)

Обычно в байесовской оптимизации для предсказания среднего μ_* и дисперсии σ_* используется регрессия на гомоскедастичных гауссовких процессах, в этой же работе для предсказания среднего и дисперсии предлагается использовать регрессию на основе гетероскедастичных гауссовских процессов. Для гетероскедастичной модели формула (1) для Expected Improvement тоже будет достаточно хорошо аппроксимировать истинный EI, так как апостериорное распределение в новой точке близко к нормальному. Типичный пример двух таких распределений приведен на Рис. 1, распределения близки, потому что апостериорное распределение в точке - бесконечная смесь нормальных распределений с одинаковым средним.

3 Экспериментальные результаты

Для анализа эффективности работы метода и сравнения его с другими использовались фреймворки GPy и GPyOpt. В качестве Acquisition Function использовалась Expected Improvement. Мы сравнивали алгоритм байесовской оптимизации на искусственных и реальных данных разной размерности. Использовались подходы на основе гетероскедастичных (далее в таблицах "гт") и гомоскедастичных (далее в таблицах "гм") гауссовских процессов.

3.1 Описание данных

Искусственная одномерная функция (первый сценарий шума).

Первая функция, на которой оценивалось качество работы - это одномерная функция $y = \sin(x)(1+0.04|x|)$. На участке (-22, -10) данная функция имеет два минимума в точке $T_1 = (-20.4531, -1.8171)$ и точке $T_2 = (-14.1802, -1.5657)$. Причем, как видно, в точке T_2 значение функции меньше, чем в T_1 . Искусственным образом в окрестность точки T_2 был добавлен сильный шум, параметр scale (масштаб) которого можно было регулировать. Таким образом функция стала гетероскедастично-зашумленной.

На Рис.2 показано, как аппроксимируют данную модель обычный и гетероскедастичный гауссовские процессы. Видно, что гетероскедастичный работает лучше, что позволяет более правильно выбрать следующую точку для вычисления целевой функции в байесовской оптимизации.

Искусственная одномерная функция (второй сценарий шума).

Далее рассматривалась та же функция, но с другим сценарием гетероскедастичного шума: зашумлялось все кроме окрестностей двух минимумов. Аппроксимация показана на Рис.3.



Рис. 2: Аппроксимация двумя алгоритмами регрессии одномерной функции с первым сценарием шума.

Искусственная двумерная функция.

Далее рассматривалась двумерная функция - $y = \sin(x)\cos(y)(1+0.04|y|)(1+0.04|x|)$. Она на прямоугольнике {-22< x_1 <-12.5, -14 < x_2 < -11} также имеет два минимума, один из которых локальный, другой глобальный. Локальный минимум был зашумлен (Рис.4), что сильно сказалось на сходимости оптимизации.

Явные формулы для оптимизируемых функций представлены в приложении А.

Реальные данные.

Agromet (Puc.5) был получен Изобель Кларк и описывает концентрацию золота в Натальском нагорье, Южная Африка, на глубине 400 метров, данные являются двумерными и состоят из 18 189 измерений.



(а) Гомоскедастичный алгоритм.



(b) Гетероскедастичный алгоритм.

Рис. 3: Аппроксимация двумя алгоритмами регрессии одномерной функции со вторым сценарием шума.





Рис. 4: Двумерная функция, используемая в экспериментах.

Рис. 5: Данные **Agromet**, используемые в экспериментах.

Brenda - это данные, полученные из медных рудников в Канаде. Каждое измерение показывает концентрацию меди, молибдена, серебра и золота в зависимости от глубины, широты и долготы. Состоят из 1856 наблюдений. В данной статье использовались данные только о концентрации меди.

3.2 Результаты

Для искусственных функций качество оценивалось с помощью двух метрик:

- количество точек, полученных в результате оптимизации, попавших в 5% окрестность точки минимума отнесенных к общему количеству запусков оптимизации. То есть, если х точка, которую выдал алгоритм после завершения оптимизации, и х ⊆ (-19.4; 21.4), то мы считаем количество таких точек и делим на общее количество запусков оптимизаций.
- 2. среднее значение целевой функции на полученных в результате оптимизации точках. То есть, если \mathbf{x}_i полученный оптимум, а n количество запусков оптимизаций, то мы считаем $\frac{\Sigma f(\mathbf{x}_i)}{n}$

Искусственная одномерная функция с первым сценарием шума.

Оптимизация запускалась 50 раз при масштабе шума равным 1.0, и по 30 раз при масштабе 0.8, 0.6, 0.4. Результаты представлены в Таблицах 1 и 2. Гетероскедастичная модель показала себя лучше в экспериментах. При увеличении масштаба разница в качестве работы увеличивалась. Чем больше был шум, тем лучше она работала по сравнению с гомоскедастичной, что связанно с тем, что гомоскедастичная модель начинала вырождаться.

Искусственная одномерная функция со вторым сценарием шума.

Оптимизация запускалась 30 раз при масштабе = 1.0. Результаты представлены в Таблице 3. Видно, что хоть по количеству попаданий в точку мини-

масштаб	Гм	Гт		масштаб	5	Гм		Гт
1.0	10%	46%		1.0	-1.4	4602 ± 0.1	801 -1.6242	$\pm 0.$
0.8	30%	57%		0.8	-1.5	5626 ± 0.1	958 -1.6517	$\pm 0.$
0.6	27%	44%		0.6	-1.5	5323 ± 0.1	733 -1.5927	$\pm 0.$
0.4	44%	50%		0.4	-1.6	6159 ± 0.1	697 - 1.6508	$\pm 0.$
Таблица	1: ŀ	Количест	BO	Таблица	2:	Среднее	значение	цел
попадани	йво	крестнос	сть	функции	на	точках п	олученных	к прі

попаданий в окрестность точки T₁ для двух алгоритмов оптимизации Таблица 2: Среднее значение целевой функции на точках полученных при оптимизации для двух алгоритмов регрессии

мума значения не сильно отличаются, но среднему значению на гетероскедастичной модели значение оказалось меньше. Более того дисперсия значений для гетероскедастичной модели оказалась меньше, что свидетельствует о меньшем разбросе значений.

Искусственная двумерная функция.

Результаты представлены в Таблице 4. Метрики качества использовались те же , масштаб шума был равен 1.0. Гетероскедастичная модель сработала лучше как по первой метрике, так и по второй.

Метрика	г Гм	Гт	Метрика	Гм	$\Gamma_{\mathbf{T}}$
1	77%	80%	1	22%	44%
2	-1.4318 ± 0.7341	-1.5405 ± 0.5914	2	-2.1318 ± 0.4180	-2.2328 ± 0.4188
T (0 D		T (4 5	

Таблица 3: Результаты экспериментов для второго сценария шума Таблица 4: Результаты экспериментов для двумерной функции



Рис. 6: **Agromet**.Регрессия на гетероскедастичном алгоритме регрессии по 40 точкам обучающей выборки.



Рис. 7: **Agromet**.Регрессия на гомоскедастичном алгоритме регрессии по 40 точкам обучающей выборки.



Рис. 8: Agromet. Сходимость байесовской оптимизации с использованием гетероскедастичной аппроксимации.



Рис. 9: Brenda. Сходимость байесовской оптимизации с использованием гетероскедастичной аппроксимации.

Реальные данные.

Данные гетероскедастичны и сильны зашумлены, и проводить байесовскую оптимизацию с использованием гомоскедастичных гауссовских процессов оказалось не возможным, так как регрессия вырождалась, в то время как регрессия на гетероскедастичных гауссовских процессах сработала правильно (Рис.6 и Рис.7). Сходимость оптимизации на гетероскедастичном алгоритме показана на Рис.8 и Рис.9.

4 Выводы

Мы рассмотрели новый подход к байесовкой оптимизации: для аппроксимации функции вместо регрессии на основе гомоскедастичных гауссовких процессов использовалась регрессия на основе гетероскедастичных процессов. Как показали эксперименты данный подход работает лучше, если оптимизируемая целевая функция имеет сильно выраженный гетероскедастичный шум. Это связанно с двумя причинами:

- Гетероскедастичные гауссовские процессы правильнее аппроксимируют данные, что позволяет точнее делать предположение о вычислении значения целевой функции для следующей точки во время байесовской оптимизации.
- На реальных данных обычные гауссовкие процессы часто вырождаются, так как не предназначены для построения регрессии на гетероскедастичных данных [10]. Поэтому для байесовской оптимизации они не годятся. В то же время как гетероскедастичные гауссовские процессы отлично справляются с этой задачей.

5 Благодарности

Работа Алексея Зайцева была частично выполнена за счет грантов Р
ФФИ 16-01-00576 А и 16-29-09649 оfi $_$ m.

Список литературы

- 1. John-Alexander M Assael, Ziyu Wang, Bobak Shahriari, and Nando de Freitas. Heteroscedastic treed bayesian optimisation. arXiv preprint arXiv:1410.7172, 2014.
- Evgeny Burnaev and Maxim Panov. Adaptive design of experiments based on gaussian processes. In *International Symposium on Statistical Learning and Data Sciences*, pages 116–125. Springer, 2015.
- GPy. GPy: A Gaussian process framework in python. http://github.com/ SheffieldML/GPy, since 2012.
- 4. Scott Kuindersma, Roderic Grupen, and Andrew Barto. Variational bayesian optimization for runtime risk-sensitive control. *Robotics: Science and Systems VIII*, 2012.
- 5. Miguel Lázaro-Gredilla and Michalis Titsias. Variational heteroscedastic gaussian process regression. 2011.
- Bobak Shahriari, Kevin Swersky, Ziyu Wang, Ryan P Adams, and Nando De Freitas. Taking the human out of the loop: A review of bayesian optimization. *Proceedings of the IEEE*, 104(1):148–175, 2016.
- Jasper Snoek, Hugo Larochelle, and Ryan P Adams. Practical bayesian optimization of machine learning algorithms. In Advances in neural information processing systems, pages 2951–2959, 2012.
- Jasper Snoek, Kevin Swersky, Rich Zemel, and Ryan Adams. Input warping for bayesian optimization of non-stationary functions. In *International Conference on Machine Learning*, pages 1674–1682, 2014.
- Yanan Sui, Vincent Zhuang, Joel W Burdick, and Yisong Yue. Stagewise safe bayesian optimization with gaussian processes. arXiv preprint arXiv:1806.07555, 2018.
- Е.В. Бурнаев, М.Е. Панов, and А.А. Зайцев. Регрессия на основе нестационарных гауссовских процессов с байесовской регуляризацией. Информационные процессы, 15(3):298–313, 2015.

Приложение

А Явные формулы для оптимизируемых функций

Первая функция

$$y = \begin{cases} \sin(x)(1+0.04|x|) + random.normal(loc = 0.0, scale = 1) \\ x \subseteq (-15, -13.3) \\ \sin(x)(1+0.04|x|) \\ x \subseteq (-22, -10) \setminus (-15, -13.3) \end{cases}$$

Вторая функция

$$y = \begin{cases} \sin(x)(1+0.04|x|) + random.normal(loc = 0.0, scale = 1) \\ x \subseteq (-22, -10) \setminus ((-21.5, -19)\&(-16, -12)) \\ \sin(x)(1+0.04|x|) \\ x \subseteq (-21.5, -19)\&(-16, -12) \end{cases},$$

Третья функция

$$y = \begin{cases} \sin(x)\cos(y)(1+0.04|y|)(1+0.04|x|) + \\ random.normal(loc = 0.0, scale = 1) \\ x \subseteq (-15.1, -13.1), y \subseteq (-13.53, -11.53) \\ \sin(x)\cos(y)(1+0.04|y|)(1+0.04|x|) \\ x \subseteq (-22, -12.5) \setminus (-15.1, -13.1), y \subseteq (-14, -11) \setminus (-13.53, -11.53) \end{cases}$$

Реализация алгоритма Grassman&Stiefel Eigenmaps на языке Python

Афонин Андрей¹ Янович Юрий^{2,3}

 $^{1}\,$ Московский физико-технический институт, Москва

² Сколковский институт науки и технологий, Москва

³ Институт проблем передачи информации им. А. А. Харкевича РАН, Москва

Аннотация Методы снижения размерности позволяют заменить многомерные описания данных на их низкоразмерные аналоги почти без потери информации и нашли широкое применение в машинном обучении. Алгоритм Grassman&Stiefel Eigenmaps (GSE) позволяет в некоторых предположениях построить оптимальную модель снижения размерности. В данной работе реализован алгоритм GSE на языке программирования Python. Для демонстрации качества работы кода проведен иллюстративный вычислительный эксперимент.

Ключевые слова: снижение размерности, оценивание многообразий

1 Введение

Данные, полученные из реальных источников, в силу наличия различных зависимостей между их компонентами и ограничений на их возможные значения, занимают, как правило, малую часть высокоразмерного пространства наблюдений, имеющую невысокую внутреннюю размерность. Следствием невысокой внутренней размерности является возможность построения низкоразмерной параметризации таких данных с минимальной потерей содержащейся в них информации. Поэтому многие алгоритмы для работы с высокоразмерными данными начинаются с решения задачи снижения размерности, результатом которого являются низкоразмерные описания таких данных.

На данный момент разработано множество алгоритмов снижения размерности: IsoMap [1], Locally-linear Embedding [2], Local Tangent Space Alignment [3], Laplacian Eigenmaps [4], Hessian Eigenmaps [5], Grassmann & Stiefel Eigenmaps (GSE) [6], и др. Однако, их теоретические свойства мало изучены [7,8,10,11]. Лишь про алгоритм GSE известно, что он асимптотически оптимален при определенных предположениях [12].

В работе рассматривается алгоритм Grassmann & Stiefel Eigenmaps. Ранее эксперименты для него производились на языке программирования Matlab и код не был открыт [13,14]. В данной статье алгоритм реализован согласно оригинальной статье на языке Python и проведены вычислительные эксперименты, иллюстрирующие его работу.

2 Алгоритм Grassman&Stiefel Eigenmaps

Задачей снижения размерности является построение отображения, которое каждому набору точек из пространства высокой размерности ставит в соответствие набор точек в пространстве меньшей размерности. Искомое решение должно сохранять относительное расстояние между точками. Рассматривается отображение из исходного пространства размерности p в пространство размерности q, q: p > q, где q известно [15,16,17]. Алгоритм GSE учитывает сохранение расстояния не только между точками, но и между их касательными пространствами.

Обозначим обучающую выборку $X_n = \{x_1, x_2, \dots, x_n\} \subset \mathbb{R}^p$.

Опишем построение низкоразмерного описания точек обучающей выборки, а также их восстановленные прообразы:

$$-h(x_1),\ldots,h(x_n)\in\mathbb{R}^q.$$

$$-g(h(x_1)),\ldots,g(h(x_n))\in\mathbb{R}^p$$

и аналогичные описания для точек x и y вне обучающей выборки h(x) и g(y).

Algorithm 1. Grassman&Stiefel Eigenmaps [6]:

2.1 Шаг 1: вычисления функции близости

- 1. Близость между точками: Для всех $x \in X_n$ определяем $U_E(x) = x' \in X_n : ||x - x'|| < \epsilon_1$, где ϵ_1 параметр алгоритма. Вычисляем евклидово ядро: $K_E(x, x') = K_0(x, x') \times exp(-\epsilon_2 \times ||x - x'||)$, где ϵ_2 - параметр алгоритма, или
 - $K_{0}(x,x^{'})=1, \; ecnu\; x^{'}\in U_{E}(x) \; u \; K_{0}(x,x^{'})=0$ unave.
- 2. Близость между касательными пространствами: В каждой точке $x \in X_n$ при помощи метода главных компонент оцениваем касательное пространство $(L_{PCA}(x))$ натянутое на столбцы матрицы Q_{PCA} , где ее столбцы - состоят из первых q главных векторов. В достаточно малой окрестности точки можем считать $L_{PCA}(x) \approx L(x)$.

Обозначим $S(x, x^{'}) = Q_{PCA}(x)^T \times Q_{PCA}(x^{'}).$ Вычисляем ядро Бине-Коши: $K_G(x, x^{'}) = K_L(L_{PCA}(x), L_{PCA}(x^{'})) = Det^2(S(x, x^{'})).$

3. Вычисляем итоговое значение функции близости $K(x, x') = K_E(x, x') \times K_G(x, x')$, которое отражает близость как между точками, так и между их между касательными пространствами к носителю данных в них.

2.2 Шаг 2: Выравнивание базисов в касательных пространствах

- 1. Приближение касательного пространства: Ищется такая матричная функция H(x), что $L_H(x) \approx L(x)$ для $\forall x \in X_n$, где $L_H(x) = Span(H(x))$ и L(x) – касательное пространство в точке x. Чтобы этого достичь матрица H(x) представляется в виде $H(x) = Q_{PCA}(x) \cdot v(x)$, где $v(x) \in STEIF - NC(q, q)$. При таком условии будет выполняться $L_H(x) = L_{PCA}(x) \approx L(x)$.
- Строится набор матриц H_n = (H_i = Q_{PCA}(x_i) × v_i, x_i ∈ X_n), nocpedством минимизации квадратичной формы: δ_H(H_n) = ½ ∑_{i,j=1}ⁿ K(x_i, x_j) × ||H_i - H_j||², учитывая нормировочное условие 1/K ∑_{i,j=1}ⁿ K(x_i, x_j) × (H_i^T × H_i) = E_q, где K = ∑_{i,j=1}ⁿ K(x_i, x_j)
 Для произвольной точки x: H(x) находится посредством минимизации
- квадратичной формы $\delta_H(H(x)) = \frac{1}{2} \sum_{j=1}^n K(x, x_j) \times ||H(x) H_j||^2$. Peweнием такой задачи является матрица $H(x) = Q_{PCA}(x) \times Q_{PCA}(x)^T \times \frac{1}{K} \sum_{j=1}^n K(x, x_j) \times H_j$, где $K = \sum_{j=1}^n K(x, x_j)$

2.3 Шаг 3: вложение

- 1. Строятся отображение-вложение h и отображение-восстановление g, удовлетворяющие условиям g(h(x)) = x и $J_g(x) = H(x)$, $\forall x \in X_n$ Для близких точек x и $x': g(h(x)) - g(h(x')) = x - x' \approx H(x) \times (h(x) - h(x'))$
- 2. Строится набор векторов $h_n = h_1, ..., h_n$ посредством минимизации функционала $\delta_h(h_n) = \frac{1}{2} \sum_{i,j=1}^n K(x_i, x_j) \times |(x_i - x_j) - H(x) \times (h_i - h_j)|^2$, учитывая естественное нормировочное условие $h_1 + ... + h_n = 0$
- 3. Учитывая построенный набор векторов h_n , для произвольной точки xвектор h(x) находится посредством минимизации функционала $\delta_h(h_n) = \frac{1}{2} \sum_{j=1}^n K(x, x_j) \times |(x - x_j) - H(x) \times (h(x) - h_j)|^2$. Решением является вектор $h(x) = \frac{1}{K(x)} \sum_{j=1}^n K(x, x_j) \times h_j + v^{-1}(x) \times Q_{PCA}^T(x) \times (x - \frac{1}{K(x)} \sum_{j=1}^n K(x, x_j) \times x_j)$. Отсюда следует, что матри-

ца Якоби отображения
$$J_h(x) = v^{-1}(x) \times Q_{PCA}^T(x)$$



Рис. 1. Исходный набор точек: Swiss-roll и S-curve. Результаты отображения вложения



Рис. 2. Исходный набор точек: половина цилиндра и четверть сферы. Результаты отображения-вложения

2.4 Шаг 4: восстановление

1. Обозначим $Y_n = [y_1..y_n] = h(X_n) = [h(x_1)..h(x_n)]$

Для близких точек x и $x^{'} \in X_{n}$ выполняется: $x-x^{'}=H(x^{'})\times(y-y^{'})=>|x-x^{'}|=|v(x^{'})\times(y-y^{'})|$

2. Torda $U_{E}(y) = [y^{'} \in Y_{n} : (v(x^{'}) \times (y - y^{'}))^{T} \times v(x^{'}) \times (y - y^{'}) < \epsilon_{1}] =>$

 $\begin{array}{l} k_{E}(y,y^{'}) \,=\, k_{0}(y,y^{'}) \times exp(-\epsilon_{2} \times v(x^{'}) \times (y-y^{'}))^{T} \times v(x^{'}) \times (y-y^{'})), \\ \text{sde } k_{0}(y,y^{'}) = 1, \text{ ecnu } y^{'} \in U_{E}(y) \text{ u } k_{0}(y,y^{'}) = 0 \text{ - unave.} \end{array}$

3. Определим окрестность $u(y) = [h^{-1}(y') : y' \in U_E(y)]$. Применяя метод главных компонент определим матрицу $q_{PCA}(y)$, столбцы которой соответствуют первым q главным векторам. Применяя Аналогичные рассуждения применяя, что и в Шаге 1, определяем функцию k(y, y'). 4. Для произвольной точки $y \in \mathbb{R}^q$ определяем матрицу $G(y) = q_{PCA}(y) \times (x_1 + y_2) + (x_2 + y_3) + (x_3 + y_3) + (x_4 + y_4) + (x_4$

$$(q_{PCA}(y))^T \times \frac{1}{k(y)} \sum_{j=1}^n k(y, y_j) \times H(x_j),$$
 которая минимизирует функцио-
нал $\delta_G(G(y)) = \frac{1}{2} \sum_{j=1}^n k(y, y_j) \times ||G(y) - H(x_j)||^2,$ где $k(y) = \sum_{j=1}^n k(y, y_j).$

5. Учитывая приблизительное равенство для близких точек у и $y': g(y') - g(y) \approx G(y) \times (y'-y)$, строим вектор $g(y) \in \mathbb{R}^p$ для произвольной точки $y \in \mathbb{R}^q$, который минимизирует функционал

$$\delta_g(g(y)) = \frac{1}{2} \sum_{j=1}^n k(y, y_j) \times ||(g(y) - x_j) - G(y) \times (y - y_j)||^2.$$

6. Решением является вектор

$$g(y) = \frac{1}{k(y)} \sum_{j=1}^{n} k(y, y_j) \times x_j + G(y) \times (y - \frac{1}{k(y)} \sum_{j=1}^{n} k(y, y_j) \times y_j).$$

3 Вычислительный эксперимент

Для демонстрации работы алгоритма взяты выборки из N = 800 обучающих точек и выборки из $N^* = 200$ тестовых точек со множеств Scurve и Swiss-roll библиотеки scikit-learn, а также со сферы и цилиндра. Все эти множества являются двумерными(q = 2) поверхностями в пространстве \mathbb{R}^3 . В качестве функции ядра использовалось индикаторное ядро k ближайших соседей с k = 15. На рисунках 1-4 приведены результаты эксперимента для этих поверхностей. Шарами обозначены точки обучающей выборки, треугольниками – тестовой выборки. Как и предполагалось, отображением-вложением является подмножество двумерной плоскости. Отображение-реконструкция с высокой точностью "повторяет" поверхность.



Рис. 3. Результаты отображения-реконструкции. Исходные данные: Swiss-roll и S-curve

Чтобы дать численную характеристику "повторению" поверхности при композиции вложения-реконструкции эксперимент проведен для различного размера обучающей выборки N = [125, 250, 500, 900]. На рисунке 5 изображен двоичный логарифм относительного корня среднего квадратичного отклонения расстояния между точкой и композицей вложения и рекон-



Рис. 4. Результаты отображения-реконструкции. Исходные данные: половина цилиндра и четверть сферы

струкции для неё. Результат был усреднен по 10 случайным генерациям для каждого размера, но разброс составил менее 1/1000 от среднего значения. На рисунке 6 изображен двоичный логарифм времени работы программы, также усредненный по 10 запускам. В данном эксперименте время работы алгоритма почти не зависит от поверхности. Видно, что опшбка реконструкции падает с огромной скоростью с ростом обучающей выборки, но время при этом сильно растет.



Рис. 5. Ошибка реконструкции в зависимости от количества обучающих точек в логарифмическом масштабе. Красный – цилиндр, Желтый – сфера, Синий – S-Curve, Зеленый – Swiss-roll



Рис. 6. Время построения модели в логарифмическом масштабе в зависимости от количества обучающих точек. Красный - цилиндр, Желтый - сфера, Синий - S-Curve, Зеленый - Swiss-roll

4 Заключение

В работе реализован алгоритм Grassman&Stiefel Eigenmaps согласно оригинальной статье на языке Python и проведен иллюстративный вычислительный эксперимент, в котором алгоритм демонстрирует себя устойчивым по входным точкам и, следовательно, может быть использован для последующего решения задач машинного обучения.

Ссылка на исходный код:

https://github.com/AfoninAndrey/ITAS/blob/master/Gse-Beg.ipynb

Список литературы

- Tenenbaum J. B., de Silva V., Langford J. A Global Geometric Framework for Nonlinear Dimensionality Reduction // Science. — 2000. — Vol. 290, no. 5500. — Pp. 2319–2323.
- Roweis S. T., Saul L. K. Nonlinear dimensionality reduction by locally linear embedding // Science. - 2000. - Vol. 290. - Pp. 2323-2326.
- Zhang Z., Zha H. Principal Manifolds and Nonlinear Dimension Reduction via Local Tangent Space Alignment // SIAM Journal on Scientific Computing. – 2004. – Vol. 26, no. 1. – Pp. 313–338.
- Belkin M., Niyogi P. Laplacian Eigenmaps for dimensionality reduction and data representation // Journal Neural Computation. - 2003. - Vol. 15, no. 6. - Pp. 1373-1396.
- Donoho D. L., Grimes C. Hessian eigenmaps: Locally linear embedding techniques for high-dimensional data // Proceedings of the National Academy of Sciences. 2003. may. Vol. 100, no. 10. Pp. 5591–5596.

- Bernstein A., Kuleshov A. P. Manifold Learning: generalizing ability and tangent proximity // International Journal of Software and Informatics. — 2013. — Vol. 7, no. 3. — Pp. 359–390.
- Gine E., Koltchinskii V. Empirical graph Laplacian approximation of Laplace-Beltrami operators: large sample results // High Dimension Probability. – 2006. – Pp. 238-259.
- Rosasco L., Belkin M., de Vito E. On learning with integral operators // The Journal of Machine Learning Research. – 2010. – Pp.905–934.
- Singer A., Wu, H.-T. Vector diffusion maps and the connection Laplacian // Communications on Pure and Applied Mathematics. – 2012. – Volume 65, Issue 8, Pp. 1067-1144.
- Yanovich Y. Asymptotic Properties of Local Sampling on Manifold // Journal of Mathematics and Statistics. – 2016. – Volume 12, Issue 3, Pp. 157-175.
- Yanovich Y. Asymptotic Properties of Nonparametric Estimation on Manifold // JMLR Workshop and Conference Proceedings. – 2017. – Volume 60, Pp. 18-38.
- Bernstein A., Kuleshov A. P., Yanovich Y. Asymptotically Optimal Method for Manifold Estimation Problem // Abstracts of the International Workshop on Statistical Learning. – 2013. – P. 325.
- Bernstein A., Burnaev E., Erofeev P., Comparative Study of Nonlinear Methods for Manifold Learning // Proc. of the conf. "Information Technologies and Systems". - 2012. - 85-91.
- Bernstein A., Kuleshov A., Yanovich Y. Manifold Learning in Regression Tasks // Lecture Notes in Computer Science. – 2015. – Volume 9047, Pp. 414-423.
- Levina E., Bickel P.J. Maximum Likelihood Estimation of Intrinsic Dimension // Advances in Neural Information Processing Systems. – 2005. – Pp. 777–784.
- Campadelli P., Casiraghi E., Ceruti C., Rozza A. Intrinsic Dimension Estimation: Relevant Techniques and a Benchmark Framework // Mathematical Problems in Engineering. – 2015. – Pp. 1-21.
- Буланов О., Янович Ю. Тестирование гипотезы о многообразии // Сборник статей конференции "Информационные технологии и системы" (ИТиС'17). – 2017. С. 1-8.

Глубокие сверточные сети в задаче классификации томограмм головного мозга, пораженного микроангиопатией и рассеянным склерозом

Якушева Софья^{1,2}, Пимкин Артем^{3,1,2}, Беляев Михаил^{3,1}

 Институт проблем передачи информации им. А.А.Харкевича РАН
Московский физико-технический институт (государственный университет)
³ Сколковский Институт Науки и Технологий yakusheva.sf@phystech.edu

Аннотация Диагностика рассеянного склероза имеет важное прикладное значение в связи с распространённостью этого заболевания. В работе предлагается метод, основанный на применении сверточных нейронных сетей к анализу снимков МРТ головного мозга пациентов, для выявления рассеянного склероза и очагов сосудистого генеза у пациента. Результаты могут быть использованы для полуавтоматической расшифровки МРТ и облегчения работы врачейрентгенологов.

Ключевые слова: сверточные сети, рассеяный склероз, очаги сосудистого генеза, микроангиопатия, глубокое обучение

1 Введение

Рассеянный склероз - распространённое заболевание нервной системы [1]. Из-за сбоев в работе иммунной системы человека разрушаются миелиновые волокна, окружающие нервные клетки. Это приводит к замедлению прохождения импульсов, ухудшению питания нейронов и их последующему отмиранию. С развитием болезни человек утрачивает способности к речи и произвольным движениям, а также теряет контроль над органами таза [2].

Основным методом диагностики рассеянного склероза является МРТ. Очаги гиперинтенсивности, появляющиеся в головном мозге, видны на томограммах как характерные белые пятна диаметром более 3 мм в больших полушариях, стволе мозга и мозжечке.

Подобные белые пятна могут быть проявлением и некоторых сосудистых заболеваний, например атеросклероза, врожденных пороков развития, инсульта и других [3].

В данной работе предлагается автоматизированный подход к расшифровке томограмм с целью выявления очагов гиперинтенсивности и классификации их как проявлений рассеянного склероза или же как следствий сосудистого генеза в случае микроангиопатии (заболевания, поражающего мелкие сосуды). Существующие методы решения используют, например, алгоритм KNN (см. [4], [5]), решающие деревья (см. [5]) и SVM (см. [5], [6]) для бинарной классификации томограмм здоровых пациентов и пациентов с рассеянным склерозом. Помимо этих подходов, используются свёрточные нейронные сети для решения задачи классификации по одному срезу (см. [7]) и задачи сегментации очагов на томограммах пациентов со склерозом (см. [8]). В данной работе на основе анализа сразу нескольких срезов, принадлежащих здоровым пациентам и пациентам с двумя указанными выше патологиями, была построена модель нейронной сети, решающая задачу трёхклассовой классификации входного изображения, которую можно использовать для ускорения и облегчения процесса постановки диагноза. Также была исследована способность модели к сегментации очагов гиперинтенсивности в отсутствие данных с предварительной сегментационной разметкой.

2 Постановка задачи

Все данные были разделены на три класса:

- 0: томограммы здоровых пациентов,
- 1: томограммы пациентов с рассеянным склерозом,
- 2: томограммы пациентов с микроангиопатией.

В качестве основной в работе решалась задача классификации входного трехмерного одноканального изображения (подробнее в параграфе 3) в двух постановках.

Бинарная задача: классифицировать снимок как принадлежащий классу 0 или не принадлежащий ему.

Тернарная задача: классифицировать снимок как принадлежащий классу 0, 1 или 2.

Помимо основных задачи, дополнительно решались бинарные задачи классификации для пар классов 0 и 1, 1 и 2, 0 и 2 с целью исследования способности модели к отделению классов друг от друга.

3 Данные

Исследования проводились на данных МРТ головного мозга 300 человек. Среди них 150 здоровых пациентов, 75 пациентов с рассеянным склерозом и 75 пациентов с сосудистыми заболеваниями (микроангиопатией). Отбирались снимки в модальности flair с временем повторения импульсов более 8000 мс и временем между возбуждением импульса и эхом более 100 мс. Данные по одному пациенту состоят из 22 сагиттальных срезов размера 256×256 каждый. Таким образом, с точки зрения обучения сверточной сети, входные изображения представляют собой одноканальные трехмерные изображения.

Благодаря большому количеству данных удалось составить как выборку из 300 изображений с равным количеством томограмм здоровых и больных



Рис. 1. Примеры томограмм, слева направо: здоровый пациент, пациент с микроангиопатией, пациент с рассеянным склерозом.

пациентов, так и выборку из 225 изображений с равным количеством томограмм здоровых пациентов, больных склерозом и больных микроангиопатией.

4 Методы

4.1 Предобработка

Поскольку на данных изображениях были заметны небольшие искажения, свойственные томограммам, была проведена предобработка изображений с помощью N4ITK [9]. Подобный подход часто используется для предобработки MPT. Но этот метод не принес существенного улучшения результатов в рамках данной задачи.

Все входные данные были нормализованы (центрированы и нормированы), это повысило точность модели на $\approx 5\%$. Также данные прошли корегистрацию (относительное выравнивание) с помощью affine + rigid регистрации (подробнее в [10]).

Кроме того в ходе экспериментов для подачи на вход модели использовались сжатые в 2 и 4 раза (размера $128 \times 22 \times 128$ и $64 \times 22 \times 64$ пикселей соответственно) вдоль осей абсцисс и аппликат.

4.2 Архитектуры

На сегодняшний день одним из наиболее популярных методов обработки изображений является применение нейронных сетей. Для решения поставленных задач в качестве базовых архитектур были выбраны ResNet, VoxCNN и DenseNet, конфигурации которых менялись для лучшего соответствия задаче.

ResNet. [11] Модель с данной архитектурой заняла 1 место в задаче классификации ILSVRC 2015. Её ключевая идея - стекинг множества блоков при наличии идентификаторов, способных пропускать один или несколько



Рис. 2. Residual block [11].

уровней (рис.2). Такие идентификаторы позволяют сети обучаться не хуже менее глубоких архитектур. В работе использовались уменьшенные версии ResNet34 с небольшими архитектурными изменениями.

DenseNet. [12] Модель с данной архитектурой показывает результат лучше, чем ResNet на наборе данных ImageNet [13]. В DenseNet все блоки сети соединены друг с другом и позволяют многократно использовать результаты работы друг друга. В экспериментах использовалась архитектура DenseNet121 и её урезанные версии вплоть до вдвое меньшей по числу слоёв.

VoxCNN. Подобная архитектура, представленная в [14], успешно используется для классификации томограмм головного мозга.

Все модели были реализованы с помощью библиотеки РуТогсћ².

В ходе экспериментов существенно менялись архитектуры моделей: увеличивалось и уменьшалось число блоков и число слоев в блоке, изменялись параметры слоев.

В ходе экспериментов лучший результат показала архитектура VoxCNN, поэтому в результирующей модели была использована именно она.

4.3 Обучение

Обучение производилось с трёхфолдовой кросс-валидацией. После экспериментов с изменением размеров батча в пределах от 2 до 20 по значениям ассигасу на кросс-валидации было выбрано значение 5 и все дальнейшие эксперименты проводились с ним. В качестве оптимизатора в ходе экспериментов использовались AdamOptimizer и SGD, в результирующей модели использован AdamOptimizer. Число эпох обучения в большинстве экспериментов равнялось 70. Параметр learning rate был выбран как кусочнопостоянная функция с начальным значением 10⁻⁴, уменьшающаяся в 10 раз на 35 и 50 эпохе. В качестве функции потерь использовалась кросс-энтропия.

² http://pytorch.org

5 Результаты

5.1 Метрики качества

Данные позволяли решать обе задачи на сбалансированных наборах данных, поэтому была выбрана метрика accuracy - доля правильных ответов.

5.2 Результаты экспериментов

В таблице 1 представлены ассигасу модели для задач, описанных в параграфе 2. Для оценки эффективности модели на основной бинарной задаче (0 vs 1&2, норма против патологии) была использована модель, решающая тернарную задачу, ответы которой были бинаризованы. Такой подход был использован в силу лучшего качества классификации у бинаризованной тернарной модели, по сравнению с моделью, обучавшейся на данных с метками "норма"и "патология". Это можно связать с дополнительной информацией, которую модель извлекает из наличия двух патологических классов и, как следствие, их лучшей отделимости. Для остальных задач (0 vs 1, 1 vs 2 и 0 vs 2) модель была обучена на данных, содержащих только метки, фигурирующие в постановке задачи. Также стоит обратить внимание на то, что все задачи, содержащие одновременно класс Нормы и Микроангиопатии, решаются существенно хуже. Подробнее об этом в параграфе 6.

Таблица 1. Ассигасу модели на трёхфолдовой кросс-валидации

Задача классификации	Accuracy
Бинарная задача	0.82 ± 0.01
Тернарная задача	0.80 ± 0.03
Норма vs Рассеянный склероз	0.95 ± 0.01
Hopma vs Микроангиопатия	0.73 ± 0.04
Склероз vs Микроангиопатия	1.00 ± 0.00

Таблица 2 представляет из себя суммарную confusion matrix по трем тестовым фолдамм кросс-валидации. Каждая ячейка таблицы указывает количество тестовых объектов, которые относятся к классу i, а модель отнесла их к классу j, где i и j - номер строки и столбца ячейки соответственно. В данной таблице также стоит заметить худшую, чем у остальных, отделимость классов Нормы и Микроангиопатии, о которой было упомянуто в описании таблицы 1.

5.3 Карта внимания модели

Особый интерес представляла проверка способности модели к выделению очагов гиперинтенсивности на томограмме, осуществляемая по карте вни-

Таблица 2. Суммарная confusion matrix по трём фолдам

	Норма	Микроангиопатия	Склероз
Норма	57	16	2
Микроангиопатия	17	54	4
Склероз	3	2	70

мания. Подобные карты позволяют понять, какие части изображения являются ключевыми для модели при принятии решения о принадлежности его к классу патологии (подробнее о картах внимания см. [15]).

Карта внимания была получена следующим образом. Было зафиксировано изображение и результат работы модели на нем - вероятность того, что изображение принадлежит пациенту с патологией. Затем небольшая часть входного изображения заменялась маской из нулевых пикселей. Использовалась маска размером $20 \times 1 \times 20$ пикселей. Маленькая размерность по оси ординат объясняется сильной дискретностью исходных изображений по данной оси. Также стоит отметить, что в силу центрированности данных (см. параграф 4.1) нулевая интенсивность близка к средней интенсивности изображения, т.е. область с наложенной маской не должна распознаваться моделью как очаг гиперинтенсивности. Попытки использовать в качестве маски соответствующий участок томограммы здорового мозга не были успешны из-за неидеального выравнивания и несовпадения размеров мозга разных пациентов. Разность между ответом модели для изображения с наложенной маской и исходным предсказанием была записана в матрицу размера $12 \times 22 \times 12$ в соответствии с местоположением маски. Далее маска перемещалась по изображению до заполнения матрицы. Таким образом, значение разности может быть интерпретировано как величина вклада участка, соответствующего маске, в ответ модели, а матрица - как карта внимания.

Поскольку для рассматриваемых патологий характерным признаком являются очаги гиперинтенсивности (белые пятна на томограммах), замена их маской со средней интенсивностью изображения должна снижать предсказываемую вероятность присутствия заболевания, то есть давать большое отклонение предсказания.

Пример полученной маски, наложенной на изображение, приведен на Рис. 3. После анализа карт врачом подтвердилось предположение о том, что модель действительно фокусируется на очагах (но, возможно, не на всех присутствующих на изображении).

6 Выводы

В ходе исследований было замечено, что принципиально разные по архитектуре модели неизменно ошибаются на некоторых изображениях. После дополнительной консультации с врачом выяснилось, что либо это были сложные случаи, в которых затруднительно поставить какой-либо диагноз, либо



Рис. 3. Пример изображения с наложенной картой внимания сети. На правом изображении хорошо видна фокусировка внимания на очаге в большом полушарии. Для постановки диагноза оказалось достаточным распознавание одного этого очага.

диагноз был поставлен после дополнительных исследований (в частности, аксиальных срезов), поскольку на сагиттальных срезах плохо видны признаки подобных заболеваний (в частности, микроангиопатии). Это объясняет плохую отделимость классов нормы и микроангиопатии друг от друга (см. Таблицы 1 и 2). Таким образом, нет возможности говорить об улучшаемости полученных результатов на текущих данных.

Тем не менее, на текущем наборе данных были получены многообещающие результаты, позволяющие говорить о продолжении исследований в данном направлении и возможном использовании подобных технологий в клинических условиях. Кроме того, достигнут важный побочный результат - карты внимания моделей сильно скоррелированы с очагами гиперинтенсивности, что позволяет выдвинуть предположение о возможном использовании подобных моделей для сегментации в отсутствие предварительно размеченных данных.

Дальнейшие исследования планируется построить вокруг сбора дополнительных данных и очистки текущих от вышеописанных проблем, что в дальнейшем может позволить, например, построить модель, накладывающую более слабые ограничения на входные данные, чем описанные в параграфе 3. Также результаты, описанные в Таблице 1 (в частности, хорошая отделимость рассеянного склероза от микроангиопатии моделью, обученной для этой задачи) позволяют говорить об ансамблировании моделей (например, стекинге).

Список литературы

- 1. Пажигова З.Б., Карпов С.М., Шевченко П.П., Бурнусус Н.И. Распространенность рассеянного склероза в мире (обзорная статья) // Международный журнал экспериментального образования. – 2014. – № 1(2). – С. 78-82.
- Рассеянный склероз: руководство для врачей / Т.Е.Шмидт, Н.Н.Яхно. 5-е изд. – М. : МЕДпресс-информ, 2016. – 272 с. : ил
- Брюхов В. В., Куликова С. Н., Кротенкова М. В. МРТ в диагностике и дифференциальной диагностике рассеянного склероза. В 2-х томах // в кн. Аутоиммунные заболевания в неврологии. Клиническое руководство // Аутоиммунные заболевания в неврологии. Клиническое руководство. — Т. 1. — РООИ Здоровье человека Москва, 2014.
- E. Tadayon, R. M. Khayati, V. Karami, S. Massood Nabavi. A Novel method for automatic classification of multiple sclerosis lesion subtypes using diffuson tensor MR images. Biomedical Engineering: Applications, Basis and Communications Vol. 28, No. 05, 1650038 (2016)
- Y. Zhang, S. Lu, X. Zhou and oth. Comparison of machine learning methods for stationary wavelet entropy-based multiple sclerosis detection: decision tree, knearest neighbors, and support vector machine. Simulation: Transactions of the Society for Modeling and Simulation International, 2016, Vol. 92(9) 861–871
- G. Kocevar, C. Stamile, S. Hannoun, F. Cotton, S. Vukusic, F. Durand-Dubief, D. Sappey-Marinier. Graph Theory-Based Brain Connectivity for Automatic Classification of Multiple Sclerosis Clinical Courses. Front. Neurosci., 25 October 2016
- S.-H. Wang, H. Cheng, P. Phillips, Y.-D. Zhang. Multiple Sclerosis Identification Based on Fractional Fourier Entropy and a Modified Jaya Algorithm. Entropy 2018, 20(4)
- S.Valverde, M.Cabezas, E. Roura, S. González-Villà, D.Pareto, J. C. Vilanova, L.Ramió-Torrentà, À. Rovira, A. Oliver, X. Lladó. Improving automated multiple sclerosis lesion segmentation with a cascaded 3D convolutional neural network approach. NeuroImage Volume 155, 15 July 2017, Pages 159-168
- Tustison NJ1, Avants BB, Cook PA, Zheng Y, Egan A, Yushkevich PA, Gee JC. : N4ITK: improved N3 bias correction. IEEE Trans Med Imaging. 2010 Jun; 29(6):1310-20.
- Avants B.B., Tustison N.J., Stauffer M., Song G., Wu B., Gee J.C. : The Insight ToolKit image registration framework. Frontiers in Neuroinformatics. 2014;8:44.
- He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition pp. 770–778 (2016)
- Huang, G., Liu, Z., Weinberger, K.Q., van der Maaten, L.: Densely connected convolutional networks 1(2), 3 (2017)
- Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. pp. 248–255. IEEE (2009)
- S. Korolev, A. Safiullin, M. Belyaev, Y. Dodonova. Residual and plain convolutional neural networks for 3D brain MRI classification. 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)
- 15. M. D. Zeiler, R. Fergus : Visualizing and Understanding Convolutional Networks. 2013 CoRR

Классификация метастаз головного мозга по анатомическим областям

Алексей Шевцов^{1,2}, Егор Кривов^{1,2}, Михаил Беляев^{3,1}

 Институт проблем передачи информации им. А.А. Харкевича РАН
Московский физико-технический институт (государственный университет)
³ Сколковский институт науки и технологий shevtsov.ae@phystech.edu

Аннотация Основным методом лечения метастаз является точечная радиохирургия, например с оборудованием «Гамма-нож». Необходимой для этого процедурой является предварительная разметка метастаз: их оконтуривание и именование. В данной работе с помощью простых методов машинного обучения решается задача автоматизации процесса именования метастаз головного мозга названием анатомической области, в которой она расположена.

Keywords: классификация, машинное обучение, МРТ, головной мозг, радиохирургия, метастазы, Гамма-нож

1 Введение

По консервативным оценкам 8 – 10% людей больных раком дополнительно страдают и от вторичного ракового образования в головном мозге — метастаз, значительно ухудшающих качество жизни заболевших. В большинстве случаев в мозгу образуется сразу несколько метастаз. Ожидаемая продолжительность жизни таких пациентов при применении исключительно поддерживающей терапии составляет 40 – 50 дней [1] [2].

Пациент, обратившийся в онкологическую клинику с проблемами головного мозга, сначала проходит процедуру МРТ, после которой ему говорят о наличии или отсутствии раковых опухолей. Обнаружив злокачественные образования (в данной статье речь пойдет про метастазы) врач принимает решения о способе их лечения.

Один из основных способов борьбы с небольшими опухолями, такими как метастазы, — удаление с помощью аппарата «Гамма-нож». Для начала работы необходимо сначала выполнить разметку, то есть указать машине местонахождение злокачественных образований. Поэтому врач должен вручную просмотреть все сделанные снимки мозга и на каждом отметить контуры метастаз, а затем дать каждому образованию название в соответствие с принятыми в больнице правилами именования. Процедура очень длительная и трудоемкая, учитывая что в среднем у нездорового пациента около 4 метастаз. В процессе можно упустить какие-либо или наоборот — отметить лишние. Существуют алгоритмы автоматического оконтуривания метастаз [3], но для их практического применения важно уметь именовать построенные контуры. Это помогает в структурировании полученных результатов. Так удобнее собирать статистики по различным видам опухолей, а врачу проще работать при последующих визитах пациента. Например, если у больного было много злокачественных образований, то доктору хочется сразу перейти к нужным снимкам определенной метастазы, а не вручную выискивать их. Еще, зачастую, программное обеспечение в одной больнице может не поддерживать работу с файлами из другой, соответственно врач, не имея доступа к изображениям, может по названию понять в каких анатомических областях расположены метастазы. Для краткости, в дальнейшем мы будем называть анатомическую область просто «локализацией».

Возникает желание ускорить и автоматизировать процесс обработки результатов МРТ, чтобы облегчить техническую сторону работы доктора, позволяя сфокусироваться на вопросах лечения пациента. И цель данной работы — создать алгоритм для автоматического именования метастаз головного мозга на основании их локализации.

2 Постановка задачи

2.1 Именование метастаз

Названия, которые присваивают врачи опухолям, характеризуют локализацию опухоли в мозгу. Название всегда содержит указание на анатомическую область (Таблица 1), часто содержит указание стороны области (левая/правая), и изредка содержит дополнительные уточнения. Таким образом название может выглядеть как: **FrSsl**, что означает расположение опухоли в лобной доле (**Fr**), слева (**S**) вверху (**s**) латерально (**l**).

Всего 20% предоставленных данных имеют дополнительные уточнения — это 6 самых распространенных областей (Рис. 1, стр. 3), которые также являются самыми большими по размеру. Но все–равно это малое количество, поэтому в данной работе мы фокусируемся только на классификации анатомической области.

Сокращение	е Анатомическая область	Сокращение	Анатомическая область
Fr	Лобная область	VL	Боковые желудочки
Ce	Мозжечок	PC	Мосто-мозжечковый угол
Pa	Теменная область	Pi	Пинеальная область
Te	Височная область	\mathbf{SC}	Хиазмально-селлярная область
Oc	Затылочная доля	V3	III желудочек
\mathbf{GB}	Базальные ганглии	V4	IV желудочек
\mathbf{BS}	Ствол		

Таблица 1. Расшифровка сокращений

2.2 Формальная постановка задачи

Для произвольной метастазы головного мозга, с известным контуром на трёхмерном снимке MPT головного мозга необходимо классифицировать анатомическую область, в которой она находится.

3 Данные

Для экспериментов использовались данные, из центра «Гамма–нож» за 2005–2011 год. Из них были отобраны пациенты только с метастазами. В итоге работа велась с 2163 опухолями 481 заболевшего (≈ 4.5 на пациента).



Рис. 1. Распределение метастаз из выборки по анатомическим областям

Всего было представлено тринадцать различных анатомических областей, из которых работа велась всего с девятью наиболее распространенными (Рис. 1). Особенность данных заключалась в неоднозначности расположения некоторых метастаз (12% среди всех случаев), например врач отнес образование одновременно к лобной и височной частям мозга. Для них в основном встречаются четыре области: теменная, лобная, затылочная и височная — 42%, 28%, 15%, 13% соответственно (Рис. 3, на стр. 5).

4 Эксперименты

В статье используются базовые методы машинного обучения: градиентный бустинг над решающими деревьями (*GBC*), k–ближайших соседей (*kNN*) и

SVM с RBF ядрами (все — библиотека Scikit-learn[4]), значительное преимущество которых состоит в простоте настройки и скорости работы.

Учитывая особенность с неоднозначностью локализации некоторых метастаз, сначала решалась задача классификации тех очагов, в которых была указана единственная локализация, а затем велась работа с неоднозначными случаями.

4.1 Однозначные локализации

Подготовка данных

Выбор начала отсчета. Основной характеристикой местонахождения метастазы являются координаты центра, которые можно вычислять как от левого угла трехмерного снимка, так и от его центра. Центром метастазы считаем центр параллелепипеда, параллельного анатомическим плоскостям, в который ее можно вписать. Таким образом было 3 варианта признаков: координаты центра метастаз, вычисленные от левого угла снимка (левый угол), координаты центра местатаз, вычисленные от центра снимка (центр) и одновременное использование их всех (левый угол и центр).

Выбор стандартизации. Обычно, предобработка признаков подразумевает преобразование значений признаков к стандартному нормальному распределению (StandardScaler), однако в данной конкретной задаче признаки расстояния в мозгу, поэтому не хочется терять метрическую интерпретацию данных. Можно выбрать признак с наибольшей дисперсией и разделить на нее значения всех координат (FeatureScaler). Признаки будут иметь дисперсию близкую к 1, при этом не потеряется суть расстояния в данных.

Разбиение выборки. Каждый пациент имеет свои особенности, поэтому разбиение выборки на обучающую и тестовую производилось по пациентам, а значит разные метастазы одного человека не могли попасть в разные подвыборки. А для проведения внутренней кросс–валидации во время подбора гиперпараметров алгоритмов, обучающая выборка была дополнительно разбита на 10 фолдов по тому же принципу.

Отбор признаков Выбор подразумевал определение комбинации «начало отсчета» — «нормализация». С одной стороны: левый угол, центр и оба сразу, а с другой: StandardScaler и FeatureScaler. В итоге качество работы алгоритмов оценивалось на всевозможных парах с помощью внутренней кросс-валидации, а подбор гиперпараметров алгоритмов производился с использованием поиска по сетке, настраиваясь на максимизацию значения accuracy.

Результаты Данные, о работе алгоритмов в тех или иных случаях приведены в Таблице 2 на стр. 5.

На их основании был принят вариант работы с расстоянием, которое отсчитывается от центра мозга, с использованием *FeatureScaler*.

Точка отсчета	StandardScaler	FeatureScaler	GBC	kNN	SVM
левый угол	-	_	82,5	81,6	83,8
левый угол	+	_	80,8	81,2	80,9
левый угол	_	+	81,4	81,1	$83,\!6$
центр	-	_	84,2	84,4	85,0
центр	+	—	83,6	$83,\!9$	82,3
центр	_	+	84,4	84,2	85,7
левый угол и центр	-	_	85,5	82,3	85,1
левый угол и центр	+	—	83,8	$83,\!5$	83,5
левый угол и центр	-	+	84,6	$83,\!1$	85,2

Таблица 2. Сравнение результатов работы алгоритмов, подсчитанных по метрике *accuracy* в различных признаковых пространствах для случая единственной локализации.

Ошибки Сопоставляя матрицы на на Рис. 2 и 3 можно понять, что алгоритм обычно не уверен в тех же областях, в которых сомневаются врачи.





Рис. 2. Матрица ошибок при прогнозировании случаев с единственной указанной локализацией *SVM* алгоритмом.

Рис. 3. Матрица, на пересечении строчек столбцов которой стоит число встречающихся пар в случаях с неоднозначной локализацией.

4.2 Неоднозначные локализации

Как уже упоминалось, в данных присутствуют метастазы, которым врач присвоил сразу несколько анатомических областей. В этих случаях почти всегда врач присваивал два местоположения, и лишь в двух случаях было указано три местоположения.

Первой идеей было свести подобные случаи к случаю одиночной локализации, а именно: научиться предсказывать хотя бы одну из двух локализаций, в которых расположена метастаза. Например можно брать только первую из локализаций, указанную в названии, полагая, что врач ставит более предпочтительную область раньше другой. Данный способ показал себя лучше предыдущего, дав результат на *SVM*. Но предположение о преимуществе локализации, указанной первой в названии над второй не имеет под собой точного основания, поэтому результат сильно зависит от выборки. Другим вариантом было сделать из одной неоднозначной записи две, у которых полностью совпадают признаковые описания, но различны значения целевой переменной, при этом никак не учитывая, что они являлись частью одной метастазы. Способ плох тем, что имея разные ответы на одинаковые признаки, алгоритм путается и дает заниженные результаты

Но можно и не разделять локализации, и не отбрасывать одну из них, а поступить следующим образом: исключить из обучающей выборки все неоднозначные случаи, чтобы не путать алгоритм, поместив их в тестовую выборку. Тогда ответ — это класс, которому была приписана наибольшая вероятность (*major*-попадание). Таким образом верным ответом считается предсказание алгоритмом одной из двух указанных врачом локализаций. Способ хорош тем, что он не зависит от числа анатомических областей и позволяет работать со всеми предоставленными случаями.

Случай	accuracy
Только однозначные	86
Только неоднозначные	91
Общий	87

Таблица 3. Результаты работы SVM алгоритма с FeatureScaler

Результаты работы модели показывают, что принятая стратегия работы с неопределенными локализациями дала отличный результат, внеся вклад в увеличение точности предсказания по всей выборке, потому как алгоритм имеет возможность допустить ошибку при классификации неоднозначных случаев, взяв за наиболее вероятный случай неверную локализацию.

График на Рис. 4 показывает, что основные ошибки алгоритма происходят на случаях, которые крайне редко встречаются в выборке



Рис. 4. Ответы алгоритма на случай предсказания однозначной локализации

(VL, PC ~ 1% каждая), поэтому алгоритм предпочитает относить их к более распространенным местам. Также видно, что кроме двух самых распространенных локализаций (Fr, Ce) правильный ответ зачастую стоит не на первом месте, а ровно на втором месте (*minor*-попадание).



4.3 Возможные пути развития

Рис. 5. Разбиение метастаз тестовой выборки по количеству локализаций и правильности прогноза. Сумма всех столбцов равна 1. На данном этапе алгоритм способен давать в качестве ответа лишь одну локализацию. Одним из вариантов улучшения является обучение его предсказанию сразу двух локализаций, в случаях, когда действительно затруднительно однозначно классифицировать метастазу.

Одним из вариантов было бы судить о наличии или отсутствии мультиклассового случая по уверенности алгоритма в своем предсказании, а именно по разнице между вероятностями приоритетной и второстепенной анатомических областей (*delta*). Графические интерпретации представлены на Рис. 6 и 7. Они показывают, что наиболее удачное значение для порога, отделяющего одноклассовый и мультиклассовый случаи находится на уровне *delta* = 0, 75. При этом

будет верно предсказана лишь половина мультиклассовых случаев, а потеряно около 20% верных ответов с однозначной локализацией, что составляет

несоизмеримое количество (Рис. 5). Поэтому стоит поискать другой способ определения порога, либо совсем не прибегать к этому методу.





Рис. 6. Распределение значений *delta* для однозначных случаев.

Рис. 7. Распределение значений *delta* для неоднозначных случаев.

5 Заключение

Задача классификации метастаз по анатомическим областям может решаться простыми методами машинного обучения с высокой точностью (86% однозначных случаев). При этом, дополнительно был продемонстрирован метод работы с ситуациями, когда анатомическая область метастазы определена врачом неоднозначно. Предсказание хотя бы одной из нескольких, указанных врачом анатомических областей, является очень качественным способом классификации в подобных случаях (91% неоднозначных случаев).

Список литературы

- J. S. Barnholtz-Sloan, A. E. Sloan, F. G. Davis, F. D. Vigneau, P. Lai, and R. E. Sawaya, "Incidence proportions of brain metastases in patients diagnosed (1973 to 2001) in the metropolitan detroit cancer surveillance system," *Journal of clinical oncology*, vol. 22, no. 14, pp. 2865–2872, 2004.
- L. J. Schouten, J. Rutten, H. A. Huveneers, and A. Twijnstra, "Incidence of brain metastases in a cohort of patients with carcinoma of the breast, colon, kidney, and lung and melanoma," *Cancer*, vol. 94, no. 10, pp. 2698–2705, 2002.
- E. Krivov, V. Kostjuchenko, A. Dalechina, B. Shirokikh, A. Denisenko, G. Makarchuk, A. Golanov, and M. Belyaev, "Tumor delineation for brain radiosurgery by a convnet and non-uniform patch generation," in 3rd International Workshop on Patch-based Techniques in Medical Imaging, p. 8p, 2018.
- F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
Синтез медицинских изображений с помощью циклических генеративно-состязательных сетей

Кондратенко Владимир $^{1,2},$ Пи
мкин Артем $^{1,2},$ Писов Максим $^{3,2},$ Беляев Миха
ил 1,3

¹ Сколковский институт науки и технологий
 ² Московский физико-технический институт (государственный университет)
 ³ Институт проблем передачи информации им. А.А. Харкевича РАН
 vladimir.kondratenko@phystech.edu

Аннотация В современной радио-хирургии возникает важная прикладная задача синтеза изображений компьютерной томографии (KT) из изображений магнитно-резонансной томографии (MPT). В данной работе рассматриваются различные вариации метода на основе циклических генеративно-состязательных сетей для синтеза KT из MPT и MPT из KT, приводится интерпретация полученных результатов, а так же озвучиваются идеи для дальнейшего исследования.

1 Введение

Магнитно-резонансная томография (MPT), как и компьютерная томография (KT), повсеместно используются в медицинской практике для исследования внутренних органов и тканей. В основе MPT лежит явление ядерного магнитного резонанса, а при KT объемное изображение получается на основании того, как рентгеновское излучение поглощается различными по плотности тканями. MPT, как и KT, имеет ряд противопоказаний, к тому же при KT человек подвергается вредному для него рентгеновскому излучению.

Зачастую, например, в радио-хирургии, возникает потребность иметь оба снимка. По MPT сегментируются опухоли, которые должны быть облучены, а KT используется для расчета доз, для того, чтобы не облучать здоровые ткани. Обе описанные процедуры требуют времени и ресурсов.

В связи со всем вышеперечисленным, возникает важная прикладная задача получения одного типа изображений из другого. В данной работе предлагается метод с использованием циклических генеративно-состязательных сетей, который позволяет синтезировать как изображения КТ головы из изображений МРТ, так и изображения МРТ из изображений КТ.

2 Постановка задачи

Формально задачу можно сформулировать следующим образом: пусть есть X - множество числовых массивов одного размера, являющихся представлениями MPT снимков головного мозга и такое же множество для снимков

КТ Y. Пусть существует $f: X \to Y$ - отображение, сопоставляющее МРТ снимку *соответствующий*¹ ему КТ снимок и обратное к нему отображение $g: Y \to X$, сопоставляющее КТ снимку *соответствующий* ему МРТ снимок. Задача состоит в том, чтобы по имеющейся *обучающей выборке* восстановить функции f и g. Следует разделить два принципиально разных случая:

Спаренные данные. Пусть имеются данные о N пациентах. Предполагается, что для любого пациента есть пара *coomsemcmsyющих* друг другу MPT и KT снимков. Иными словами: $D = \{(x_i, y_i)\}_{i=1}^N \subset X \times Y : \forall k \in \overline{1, N} \hookrightarrow f(x_k) = y_k, g(y_k) = x_k$, где x_i и y_i являются *coomsemcmsyющими* друг другу. Множество D называется *обучающей выборкой*.

Неспаренные данные. Пусть имеются N КТ снимков различных пациентов, и M МРТ снимков различных пациентов: $A = \{x_i\}_{i=1}^N \subset X, B = \{y_i\}_{i=1}^M \subset Y$, При этом может оказаться, что $\forall x \in A, y \in B \hookrightarrow f(x) \neq y, g(y) \neq x$. В этом случае обучающей выборкой называется множество пар $D = \{(x, I_A(x)), x \in A \cup B\}$, где I_A - индикаторная функция множества A.

3 Существующие подходы

3.1 Спаренные данные

Одним из зарекомендовавших себя методов решения задачи синтеза изображений КТ из МРТ с использованием спаренных данных является метод на основе патчей²: из пар снимков вырезаются трехмерные патчи и составляется большая "база данных" пар патч из МРТ - патч из КТ. При синтезе из исходного изображения выделяются патчи того же размера, затем ищутся несколько наиболее подходящих ему по структуре, и в качестве сгенерированного патча выдается их взвешенная сумма. Подробнее этот метод изложен в [1].

В связи с бурным развитием сверточных нейронных сетей появился ряд работ, в которых применялись полносверточные сети (fully convolutional networks) [2,3]. Причем, в [3] применялся подход на основе генеративносостязательных сетей (Generative Adversarial Networks [4]). О том, что такое GAN, будет рассказано ниже.

3.2 Неспаренные данные.

Обучение на спаренных данных может иметь ряд недостатков. Во-первых, в реальных условиях трудно добиться того, чтобы снимки были хорошо выровнены, из-за этого исходный снимок и соответствующий ему будут соотноситься неидеально[5], что может существенно снизить качество обучаемых

¹ Как будто пациенту делают снимок КТ вместо данного снимка МРТ, то есть тот же человек на снимке КТ находится в такой же позе, состоянии и т. д.

 $^{^2}$ патч (patch) - вырезанное из исходного изображение меньшего размера

моделей. Во-вторых, парных данных довольно мало (все статьи, перечисленные в предыдущем разделе имели в обучающей выборке не более 40 пар снимков).

С более общей задачей неспаренного отображения одного типа изображений в другой (unpaired image-to-image translation) справилась модель CycleGAN, предложенная в [6]. Подход на основе этой модели был предложен в [5], где авторы показали, что обучение на неспаренных данных при помощи CycleGAN дает лучший результат, чем обучение на спаренных данных.

4 Данные

Набор данных состоит из 239 MPT снимков и 245 KT снимков. MPT снимки имеют диапазон значений от 0 до 337, KT снимки - от 0 до 4095.

Для оценки качества полученной модели мы использовали *тестовую* выборку, состоящую из пар МРТ и КТ снимков для 8 пациентов, которые не участвовали в процессе обучения модели. Для того, чтобы их выровнять, применялся алгоритм аффинной корегистрациииз библиотеки ANTs[7].

5 Предложенный метод

5.1 Предобработка данных

Для того, чтобы избавиться от всех артефактов на снимках КТ (цилиндр рабочей зоны, рамки и т. д.), были выполнены следующие шаги: сначала было произведено обнуление всех значений меньше 750 (tresholding), затем были удалены крупные (>1000 вокселей) компоненты связности, кроме самой головы, и было произведено посрезовое³ открытие (эрозия с последующим наращиванием), в конце посрезово бралась выпуклая оболочка. В результате получилась трехмерная бинарная маска, по которой потом из исходного снимка удалялось все лишнее. Для части этапов предобработки использовалась библиотека OpenCV.

На следующем шаге из обучающей выборки был выбран снимок МРТ, на который при помощи алгоритма корегистрации были выровнены все остальные снимки МРТ и КТ. Затем исходный тензор линейно отображался в [-1,1], после этого выделялся bounding box, разрешение которого изменялось до 256 × 256 пикселей по длине и ширине, из которого затем выбирался случайный срез. Итого после предобрабоки получилось 239 МРТ снимков размера $256 \times 256 \times z$, где $z \in [41, 150]$ (z < 70 на 10 снимках) и 245 КТ снимков размера $256 \times 256 \times z56 \times z$, где $z \in [74, 147]$.

³ Здесь и далее брались аксиальные (по оси z) срезы, если не оговорено обратное

5.2 Оценка качества

Для оценки качества полученных моделей, как правило, используют Mean Absolute Error (MAE): $MAE(y,y') = \frac{1}{N} \sum_{i=1}^{N} |y(i) - y'(i)|$, где N - количество вокселей в y (y' имеет такой же размер) или в бинарной маске, по которой считается MAE.

Для оценки качества полученной модели выполнялись следующие шаги:

- 1. Для МРТ из тестовой выборки предсказывалось КТ
- 2. Спаренное КТ корегистрировалось на МРТ из шага 1
- МАЕ считалось по пересечению бинарных маскок предсказанного и парного КТ, чтобы исключить из рассмотрения воздух вокруг головы, аналогично [5]

5.3 Архитектура модели

В нашей работе мы использовали различные вариации CycleGAN с архитектурами нейронных сетей, описанными в [6], которые так же использовались авторами [5]. Остановимся на том, что такое GAN и CycleGAN подробнее.

GAN Generative Adversarial Networks - генеративно-состязательные сети, впервые были предложены в [4].

Рассмотрим следующую задачу: пусть есть множество портретных фотографий людей и мы хотим научиться генерировать такие же изображения из случайного шума. GAN состоит из двух нейронных сетей - генератора G и дискриминатора D. Опуская технические детали, можно сказать, что обучение модели происходит следующим образом - D учится различать реальные фото от того, что генерирует G, а G учится генерировать из случайного шума изображения так, чтобы D классифицировал их как реальные фотографии людей.

CycleGAN Пусть есть два множества неспаренных изображений: X и Y. Задача состоит в том, чтобы научиться изменять изображение из одно множества так, будто это изображение было в другом множестве.

СусleGAN состоит из двух GANoв. Один GAN учится "переносить" изображения из первого множества во второе, а другой GAN - наоборот: из второго множества в первое. Обозначим генератор первого GANa как F_x (принимает на вход изображения из первого множества $x \in X$), а дискриминатор D_y (различает сгенерированные изображения из второго множества $y \in Y$). Аналогично обозначим генератор и дискриминатор второго GANa как F_y и D_x соответственно⁴ Главной особенностью СусleGAN является то, что в минимизируемые при обучении F_x и F_y функционалы добавлен *cycle consistency*

⁴ В оригинальной статье [6] F_x , F_y называются G и F соответственно.

$$||F_y(F_x(x)) - x||_1, ||F_x(F_y(y)) - y||_1$$
(1)

который отвечает за то, чтобы оба преобразования были обратимыми и согласовывались друг с другом. В роли X и Y в нашем случае выступают МРТ и КТ снимки.

Алгоритм обучения модели был аналогичен алгоритмам обучения, предложенным в [6,5].

Оптимизируемый при обучении генераторов F_x, F_y функционал ошибки:

$$L_{generator} = W_{gen} \cdot L_{gen} + W_{cycle} \cdot L_{cycle} = W_{gen} \cdot ((1 - D_x(F_x(x)))^2 + (1 - D_y(F_y(y)))^2) + W_{cycle} \cdot (||x - F_y(F_x(x))||_1 + ||y - F_x(F_y(y))||_1)$$
(2)

где W_{gen}, W_{cycle} - веса перед соответствующими функционалами ошибок.

Для обучения D_x использовался следующий функционал ошибки:

$$L_{D_x} = W_{disc} \cdot (D_x^2(x) + (1 - D_x(x_{fake}))^2)$$
(3)

и аналогичный для D_y с заменой x на y, где x_{fake} , y_{fake} - это одно из сгенерированных F_y и F_x соответственно изображений.

5.4 Особенности предлагаемой модели

В данной работе сравниваются модели, базирующиеся на трех ключевых идеях.

Во-первых, изначально веса всех нейронных сетей инициированы случайно, поэтому возникает вопрос: действительно ли генератору необходимо учиться "обманывать" слабый дискриминатор, или же можно сначала сместить акцент на обратимость преобразования, а затем, уже учить генератор обманывать дискриминатор?

В статье [8] при помощи подхода на основе GAN решается задача генерации изображений с изменением атрибутов (например, сделать человека на фото улыбающимся, в роли атрибута - бинарная величина улыбается человек или нет). Вес слагаемого, аналогичному W_{gen} в (2) рос линейно от 0 и до определенного порога (изначально генератор не пытался обмануть дискриминатор). Это существенно улучшило качество модели. Аналогичный подход мы применяем и в нашей работе.

Во-вторых, чтобы избавиться от случайной инициализации весов нейронных сетей мы предобучали генераторы осуществлять тождественное преобразование: $F_x(x) \approx x$, $F_y(y) \approx y$, а дискриминаторы - отличать реальное КТ от реального МРТ. Так же мы комбинировали этот подход с увеличенным в начале обучения W_{gen} , который линейно уменьшался до 1 в процессе обучения.

Третья идея заключалась в том, чтобы довольно грубо "направлять" генераторы в начале обучения. Для этого мы добавили в (2):

$$L_{map} = ||F_x(x) - y||_1 + ||F_y(y) - x||_1$$
(4)

loss:

с соответствующим весом W_{map} , который в процессе обучения линейно убывал до 0. В тех численных экспериментах, где W_{map} бывает отличным от 0, требуется, чтобы x и y были похожи на парные. Для этого они выбираются из одного места на снимках.

Так же в ходе экспериментов мы увеличили количество сохраняемых сгенерированных изображений - пула изображений, откуда берутся x_{fake}, y_{fake} в (3) до 1000, вместо 50, как в [6,5] и модифицировали стратегию изменения коэффициента обучения (learning rate policy).

5.5 Численные эксперименты

Модель была реализована на языке программирования Python с использованием библиотек Pytorch и DeepPipe. Во всех экспериментах использовался алгоритм оптимизации Adam, размер батча 1, так же, как и в [6,5]. Обучение длилось 150 эпох, в каждой эпохе 1500 шагов оптимизации, коэффициент обучения в начале обучения равнялся 0.0001, затем уменьшался в 10 раз каждые 40 эпох, $W_{gen} = 0.5, W_{disc} = 0.5, W_{cycle} = 10, W_{map} = 0$, если не оговорено обратное.

Мы провели следующие эксперименты:

- baseline эксперимент, описанный в [5]: коэффициент обучения (learning rate) равнялся 0.0002 в первые 100 эпох, затем линейно уменьшался до 0 в последующие 100 эпох. Обучение длилось 200 эпох.
- lr policy "стандартный" эксперимент, описанный выше.
- gen loss W_{qen} линейно менялся от 0 до 0.5 в первые 15 эпох.
- map loss W_{map} линейно менялся с 10 до 0 в первые 10 эпох.
- gen map loss эксперимент аналогичен gen loss, только W_{map} начиная с 15ой эпохи в течении 10 эпох линейно уменьшается до 0. Первые 15 эпох $W_{map} = 0$.
- pretrained генераторы предобучены осуществлять тождественное преобразование, а дискриминаторы - осуществлять бинарную классификацию МРТ или КТ снимок.
- pretrained gen loss эксперимент аналогичный предыдущему, только *W_{gen}* линейно менялся от 1 до 0.5 в первые 15 эпох обучения.

6 Результаты

Как видно из таблицы 1 лучшие результаты показывает lr policy.

В [5] авторы получили $MAE = 73.7 \pm 2.3$. Снимки КТ в их работе имеют диапазон значений [-600, 1400], в нашей работе - [0, 4095]. Модель, используемая в [5] на наших данных, показала $MAE = 167.49 \pm 18.78$.

Большая ошибка в моделях, отличных от *lr policy* возникает по двум причинам: модели "рисуют" кость, во-первых, не в том месте, во-вторых, ярче, чем она есть на самом деле (~ 2500 против ~ 3000) (Рис. 1).



Рис. 1: а) Снимки, сгенерированые моделями 1. lr policy, 2. gen map loss, 3. baseline и 4. соответствующий им КТ снимок из обучающей выборки. Модель gen map loss обучалась 130 эпох и показывает на тестовой выборке среднее MAE = 199.42, baseline обучалась 50 эпох, MAE = 167.9(см. Таблицу 1). На этом срезе у lr policy MAE = 92.49, у gen map loss MAE = 140.77, у baseline MAE = 130.3 b) Снимки КТ для наглядности наложены на соответствующие им МРТ. На втором и третьем снимках, сгенерированных gen map loss и baseline, виден "зазор" между костью и мозгом, которого нет на на четвертом снимке. с) Попиксельная разность между сгенерированным и реальным КТ. Видно, что gen map loss и baseline некорректно генерируют кость около мозга, при этом gen map loss может генерировать лишнюю костную ткань, a baseline может генерировать кость не полностью.

7 Обсуждение

Во всех приведенных в разделе 3 статьях использовались *парные* данные, а в [5], в обучающей выборке были выровненные пары МРТ и КТ снимков одних пациентов. В нашей же работе показано, что в качестве обучающей выборки можно использовать МРТ и КТ снимки разных людей и получать модели не уступающие по качеству. Так же рассмотрен ряд более общих моделей, в которых функционал ошибки меняется в ходе обучения. В дальнейшем эти подходы могут быть улучшены за счет динамического изменения весов на основании полученных в ходе обучения значений функционалов ошибок.

В отличии от [5], где брались сагиттальные (по оси y) срезы, мы брали аксиальные срезы. Они сильнее отличаются друг от друга (например, в верхней и нижней частях головы), чем сагиттальные срезы. Этим можно объяснить трудности генератора в преобразовании нижней части головы



Таблица 1: Для каждого эксперимента была выбрана эпоха (число в строке $\Im noxa$) с наименьшим средним MAE (MAE) на тестовой выборке. Стандартное отклонение (Std) считалось по снимкам.

Рис. 2: График МАЕ для одного снимка из тестовой выборки.

(Рис. 2). Для решения этой проблемы в будущем можно использовать 2 модели: одну для нижней части головы, другую для верхней.

В нашей работе мы не рассматривали нейронные сети, работающие с трехмерными данными. Обучение таких моделей является гораздо более трудной с точки зрения вычислений задачей. Для обучения таких моделей в будущем можно вместо срезов вырезать патчи из исходного снимка, или же использовать "объемные" срезы, состоящие из 3-5 двумерных срезов.

Список литературы

- Daniel Andreasen, Koen Van Leemput, Rasmus H. Hansen, Jon A. L. Andersen, and Jens M. Edmund. Patch-based generation of a pseudo CT from conventional MRI sequences for MRI-only radiotherapy of the brain. *Medical Physics*, 42(4):1596–1605, mar 2015.
- Xiao Han. MR-based synthetic CT generation using a deep convolutional neural network method. *Medical Physics*, 44(4):1408–1419, mar 2017.
- D. Nie, R. Trullo, C. Petitjean, S. Ruan, and D. Shen. Medical Image Synthesis with Context-Aware Generative Adversarial Networks. *ArXiv e-prints*, December 2016.
- I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative Adversarial Networks. ArXiv e-prints, June 2014.
- Jelmer M. Wolterink, Anna M. Dinkla, Mark H. F. Savenije, Peter R. Seevinck, Cornelis A. T. van den Berg, and Ivana Isgum. Deep MR to CT synthesis using unpaired data. *CoRR*, abs/1708.01155, 2017.

- Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. *CoRR*, abs/1703.10593, 2017.
- 7. Brian B Avants, Nick Tustison, and Gang Song. Advanced normalization tools (ants). *Insight j*, 2:1–35, 2009.
- 8. G. Lample, N. Zeghidour, N. Usunier, A. Bordes, L. Denoyer, and M. Ranzato. Fader Networks: Manipulating Images by Sliding Attributes. *ArXiv e-prints*, June 2017.

Ensembling Convolutional Neural Networks for Ischemic Stroke Lesion Segmentation Based on CT Images

Mikhail Goncharov^{1,2}, Maxim Pisov^{1,2}, and Mikhail Belyaev^{3,1}

¹ HTP RAS, Moscow, Russia ² MIPT, Moscow, Russia ³ Skoltech, Moscow, Russia goncharov.myu@phystech.edu, maksim.pisov@phystech.edu, m.belyaev@skoltech.ru

Abstract. The problem of segmentation of lesion tissue in acute stroke based on CT images is critically important for clinicians. In this paper we describe a method of solving this problem exploiting state-of-art convolutional neural networks, which are now widely used to segment medical images. The final prediction is obtained as a result of averaging the predictions of 3 different U-Net-based models. The method was developed within the framework of the ISLES-18 challenge and showed quite good results on challenge's dataset.

Keywords: Convolutional Networks, Stroke Lesion Segmentation

1 Introduction

Ischemic stroke occurs when the blood supply to an area of the brain is disrupted, resulting in cells' death. Defining the location of the affected area is a key part of the decision-making process in acute stroke. At present, actually, clinicians manually outline core lesions on MRI images. Comparing with MRI, CT is more preferable due to its speed, availability and lack of contraindications. However, now there are no accurate methods, either manual or automated, defining regions of damaged tissue based on CT images. That is why *Ischemic Stroke Lesion Segmentation Challenge* (ISLES-18) [1] asks for advanced data analysis techniques that could help to define these regions on perfusion CT images. In recent years deep learning methods performed well in solving a wide variety of image processing tasks [5,6], including medical ones [8,3]. In this work we propose a method which combines inferences of several convolutional neural networks in order to decrease variability of the result of the segmentation.

2 Data and problem formulation

ISLES-18 dataset contains image data for 63 stroke patients. For some patients 2 slabs are provided and in total there are 94 instances. For each instance CT

brain image and binary segmentation mask of the stroke lesion are provided. CT images have shape $5 \times 256 \times 256 \times s$ where 5 stands for a number of perfusion CT modalities and s is a number of slices, which varies from 2 up to 22. The provided ground-truth segmentation masks were manually drawn on DW-MRI images and have corresponding spatial shapes.



Fig. 1: An example of one slice of CT brain image and the corresponding segmentation mask.

Our task was to develop an automated method for segmentation stroke lesion based on CT images, which is quite challenging, taking into account the relatively small amount of data, its anisotropy, different ways of obtaining input data and target data (CT and MRI respectively).

3 Method

Since most of the images consist of 2 or 4 slices the use of 3D convolutional networks is unfeasible. That is why in our method we use 2D convolutional networks (the training and inference are performed on axial slices).

U-Net [8] is a fully-convolutional neural network architecture for segmentation, which is very popular in medical imaging. U-Net consists of a downsampling branch which alternates convolutional and pooling layers and a symmetric upsampling branch which alternates convolutional and upsampling layers. Outputs from layers of the downsampling branch are stacked with inputs of layers of the upsampling branch. Thus, the network is able to combine patterns from different scales in order to yield a more precise segmentation.

T-Net [7] is a variation of the U-Net architecture with convolutional layers inserted to the connections between the branches, which slightly increase the amount of processing at each scale.

In our method we ensemble both these architectures as well as a modification of T-Net aimed at yielding a better result according to one of three quality metrics considered in the challenge. In each network we replace simple convolutional layers with residual blocks [2] and add initial convolutional layers, which is a standard approach to improve learnability of deep networks.

Preprocessing As it turned out, there are slices in the data, on which there is almost no brain tissue, and therefore, no stroke. In order not to learn on such untypical edge cases we leave out the slices with background area greater than a fixed threshold and predict an empty mask for them; the threshold was chosen



Fig. 2: T-Net architecture

as the maximal background area among the slices with lesions in the train set. Then we crop all images to their 3-dimensional bounding boxes and rescale them to the shape 256×256 in the axial plane.

Setup In our experiments we used Adam optimizer [4] with a constant learning rate of 10^{-3} or 10^{-4} .

To test the models' performance we used group 5-fold cross validation. In order to avoid overfitting the splits were made so that all images belonging to the same patient were presented in the same fold. We used the following quality metrics: Dice coefficient, Hausdorff distance and average symmetric surface distance (from the ground truth), which were proposed by the organizers of the challenge and also precision and recall coefficients. Each metric highlights different aspects of the segmentation quality, e.g. Hausdorff metric is very sensitive to false positives that are located far from the ground truth, whereas Dice metric is much more robust and estimates the fraction of the intersection of our prediction and the ground truth.

As loss functions we used simple binary cross entropy (BCE-loss) and also weighted binary cross entropy (WCE-loss)

$$WCE(x,y) = -\sum_{i} y_i \log(\sigma(x_i)) + w_i(1-y_i) \log(1-\sigma(x_i)),$$

where x is the network's output, y is the target, σ is the sigmoid function, which transforms network's outputs into probabilities; $w_i = 1 + 0.1d(i, y)$ weights for background pixels, d(i, y) is the distance from the *i*-th pixel to the lesioned region, i.e. this loss strongly penalizes distance from the ground truth false positives. We exploited this fact in order to train model that would yield a good result according to the Hausdorff metric.

3

Also, in order to overcome the dataset size limitations we use data augmentation techniques: random flips with respect to the sagittal plane and random rotations in the axial plane.

Models In our method we use 3 models

- U-Net with residual blocks operating on images downsampled by a factor of 4 along each dimension. During training *BCE*-loss was optimized.
- T-Net (Fig. 2). This model differs from U-Net only by the presence of additional convolutional layers between the branches, however, this change significantly influences the output, as can be seen in Fig. 3.
 Intuitively, both of the above models make relatively coarse predictions.
- A model similar to T-Net, but deeper and operating on images in the original scale (T-Net HD). Due to high resolution of the input and network's depth this model makes more localized predictions. During training WCE-loss was optimized, which, as expected, results in improvement of the Hausdorff score by 10% (comparing to the same model but trained with simple BCE-loss, see Tab. 1).

The final segmentation, which is equivalent to a pixelwise binary classification, is performed according to the averaging decision rule

$$z_i = \mathbb{1}(\frac{\sigma(x_i^1) + \sigma(x_i^2) + \sigma(x_i^3)}{3} \ge 0.5)$$

where x^1, x^2, x^3 are the outputs of the U-Net, T-Net and T-Net HD respectively.

4 Results

Table 1 shows the models' performances according to considered metrics. As seen, averaging improves both Dice score and Hausdorff distance, which is expected, as averaging unites confident predictions and effectively filters out random outliers.

Model	Mean Dice	Mean HD	Mean ASSD	Mean Precision	Mean Recall
U-Net	0.52 ± 0.01	29.65 ± 0.49	3.19 ± 0.26	0.60 ± 0.01	0.53 ± 0.01
T-Net	0.51 ± 0.01	28.45 ± 0.34	3.36 ± 0.27	0.59 ± 0.01	0.52 ± 0.01
T-Net HD	0.51 ± 0.01	24.75 ± 0.01	2.51 ± 0.27	0.64 ± 0.01	0.50 ± 0.01
T-Net HD-BCE	0.51 ± 0.01	27.58 ± 1.32	2.80 ± 0.38	0.63 ± 0.01	0.49 ± 0.01
Averaging	0.53 ± 0.01	23.29 ± 0.02	2.44 ± 0.32	0.65 ± 0.01	0.52 ± 0.01

Table 1: Segmentation results. T-Net HD-BCE denotes the same model as T-Net HD but trained with simple BCE.

4



Fig. 3: An example of the final segmentation. From left to right: predicted by U-Net, T-Net, T-Net HD and averaged probability maps, binary segmentation mask and the ground truth.

5 Conclusion

From the very beginning of the challenge to the moment of writing this paper, we tested a large number of models with different setups and, finally, averaged predictions of the best of them. In most cases this method makes visually adequate predictions and yields quite good quality (average values of the quality metrics suffer due to several complex cases, our method can not cope with). Now we can not say with certainty whether the method is applicable in practice, since it could be possible to identify a particular group of people for whom the method yields acceptable quality or it can be useful in another formulation of the problem, for example, in determining the volume of the affected area. Unfortunately now the competition is not over yet and we can not compare our results with other participants, but it may well be impossible to achieve a significantly better result using deep learning methods. In the future work, in order to improve the method one can try to get a final prediction by ensembling more models and applying more sophisticated postprocessing techniques.

References

- 1. Web page of the challenge: http://www.isles-challenge.org/
- He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition (2015)
- Kamnitsas, K., Ferrante, E., Parisot, S., Ledig, C., Nori, A.V., Criminisi, A., Rueckert, D., Glocker, B.: Deepmedic for brain tumor segmentation. In: International Workshop on Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. pp. 138–149. Springer (2016)
- Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
- Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems. pp. 1097–1105 (2012)
- Long, J., E., S., Darrell, T.: Fully convolutional networks for semantic segmentation arXiv:1411.4038 (2014)
- Makarchuk, G., Kondratenko, V., Pisov, M., Pimkin, A., Krivov, E., Belyaev, M.: Ensembling neural networks for digital pathology images classification and segmentation. arXiv preprint arXiv:1802.00947 (2018)
- 8. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation (2015)

Предсказание дебюта болезни Альцгеймера методами обучения представлений и анализа выживаемости

Домрачев Алексей^{1,2}, Шмулев Ярослав¹, Беляев Михаил^{1,3}

¹ Институт проблем передачи информации им. А.А. Харкевича РАН
² Московский физико-технический институт (государственный университет)
³ Сколковский институт науки и технологий domrachev.av@phystech.edu

Аннотация Задача прогноза месяца, в котором пациент подвергнулся болезни Альцгеймера носит важный прикладной характер. Эта проблема в большинстве работ рассматривается как задача классификации, либо регрессии. В данном исследовании сравниваются эти подходы. Классификационный алгоритм на данный момент показывает более низкое качество решения относительно применения регрессии, но в работе предлагаются шаги для усовершенствования показанного метода. В настоящей статье также показано, что информация, полученная из МРТ снимков, помогает повысить качество долгосрочного прогноза.

1 Введение

Болезнь Альцгеймера (БА) — это одна из наиболее распространенных форм деменции, на которую приходится 60–80% всех случаев. На данный момент она считается неизлечимым заболеванием и только применение специальных методов помогают отсрочить наступление недуга или замедлить его прогрессию, поэтому ранний прогноз развития данной болезни может быть востребованным и важным шагом для того, чтобы повысить качество жизни возможных пациентов [7].

При исследованиях болезни Альцгеймера собираются разнообразные типы данных, включая:

- когнитивные тесты и различные клинические показатели такие как пульс, возраст и т.п;
- магнитно-резонансную томографию (MPT).

Для диагностики деменции (в том числе БА) основное значение имеют специализированные когнитивные тесты, однако согласно некоторым исследованиям структурные изменения, видимые на МРТ, могут быть предиктором развития болезни [8].

2 Данные

Для экспериментов мы использовали набор данных, полученный из Alzheimer's Disease Neuroimaging Initiative (ADNI) [1]. ADNI — это глобальное, продолжительное исследование, предназначенное для сбора и анализа клинических, генетических и биохимических маркеров пациентов из группы риска болезни Альцгеймера.

Первоначальной целью данного исследования являлось раннее выявление и отслеживание БА. Оно началось в 2004 году и в настоящий момент насчитывает более 1800 участников. Из этого набора данных мы отобрали только пациентов с диагнозом нормального контроля (NC), умеренных когнитивных нарушений (MCI) и болезни Альцгеймера (AD).



Рис. 1. Гистограмма распределения значений месяца дебюта БА в рассматриваемой выборке

Для каждого пациента было проведено несколько сеансов с периодом 6 месяцев. Мы рассматривали только те случаи, для которых доступны как MPT-изображения, так и клинические данные, поэтому для некоторых пациентов указанная выше периодичность нарушилась. Также мы не рассматривали субъектов, у которых болезнь дебютировали до их базового сеанса, так как наша задача будет заключаться в предсказании месяца дебюта недуга, а для таких участников узнать его возможным не представляется. Распределение месяца дебюта, представлено на рис. 1.

Клинические данные. Полученный из ADNI набор данных содержит разнообразную информацию об исследуемых пациентах, включая демографию, результаты физиологических, неврологических и когнитивных тестов. Мы также добавили информацию о генетике и различных анализах пациента. Полный список атрибутов доступен на официальном сайте ADNI [1].

Данные, полученные из нейровизуализации. В данной работе мы используем низкоразмерное векторное представление MPT снимков головного мозга, полученное в результате решения задачи вложения, как было предложено в [3]. Данное векторное представление имеет размерность 64, что намного упрощает работу с MPT снимками и, в тоже время, сохраняет основные свойства исходного трехмерного изображения.

3 Постановка задачи

Целью нашего исследования является прогноз развития болезни с полугодовым периодом. Данное ограничение возникает из-за периода сбора данных о пациентах в ADNI. Оно не влияет на качество решения задачи, потому что предполагается, что болезнь Альцгеймера развивается на протяжении многих лет и за 6 месяцев не происходит кардинальных изменений. Одной из задач является проверка того, что предложенный метод дает приемлемую ошибку. Также мы проведем данный эксперимент на нескольких наборах признаков:

- 1. Клинические данные;
- 2. Информация, извлеченная из МРТ снимков;
- 3. Оба набора.

и проверим несут ли в себе представления MPT какую-либо полезную информацию.

4 Схема эксперимента

4.1 Предобработка данных

В ADNI для каждого пациента снимались значения маркеров каждые 6 месяцев. Так как мы рассматривали только те сессии, в которых присутствуют и MPT, и клинические данные, а также возможны пропуски из-за неявки на обследование, то для некоторых пациентов периодичность будет нарушена. Например, для определенного пациента известны только данные на момент 0, 6, 12 и 24 месяцев, а на момент 18 месяца данные отсутствуют. Для восстановления периодичности, мы предприняли следующее: пропущенные полугодовые периоды заполнили путем копирования данных из предыдущего известного обследования этого пациента и ввели новый признак сору label. Он равен 0, если данные были скопированы, а если они присутствовали в наборе данных сразу, то он равен 1. Также введем признак time. Его значение равно месяцу, в котором были сняты данные о каждом из сеансов каждого пациента. Начальным моментом времени для пациента, является его базовый сеанс, т.е. значение признака time в этот момент — 0. Далее мы разделили выборку на обучающую и тестовую в соотношении близком к 8:2, причем сеансы одного и того же пациента не могут находится и в обучающей и в тестовой выборке. Далее мы получаем низкоразмерное векторное представление МРТ снимков головного мозга, полученное в результате решения задачи вложения, как было предложено в [3], обучая модель на нашей обучающей выборке, чтобы не подсказывать значение целевой метки через вложение. Необходимо упомянуть, что мы не рассматривали сеансы пациентов, которые следуют после дебюта болезни, потому что в реальной ситуации, когда врач будет прогнозировать месяц дебюта, они, конечно, не будут известны.

Чтобы принять во внимание все ситуации, с которыми могут столкнуться медики при предсказании, предпримем следующее. Если у пациента было 3 сеанса, то есть в 0, 6 и 12 месяцы. Введем переменную time_real и будем делать для субъекта предсказание трижды: основываясь сначала только на данных о базовом сеансе (time_real = 0), далее о первых двух (time_real = 6) и о всех (time_real = 12). Таким образом мы сможем понять как ведут себя наши модели в разных ситуациях и добиться более точной валидации.

4.2 Регрессионный алгоритм

Рассматриваемую задачу можно решать как регрессионную, в которой целевой меткой является количество месяцев с момента рассматриваемого обследования пациента, на момент окончания которых БА дебютировала. Кратко опишем этот алгоритм:

- Сначала уберем из датасета всех пациентов, у которых болезнь еще не дебютировала, так как мы не можем выставить им значение целевой метки.
- Для каждого сеанса целевая метка будет равна разности месяца дебюта болезни и месяца обследования, то есть если болезнь дебютировала в 12 месяце, а мы рассматриваем сеанс на момент 6 месяца, то целевая метка для этого обследования равна 6.
- 3. С помощью метода GroupKFold из библиотеки sklearn [4], в котором группами будут служить ID пациентов, а число разбиений будет равно 5, разделим отложенную раннее выборку на части для обучения и валидации. С помощью жадного поиска по сетке выберем наилучшие гиперпараметры для обучения.
- Обучим модель и сделаем предсказание для каждого посещения каждого пациента из тестовой выборки.
- Для каждого сеанса сложим месяц, в котором он был проведен, и значение целевой метки, чтобы получить месяц прогнозируемого дебюта относительно базового обследования.
- Для каждого субъекта итоговым прогнозом выберем среднее значение прогнозов от всех сеансов, далее округлим это значение до ближайшего числа кратного шести.

Классификационный алгоритм 4.3

Теперь покажем предлагаемый нами способ решения этой задачи. Мы воспользуемся методом описанным в [2] и чуть усовершенствуем его. Сначала выпишем алгоритм работы, а ниже опишем некоторые шаги подробнее:

- 1. Введем новую метку АD. Для каждого сеанса, на момент которого пациент здоров, т.е. имеет диагноз NC или MCI, она равна 0, а в остальных случаях соответственно 1.
- 2. Для каждого пациента и для каждого значения time_real у рассматриваемого субъекта сделаем следующее. Будем копировать данные о сеансе, месяц которого равен time_real, меняя при этом месяц сеанса на 6, до тех пор, пока месяц скопированного сеанса не будет равен максимальному time_real из обучающей выборки. Для всех скопированных сеансов выставим значение сору label равное 0. Это сделано для того, чтобы в последствии строить функцию вероятности заболевания, что будет объяснено далее.
- 3. Выполним шаг 3 из регрессионного алгоритма, а далее разделим валидационную выборку пополам, одну из половин будем использовать для валидации, а вторую для калибровки. С помощью жадного поиска по сетке подберем наилучшие гиперпараметры.
- 4. На полученном наборе данных при выбранных параметрах, обучим классификатор, целевой меткой будет являться AD, а функцией потерь взвешенный log loss, так как наше решение строится на прогнозе вероятностей, что будет показано в последующих пунктах.
- 5. Проведем калибровку классификатора, с помощью метода Platt's scalling, который описан в статье [6].
- 6. Произведем предсказание вероятности принадлежности классу 1 для всех сеансов всех пациентов из тестовой выборки.
- 7. Построим функцию вероятности заболевания в месяц k, при условии того, что пациент здоров до месяца k, путем несложных математических преобразований, описанных в пункте 4.6.



4.4 Баланс классов

После преобразований мы построили гистограмму распределения классов (рис. 2), из которой видно, что они не сбалансированы. Поэтому функция, которую нам следует минимизировать это взвешенный log loss:

$$-\frac{1}{N}\sum_{i=1}^{N} y_i \log(p_i) + \frac{class_1}{class_0} (1-y_i) \log(1-p_i),$$

Рис. 2. Гистограмма распределение значений целевой метки в обучаю- где n — размер обучающей выборки, щей выборке

 y_i — значения целевой метки, а p_i — прогноз, $class_1$ и $class_0$ количество представителей классов 1 и 0 из обучающей выборки соответственно.

4.5 Обучение и калибровка

Авторы статьи [5] показали, что прогнозы решающих деревьев не откалиброваны, то есть получение прогноза равного 0.8 для определенного элемента, не эквивалентно утверждению о том, что похожие на этот элемент векторы с вероятностью 80% принадлежат 1 классу. Так как наш метод строится на прогнозируемых вероятностях, что будет объяснено в следующем пункте, мы разделили обучающую выборку в отношении 8 к 1 к 1, бо́льшая часть использовалась для обучения, две следующие – для валидации во время обучения и калибровки.

Мы производили калибровку с помощью метода *Platt's scalling*, который описан в статье [6] и заключается в следующем: полученные прогнозы для калибровочной выборки подаются как значения признаков методу логистической регрессии, а целевая метка сохраняется, после обучения алгоритма логистической регрессии мы получаем модель откалиброванного классификатора.

4.6 Прогноз

Обозначим функцию полученной модели как h(X, t). Она принимает на вход вектор описывающий состояние пациента и значение признака time, которое тут вынесено для удобства. С помощью данной функции для тестовой выборки мы получили вероятность того, что человек был здоров до месяца t, а заболел в этом месяце таким образом:

$$\mathbb{P}(Y = t|X) = h(X, t) \prod_{i=1}^{t-1} (1 - h(X, i)),$$

Далее для каждого пациента из тестовой выборки была составлена функция вероятности заболевания от первого месяца до последнего известного нам из обучающей выборки. Взяв argmax от функции мы получили прогноз месяца заболевания.

5 Результаты

Мы получили три различных разделения ADNI на обучение и тест и для них произвели эксперимент, показанный выше, на трех наборах данных:

- 1. Клинические данные;
- 2. Признаки, извлеченные из МРТ;
- 3. Объединение клинических данных и признаков из МРТ.

В таблице 1 представлены полученные после экспериментов значения метрик для оценки классификации состояния пациента на момент каждого сеанса. В таблице 2 указанны, соответственно, значения метрик оценивающие прогноз месяца дебюта болезни. Можно заметить, что небольшие изменения в результате работы классификатора сильно влияют на качество предсказания месяца дебюта. Также в обеих таблицах видно, что добавление представления МРТ в пространство признаков к клиническим данным немного улучшает качество предсказания, как для регрессионной, так и для классификационной модели. Для любых рассматриваемых наборов признаков регрессионный метод показывает более высокое качество, чем алгоритм классификации.

Данные / Метод ACC ROC AUC FNR FPR Все данные / XGBoost Classifier $|.916 \pm .003 |.969 \pm .001$ 14.1 ± 2.6 $6.8 \pm .9$ Клинические данные / XGBoost $.914 \pm .004$ $.968 \pm .002$ 15.0 ± 2.0 Classifier 6.8 ± 1.0 MPT / XGBoost Classifier $.849 \pm .014$ $.890 \pm .007$ 40.3 ± 5.2 8.2 ± 2.9

Таблица 1. Результаты оценивающие классификацию состояния пациента на момент каждого сеанса

Таблица 2. Результаты оценивающие прогноз месяца дебюта болезни

Данные / Метод	RMSE	R2
Все данные / XGBoost Classifier	28.6 ± 4.6	$.02 \pm .12$
Все данные / XGBoost Regressor	20.8 ± 3.4	$.48 \pm .08$
Клинические данные / XGBoost Classifier	29.0 ± 4.4	$.00 \pm .11$
Клинические данные / XGBoost Regressor	21.0 ± 1.8	$.47 \pm .03$
MPT / XGBoost Classifier	36.2 ± 5.9	$60 \pm .51$
MPT / XGBoost Regressor	25.0 ± 4.8	$.25 \pm .17$

Также график (рис. 3) показывает изменение значения RMSE от максимального месяца заболевания в тестовой выборке. Этот график построен следующим образом: по значению переменной month_bound отделяются только те пациенты из тестовой выборки, месяц заболевания которых меньше значения переменной, для данной выборки делается предсказание и подсчитывается значение RMSE, полученная точка наносится на график. Необходимо заметить, что MPT дает небольшой выигрыш в предсказаниях почти на всей временной прямой.



Рис. 3. График зависимости RMSE от максимального месяца в выборке для каждого метода и набора данных



Рис. 4. График распределения предсказанных и фактических месяцев дебюта при обучении регрессионного алгоритма на всех признаках

График (рис. 4) отображает распределение предсказанных и фактических месяцев дебюта при прогнозе с помощью XGBoost Regressor, который обучен на всех признаках. Размер точки зависит от количества элементов попавших в нее. Наши предсказания до 42 месяца дебюта скапливаются в окрестности красной прямой, которая отражает работу идеальной модели. Отсюда можем сделать вывод о том, что до 42 месяца наша модель работает намного лучше, чем на более дальних промежутках. Это может быть связанно с уменьшением количества образцов в выборке при увеличении месяца дебюта, что видно из рис. 1.

6 Выводы и заключение

В данной работе был предложен метод прогноза месяца дебюта болезни Альцгеймера с помощью классификационных моделей. Такой подход помог извлекать полезную информацию из данных здоровых пациентов и учитывать историю болезни субъектов при прогнозе, но не оправдал ожиданий и показал результаты не достигающие качества регрессионных алгоритмов. Мы считаем, что это связано с большим объемом скопированных сеансов с изменением месяца, так как из-за этого распределение нарушается, и модель хуже понимает как развивается болезнь. Поэтому логичным шагом для развития данного метода будет создание способа генерировать результаты последующих сеансов на основе известных.

Интересно заметить следующее: не смотря на то, что прогнозы полученные на основе снимков МРТ имеют не лучшее качество, они способны привнести полезную информацию и немного улучшить результат прогноза при использовании их вместе с клиническими признаками. Эффект от представления МРТ виден при долгосрочном прогнозе.

Список литературы

- Alzheimer's Disease Neuroimaging Initiative: http://adni.loni.usc.edu/ (2003), [Online; accessed 22-May-2018]
- 2. Ben Khun: https://www.benkuhn.net/survival-trees (2014)
- Shmulev, Y., Belyaev, M.: Predicting conversion of mild cognitive impairments to alzheimer's disease and exploring impact of neuroimaging. In: Beyond MIC: integrating imaging and non-imaging modalities for healthcare challenges. p. 8p (2018)
- $4. Sklearn: http://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GroupKFold.html = 1.00\% (MeV) = 1.0\% (MeV) = 1.0$
- 5. Zadrozny, B., Elkan, C.: Obtaining calibrated probability estimates from decision trees and naive bayesian classifiers. In: Icml. vol. 1, pp. 609–616. Citeseer (2001)
- Zadrozny, B., Elkan, C.: Transforming classifier scores into accurate multiclass probability estimates. In: Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 694–699. ACM (2002)
- 7. Дудук, : К вопросу о диагностике болезни Альцгеймера. Журнал Гродненского государственного медицинского университета (1 (25)) (2009)
- Емелин, : Новые критерии диагностики болезни Альцгеймера. Неврология, нейропсихиатрия, психосоматика (4) (2011)

Прогнозирование клинически значимого ухудшения функционального статуса у носителей мутации болезни Гентингтона на основании данных наблюдательного исследования Enroll-HD

Игорь Медведев^{1,2}, Артём Борзов^{1,2}, Юрий Селивёрстов⁴, Михаил Беляев^{3,1}

¹ Институт проблем передачи информации им. А.А.Харкевича РАН
 ² Московский физико-технический институт (государственный университет)
 ³ Сколковский институт науки и технологий

⁴ Научный центр неврологии (ФГБНУ НЦН) medvedev.is@phystech.edu

Аннотация В работе представлено решение задачи предсказания прогрессирования болезни Гентингтона по данным о состоянии больного на текущий момент. Используемый нами алгоритм решения правильно прогнозирует 69% людей, состояние которых ухудшится, и 86% людей, которые останутся на том же уровне. В ходе исследования было выявлено, что для предсказания достаточно лишь признака TFC score — важнейшей характеристики, оценивающей стадию болезни Гентингтона. Также были исследованы закономерности, на которые опирается алгоритм при прогнозировании.

1 Введение

Болезнь Гентингтона (БГ) — наследственное нейродегенеративное заболевание, развивающееся вследствие увеличения числа CAG-повторов в первом экзоне гена гентингтина [1]. Возраст начала БГ варьирует в зависимости от выраженности CAG-экспансии, а само заболевание характеризуется неуклонным прогрессированием с постепенной утратой функциональной независимости человека по мере нарастания у него выраженности двигательных, когнитивных и психопатологических нарушений.

БГ является относительно редким заболеванием, в связи с чем недостаточно внимания уделяется вопросам помощи таким пациентам и членам их семей. Между тем, нарастающая инвалидизация человека вследствие прогрессирования БГ требует адаптации или смены условий пребывания пациента, а также обеспечения должного ухода за ним. Эти обстоятельства нередко сопряжены с затратами времени и материальных ресурсов семьи, что требует долгосрочного планирования. В этой связи представляется важным заблаговременное прогнозирование ухудшения функционального статуса носителей мутации БГ. В нашей работе мы поставили задачу разработать алгоритм прогнозирования клинически значимого ухудшения функционального статуса носителей мутации БГ на основании текущих демографических и клинических характеристик человека. С этой целью мы обратились к данным крупнейшего на сегодняшний день международного многоцентрового проспективного наблюдательного исследования БГ — Enroll-HD [2].

2 Постановка задачи

Обследование в рамках Enroll-HD подразумевает всестороннюю оценку демографических и клинических (двигательные, когнитивные, психопатологические и иные нарушения) характеристик человека. Одним из основных инструментов оценки функционального статуса носителей мутации БГ является подраздел TFC (Total Functional Capacity) шкалы UHDRS [3]. Оценка по TFC характеризует способности человека к трудовой деятельности, выполнению финансовых и домашних обязанностей и самообслуживанию. Эта величина может принимать значения от 0 до 13 баллов.

Выделяют 5 стадий БГ, которые оцениваются по величине балла TFC. На стадиях 1 и 2 балл TFC принимает значения в интервалах 13-11 и 10-7 соответственно. Это ранние стадии, на которых у человека еще сохраняется относительная функциональная независимость. На стадиях 3, 4 и 5 балл TFC принимает значения 6-3, 2-1 и 0 соответственно. Переход в эти стадии сопряжен с прогрессирующей утратой человеком способности жить самостоятельно и необходимостью обеспечения специального ухода. В этой связи представляется целесообразным прогнозирование перехода человека из стадий 1 и 2 к стадиям 3, 4 и 5.

Учитывая, что в Enroll-HD участники проходят регулярное обследование с частотой около раза в год, мы поставили задачу спрогнозировать на основании данных текущего визита человека его функциональный статус на момент следующего визита. При этом мы также решили выяснить, какие признаки являются индикаторами изменения ухудшения функционального состояния пациента.

Опишем формально нашу задачу. Условимся понимать под группой 1 людей, у которых TFC не меньше 7, а под группой 2 — тех, у кого он меньше 7. Обозначим людей из первой группы, которые к следующему визиту останутся в ней, меткой 0, а тех, кто перейдет из первой во вторую, меткой 1. Наша задача состоит в том, чтобы научиться предсказывать перейдет ли человек к следующему визиту из группы 1 в группу 2 или нет, основываясь на данных текущего посещения.

3 Данные

Для экспериментов мы использовали данные исследования Enroll-HD, включающего информацию о более чем 8700 человек. В рамках Enroll-HD происходит систематический сбор данных как о носителях мутации БГ, так и о людях без этой мутации, то есть контрольных субъектах.

Для формирования выборки, которую мы использовали в нашей работе, мы выбрали людей, у которых доступно для анализа более одного посещения. Мы отбросили те визиты, где не зафиксировано значение TFC, так как мы использовали эту характеристику, чтобы присвоить пациентам метку класса. Далее из оставшихся субъектов мы выбрали тех, кто находится в группе 1. После этого мы отобрали носителей мутации БГ, то есть тех, у кого число CAG-повторов в гене гентингтина составляет 36 и более. Если у человека были посещения с меткой 1, то мы выбирали первое из них. Если же все посещения имели метку 0, то мы выбирали первое из них. Если из них. В конце концов, наша выборка составила 2848 человек. При этом баланс классов таков, что доли людей с метками 0 и 1 составляют 87.3% и 12.7% соответственно. Необходимо отметить, что данные содержат значительное количество категориальных признаков и пропущенных значений, что усложняет работу с ними.

Из всех представленных признаков мы выбрали те, которые используются в работе [4], так как они в наибольшей степени характеризуют клиническое и функциональное состояние носителя мутации БГ.

4 Методы

В качестве базового алгоритма предсказания мы выбрали градиентный бустинг над решающими деревьями. В нашей работе была использована его реализация в библиотеке XGBoost [5], [6]. Причина, по которой мы выбрали этот подход, состоит в том, что данная модель способна работать с категориалными признаками, а также с пропущенными значениями. Параметры для него подбираются автоматически при помощи функции GridSearch из библиотеки Scikit-Learn [7].

Для оценки качества алгоритма мы выбрали precision score, f1-score, sensitivity, specificity. Данные метрики являются стандартным выбором в задачах бинарной классификации на несбалансированных выборках.

Так как размер нашей выборки мал, а классы несбалансированы, оценка одного классификатора является неустойчивой. По этой приине мы используем метод агрегации алгоритмов – бэггинг. В данном разделе мы кратко опишем те методы, которые использовали в нашем исследовании.

4.1 Бэггинг

Идея бэггинга алгоритма подробно изложенна в [8]. Мы постараемся сжато изложить принципы его работы при помощи псевдокода Algorithm 1.

4.2 Модификация

Недостатком описанного метода является то, что необходимо задавать число итераций заранее. Часто бывает непонятно, как нужно подбирать этот

Algorithm 1 Алгоритм бэггинг

- 1: n- подаваемое на вход число итераций алгоритма.
- 2: *data* наш набор данных.
- 3: person = choose_one(data) выбранный из data человек, для которого мы хотим сделать предсказание. В обучении он участие не принимает.
- 4: *labels*[] массив для сохранения меток для person в цикле.
- 5: for i = 0 to n do
- 6: $cur_data = get_random_part(data, 4/5)$ случайным образом выбираем подвыборку фиксированного размера рамера 4/5 от общего.
- 7: best_params = gridsearch(cur_dataf1 score) автоматический подбор параметров классификатора с оптимизацией f1-score.
- 8: *label = predict(person, best_params)* предсказание на наилучших параметрах.
- 9: labels[i] = label сохраняем метку, предсказанную на текущей итерации. 10: end for

11: prob = sum(labels)/n — искомая вероятность принадлежности person к классу.

параметр, чтобы произвести устойчивую оценку вероятностей.

Чтобы решить эту проблему мы воспользуемся идеей модификации бэггинга, которая предложена в [9]. Она основывается на том, что функция вероятности, описанная выше, сходится по параметру n. Представим ее краткое описание.

Обозначим для фиксированного субъекта вероятность через n шагов p_n , а предсказанную на *i*-м шаге метку класса y_i . Формула для p_n имеет вид:

$$p_n = \frac{\sum_{i=1}^n y_i}{n} \tag{1}$$

Формула для p_{n+1} имеет вид:

$$p_{n+1} = \frac{\sum_{i=1}^{n+1} y_i}{n+1} = \frac{\sum_{i=1}^n y_i}{n+1} + \frac{y_{n+1}}{n+1} = p_n \frac{n}{n+1} + \frac{y_{n+1}}{n+1}$$
(2)

Разница между вероятностями на шагах n и n + 1:

$$|p_{n+1} - p_n| = \left|\frac{p_n}{n+1} - \frac{y_{n+1}}{n+1}\right| \le \left|\frac{p_n}{n+1}\right| + \left|\frac{y_{n+1}}{n+1}\right| \le \left|\frac{2}{n+1}\right| \tag{3}$$

Как видно из формулы 3, алгоритм достигает конечной точности оценки вероятности за конечное число шагов.

После модификации алгоритма бэггинг у нас отпадает необходимость заранее зададвать число итераций. Вместо этого мы задаем точность, которой нам надо достигнуть для каждого человека. Описание модифицированного алгоритма представлено в Algorithm 2.

Чтобы к началу работы модифицированной части у нас было некое значение вероятности, от которой алгоритм может отталкиваться, мы используем теплый старт. Если же отказаться от него, то модификация может работать некорректно.

Algorithm 2 Модицированный алгоритм

- 1: n подаваемое на вход число итераций теплого старта
- 2: *accuracy* подаваемое на вход значени точности, с которой должен сойтись алгоритм
- 3: *data* наш набор данных
- 4: person = choose_one(data) выбранный из data человек, за которым мы следим далее
- 5: labels[] массив для сохранения меток для person в цикле
- 6: for i = 0 to n do Теплый старт:
- 7: Такой же цикл, как в Algorithm .
- 8: end for
- 9: *cur_prob* = *sum*(*labels*)/*n* вероятность принадлежности к классу после *n* итераций теплого старта
- 10: repeat Модифицированная часть:
- 11: prev_prob = cur_prob фиксируем вероятность из прошлой итерации
- 12: Такой же цикл, как в Algorithm делаем предсказание метки для person
- 13: $cur_prob = sum(labels)/(i+1)$ вероятность на текущем шаге
- 14: **until** $prev_prob cur_prob \ge accuracy$
- 15: cur_prob искомое значение вероятности для person

5 Результаты

5.1 Предсказание на всем наборе признаков

Свои опыты мы начинаем на всех предоставленных признаках. Как было отмечено выше, на каждой итерации бэггинга мы откладываем случайным образом выборку размера 4/5 от общего для обучения алгоритма. Подбор параметров происходит на ней автоматически с помощью GridSearch на пятифолдовой кросс-валидации с максимизацией f1-score.

Метрика качеств	а Результат
F1-score	0.56 ± 0.04
Sensitivity	0.72 ± 0.10
Specificity	0.87 ± 0.02
Precsion score	0.45 ± 0.03

Таблица 1. Оценка качества алгоритма на всех принаках.

Проанализируем результаты работы, представленные в Таблице 1. Высокое значение Sensitivity, дает нам понимаение того, что мы с достаточно большой точностью предсказыеваем людей с меткой 1. А величина Specificity показывает, что мы выявляем почти всех людей с меткой 0. Таким образом, мы видим, что классификатор показывает хорошие результаты.

Значения метрики precision близко к 0.5, то есть доля верных срабатываний алгоритма. Выясним, почему эта велична имеет такое значение. Для этого визуализируем распределение вероятностей для людей с меткой 0 и меткой 1 так, чтобы наглядно оценить предсказание (см. рис. 1).



Рис. 1. Распределение вероятностей при предсказании на всех признаках.

5.2 Анализ признаков

При анализе рисунка 1 мы видим, что гистограмма имеет довольно тяжелые хвосты. Разберемся, какие признаки вляют на их появление.

Для поиска признаков, влияющих на качество классификации, введем некоторые определения. Плохо классифицируемыми людьми с меткой 0 считаем тех, у кого значение вероятности больше 0.66, хорошо классифицируемыми — со значением меньше 0.33. Для людей с меткой 1 – обратно.

Мы хотим сравнить признаки для этих групп. Чтобы выявить статистически значимые различия у хорошо и плохо классифицируемых людей, мы используем критерий Манна-Уитни [10] для непрерывных признаков и точный тест Фишера [11] для остальных. На выходе мы получаем набор значений p-value для всего набора характеристик. После этого вносим поправку Холма [12], для исправления статистической ошибки множественного тестирования.

В нашей задаче число хорошо классифицируемых людей заметно превышает число плохо классифицируемых. Поэтому нам необходимо прибегнуть к андерсэмплингу, то есть отбрасыванию людей более численной группы, чтобы в сравниваемых подвыборках было одинаковое количество человек. Сэмплирование мы делаем 10 раз случайным образом для хорошо классифицируемой группы. Для каждого раза делаем статистические тесты и усредняем результат. Таким образом, выбирая признаки, у которых p-value меньше 0.05, мы получаем набор характеристик, по которым статистически значимо отличаются группы хорошо и плохо классифицируемых людей.

Список признаков, в которых есть статистически значимые различия в сравниваемых группах, представлен в Таблице 2. Он состоит по большей ча-

Люди с меткой 0	Люди с меткой 1
tfcscore, indepscl	tfcscore, indepscl
fascore, sdmt1, sit1	sdmt1, cognitive1
cognitive1, motscore	verfct5, motscore
verfct5, verflt05	sit1, verflt05
apathy, dis_duration	
caghigh, jobclas	

Таблица 2. Признаки, по которым различаются хорошо и плохо классифицируемые люди с метками 0 и 1.

сти из характеристик, от которых зависит значение балла TFC. Рассмотрим детальней эту характеристику.

5.3 Поиск различий в признаке TFC score

Посмотрим на то, как влияет значение TFC на переход в другую группу. Для этого построим Таблицу 3 зависимости положения человека через год от текущего состояния.

	Останется в группе 1	Перейдет в группу 2
7	139	155
8	200	94
9	249	49
10	241	33
11	260	19
12	307	6
13	1090	6

Таблица 3. Зависимость функционального статуса носителя мутации БГ через год от текущего балла TFC (по вертикальной оси отложены значения TFC текущего визита, по горизотальной — состояние через год)

Из Таблицы 3 видно, что наибольшее скопление людей с меткой 1 находится у границы раздела второй и третьей стадии (балл TFC равняется 7-8).

Для того, чтобы удостоверится в этом, вручную предскажем метки и сравним их с фактическими. Тем, у кого балл TFC на данный момент меньше 9, присвоим 1, а тем, у кого больше 8, соответственно — 0. Оценим качество такого предсказания и отразим его в Таблице 4.

При сравнении метрик sensitivity и specificity в таблицах 1 и 4 становится понятно, что исходный алгоритм отмечает тех людей, у кого значение TFC близко к границе между первой и второй группами.

Метрика качества	Результат
F1-score	0.52
Sensitivity	0.69
Specificity	0.86
Precision	0.42

Таблица 4. Оценка качества алгоритма, построенного только на значении TFC

6 Заключение

С помощью базы данных Enroll-HD мы научились с достаточно высокой точностью предсказывать клинически значимое ухудшение функционального статуса носителей мутации БГ. По ходу опытов мы столкнулись с проблемой того, что данные несбалансированы и содержат множество пропусков. Несмотря на это, нам удалось получить хорошие оценки качества работы алгоритма.

Также в ходе работы было обнаружено, что для предсказания достаточно только значения TFC. Причина этого заключается в том, что та информация, которую дают другие признаки для предсказания, косвенно уже содержится в TFC. По этой причине их добавление не дает прироста в качестве классификации. Также мы выяснили, что алгоритм определяет людей с меткой 1 по их близости к границе групп 1 и 2, а именно: относит к классу 1 в основном тех, у кого значение TFC составляет 7 или 8.

Список литературы

- 1. The Huntington's Disease Collaborative Research Group. A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. Cell, 72(6):971–983, 1993.
- 2. Enroll-HD. https://www.enroll-hd.org.
- Starosta-Rubinstein S. Young A.B., Penney J.B. Pet scan investigations of huntington's disease: cerebral metabolic correlates of neurological features and functional decline. Ann. Neurol. 20(3), pages 296–-303, 1986.
- 4. Юрий Селивёрстов Юлия Додонова Георг Бернхард Ландвермайер Артём Борзов, Михаил Беляев. Предсказание развития суицидальных мыслей при болезни Гентингтона с помощью анализа данных многоцентрового наблюдательного исследования enroll-hd. 2017.
- 5. Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. <u>arXiv</u> preprint arXiv:1603.02754, 2016.
- $6. \ \overline{\rm XGBoost. \ https://github.com/dmlc/xgboost.}$
- 7. <u>Scikit-Learn</u>. http://scikit-learn.org.
- 8. Leo Breiman. Bagging predictors. Technical Report No. 421, 1994.
- Leo Breiman. Pasting small votes for classification in large databases and on-line. Machine Learning 36, pages 85–103, 1999.

- Whitney D. R. Mann H. B. On a test of whether one of two random variables is stochastically larger than the other. <u>Annals of Mathematical Statistics. — № 18</u>, pages 50–60, 1947.
- 11. R. A. Fisher. On the interpretation of 2 from contingency tables, and the calculation of p. Journal of the Royal Statistical Society 85(1), pages 87–94, 1922.
- 12. S. Holm. A simple sequentially rejective multiple test procedure. <u>Scandinavian</u> Journal of Statistics 6 (2), pages 65–70, 1979.

Построение вложений вершин графов с помощью максимизации расстояния Вассерштейна

Станислав Цепа^{1,2,3}, Константин Славнов^{1,3}, Максим Панов^{1,3}

 ¹ Сколковский институт науки и технологий, Россия
 ² МФТИ (ГУ), Долгопрудный, Россия
 ³ Институт проблем передачи информации РАН, Большой Каретный пер. 19, 127051, Москва, Россия

Аннотация В данной работе мы рассматриваем способы представления вершин графов в виде числовых признаков для дальнейшего использования в задачах машинного обучения. Мы вводим специальную функцию потерь, которая оценивает качество вложения по тому, насколько соседние вершины в графе близки в пространстве вложения и насколько остальные пары вершин далеки. На реальных графах в задаче предсказания ребер наш алгоритм показал результаты, превосходящие результаты известных алгоритмов, таких как deepwalk и hope.

Ключевые слова: Вложение графов, расстояние Вассерштейна, предсказание ребер.

1 Введение

Большинство алгоритмов машинного обучения основано на использовании признаков, являющихся числами (а также иногда категориями). Однако, далеко не все сущности, которые хотелось бы использовать в качестве признаков, можно легко перевести в числовой вид. Примерами таких объектов являются слова в тексте или вершины в графе. В таком случае возникает задача вложения – как каждой отдельной сущности сопоставить точку вещественного *d*-мерного пространства, или, проще говоря, *d* вещественных чисел, чтобы дальше можно было использовать их как признаки в задачах машинного обучения.

Мы рассматриваем задачу вложения графов, то есть определения таких точек в *d*-мерном пространстве, чтобы взаимное расположение этих точек лучшим образом отображало структуру графа. В нашей работе мы сосредоточились на самом простом и естественном варианте вложения вершин

⁴ Термин «вложение» (*embedding*) означает что мы «вкладываем» сущность, не имеющую явных числовых признаков, в числовое пространство.

 $^{^5}$ Чаще всего предполагается, что $d \ll n,$ где n – количество объектов.

графа – основанном на их локальном соседстве. От нашего алгоритма вложения мы ожидаем, что если две вершины имеют между собой ребро или же общего соседа, то они будут близко располагаться в пространстве вложения. Если же вершины не связаны между собой, то лучшим вариантом будет, если в пространстве вложения они будут ортогональны.

Ранее мы уже представляли алгоритм **5**, идея которого взята из **4**, где похожим методом классифицируются изображения. Данная работа описывает его улучшенную версию адаптированную для работы с графами.

Мы сравниваем наш алгоритм с другими известными алгоритмами вложения. В задаче предсказания ребер на реальных графах наш алгоритм показывает лучший результат.

2 Постановка задачи

В данном разделе мы вводим обозначения, используемые далее.

Пусть дан невзвешенный граф G = (V, E), n = |V|. Его матрица смежности $A \in \mathbb{R}^{n \times n}, A_{ij} \in \{0, 1\}.$

Также вводится матрица сходства $S, S_{ij} \in \mathbb{R}_+$, каждый элемент которой оценивает некое "сходство"двух вершин. Обычно, матрицу сходства получают из матрицы смежности таким образом, чтобы учитывалось соседство второго и более высших порядков. Стоит заметить, что не во всех алгоритмах возможно использовать небинарную матрицу, например, в нашем алгоритме это возможно только при некоторых усложнениях, поэтому, хоть далее мы и будем для общности везде указывать матрицу S, в основном будет подразумеваться S = A.

Алгоритм, строящий вложение графа по этим данным, обычно состоит из следующих частей:

- функция кодирования (encoder) строит вложение графа,

enc:
$$V \to \mathbb{R}^{n \times d}$$
,

далее будем обозначать вложение *i* вершины $\mathcal{E}_i = \operatorname{enc}(v_i)$, аналогично вложение всех вершин будем обозначать \mathcal{E} ;

– функция декодирования (decoder) оценивает сходство двух элементов вложения, в идеальном случае $dec(\mathcal{E}_i, \mathcal{E}_j) = S_{ij}$,

$$\operatorname{dec} \colon \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R};$$

 – функция потерь (loss) *L* оценивает, насколько хорошо функции декодирования удалось восстановить элементы исходной матрицы сходства.

Таким образом, задача вложения графа сводится к поиску подходящего функционала кодирования и выбору правильной функции потерь. После этого решается задача оптимизации (при помощи градиентного спуска или аналитически).

3 Алгоритмы

Deepwalk [3]. Для данного графа и стартовой точки на каждом шаге мы случайно выбираем соседнюю вершину и перемещаемся в нее, продолжая так определенное количество итераций. Запоминая результат, мы получаем последовательность вершин, которая называется случайным блужданием на графе. Далее каждая вершина интерпретируется как слово, а одно случайное блуждание как предложение. Используя механизм поиска скрытых представлений для слов word2vec, методы получают искомое вложение для вершин графа. В данном алгоритме переходы в любую соседнюю вершину равновероятны.

НОРЕ [2]. В алгоритме НОРЕ вложение получается за счет минимизации

$$\mathcal{L} = \|S - \mathcal{E}^s \mathcal{E}^{t^T}\|.$$

Этот алгоритм изначально создавался для ориентированных графов, поэтому в нем появляется два вложения: \mathcal{E}^s и \mathcal{E}^t , source и target соответственно. Если ребро идет только из *i* в *j*, то \mathcal{E}^s_i должно быть близко к \mathcal{E}^t_j , а \mathcal{E}^t_i наоборот далеко от \mathcal{E}^s_j .

В алгоритме НОРЕ производится сингулярное разложение матрицы S,

$$S = \sum_{i=1}^{n} \sigma_i \mathbf{v}_i^s {\mathbf{v}_j^t}^T.$$

Оптимальное вложение размерности *d* получается, если взять *d* самых больших сингулярных чисел.

$$\begin{aligned} \boldsymbol{\mathcal{E}}^s &= [\sqrt{\sigma_1} \mathbf{v}_1^s, \dots, \sqrt{\sigma_d} \mathbf{v}_d^s], \\ \boldsymbol{\mathcal{E}}^t &= [\sqrt{\sigma_1} \mathbf{v}_1^t, \dots, \sqrt{\sigma_d} \mathbf{v}_d^t]. \end{aligned}$$

Этот алгоритм примечателен тем, что в нем может использоваться любая матрица сходства, в том числе для взвешенного и ориентированного графа.

4 Функция потерь

Основная идея нашего исследования заключалась в том, чтобы найти такую статистическую характеристику вложения, оптимизируя которую (при помощи градиентного спуска), мы бы получили вложение хорошего качества. Ниже приведена финальная реализация нашего алгоритма.

4.1 Описание алгоритма

Для каждой пары вершин *u*, *v* определим, соединены ли они ребром (положительная пара) или нет (отрицательная пара). Определим

$$a_{uv} = \begin{cases} 1, & \text{если } (u, v) \in E, \\ 0, & \text{если } (u, v) \notin E. \end{cases}$$

Легко заметить что a_{uv} есть не что иное как элементы матрицы смежности (в случае невзвешенного графа). То есть положительные пары для графа G = (V, E) это все ребра $(u, v) \in E$, отрицательные – ребра $(u, v) \notin E$ соответственно.

Пусть нам известно некоторое вложение *E*. Произведем попарное вычисление расстояний между вершинами для каждого элемента вложения

$$l_{uv} = \det(\mathcal{E}_u, \mathcal{E}_v).$$

В качестве функции dec решено было выбрать корреляцию

$$\det(\mathcal{E}_i, \mathcal{E}_j) = \frac{\mathcal{E}_i^T \mathcal{E}_j}{|\mathcal{E}_i||\mathcal{E}_j|}.$$
(1)

В этом случа
е $l_{uv} \in [-1,1],$ что будет удобно для дальнейшего использования.

И наконец, разобъем подсчитанные l_{uv} на два множества в соответствии с тем к какой паре, положительной или отрицательной, относится (u, v).

$$L^{+}(\mathcal{E}) = \{l_{uv} = \operatorname{dec}(\mathcal{E}_u, \mathcal{E}_v) \mid a_{uv} = 1\},\$$

$$L^{-}(\mathcal{E}) = \{l_{uv} = \operatorname{dec}(\mathcal{E}_u, \mathcal{E}_v) \mid a_{uv} = 0\}.$$

В итоге мы получили два множества значений попарных корреляций: $L^+(\mathcal{E})$ и $L^-(\mathcal{E})$.

Наш подход основан на эвристике, что вложение будет хорошим, если значения внутри $L^+(\mathcal{E})$ будут максимально близки к 1, а внутри $L^-(\mathcal{E})$ максимально близки к 0^{6} .

Чтобы формализовать эту эвристику, мы прибегаем к следующему подходу: для множеств $L^+(\mathcal{E})$ и $L^-(\mathcal{E})$ построим две плотностных гистограммы при помощи методов непараметрической оценки плотности: $H^+(\mathcal{E})$, $H^-(\mathcal{E})$. Поскольку в качестве функции dec использована корреляция, то значения в $H^+(\mathcal{E})$ и $H^-(\mathcal{E})$ находятся в отрезке [-1,1]. Поэтому их легко представить как две гистограммы с одинаковым количеством бинов *bin*. Построение плотностных распределений на основе множеств $L^+(\mathcal{E})$ и $L^+(\mathcal{E})$ подробно описано в работе [4], ключевым моментом там является использование непрямоугольного (треугольного или гауссовского) ядра при построении гистограммы. Как показано в работах [4], [5], качество вложения почти

⁶ Мы выбираем не -1 (минимальное значение корреляции), а 0 потому что стремимся получить ортогональные вложения для отрицательных пар.
не зависит от значения bin и ширины ядра, в нашей работе было выбрано bin = 64.

Определим нашу функцию потерь как

$$\mathcal{L}(\mathcal{E}) = D(H^{-}(\mathcal{E}), H^{+}(\mathcal{E})),$$

где *D* – функция вычисления расстояния между распределениями (*divergence*). Отличие нового алгоритма от [4] и нашей предыдущей работы [5] в том, что в качестве *D* мы используем усовершенствованное расстояние Вассерштейна [1] для одномерного случая.

Одномерное расстояние Вассерштейна (также называется earth mover distance, EMD) определяется как

$$EMD(H_1, H_2) = \sum_{i=1}^{bin} |\varphi_i|, \text{ где } \varphi_i = \sum_{j=1}^{i} \left(\frac{H_{1j}}{\|H_1\|_1} - \frac{H_{2j}}{\|H_2\|_1} \right).$$
(2)

Однако, в нашем случае такая формула не подходит, поскольку $EMD(H_1, H_2) = EMD(H_2, H_1)$. Из-за этого иногда градиентный спуск попадал в локальный минимум когда H^+ левее H^- . Такая ситуация не симметрична, а наоборот, антисимметрична. Оказывается, если отказаться в (2) от модуля, то новая метрика как раз будет антисимметричной.

$$EMD_{asym}(H_1, H_2) = \sum_{i=1}^{bin} \varphi_i, \text{ где } \varphi_i = \sum_{j=1}^{i} \left(\frac{H_{1j}}{\|H_1\|_1} - \frac{H_{2j}}{\|H_2\|_1} \right), \quad (3)$$
$$EMD_{asym}(H_1, H_2) = -EMD_{asym}(H_2, H_1).$$

Расстояние Вассерштейна было выбрано потому, что его оптимизация не только уменьшает пересечение между двумя распределениями, но и стремится максимально «раздвинуть» их.

У этого свойства EMD есть и недостаток: H^- слишком сильно «уезжает» влево к значению -1. Это означает, что отрицательные пары вершин будут сильно отрицательно коррелировать, и в итоге все вложение займет подпространство очень маленькой размерности. В оптимальном вложении корреляции между отрицательными парами должны быть близки к 0. Чтобы обойти эту проблему мы решили отбрасывать из H^- все значения с x < 0. Таким образом в подсчете расстояния Вассерштейна у нас участвовала плотностная гистограмма $H^-_{cut}(\mathcal{E})$, получаемая из гистограммы $H^-(\mathcal{E})$ отбрасыванием части меньше нуля. В итоге к нулю смещались только те значения, которые были больше нуля, а остальные оставались зафиксированы.

Итак,

$$\mathcal{L}(\mathcal{E}) = D(H_{cut}^{-}(\mathcal{E}), H^{+}(\mathcal{E})) = EMD_{asym}(H_{cut}^{-}(\mathcal{E}), H^{+}(\mathcal{E})).$$
(4)

Видно, что функция \mathcal{L} для заданного графа зависит только от \mathcal{E} . Чтобы построить задачу оптимизации, запишем соотношение связывающее \mathcal{E} с оптимизируемыми параметрами. Случайным образом инициализируем матрицы W и b и вычислим

$$\mathcal{E} = \mathbf{W}S + \mathbf{b}.$$

В итоге

$$\mathcal{L}(\mathcal{E}) = \mathcal{L}(\mathbf{W}, \mathbf{b})$$

Запишем итоговую задачу оптимизации

$$\mathbf{W}_{opt}, \mathbf{b}_{opt} = \operatorname*{arg\,min}_{\mathbf{W}, \mathbf{b}} \mathcal{L}(\mathbf{W}, \mathbf{b}), \tag{5}$$
$$\mathcal{E}_{opt} = \mathbf{W}_{opt} S + \mathbf{b}_{opt}.$$

5 Эксперименты

В данном разделе приведено экспериментальное исследование нашего алгоритма. Оптимизация (5) производилась при помощи библиотеки Tensorflow методом стохастического градиентного спуска. Матрицы **W** и **b** инициализировались случайными весами и были заданы как параметры оптимизации.

Также применялось несколько улучшений алгоритма. Например, на каждом шаге оптимизации вычисление градиентов производилось на случайном подмножестве вершин, то есть изменялась только часть матриц W и b, этот метод является аналоигей «батчинга» (batching). К сожалению, «батчинг» не подходит для графов размером $|V| > 10^5$, поскольку становится очень маловероятно, что вершины в одном «батче» будут связаны, в таком случае гистограмма H^+ не содержит ни одного элемента. Также учитывался тот факт, что практически в любом графе $|L^-| \gg |L^+|$. В связи с этим на каждой итерации множество L^- сэмплировалось до размера L^+ . Эксперименты показали, что применение этих оптимизаций не повлияло на качество итогового вложения, но позволило значительно ускорить работу алгоритма и уменьшить потребление оперативной памяти.

Код всех экспериментов доступен на github.com/premolab/GraphEmbeddings.

5.1 Данные для экспериментов

В качестве данных для экспериментов были использованы графы размером от 100 до 4000 вершин. Все графы являются либо модельными, либо реальными графами, часто используемыми для исследования алгоритмов. Данные о графах приведены в таблице []. Отдельно стоит описать граф SBM (Stochastic Block Model). Он является модельным графом с 900 вершинами. Сперва вершины разбиваются на три группы, а затем ребра добавляются случайным образом: между вершинами одной группы – с вероятностью p_{in} , между вершинами разных групп – с вероятностью p_{out} . Значения p_{in} и p_{out} в наших экспериментах: $p_{in} = 0.1$, $p_{out} \in \{0.01, 0.03\}$.

Название	Количество вершин	Описание
American college football	115	реальный
Books about US politics	105	реальный
Facebook	4039	реальный
Stochastic Block Model	900	модельный

Таблица 1: Список графов

5.2 Описание экспериментов

Задача предсказания ребер Задача предсказания ребер (link prediction) является наиболее универсальным способом оценить качество вложения, поскольку не требует целевой переменной (например меток классов в вершинах графа). В ней вложение подсчитывается только на части ребер графа, а затем проверяется насколько хорошо по этому вложению можно предсказать пропущенные ребра.

- 1. Разбить случайным образом множество ребер E на две части E_{train} и E_{test}
- 2. построить вложение $\mathcal{E} = \mathcal{E}(V, E_{train});$
- 3. выбрать примеры ребер из множеств E_{train} и E_{test} , а также отрицательные примеры такие что $(u, v) \notin E$;
- 4. определить переменные

$$y(u,v) = \begin{cases} 1, & \text{если } (u,v) \in E, \\ 0, & \text{если } (u,v) \notin E, \end{cases}$$
$$X(u,v) = (\mathcal{E}_u, \mathcal{E}_v),$$

где (\cdot, \cdot) означает конкатенацию;

5. решить задачу обучения с учителем для данных X и целевой переменной y, подсчитанных на примерах из пункта 39.

5.3 Результаты экспериментов

Мы проводили сравнение работы нашего алгоритма с алгоритмами двух других семейств: алгоритмов на случайных блужданиях и алгоритмов с использованием матричной факторизации. Были выбраны алгоритмы показывающие лучшие результаты внутри своих групп: deepwalk и HOPE coorветственно.

 8 В нашем случае был выбран такой коэффициент разбиения, что $|E_{train}|=|E_{test}|=\frac{|E|}{2}$

⁷ Чаще всего накладывается дополнительное требование чтобы $G_{train} = (V, E_{train})$ был связным графом, многие алгоритмы вложения работают только со связными графами.

 $^{^9}$ Мы использовали логистическую регрессию, обучаясь на данных из E_{train} и подсчитывая ROC-AUC на данных из E_{test}

Задача предсказания ребер Результаты представлены на рис. П. Видно, что на всех реальных графах наш алгоритм превосходит другие. Это может быть связано со спецификой работы метрики гос-аис, используемой при оценке качества предсказания ребер. Метрика гос-аис оценивает разделимость двух классов в данном случае положительных и отрицательных пар вершин. В то же время, оптимизация (5) пытается «развести» положительные и отрицательные пары вершин как можно дальше, что и приводит к хорошей разделимости этих классов.

6 Выводы

В данной работе структурная функция потерь из нашей прошлой работы Б была улучшена и адаптирована к задаче вложения графов. В отличие от предыдущих работ, результаты экспериментов показали, что качество работы алгоритма сравнимо с качеством существующих методов, и что функция потерь, основанная на расстоянии Вассерштейна, может считаться хорошим алгоритмом для вложения. Были выделены основные направления для дальнейшей работы: улучшение масштабируемости нашего алгоритма и улучшение работы с небинарными матрицами сходства.

Сейчас время работы алгоритма сильно увеличивается при увеличении размера графа, также построение гистограмм распределений может занимать большое количество оперативной памяти. Кроме того, наш алгоритм не рассчитан на небинарные значения s_{ij} , поскольку он строится на идее, что все ребра можно разделить на два класса, где s_{ij} принимает значение 0 или 1. Были осуществлены попытки бинаризации небинарной S, например считать $s_{ij} > threshold \approx 0.1$ положительными примерами, а остальные отрицательными, но это не дало прироста в качестве работы.

Список литературы

- 1. Manuel Martinez, Makarand Tapaswi, and Rainer Stiefelhagen. A closed-form gradient for the 1d earth mover's distance for spectral deep learning on biological data.
- Mingdong Ou, Peng Cui, Jian Pei, Ziwei Zhang, and Wenwu Zhu. Asymmetric transitivity preserving graph embedding. In *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, pages 1105–1114, New York, NY, USA, 2016. ACM.
- Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. Deepwalk: Online learning of social representations. CoRR, abs/1403.6652, 2014.
- Evgeniya Ustinova and Victor Lempitsky. Learning deep embeddings with histogram loss. In Advances in Neural Information Processing Systems, pages 4170– 4178, 2016.
- 5. Максим Панов Станислав Цепа, Константин Славнов. Анализ применимости структурной функции потерь для построения вложений графов. 2014.

¹⁰ По одному из определений ROC-AUC = $P(y_i^{pred} > y_i^{pred} | y_i = 1, y_j = 0)$



Рис. 1: Link prediction

Влияние предобработки и аугментации данных на качество сегментации гиперинтенсивности белого вещества методами глубокого обучения

Борис Широких^{1,2,3} и Михаил Беляев^{1,2}

¹ Сколковский институт науки и технологий, Москва, Россия,

 $^{2}\,$ Институт проблем передачи информации РАН, Москва, Россия,

³ Московский физико-технический институт, Москва, Россия boris.shirokikh@phystech.edu, m.belyaev@skoltech.ru

Аннотация В работе решается задача сегментации трехмерных MPT снимков головного мозга человека. Наиболее популярными и эффективными подходами являются архитектуры трехмерных полносверточных нейронных сетей: U-Net и DeepMedic. Но область предобработки данных остается малоизученной, несмотря на ее важность. В работе мы демострируем влияние предварительной обработки и аугментации данных на качество сегментации. Эксперименты проводятся на данных конкурса White Matter Hyperintensivity Segmentation Challenge.

1 Введение

Задача анализа и обработки трехмерных медицинских изображений является одной из важнейших для диагностики и ведения наблюдения за множеством заболеваний. Например, анализ такого рода требуется при диагностике нейродегенеративных заболеваний и в радиохирургии. В нем используются изображения, полученные с помощью *магнитно-резонансной томо*графии (MPT). Основным этапом анализа является сегментация изображения. Сегментация, например, опухолей на MPT снимке позволяет понять, на какой стадии находится заболевание, делать соответствующие прогнозы и назначать лечение.

На данный момент основной подход к сегментации МРТ изображений - это неавтоматизированная обработка, осуществляемая специалистомврачом. Другие подходы, которые являются автоматизированными [9], [10], находятся в основном на стадии разработки и исследований, но не используются на практике. Однако, переход к автоматизированному анализу дает значительный выигрыш во времени и независимость сегментации от конкретного эксперта-врача и таких человеческих факторов, как усталость или невнимательность.

Создание такого алгоритма сопряжено с целым набором серьезных проблем. В первую очередь, алгоритмы являются неустойчивыми к вариабельности данных, вызванной различными настройками аппаратов (значения интенсивностей на MPT снимке или масшаб получаемого изображения). Такая проблема может возникать из-за недостатка данных, когда их приходится собирать с нескольких различных томографов.

Одной из существующих задач анализа MPT изображений головного мозга является сегментация *гиперинтенсивности белого вещества* (ГБВ), которое может является предвестником или следствием таких болезней как болезнь Альцгеймера, рассеянный склероз, инсульт. Подсчет объема ГБВ, точное определение его локализации и формы необходимо для построения плана лечения пациента.

В данной работе мы поставили перед собой задачу создать автоматиеский алгоритм сегментации ГБВ на основе методов машинного обучения. Более того, от алгоритма требуется устойчивость к аппаратной вариабельности, поэтому мы реализовали и исследовали различные подходы к предобработке и аугментации данных.

2 Описание задачи и данных

В работе рассматривается построение алгоритма, обрабатывающего 3D изображения, каждый воксель которых несет в себе информацию об интенсивности. На выходе должно быть размеченное изображение, где в соответствие каждому вокселю исходного изображения поставлена метка о наличии или отсутствии в нем интересующих врача структур (например, ГБВ). Другими словами, алгоритм решает задачу сегментации трехмерного изображения.

В данном случае автоматический алгоритм будет решать задачу выделения ГБВ как области или совокупности областей на трехмерном МРТ изображении. Для обучения и тестирования алгоритма мы воспользовались данными с конкурса White Matter Hyperintensivity Segmentation Challenge конференции MICCAI 2017 [4].

Данные включают в себя 3D-изображения с трех различных сканеров из трех госпиталей в Нидерландах и Сингапуре (таблица 1). Это служит отличным примером данных, в которых имеется аппаратная вариабельность. Основная особенность набора данных - снимки, полученные с различных аппаратов с разными настройками, отличаются распределением интенсивностей и разрешением изображений.

	циентов	
3 T Philips Achieva	20	(240, 240, 48)
3 T Siemens TrioTim	20	(132, 256, 83)
3 T GE Signa HDxt	20	(256, 232, 48)
•	3 T Philips Achieva 3 T Siemens TrioTim 3 T GE Signa HDxt	3 T Philips Achieva203 T Siemens TrioTim203 T GE Signa HDxt20

Таблица 1: Описание данных

Всего выборка содержит 60 пациентов с ГБВ. Для каждого пациента сделано 2 МРТ снимка модальностей Т1 и FLAIR. Также была предостав-

лена разметка заболевания - маска. Это размеченное врачами-экспертами 3D-изображение, на котором указано местоположение ГБВ (рисунок 1).



Рис. 1: Модальности FLAIR, Т1 и маска с заболеванием

3 Методы

В работе выделяются две основные составляющие алгоритма для сегментации трехмерных MPT изображений.

Ключевой задачей без всяких сомнений является выбор модели машинного обучения. Описание и комментарии по архитектурам (в основном речь пойдет о полносверточных нейронных сетях) даны ниже, в подразделе 3.1.

Не менее важным является этап предобработки и аугментации данных. Соответствующая обработка данных направлена на исправление ошибок, удаление «лишней» информации и искуственное увеличение выборки, что упрощает задачу алгоритму машинного обучения и улучшает результаты его работы. О чем пойдет речь в подразделах 3.2 и 3.3 соответственно.

3.1 Архитектуры

В работе исследовались только подходы, использующие сверточные нейросети, так как они являются state-of-the-art подходами.

Как было показано в нашей предыдущей работе [11], модели, использующие двумерные свертки или полносвязные слои, работают значительно хуже трехмерных полносверточных моделей. Поэтому для сегментации использовались две трехмерные полносверточные нейронных сети: 3D U-Net [10] и DeepMedic [9].

DeepMedic Одной из основных архитектур для исследования был выбран DeepMedic [9] как победитель конкурса по сегментации областей, пораженных ишемическим инсультом (SISS-ISLES 2015 [2]), а также показавший наилучшие результаты при сегментации опухолей головного мозга на датасетах BRATS 2015, 2016 [3].

DeepMedic использовался в экспериментах практически без модификаций. Единственное изменение в архитектуре - на каждом из слоев было поставлено больше фильтров.



Рис. 2: Архитектура DeepMedic, изображение из оригинальной статьи [9]. Слой L включает в себя Batch Normalization [8], активацию ReLU и свертку.

3D U-Net U-Net - это широко известная архитектура для сегментации изображений, впервые предложенная в работе [10]. При этом, она может быть легко модифицирована для сегментации трехмерных изображений [6]. Соответственно, в экспериментах была использована трехмерная версия U-Net.

Во всей модели, за исключением первых двух сверточных слоев, все свертки были заменены на Residual Blocks [7], так как это значительно подняло качество сегментации в наших экспериментах. Также, из-за больших размеров изображений, было незначительно уменьшено количество фильтров в слоях.

3.2 Предобработка данных

Несмотря на важность выбора архитектуры для сегментации изображений, не менее релевантной темой является предобработка данных. Зачастую, без соответствующей обработки алгоритмы машинного обучения могут показывать крайне низкие результаты. Большинство задач, в которых применялись описанные выше архитектуры, решались на данных, в которых отсутствует инструментальная вариабельность, поэтому тема осталась мало изученой.

Удаление черепа с MPT снимка Одним из основных подходов является удаление черепа с MPT снимка. Соответствующая этому операция поиска маски головного мозга (рисунок 3) позволяет занулить большую часть вокселей изображения. Очень часто интенсивность самого черепа может быть близкой к интенсивности ГБВ. Зануление вокселей на всех этих участках помогает избежать явных ошибок нейросетей во время предсказания.

В работе эта операция осуществлена при помощи утилит fsl-BET [1] и ANTs [5].

Нормализация и скейлинг При использовании алгоритмов глубокого обучения принято нормализовать входные данные или приводить их в отрезок от 0 до 1. Такой подход значительно улучшает результаты или даже просто делает процесс обучения возможным.



Рис. 3: Схема процедуры получения маски головного мозга

В связи с различными распределениями интенсивностей на снимках с разных томографов мы предлагаем использовать «робастный скейлинг» вместо нормализации. Робастный скейлинг интенсивностей - скейлинг к отрезку от 0 до 1, но 1 будет соответствать 95-квантиль распределения интенсивностей на конкретном MPT снимке.

Робастный скейлинг считается согласно следующей формуле:

$$X_{scaled} = \frac{X - \min\left(X\right)}{q_{95}\left(X\right)}$$

3.3 Аугментация

Отличный способ исскуственно увеличить выборку для обучения - это использовать аугментацию. Это позволяет добавить к данным новые примеры для обучения, которые нейросеть еще не видела.

Преобразования симметрии Один из простых в реализации и надежных методов аугментации в случае обработки изображений - симметричные преобразования исходного снимка и соответствующей ему маски. Во время оучения нейросети мы отражали выделенный патч изображения с вероятностью 0.5 относительно плоскости, проходящей через центр патча и параллельной одной из граней патча.

Вариация интенсивностей Исходя из мотивации различия МРТ изображений по распределениям интенсивностей в следствие аппаратной вариабельности, предлагается подход аугментации с варьированием значений интенсивностей. Вариация интенсивности во время обучения происходит на этапе подготовки патчей, выбранных из МРТ-снимков. Каждое значение интенсивности в патче домножается на случайное число, выбраное из равномерного распределения на отрезке $[1 - \alpha; 1 + \alpha]$. Ожидается, что это не даст нейросети переобучиться под поиск опухолей по конкретным значениям интенсивности.

4 Постановка эксперимента

4.1 Метрики качества

Стандартной метрикой качества для задач сегментации медицинских изображений является Dice Score, также являющейся основной метрикой на соревновании WMH Segmentation Challenge. Соответственно, ее удобно использовать для валидации моделей на данных с ГБВ.

Dice Score рассчитывается по формуле:

$$DiceScore = \frac{2 \cdot TP}{2 \cdot TP + FN + FP} ,$$

где TP - число правильно предсказанных вокселей, FN - число ложно предсказанных как отрицательные, FP - число ложно предсказанных как положительные воксели.

4.2 Кросс-валидация

В общем случае, для обучения и измерения качества модели используется схема кросс-валидации для 5 разбиений (cv5). Дополнительно используется схема leave-one-scanner-out (loo): обучаемся на снимках с жвух томографов, предсказываем на снимках с третьего.

4.3 Детали обучения

От тренировочной выборке было отделено 5 снимков для валидации результатов во время обучения. Каждая модель обучалась, пока метрики на валидации не стабилизировались. Параметр обучения (learning rate) был выставлен равным 0.1 и уменьшался вдвое каждый раз, когда функция потерь на валидации переставала уменьшаться. Размер батча для DeepMedic составляет 64, для U-Net - 4, ввиду намного большего размера входного патча.

При обучении минимизировалась бинарная кросс-энтропия. В качестве метода обучения использовался метод стохастического градиентного спуска с моментумом Нестерова.

5 Эксперименты

Как показали предварительные эксперименты, архитектура U-Net незначительно, но всегда уступает по результатам архитектуре DeepMedic. Поэтому представленные ниже результаты указаны для архитектуры DeepMedic.

В таблице 2 продемонстрировано влияние основных этапов предобработки данных на результат сегментации. Влияние аугментации отражениями и аугментации интенсивностей, направленных на борьбу с аппратной вариабельностью, продемонстрированно в таблице 3.

Кросс-валидация	Без масок	Удаление черепа	Скейлинг
cv5	0.737 ± 0.002	0.748 ± 0.003	0.751 ± 0.002
loo	0.328 ± 0.010	0.680 ± 0.003	0.702 ± 0.009

Таблица 2: Результаты модели DeepMedic без и с применением маски черепа, затем с заменой нормализации на скейлинг.

Kpocc-	Без улучшений	Аугментация	Аугментация	Оба подхода
валидация		отражениями	интенсивностей	
cv5	0.751 ± 0.002	0.773 ± 0.002	0.776 ± 0.004	0.780 ± 0.003
loo	0.702 ± 0.009	0.715 ± 0.004	0.711 ± 0.006	0.721 ± 0.003
- 0 D		D 16 11		

Таблица 3: Результаты модели DeepMedic на снимках с удаленным черепом при применении различных способов аугментации.

Также были проведены эксперименты с различными диапазонами аугментации интенсивностей. При значениях α от 0.05 до 0.10 наблюдаются наилучшие результаты (в пределах погрешности), представленные одним числом для $\alpha = 0.10$ в таблице 3. При α большем 0.20 уже наблюдается значительное ухудшение результата.

Примеры предсказаний можно увидеть на рисунке 4



Рис. 4: Четыре примера предсказаний нашей модели. Каждый пример содержит в себе: модальность; маску, размеченную экспертом-врачом и предсказание модели.

6 Выводы

Прежде всего стоит отметить, что работа направлена на прикладную применимость автоматических алгортимов для сегментации медицинских изображений. Для этого в первую очередь ставились и решались задачи сегментации в условиях аппаратной вариабельности и достижения алгоритмами высоких значений метрик.

Результаты экспериментов на данных с ГБВ при кросс-валидации типа leave-one-scanner-out отлично демонстрируют, что алгоритм, предложенный в работе, обладает достаточной обобщающей способностью. Скейлинг и аугментация данных помогли достичь инвариантности работы алгоритмов относительно различного распределения интенсивностей, а также размеров снимков. Предложенный метод способен сегментировать очаги заболевания на MPT снимках с неизвестных томографах практически также хорошо, как и с известных томографов: 0.72 Dice Score на снимках с «новых» сканеров в сравнении с 0.78 Dice Score на снимках с «известных» сканеров.

На веб-сайте соревнования WMH Segmentation Challenge [4] можно сравнить результаты работы нашего алгоритма в сравнении с аналогами других участников.

Список литературы

- 1. fsl BET. https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/BET/.
- 2. ISLES: Ischemic Stroke Lesion Segmentation Challenge. http://www.isleschallenge.org/.
- 3. MICCAI BRATS The Multimodal Brain Tumor Segmentation Challenge. http://braintumorsegmentation.org/.
- 4. WMH Segmentation Challenge Grand Challenge at MICCAI 2017. http://wmh.isi.uu.nl/.
- 5. Brian B Avants, Nicholas J Tustison, Gang Song, Philip A Cook, Arno Klein, and James C Gee. A reproducible evaluation of ants similarity metric performance in brain image registration. *Neuroimage*, 54(3):2033–2044, 2011.
- Özgün Çiçek, Ahmed Abdulkadir, Soeren S Lienkamp, Thomas Brox, and Olaf Ronneberger. 3d u-net: learning dense volumetric segmentation from sparse annotation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 424–432. Springer, 2016.
- Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 770–778, 2016.
- Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. arXiv preprint arXiv:1502.03167, 2015.
- Konstantinos Kamnitsas, Enzo Ferrante, Sarah Parisot, Christian Ledig, Aditya V Nori, Antonio Criminisi, Daniel Rueckert, and Ben Glocker. Deepmedic for brain tumor segmentation. In *International Workshop on Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, pages 138–149. Springer, 2016.
- Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 234–241. Springer, 2015.
- Б. Широких, А. Сафиуллин, А. Мусабаева, and М. Беляев. Сегментация гиперинтенсивности белого вещества методами глубокого обучения. In *Тр. конф.* Информационные Технологии и Системы, раде 9, 2017.

Восстановление матриц в задаче предсказания лекарственной устойчивости бактерий туберкулеза

Мария Буркина^{2,3}, Геннадий Федонин^{2,4}, Максим Панов^{1,2}

¹ Сколковский институт науки и технологий, Москва, Россия,

 $^{2}\,$ Институт проблем передачи информации РАН, Москва, Россия,

³ Московский физико-технический институт, Москва, Россия,

⁴ ФБУН Центральный НИИ эпидемиологии Роспотребнадзора, Москва mariya.burkina@phystech.edu

Аннотация Задача предсказания ЛУ бактерий туберкулеза состоит в том, чтобы предсказать реакцию (устойчивость или чувствительность) штамма туберкулеза на каждое из 13 существующих лекарств. С биологической точки зрения, среди множества всех генетических мутаций штаммма существует небольшое число мутаций, наличие которых делает штамм устойчивым к лекарству (эталонный штамм чувствителен ко всем лекарствам). Целью работы было проверить, улучшает ли совместное обучение качество предсказания и качество отбора важных признаков. В качестве метода был выбран алгоритм индуктивного восстановления матриц SGIMC, способный оставлять только информативные признаки. Предложенный метод сравнивался с LASSO-регрессией для каждого лекарства и эффективным существующим методом. Совместное рассмотрение лекарств не дало преимущества.

Ключевые слова: алгоритм индуктивного восстановления матриц, многозначная классификация, отбор признаков, лекарственная устойчивость туберкулеза.

1 Введение

Лекарственная устойчивость бактерий – это способность бактерий сохранять жизнедеятельность, несмотря на контакт с химиопрепаратами. Противотуберкулезные лекарства принято делить на два ряда: препараты первого ряда (5 лекарств) оказывают максимальный эффект при минимальной токсичности и препараты второго ряда (8 лекарств) оказывают более слабое воздействие и более токсичны, применяются тогда, когда у больного определяется устойчивость микобактерий туберкулеза к препаратам первого ряда. Штаммы туберкулеза, резистентные к препаратам первого ряда, являются серьезным препятствием на пути лечения туберкулеза.

Обычно, для того, чтобы подобрать эффективную схему терапии, используется фенотипирование туберкулеза, непосредственно определяющее способность штамма, взятого на анализ, размножаться в среде, содержащей лекарственные препараты. Фенотипирование Mycobacterium tuberculosis занимает недели, а доступ к необходимым лабораториям, особенно в развивающихся странах, часто ограничен. Поэтому разрабатываются ускоренные молекулярно-генетические методы анализа, использующие информацию о мутациях в геноме штамма. Их можно описать следующим образом: сначала по выборке штаммов определяются статистически наиболее вероятные связанные с лекарственной устойчивостью мутации, а затем на этапе анализа они ищутся в образце. Таким образом, качество клинического анализа зависит от полноты каталога этих мутаций.

Известные реакции штаммов на лекарства распределены неравномерно: для препаратов первого ряда известно много реакций, потому что штаммы проверяют на ЛУ в первую очередь к этим препаратам, а для некоторых препаратов второго ряда размер известных данных может быть недостаточен для хорошего обучения обычного классификатора. Также известно, что реакции штамма на разные лекарства коррелируют. Поэтому мы предположили, что алгоритм восстановления матриц даст преимущество перед алгоритмами, обучающимися для каждого лекарства отдельно.

Задача восстановления матрицы - задача восстановления недостающих элементов частично заполненной матрицы. Алгоритмы восстановления матриц широко используются в различных приложениях, таких как рекомендательные системы, кластеризация, многозначная классификация, обработка сигналов, компьютерное зрение [10], [14], [1], [13].

Задача предсказания лекарственной устойчивости бактерий естественно организована в матричной форме. Алгоритмы восстановления матриц делятся на два широких класса: с использованием дополнительной информации (inductive) и без нее (transductive). В данной работе проводится экспериментальное исследование эффективности предложенного метода восстановления матриц на данной задаче многозначной классификации, а также его сравнение с уже существующим каталогом мутаций и логистической ресрессией.

2 Алгоритмы восстановления матриц

В отстутствие дополнительной информации в задаче восстановления матрицы (transductive setting), предсказание строится только на основе наблюдаемых значений, при этом накладываются ограничения на ранг матрицы, так как предполагается, что существует некоторое пространство факторов небольшой размерности.

Самым распространенным способом поиска подходящей матрицы в таком случае является добавление к целевой функции штрафа на следовую норму (trace norm, nuclear norm) матрицы. Следовая норма является хорошо известным выпуклым приближением ранга матрицы [11].

Другим подходом к решению задачи являются представление искомой матрицы в виде $M = \tilde{U}\tilde{V}^{\mathrm{T}}$, где $\tilde{U} \in \mathbb{R}^{n_1 \times k_1}$, $\tilde{V} \in \mathbb{R}^{n_2 \times k_1}$ и добавление штрафа на нормы Фробениуса этих матриц.

2.1 Индуктивное восстановление матриц (IMC)

В нашей задаче, помимо матрицы реакций, есть важная дополнительная информация - список однонуклеотидных мутаций в геноме каждого штамма *M tuberculosis*, и важно не просто предсказать реакцию каждого штамма на лекарство, но и связать эту реакцию с генотипом, так как основной биологической задачей является исследование генетических механизмов лекарственной устойчивости туберкулеза.

Пусть $M \in \mathbb{R}^{n_1 \times n_2}$ – бинарная матрица реакций бактерий на лекарства, которую мы собираемся восстанавливать, и пусть $\Omega \subset \{1, \ldots, n_1\} \times \{1, \ldots, n_2\}$ – множество известных значений этой матрицы.

Пусть также известны матрицы векторов признаков $X \in \mathbb{R}^{n_1 \times d_1}$ и $Y \in \mathbb{R}^{n_2 \times d_2}$ для строк (в нашей задаче, бактерий туберкулеза) и столбцов (лекарств) матрицы M соответственно.

Цель индуктивного восстановления матриц – восстановить неизвестные значения M с помощью уже известных Ω и дополнительной информации об объектах строк и столбцов X и Y.

Предполагается, что значения M можно представить в виде $M_{ij} \sim \mathbf{x}_i^{\mathrm{T}} W \mathbf{y}_j$, где $W \in \mathbb{R}^{d_1 \times d_2}$ – некоторая матрица . Вместе с этим, обычно предполагается, что W имеет ранг $k < \min(d_1, d_2)$ и мы будем достигать этого с помощью факторизации матрицы $W = UV^{\mathrm{T}}$, где $U \in \mathbb{R}^{d_1 \times k}$ и $V \in \mathbb{R}^{d_2 \times k}$.

Пусть $\mathcal{L}: \mathbb{R} \times \mathbb{R} \mapsto \mathbb{R}$ - выпуклая функция потерь, тогда задача индуктивного восстановления матриц сводится к следующей задаче:

$$J(U,V) = \sum_{(i,j)\in\Omega} \mathcal{L}(M_{ij}, (XUV^{\mathrm{T}}Y^{\mathrm{T}})_{ij}) \to \min_{U,V} .$$
(1)

В качестве функции потерь \mathcal{L} возьмем стандартную для задачи бинарной классификации функцию $\mathcal{L}(y,p) = \log(1 + e^{-yp}) - \log$ -loss, где $p_{ij} = \mathbf{e}_i^{\mathrm{T}} X U V^{\mathrm{T}} Y^{\mathrm{T}} \mathbf{e}_j$.

2.2 Sparse Group Inductive Matrix Completion (SGIMC)

Прямое использование всей дополнительной информации обычно неэффективно - чаще всего не все признаки информативны и желательно, чтобы алгоритм восстановления матрицы был робастен к неинформативным шумовым признакам. В таком случае к целевой функции (1) добавляется некоторая функция штрафа:

$$J(U,V) = \sum_{(i,j)\in\Omega} \mathcal{L}\left(M_{ij}, (XUV^{\mathrm{T}}Y^{\mathrm{T}})_{ij}\right) + \lambda_U R(U) + \lambda_V R(V) \to \min_{U,V} .$$
(2)

Различные априорные знания о виде решения задачи влекут за собой различные виды накладываемых ограничений [6].

В случае признаков - мутаций генов мы предполагаем, что только небольшая часть мутаций влияет на лекарственную устойчивость образца и поэтому в качестве функции штрафа будем использовать групповую норму матрицы:

$$R(Z) = \|Z\|_{2,1} = \sum_{j=1}^{d} \|\mathbf{e}_{j}^{\mathrm{T}}Z\|_{2}, \qquad (3)$$

где \mathbf{e}_i - *i*-й столбец единичной матрицы.

Тогда в процессе решения задачи некоторые строки матрицы коэффициентов будут вынуждены стать нулевыми, и таким образом в решении не будут учитываться соответствующие признаки - мутации.

Оптимизационная задача индуктивного восстановления матриц со Sparse-Group штрафом (алгоритм оптимизации основан на методе ADMM [2]):

$$J(U,V) = \sum_{(i,j)\in\Omega} \mathcal{L}\left(M_{ij}, (XUV^{\mathrm{T}}Y^{\mathrm{T}})_{ij}\right) + \lambda_U \sum_{i=1}^{a_1} \|\mathbf{e}_i^{\mathrm{T}}U\|_2 + \lambda_V \sum_{j=1}^{a_2} \|\mathbf{e}_j^{\mathrm{T}}V\|_2 \to \min_{U,V}$$
(4)

3 Данные для экспериментов

Данные о чувствительности и устойчивости штаммов к лекарствам и отсеквенированные геномы - объединенный набор данных, использовавшихся в статьях [3], [12], [4], [8]. Также в каждой из этих статей получен свой список мутаций, ассоциированных с устойчивостью к каждому из лекарств. Список мутаций, полученный в статье [12], мы использовали для дополнительной валидации, позволяющей оценить адекватность отбора признаков нашей моделью.

3.1 Матрица реакций

Предсказываются реакции 4734 штаммов туберкулеза на следующие 13 лекарств:

- Препараты первого ряда:
- Изониацид, Этамбутол, Рифампицин, Пиразинамид, Стрептомицин.
- Препараты второго ряда:
 - Фторхинолоны: Офлоксацин, Моксифлоксацин, Ципрофлоксацин;
 - Аминогликозиды: Амикацин, Канамицин;
 - Капреомицин, Протионамид, Этионамид.

Для каждого штамма известна реакция хотя бы на одно лекарство. Как и во многих других задачах, связанных с медицинской диагностикой, выборка сильно несбалансирована: количество штаммов, чувствительных к данному антибиотику, существенно превосходит количество штаммов, устойчивых к нему. На рис. 1 приведена диаграмма, показывающая общее количество известных реакций на лекарство и количество резистентных среди них.

Таким образом, известно 0.46 всех значений матрицы реакций $M \in \mathbb{R}^{4734 \times 13}$. Из 4734 штаммов 1572 устойчивы хотя бы к одному лекарству, из них 1138 устойчивы более чем к одному лекарству.



Рис. 1: Количество известных реакций на лекарство и количество резистентных среди них.

3.2 Матрицы признаков строк и столбцов

- $X \in \mathbb{R}^{n_1 \times d_1}$ матрица признаков бактерий. Известны следующие мутации, встречающиеся среди данных 4734 штаммов туберкулеза:
 - Мутации в генах (в том числе в кодирующей части гена и в промоторе) всего 177185 различных мутаций в 4099 генах, среди них SNP (однонуклеотидный полиморфизм отличие последовательности ДНК размером в один нуклеотид (A, T, G, C) в геноме штамма от генома штамма, взятого за образец) и indels (удаление или вставка одного нуклеотида в гене).
 - Мутации в рибосомных, транспортных и некодирующих РНК.
 - Мутации в некодирующих участках ДНК.

Также имеются признаки - индикаторы того, что данный ген изменился, ген перестал воспроизводиться и индикаторы того, в каком месте гена произошла мутация, без остальной информации о ней.

Таким образом, имеем $n_1 = 4734$ штаммов *M* tuberculosis и для каждого известен бинарный вектор признаков размерности $d_1 = 355709$, каждое вхождение единицы в который обозначает наличие одной из перечисленных выше мутаций. Также, многие мутации встречаются стреди штаммов очень редко, и если удалить признаки, встречающиеся 1 раз, в итоге останется $d_1 = 130893$ признаков.

– $Y \in \mathbb{R}^{n_2 \times d_2}$ - матрица признаков лекарств. Для каждого из $n_2 = 13$ лекарств известен бинарный вектор признаков размерности $d_2 = 28$, все

признаки - индикаторы принадлежности к некоторой группе (химической, группе антибиотиков).

4 Описание экспериментов

4.1 Схема кросс-валидации

Наибольший практический интерес представляет алгоритм, с помощью которого для нового штамма, попавшего в лабораторию, на основе его генотипа можно предсказать реакции на все препараты. В литературе, посвященной изучению связи генотипа и лекарственной устойчивости, качество построенных алгоритмов проверяется на штаммах, которые не присутствовали в обучающей выборке. Для задачи восстановления матриц это называется проблемой холодного старта - никакие взаимодействия нового объекта не известны и предсказание возможно только на основе дополнительной информации.

Поэтому, чтобы объективно сравнить наш алгоритм с уже существующими подходами, нужно разделить штаммы на две группы: на обучающую и тестовую выборку. Но разделить образцы нужно таким образом, чтобы в обеих выборках были представлены реакции на все лекарства и чтобы соотношение чувствительных и резистентных реакций на каждое лекарство было одинаковым (тогда обучающая и тестовая выборки будут хорошим приближением полных данных), в противном случае мы можем получить большую погрешность в оценке качества работы алгоритма [7].

Желаемое разбиение достигается следующим образом:

Алгоритм 1: Алгоритм разбиения штаммов на обучающую и тестовую
выборки
Вход: Строки матрицы $M, M_{ij} \in \{S, R, nan\}$
для $i om 0 \partial o n_1$ выполнять
Представляем каждую строку матрицы M как строку $\in \{0, 1, 2\}^{n_2}$
С помощью хеш-функции отображаем полученную строку в $(0, 2^7)$ и
получаем коллизии - теперь нескольким изначально разным строкам
соответствуют одинаковые значения
конец
Полученные целочисленные представления реакций штаммов делим на
обучающую и тестовую выборку с помощью стандартной функции
StratifiedShuffleSplit.

4.2 Существующий подход

В статье [12], с результатами которой мы будем сравнивать наш алгоритм, рассматривались мутации только в 23 генах, ранее ассоциированных с ле-

карственной устойчивостью и, помимо этого, только несинонимичные мутации (мутации, переключающие кодон на кодирование другой аминокислоты), поскольку они с большей вероятностью кодируют функциональные белковые изменения. После того, как синонимичные мутации были помечены как не влияющие на резистентность и отброшены, для каждой из оставшихся мутаций проверялась гипотеза относительно ее связи с лекарственной устойчивостью (для каждого лекарства по отдельности). Мутация характеризовалась как определяющая резистентность, если она являлась единственной мутацией среди всех возможных в 23 исследовавшихся генах по крайней мере в одном резистентном штамме. Если мутация встречалась только в чувствительных к лекарству образцах или некотором чувствительном образце это была единственная мутация, то она определялась как не влияющая на резистентность и отбрасывалась. После этого алгоритм повторялся снова.

4.3Результаты

Наш алгоритм на выходе дает вероятность штамма быть устойчивым к соответствующему лекарству, поэтому в качестве метрики оценки качества была взята площадь под ROC-кривой AUC (Area Under Curve). Сравнивались пять моделей:

- Логистическая регрессия с l1-штрафом, обучающаяся на $d_1 = 355709$ признаках;
- Логистическая регрессия с l1-штрафом, обучающаяся на $d_1 = 130893$ признаках;
- SGIMC, обучающийся на $\tilde{d}_1 = 355709$ признаках; SGIMC, обучающийся на $d_1 = 130893$ признаках;
- Словарный алгоритм (тестовый образец определяется резистентным, если в нем есть мутация, отвечающая за устойчивость к данному лекарству). Словарь составляют признаки, отобранные в статье [12] (всего 4760 мутаций).

На рис. 2 показаны среднее, минимальное и максимальное значения метрики ROC-AUC по 10 различным разбиениям 50/50 на обучающую и тестовую выборки, согласно методу, описанному в разделе 4.1. Видно, что отбрасывание > 100 тысяч редких признаков положительно влияет на качество работы алгоритма SGIMC. В целом, SGIMC на всех лекарствах, кроме Ципрофлоксацина, проигрывает логистической регрессии, но выигрывает у словарного алгоритма.

Переходя к признакам, отобранным алгоритмами, у SGIMC без удаления редких признаков, остается в среднем 6260 ненулевых строк матрицы U, из соответствующих признаков содержатся в словаре около 270, SGIMC с удалением редких признаков, отбирает в среднем 9000 признаков, из которых содержатся в словаре 320. Логистическая регрессия, независимо от того, на каком наборе признаков она работает, отбирает около 1770 мутаций, из которых в среднем в словаре содержатся 80.



Рис. 2: Качество работы различных алгоритмов.

На рис. 3 приведены результаты аналогичного сравнения двух алгоритмов восстановления матриц: стандартного IMC [9] и SGIMC [5]. Хотя алгоритм SGIMC не превосходит в эффективности IMC, он достигает хорошего качества, используя для предсказания 6000 мутаций-признаков, в то время как IMC использует все 355709 признаков и, следовательно, не может помочь в исследовании генетических механизмов лекарственной устойчивости туберкулеза.

4.4 Выводы

Преимущество логистической регрессии с *l*1-штрафом перед алгоритмом SGIMC состоит в том, что LASSO-регрессия обладает возможностью фильтровать признаки для каждого лекарства по отдельности. И хотя ни алгоритму восстановления матрицы, ни логистической регрессии не удалось отобрать важные признаки так же эффективно, как существующим алгоритмам, в терминах метрики качества ROC-AUC они работают лучше словарного алгоритма с отобранным словарем признаков.

5 Заключение

Для решения задачи предсказания резистентности бактерий в данной работе был предложен индуктивный алгоритм восстановления матриц, способный



Рис. 3: Сравнение качества работы двух разных алгоритмов восстановления матриц.

отбирать важные признаки. Для тестирования алгоритма была предложена подходящая для этой задачи схема разбиения выборки на обучающую и тестовую. Качество предсказаний алгоритма и его способность отфильтровывать информативные признаки оценивались в сравнении с существующим словарем резистентных мутаций и логистической регрессией с *l*1-штрафом. И алгоритм матричного восстановления, и LASSO-регрессия отбирают лишь малую часть важных мутаций. В качестве наиболее подходящей для сравнения метрики качества была выбрана площадь под ROC-кривой. В терминах этой метрики алгоритм матричного восстановления работает хуже логистической регрессии, но лучше словарного алгоритма.

Список литературы

- Andreas Argyriou, Theodoros Evgeniou, and Massimiliano Pontil. Convex multitask feature learning. *Machine Learning*, 73(3):243–272, 2008.
- Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3(1):1–122, January 2011.
- 3. Coll et al. Genome-wide analysis of multi- and extensively drug-resistant mycobacterium tuberculosis. In *Nature Genetics*, pages 307–316, 2018.
- Pankhurst et al. Rapid, comprehensive, and affordable mycobacterial diagnosis with whole-genome sequencing: a prospective study. In *The Lancet Respiratory Medicine*, pages 49–58, 2016.
- 5. Maria Burkina Gennady Fedonin Maxim Panov Ivan Nazarov, Boris Shirokikh. Sparse group inductive matrix completion.
- Sham M Kakade, Shai Shalev-Shwartz, and Ambuj Tewari. Regularization techniques for learning with matrices. *Journal of Machine Learning Research*, 13(Jun):1865–1890, 2012.
- Ron Kohavi. A study of cross-validation and bootstrap for accuracy estimation and model selection. *IJCAI*, 1995.
- Karen Kieser et al. Maha Farhat, Jesse Shapiro. Genomic analysis identifies targets of convergent positive selection in drug resistant mycobacterium tuberculosis. In *Nature Genetics*, pages 1183—1189, 2013.
- Nagarajan Natarajan and Inderjit S Dhillon. Inductive matrix completion for predicting gene-disease associations. *Bioinformatics*, 30(12):i60–i68, 2014.
- Jasson DM Rennie and Nathan Srebro. Fast maximum margin matrix factorization for collaborative prediction. In *Proceedings of the 22nd international conference* on Machine learning, pages 713–719. ACM, 2005.
- Nathan Srebro and Adi Shraibman. Rank, trace-norm and max-norm. In International Conference on Computational Learning Theory, pages 545–560, 2005.
- 12. Shaheed V Omar Jessica Hedge et al. Timothy M Walker, Thomas A Kohl. Whole-genome sequencing for prediction of mycobacterium tuberculosis drug susceptibility and resistance: a retrospective cohort study. In *The Lancet Infectious Disease*, pages 1193—1202, 2015.
- Zhiyuan Weng and Xin Wang. Low-rank matrix completion for array signal processing. In Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on, pages 2697–2700. IEEE, 2012.
- Jinfeng Yi, Tianbao Yang, Rong Jin, Anil K Jain, and Mehrdad Mahdavi. Robust ensemble clustering by matrix completion. In *Data Mining (ICDM), 2012 IEEE* 12th International Conference on, pages 1176–1181. IEEE, 2012.

Towards Hypothesis Testing for Random Graphs with Community Structure

Valentina Shumovskaia¹² and Maxim Panov¹²³

¹ Skolkovo Institute of Science and Technology (Skoltech) ² Higher School of Economics

³ Institute for Information Transmission Problems of Russian Academy of Sciences

Abstract. The analysis of random graphs and network analysis recently become an active area of research. In this paper we study the problem of testing between two populations of inhomogeneous random graphs defined on the same set of vertices, where the null hypothesis means that the underlying edge probability matrices coincide. We propose a new approach of random graphs testing based on community detection procedure: we introduce a structural assumption, which is that our graphs have community structures with k communities. A graph has such structure if we can divide its vertices into k non overlapping clusters, every cluster has its own probability of edges appearing and there is some probability of appearing of edges between clusters. Under this assumption the null hypothesis means that clustering structures and probabilities inside and outside clusters coincide for both samples. We propose a test statistic T_{CD} , prove its theoretical properties and provide numerical experiments with it.

1 Introduction

The analysis of random graphs has become an active area of research in recent years, there is progress in problems related to network modelling [2], [10], community detection [5], [1], network dynamics [3], the problem of hypothesis testing [7], [4]. In this paper we consider the problem of testing between two populations of inhomogeneous random graphs defined on the same set of vertices. Suppose that two (or more) random graphs over the same set of vertices are given, then the null hypothesis means that the underlying edge probability matrices coincide (since we always can represent a graph as an adjacency matrix), the test procedure can compare the structure for both matrices. Such tests can be useful in different areas. For example, in functional neuroimaging [8] one may consider a graph, where each vertex corresponds to some region of interest (ROI) in the brain. Each edge in this graph denotes a measure of association between corresponding ROIs. So we can be interested in questions: "Do two collections of networks differ on average?", "Has there been a change in the networks for a given subpopulation from yesterday to today?", and the answers can be found by two sample test for random graphs. The other applied problem is testing between omics networks [9], where large numbers of molecular components operate simultaneously through a network of interactions to generate cellular functions and phenotypic states. Also, such tests can be applied in social networks research. For example, we can find the time point when one of communities split into two or when two communities merged.

1.1 Generalized problem statement and notations

Let $G = (V, E_G)$ be an undirected graph, where $V = \{1, \ldots, n\}$ is a set of vertices and E^G is a set of undirected edges. Denote by $A^G \in \{0, 1\}^{n \times n}$ an adjacency matrix of G: $A_{i,j}^G = 1$ if $(i, j) \in E_G$ and 0 otherwise. Let $\mathbb{M}_n \subset [0, 1]^{n \times n}$ be the set of symmetric matrices with zero diagonal and off-diagonal entries in [0, 1]. For any $P \in \mathbb{M}_n$, we say that G is an IER (Inhomogeneous Erdos-Renyi) graph, denoted by $G \sim \operatorname{IER}(P)$, if the adjacency matrix A^G is a symmetric random matrix such that $A_{ij}^G \sim \operatorname{Bernoulli}(P_{ij})$ and they are independent.

Fix $P, Q \in \mathbb{M}_n$, suppose there are two samples $G_1, \ldots, G_m \sim_{iid} IER(P)$ and $H_1, \ldots, H_m \sim_{iid} IER(Q)$. We would like to test

$$\mathcal{H}_0: P = Q \text{ against } \mathcal{H}_1: P \neq Q. \tag{1}$$

For any n, m and given graphs the test Ψ is a binary function of 2m adjacency matrices. $\Psi = 0$ if the test accepts the null hypothesis, $\Psi = 1$ otherwise. Also, let Ω_0, Ω_1 be the subsets of $\mathbb{M}_n \times \mathbb{M}_n$ such that \mathcal{H}_0 holds if $\theta \in \Omega_0, \mathcal{H}_1$ holds if $\theta \in \Omega_1$.

We are interested in maximum (or worst-case) risk of a test:

$$\sup_{\theta \in \Omega_0} \mathbb{P}_{\theta}(\Psi = 1) + \sup_{\theta \in \Omega_1} \mathbb{P}_{\theta}(\Psi = 0)$$

which is the sum of Type-I and Type-II errors incurred by the test, and to study the optimality of the test, we are interested in the minimax risk of the problem:

$$\inf_{\Psi} \Big(\sup_{\theta \in \Omega_0} \mathbb{P}_{\theta}(\Psi = 1) + \sup_{\theta \in \Omega_1} \mathbb{P}_{\theta}(\Psi = 0) \Big).$$

1.2 Related work

There already exist two sample tests for random graphs [7], [4], let's focus on [7], where authors propose a test based on Frobenius norm.

First, we introduce quantities $S(\theta)$ which corresponds to the separation criterion and $C(\theta)$ which controls the complexity of the problem:

$$S(\theta) = \frac{\|P - Q\|_F}{\sqrt{\|P + Q\|_F}},$$
$$C(\theta) = \|P + Q\|_F,$$

where $P, Q \in \mathbb{M}_n$, $\|\cdot\|_F$ is Frobenius norm. In terms of these quantities we transform Ω_0 and Ω_1 :

$$\Omega_0(n, m, \mathscr{C}_{n,m}) = \{ \theta \in \mathbb{M}_n \times \mathbb{M}_n \colon \mathcal{S}(\theta) = 0, \mathcal{C}(\theta) \ge \mathscr{C}_{n,m} \},\$$

 $\Omega_1(n,m,\mathscr{S}_{n,m}\mathscr{C}_{n,m}) = \{ \theta \in \mathbb{M}_n \times \mathbb{M}_n \colon \mathcal{S}(\theta) > \mathscr{S}_{n,m}, \mathcal{C}(\theta) \ge \mathscr{C}_{n,m} \},\$

where $\mathscr{S}_{n,m}, \mathscr{C}_{n,m}$ are some non-negative scalars that may depend only on n, m. Hence, we test

$$\mathcal{H}_0: \theta \in \Omega_0(n, m, \mathscr{C}_{n,m}) \text{ against } \mathcal{H}_1: \theta \in \Omega_1(n, m, \mathscr{S}_{n,m} \mathscr{C}_{n,m}).$$

Define the following test statistic, assuming m is even, $G_1, \ldots, G_m \in IER(P)$, $H_1, \ldots, H_m \in IER(Q)$:

$$T_{Fro} = \frac{\sum_{i,j=1,i m/2} A_{ij}^{G_k} - A_{ij}^{H_k}\right)}{\sqrt{\sum_{i,j=1,i m/2} A_{ij}^{G_k} + A_{ij}^{H_k}\right)}}.$$
 (2)

We can say that numerator in (2) is estimator of $||P - Q||_F$ up to some scaling and it is normalized by a rough estimate of its standard derivation.

Theorem 1 (Sufficient conditions for detecting separation; Theorem 3.1 in [7]).

 $\forall \eta \in (0,1), \text{ if } m \ge 2, \mathscr{C}_{n,m} \ge 16\sqrt{6\ln\frac{4}{\eta}}, \mathscr{S}_{n,m} \ge \frac{8\sqrt{6}}{\sqrt{m}} \sqrt[4]{\ln\frac{4}{\eta}}, \text{ then the maximum risk of the test } \Psi = \mathbbm{1}\left(T_{Fro} > 8\sqrt{6\ln\frac{4}{\eta}}\right) \text{ is smaller than } \eta.$

We interpret this result as follows: if the expected number of edges in one of the graph models grow at least linearly with n, then the test (2) can distinguish between two graphs that are separated by a certain factor, and a smaller separation can be identified with more samples.

Theorem 2 (Necessary conditions for detecting separation; Theorem 3.2 in [7]).

Consider the testing problem with $\mathscr{C}_{n,m} \leq \frac{n}{\sqrt{8}}$. For any $\eta \in (0,1)$ the minimax risk is at least η if either one of the following conditions hold:

1.
$$\mathscr{S}_{n,m} < \frac{\sqrt[4]{\ln(1+4(1-\eta)^2)}}{\sqrt{6m}}, \text{ or }$$

2. $m = 1 \text{ and } \mathscr{S}_{n,m} < \sqrt{\frac{n}{12}}.$

It is clear that necessary and sufficient conditions are quite close in terms of \mathscr{S} but there is a significant gap for allowed sparsity \mathscr{C} .

2 Structural assumption and new approach

We suppose that if graphs have community structure (the strict definition will be given below), it is natural to apply any community detection algorithm and test these graphs in terms of communities. As before, suppose we are given $G_1, \ldots, G_m \in \text{IER}(P), H_1, \ldots, H_m \in \text{IER}(Q)$, where $P, Q \in \mathbb{M}_n, m \ge 1$. Additionally, we assume that our graphs have community structures with k communities, so additional restrictions on P and Q appear. P has community structure if we can define k non-overlapping sets of each community indices $\mathcal{C}_G^1,\ldots,\mathcal{C}_G^k$ and

$$\begin{split} (P)_{ij} &= p_{aa}, \text{if } i \in \mathcal{C}_G^a, j \in \mathcal{C}_G^a, a = 1, \dots, k, \\ (P)_{ij} &= p_{ab}, \text{if } i \in \mathcal{C}_G^a, j \in \mathcal{C}_G^b, a, b = 1, \dots, k, a \neq b. \end{split}$$

At this point we can suggest to slightly change our hypothesises:

$$\begin{aligned} \mathcal{H}_0: \quad p_{ab} &= q_{ab} = \theta_{ab}, a, b = 1, \dots, k \quad \text{against } \mathcal{H}_1: \quad \exists a, b : p_{ab} \neq q_{ab} \text{ or} \\ \mathcal{C}^a_G &= \mathcal{C}^a_H, a = 1, \dots, k, \quad \qquad \exists a : \mathcal{C}^a_G \neq \mathcal{C}^a_H. \end{aligned}$$

Similarly to [6] let us define the set $\mathcal{Z}_{n,k} = \{z : [n] \to [k]\}$ as a collection of all possible mappings from $[n] := \{1, \ldots, n\}$ to [k], so every z defines a clustering structure of n nodes. For a given matrix $\{\eta_{ij}\} \in \mathbb{R}^{n \times n}$ we can define such quantities:

$$\bar{\eta}_{ab}(z) = \frac{1}{|z^{-1}(a)||z^{-1}(b)|} \sum_{\substack{i \in z^{-1}(a) \\ j \in z^{-1}(b)}} \eta_{ij}, \text{ where } a \neq b \in [k], |z^{-1}(a)| > 1,$$
$$\bar{\eta}_{aa}(z) = \frac{1}{|z^{-1}(a)|(|z^{-1}(a)| - 1)} \sum_{i \neq j \in z^{-1}(b)} \eta_{ij}, \text{ where } a \in [k].$$

So, we can construct $\bar{\eta}(z) \in \mathbb{R}^{k \times k}$:

$$\bar{\eta}_{ij}(z) = \bar{\eta}_{ij}, i \neq j.$$

We will use edge probabilities estimated by community detection procedure in our test statistic, so we should take into account how well they are estimated. Let's find the convergence rate for parameter estimation. Following [6], for the first sample we want to minimize the quantity (and similarly for the second):

$$L(Q, z) = \frac{1}{m} \sum_{s=1}^{m} \sum_{a, b \in [k]} \sum_{(i,j) \in z^{-1}(a) \times z^{-1}(b)} (A_{ij}^{G_s} - Q_{ab})^2,$$

where $z \in Z_{n,k}, \ Q \in \mathbb{R}^{k \times k}$.

Let us denote

$$(\hat{\mathcal{Q}}, \hat{z}) \in \underset{\mathcal{Q} \in \mathbb{R}^{k \times k}, z \in \mathbb{Z}_{n,k}}{\operatorname{argmin}} L(\mathcal{Q}, z).$$
(3)

Thus, we can define the estimator of p_{ab} as $\hat{p}_{ab} = \hat{\mathcal{Q}}_{\hat{z}(a)\hat{z}(b)}, a, b \in [k]$.

Proposition 3. For any minimizer $(\hat{\mathcal{Q}}, z)$, the entries of $\hat{\mathcal{Q}}$ has representation $\hat{\mathcal{Q}}_{ab} = (\bar{A}^G(\hat{z}))_{ab} := \frac{1}{m} \sum_{s=1}^m (\bar{A}^{G_s}(\hat{z}))_{ab} \ \forall a, b \in [k].$

Under our assumptions each node $i \in [n]$ is associated with label $a \in [k]$ and the edge $A_{ij}^{G_s} \sim \text{Bernoulli}(P_{ij})$. Denote $\theta_{ij} := P_{ij}$. So, we assume that $\{\theta_{ij}\}$ is from the parameter space

$$\Theta_k = \{\{\theta_{ij}\} \in [0,1]^{n \times n} : \theta_{ii} = 0, \theta_{ij} = \mathcal{Q}_{ab} = \mathcal{Q}_{ba}$$

for $(i,j) \in z^{-1}(a) \times z^{-1}(b)$ for some $Q_{ab} \in [0,1]$ and $z \in Z_{n,k}\}$

In the Theorem 4 we claim the convergence rate for estimating θ_{ij} by estimator (3), where $\hat{\theta}_{ij} = \hat{Q}_{\hat{z}(i)\hat{z}(j)}$.

Theorem 4. For any constant C' > 0, there exists a constant C > 0 only depending on C', such that

$$\frac{1}{n^2} \sum_{i,j \in [n]} (\hat{\theta}_{ij} - \theta_{ij})^2 \le C \Big(\frac{k^2}{n^2 m} + \frac{\log k}{nm} \Big)$$

with probability at least $1 - \exp(-C'k \log n)$, uniformly over $\theta \in \Theta_k$. Furthermore, we have

$$\sup_{\theta \in \Theta_k} \mathbb{E} \left(\frac{1}{n^2} \sum_{i,j \in [n]} (\hat{\theta}_{ij} - \theta_{ij})^2 \right) \le C_1 \left(\frac{k^2}{n^2 m} + \frac{\log k}{nm} \right)$$

for all $k \in [n]$ with some universal constant $C_1 > 0$.

At this point we are returning to our test procedure. First, in order to find clustering structures of our graphs we apply a community detection procedure (for example, spectral clustering) to our adjacency matrices: we get clustering structures z_G and z_H of the first and the second sample correspondingly. And then we calculate $\bar{A}^{G_1}(z_G), \ldots, \bar{A}^{G_m}(z_G), \bar{A}^{H_1}(z_H), \ldots, \bar{A}^{H_m}(z_H) \in \mathbb{R}^{k \times k}$. As in Proposition 3, we denote

$$\bar{A}^G = \frac{1}{m} \sum_{j=1}^m \bar{A}^{G_j},$$
$$\bar{A}^H = \frac{1}{m} \sum_{j=1}^m \bar{A}^{H_j}.$$

Now let us define new matrices $\widehat{P}, \widehat{Q} \in \mathbb{R}^{n \times n}$ which are matrices of probabilities estimated by community detection:

$$\widehat{P} = \{\widehat{P}_{ij} = \bar{A}^{G}_{z_{G}(i)z_{G}(j)}\}_{i,j=1}^{n}, \\ \widehat{Q} = \{\widehat{Q}_{ij} = \bar{A}^{H}_{z_{H}(i)z_{H}(j)}\}_{i,j=1}^{n}.$$

We propose the following test statistic:

$$T_{CD} = \sum_{i,j} \mathcal{K}(\widehat{P}_{ij}, \widehat{Q}_{ij}), \tag{4}$$

where $\mathcal{K}(p,q)$ is Kullback-Leibler divergence between Bernoulli distributions with probabilities $p, q \in (0, 1)$:

$$\mathcal{K}(p,q) = p \log \frac{p}{q} + (1-p) \log \frac{1-p}{1-q}.$$
(5)

Consider a restriction on \mathcal{H}_0 : $q_{ij} = q_0, p_{ij} = p_0, \forall i \neq j \in [n]$. We are interested in behaviour of T_{CD} when \mathcal{H}_0 is true. Suppose that true size of communities are n_1, \ldots, n_k , then the following result holds:

Theorem 5. If \mathcal{H}_0 is true, then $T_{CD} \sim \chi^2(k+1)$ with $N_1, \ldots, N_k, N_0 \to \infty$, where $N_1 = \frac{n_1(n_1-1)}{2}, \ldots, N_k = \frac{n_k(n_k-1)}{2}, N_0 = \frac{n(n-1)}{2} - \sum_{j=1}^k n_k$.

With Theorem 5 we can easily define quantiles and start to test the data. In further work we plan to find behaviour of T_{CD} using the result from Theorem 4 without a restriction on probabilities between clusters.

3 Experiments

3.1 Setup

For each experiment we consider fixed community sizes N_1, N_2 , fixed probabilities p_{11}, p_{22}, p_{12} for the first sample and probabilities $q_{11} = p_{11} + \delta, q_{22} = p_{22} + \delta, q_{12} = p_{12}$ for the second sample. Let us conduct 100 experiments for every δ and show the ratio of \mathcal{H}_0 acceptance of T_{CD} and T_{Fro} . In every experiment we apply a spectral clustering algorithm (since we suppose a number of communities to be known) to one of the graphs with 2 clusters and take $\eta = 0.05$. Our experiments are:

1. See Fig. 1. $N_1 = 100, N_2 = 100$. Each sample consists of 2 graphs but T_{CD} uses only one of each. Find p_{11}, p_{22}, p_{12} in the figure description.



Fig. 1. $N_1 = 100, N_2 = 100$. Each sample consists of 2 graphs but T_{CD} uses only one of each.



Fig. 2. $N_1 = 100, N_2 = 100$. Each sample consists of 2 graphs.

- 2. See Fig. 2. $N_1 = 100, N_2 = 100$. Each sample consists of 2 graphs. Find p_{11}, p_{22}, p_{12} in the figure description.
- 3. See Fig. 3. $N_1 = 100, N_2 = 100$. Each sample consists of 20 graphs. Find p_{11}, p_{22}, p_{12} in the figure description.
- 4. See Fig. 4. $N_1 = 100, N_2 = 100$. Each sample consists of 100 graphs. Find p_{11}, p_{22}, p_{12} in the figure description.
- 5. See Fig. 5. $N_1 = 150$, $N_2 = 70$, $N_3 = 30$, $p_{11} = 0.75$, $p_{22} = 0.6$, $p_{33} = 0.7$, $p_0 = 0.2$. Find number of samples in the figure description.
- 6. See Fig. 6. $N_1 = 100, N_2 = 80, N_3 = 60, N_4 = 40, N_5 = 20, p_{11} = 0.65, p_{22} = 0.6, p_{33} = 0.7, p_{44} = 0.65, p_{55} = 0.8, p_0 = 0.2$. Find number of samples in the figure description.

3.2 Results of Experiments

Usually when we work with tests, we are interested in I and II types of errors. In our experiments, for T_{Fro} the I type of error is zero while the II type of error is much more than $\eta = 0.05$. The II type of error of this test gets less with number of graphs growing but anyway in every experiment it is more than for T_{CD} . Also, we may notice that when the network structure is poor (e.g. $p_{11} = p_{22} = 0.2, p_{12} = 0.2$ and $p_{11} = p_{22} = 0.4, p_{12} = 0.2$), T_{CD} works badly. It can be explained: even if spectral clustering algorithm finds clusters, they are not



Fig. 3. $N_1 = 100, N_2 = 100$. Each sample consists of 20 graphs.



Fig. 4. $N_1 = 100, N_2 = 100$. Each sample consists of 100 graphs.

found correctly. Thus, the test we proposed should be used in case with "visible" community structure. Also, we see that our test works fine with a small number of samples: in the experiments provided in the previous section m = 2 is enough. We can tell that T_{CD} works fine even in the case with 1 graph in each sample but community sizes should be big enough (in Theorem 5 clusters sizes tend to infinity), see Fig. 7: the more community sizes are, the less the I type of error is.



(a) Each sample consists of 2 graphs (b) Each sample consists of 2 graphs. but T_{CD} uses only one of each.



(c) Each sample consists of 10 graphs. (d) Each sample consists of 100 graphs.

Fig. 5. $N_1 = 150, N_2 = 70, N_3 = 30, p_{11} = 0.75, p_{22} = 0.6, p_{33} = 0.7, p_0 = 0.2.$

4 Conclusions and Outlook

We proposed a new approach based on community detection procedure for comparison of random graphs. We showed the convergence rate of parameters estimated by community detection, showed the behaviour of test statistic when null hypothesis holds, provided experiments that show its advantage over the test based on Frobenius norm. In further work we plan to estimate minimax risks of our test and conduct experiments on the real data.

References

- Emmanuel Abbe and Colin Sandon. Achieving the KS threshold in the general stochastic block model with linearized acyclic belief propagation. 29th Conference on Neural Information Processing Systems (NIPS 2016), (1):1–9, 2016.
- Réka Albert and Albert-László Barabási. Statistical mechanics of complex networks. Reviews of Modern Physics, 74(1):47–97, 2002.
- [3] Noam Berger, Christian Borgs, Jennifer T. Chayes, and Amin Saberi. On the spread of viruses on the internet. Proceedings of the sixteenth annual ACM-SIAM symposium on Discrete algorithms, pages 301–310, 2005.
- [4] Guy Bresler and Dheeraj Nagaraj. Optimal Single Sample Tests for Structured versus Unstructured Network Data. pages 1–42, 2018.



(a) Each sample consists of 2 graphs (b) Each sample consists of 2 graphs. but T_{CD} uses only one of each.



(c) Each sample consists of 10 graphs. (d) Each sample consists of 100 graphs.

Fig. 6. $N_1 = 100, N_2 = 80, N_3 = 60, N_4 = 40, N_5 = 20, p_{11} = 0.65, p_{22} = 0.6, p_{33} = 0.7, p_{44} = 0.65, p_{55} = 0.8, p_0 = 0.2.$



Fig. 7. $p_{11} = 0.7, p_{22} = 0.7, p_{12} = 0.2, H_0$ is true, each sample consist of 1 graph.

- [5] Aurelien Decelle, Florent Krzakala, Cristopher Moore, and Lenka Zdeborová. Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 84(6):1–25, 2011.
- [6] Chao Gao, Yu Lu, and Harrison H. Zhou. Rate-optimal graphon estimation. Annals of Statistics, 43(6):2624–2652, 2015.
- [7] Debarghya Ghoshdastidar, Maurilio Gutzeit, Alexandra Carpentier, and Ulrike von Luxburg. Two-sample Hypothesis Testing for Inhomogeneous Random

Graphs. pages 1–46, 2017.

- [8] Cedric E. Ginestet, Jun Li, Prakash Balachandran, Steven Rosenberg, and Eric D. Kolaczyk. Hypothesis Testing For Network Data in Functional Neuroimaging. pages 1–34, 2014.
- [9] Daniel R. Hyduke, Nathan E. Lewis, and Bernhard Ø. Palsson. Analysis of omics data with genome-scale models of metabolism. *Mol. BioSyst.*, 9(2):167–174, 2013.
- [10] László Lovász. Large networks and graph limits. 2012, page 487, 2012.

A Proofs

A.1 Proof of Theorem 5

Proof. First, rewrite (5) in the following way:

$$\mathcal{K}(p,q) = -p \log \left(1 - \frac{p-q}{p}\right) - (1-p) \log \left(1 + \frac{p-q}{1-p}\right).$$

By Taylor's theorem:

$$-\log(1-x) = x + \frac{x^2}{2} + \frac{1}{(1-\xi)^3} \frac{x^3}{3}, \xi \in [0;x],$$

$$-\log(1+y) = -y + \frac{y^2}{2} - \frac{1}{(1-\eta)^3} \frac{y^3}{3}, \eta \in [0;y].$$

Let us denote

$$\begin{split} \hat{p}_0 &= \hat{P}_{ij}, \ i \in |z_G^{-1}(i)|, j \in |z_G^{-1}(j)|, \ i \neq j, \\ \hat{p}_m &= \hat{P}_{ij}, \ i = j \in |z_G^{-1}(m)|, \ m \in [k], \\ \hat{q}_0 &= \hat{Q}_{ij}, \ i \in |z_H^{-1}(i)|, j \in |z_H^{-1}(j)|, \ i \neq j, \\ \hat{q}_m &= \hat{Q}_{ij}, \ i = j \in |z_H^{-1}(m)|, \ m \in [k]. \end{split}$$

Obviously if \mathcal{H}_0 is true, we can represent our test statistic:

$$T_{CD} = 2\sum_{m=0}^{k} N_m \mathcal{K}(\hat{p}_m, \hat{q}_m).$$

Then for any $m = 0, \ldots, k$ we can rewrite:

$$2N_m \mathcal{K}(\hat{p}_m, \hat{q}_m)$$

=2 $N_m \Big[-\hat{p}_m \log \Big(1 - \frac{\hat{p}_m - \hat{q}_m}{\hat{p}_m} \Big) - (1 - \hat{p}_m) \log \Big(1 + \frac{\hat{p}_m - \hat{q}_m}{1 - \hat{p}_m} \Big) \Big]$
= $N_m \Big(\frac{(\hat{p}_m - \hat{q}_m)^2}{\hat{p}_m (1 - \hat{p}_m)} \Big) + R_m^{(1)} + R_m^{(2)},$

where

$$\begin{split} R_m^{(1)} &= \frac{2N_m}{3} \frac{1}{(1-\xi)^3} \frac{(\hat{p}_m - \hat{q}_m)^3}{(\hat{p}_m)^2} = N_m (\hat{p}_m - \hat{q}_m)^3 \frac{2}{3(1-\xi)^3} \frac{1}{(\hat{p}_m)^2},\\ R_m^{(2)} &= -\frac{2N_m}{3} \frac{1}{(1-\eta)^3} \frac{(\hat{p}_m - \hat{q}_m)^3}{(1-\hat{p}_m)^2} \end{split}$$

are residuals with random variables $\xi \in \left[0, \frac{\hat{p}_m - \hat{q}_m}{\hat{p}_m}\right], \eta \in \left[0, \frac{\hat{p}_m - \hat{q}_m}{1 - \hat{p}_m}\right].$

Let us show that $R_m^{(1)}, R_m^{(2)} \xrightarrow{\mathbb{P}} 0$. Firstly, consider the case when $\hat{p}_m, \hat{q}_m \in [\theta_m - \delta, \theta_m + \delta]$ for some δ . If $\hat{p}_m > \hat{q}_m$, then

$$\frac{2}{3(1-\xi)^3} \frac{1}{(\hat{p}_m)^2} \le \frac{2}{3(1-\frac{\hat{p}_m-\hat{q}_m}{\hat{p}_m})^3} \frac{1}{(\hat{p}_m)^2} = \frac{2\hat{p}_m}{3(\hat{q}_m)^3} \le \frac{2}{3} \frac{\theta_m + \delta}{(\theta_m - \delta)^3} < \frac{2}{3} \frac{\frac{3}{2}\theta_m}{(\frac{1}{2})^3 \theta_m^3} = \frac{8}{\theta_m^2}.$$

If $\hat{p}_m \leq \hat{q}_m$, then

$$\frac{2}{3(1-\xi)^3} \frac{1}{(\hat{p}_m)^2} \le \frac{2}{3} \frac{1}{(\theta_m - \delta)^2} < \frac{2}{3} \frac{4}{\theta_m^2} < \frac{8}{\theta_m^2}$$

For any $\varepsilon > 0$ denote an event $\mathcal{A} = \{ \left| N_m (\hat{p}_m - \hat{q}_m)^3 \right| \cdot \left| \frac{2}{3(1-\xi)^3} \frac{1}{(\hat{p}_m)^2} \right| > \varepsilon \}$, then

$$\mathbf{P}(\mathcal{A}) = \mathbf{P}\left(\mathcal{A} \middle| \hat{p}_{m}, \hat{q}_{m} \in [\theta_{m} - \delta, \theta_{m} + \delta]\right) \cdot \mathbf{P}\left(\hat{p}_{m}, \hat{q}_{m} \in [\theta_{m} - \delta, \theta_{m} + \delta]\right) \\
+ \mathbf{P}\left(\mathcal{A} \middle| \hat{p}_{m}, \hat{q}_{m} \notin [\theta_{m} - \delta, \theta_{m} + \delta]\right) \cdot \mathbf{P}\left(\hat{p}_{m}, \hat{q}_{m} \notin [\theta_{m} - \delta, \theta_{m} + \delta]\right) \\
\leq \mathbf{P}\left(\mathcal{A} \middle| \hat{p}_{m}, \hat{q}_{m} \in [\theta_{m} - \delta, \theta_{m} + \delta]\right) + \mathbf{P}\left(\hat{p}_{m}, \hat{q}_{m} \notin [\theta_{m} - \delta, \theta_{m} + \delta]\right).$$

Since δ is fixed, $\mathbb{P}(\hat{p}_m, \hat{q}_m \notin [\theta_m - \delta, \theta_m + \delta]) \xrightarrow{\mathbb{P}} 0$ with $N_m \to \infty, m \ge 0$ by the law of large numbers. And

$$\begin{split} & \mathbb{P}\left(\mathcal{A}\Big|\hat{p}_{m},\hat{q}_{m}\in[\theta_{m}-\delta,\theta_{m}+\delta]\right) \\ =& \mathbb{P}\left(\left|N_{m}(\hat{p}_{m}-\hat{q}_{m})^{3}\right|\cdot\Big|\frac{2}{3(1-\xi)^{3}}\frac{1}{(\hat{p}_{m})^{2}}\Big|>\varepsilon\Big|\hat{p}_{m},\hat{q}_{m}\in[\theta_{m}-\delta,\theta_{m}+\delta]\right) \\ \leq& \mathbb{P}\left(\left|N_{m}(\hat{p}_{m}-\hat{q}_{m})^{3}\right|\cdot\frac{8}{\theta_{m}^{2}}>\varepsilon\Big|\hat{p}_{m},\hat{q}_{m}\in[\theta_{m}-\delta,\theta_{m}+\delta]\right) \\ \leq& \frac{\mathbb{P}\left(\left|N_{m}(\hat{p}_{m}-\hat{q}_{m})^{3}\right|\cdot\frac{8}{\theta_{m}^{2}}>\varepsilon\right)}{\mathbb{P}\left(\hat{p}_{m},\hat{q}_{m}\in[\theta_{m}-\delta,\theta_{m}+\delta]\right)}\leq2\mathbb{P}\left(\left|N_{m}(\hat{p}_{m}-\hat{q}_{m})^{3}\right|\cdot\frac{8}{\theta_{m}^{2}}>\varepsilon\right) \\ =& 2\mathbb{P}\left(\left|(\hat{p}_{m}-\hat{q}_{m})\right|>\left(\frac{\varepsilon\theta_{11}^{2}}{8N_{m}}\right)^{\frac{1}{3}}\right) \\ \leq& 2\mathbb{P}\left(\left|(\hat{p}_{m}-\theta_{m})\right|>\left(\frac{\varepsilon\theta_{11}^{2}}{8N_{m}}\right)^{\frac{1}{3}}\right)+2\mathbb{P}\left(\left|(\hat{q}_{m}-\theta_{m})\right|>\left(\frac{\varepsilon\theta_{11}^{2}}{8N_{m}}\right)^{\frac{1}{3}}\right)\xrightarrow{\mathbb{P}}0 \end{split}$$

with $N_m, N_{12}, N_{22} \to \infty$ by the Hoeffding's inequality. Thus, $R_m^{(1)} \xrightarrow{\mathbb{P}} 0$. Similarly one can show that $R_m^{(2)} \xrightarrow{\mathbb{P}} 0$ with $N_m \xrightarrow{\mathbb{P}} 0, m \ge 0$. Also, $N_m \left(\frac{(\hat{p}_m - \hat{q}_m)^2}{\hat{p}_m (1 - \hat{p}_m)} \right) - N_m \left(\frac{(\hat{p}_m - \theta_m)^2}{\theta_m (1 - \theta_m)} \right) \xrightarrow{\mathbb{P}} 0$ with $N_m \to \infty$ (by the same way).

So, we obtain that $2N_m \mathcal{K}(\hat{p}_m, \hat{q}_m) \xrightarrow{\mathbb{P}} N_m \left(\frac{(\hat{p}_m - \theta_m)^2}{\theta_m (1 - \theta_m)}\right)$ with $N_m \to \infty$. It is a squared standard normal random variable since $\mathbb{E}\hat{p}_m = \theta_m$ and $\operatorname{Var} \hat{p}_m = \sqrt{\theta_m (1 - \theta_m)}$.

Hence,

$$T_{CD} \xrightarrow{\mathbb{P}} \sum_{m=0}^{k} N_m \Big(\frac{(\hat{p}_m - \hat{q}_m)^2}{\theta_m (1 - \theta_m)} \Big),$$

which is the sum of m + 1 independent squared standard normal random variables.
Сравнительный анализ: оценка влияния шизофрении на форму субкортикальных структур головного мозга

Аягоз Мусабаева¹, Алексей Крошнин^{1,2}, Анвар Курмуков^{1,2}, Юлия Денисова¹ и Борис Гутман^{1,3}

¹ Институт проблем передачи информации им. А.А. Харкевича РАН
 ² Национальный исследовательский университет «Высшая школа экономики»
 ³ Иллинойский технологический институт

amusabaeva@edu.hse.ru

Аннотация В работе продемонстрирован метод оценивания влияния болезни шизофрении на изменение субкортикальных структур головного мозга. Один из современных способов визуализировать эти изменения - это использование методов регистрации изображений, таких как Large Deformation Diffeomorphic Metric Mapping (LDDMM). Суть метода LDDMM заключается в том, что он индуцирует Риманову метрику на пространстве диффеоморфизмов, что позволяет получить ядро и использовать его в контексте машинного обучения. Для простоты в данной работе в качестве алгоритма машинного обучения был выбран Линейный Ядерный Дискриминантный Анализ Фишера (KLDA). Использовалась оптимизация параметров ядра, индуцирующего норму дифференциального оператора LDDMM, при решении задачи групповой регистрации и классификации в рамках ЕМ-алгоритма. Данная модель продемонстрировала, что самый значительные изменения и максимальная информативность в смысле задачи классификации наблюдаются у левого гиппокампа и правого путамена субкортикальной структуры головного на основе ROC AUC.

1 Введение

Методы машинного обучения всё чаще применяются в современных нейронауках. В частности, естественным образом в этой предметной области возникает задача классификации - автоматизированного различения групп здоровых людей и пациентов с различными патологиями, в частности нейродегенеративными и психиатрическими заболеваниями. В качестве входных данных в такого рода анализе используются тем или иным образом предварительно обработанные данные неинвазивной нейровизуализации.

Обработка данных нейровизуализации, поступающих на вход классификационного алгоритма, может различаться существенным образом. В простейшем случае классификация строится непосредственно на основе предобработанных медицинских изображений мозга той или иной модальности. Регистрация изображений является основным шагом предобработки данных для анализа в медицинской визуализации. Это базовый шаг анализа медицинских изображений, и различные методы регистрации активно рассматривалась в течение последних десятилетий, в результате чего было предложено несколько разновидностей надежных алгоритмов [1]. Ряд подходов к регистрации рассматривает метрическое пространство, состоящее из геометрических объектов, которые представляют интерес, как например анатомические формы, диффузионные тензоры, изображения и т. д. Важное место среди известных сегодня методов занимает Large Deformation Diffeomorphic Metric Mapping (LDDMM) [2]. Вместо того, чтобы рассматривать изображения как объекты, представляющие интерес непосредственно, LDDMM строит пространство диффеоморфизмов на многообразии с использованием Римановой метрики на полях скоростей. Эта метрика индуцируется дифференциальным оператором, который сразу контролирует природу метрического пространства и регулирует регистрацию.

Структура такого пространства представляет собой многообразие гладких отображений с четко определенными геодезическими, что позволяет обобщить несколько стандартных методов статистического анализа. Эти методы неоднократно показывали, что они улучшают чувствительность и способность различать динамику изменения структуры по сравнению с проецированием данных на евклидово пространство. Работы в этой области включают вычисление геометрической медианы и метрической оптимизации для надежной оценки атласа головного мозга [3][4], геодезической регрессии временного ряда [5] и анализ главных геодезических компонент [6]. За исключением [4], в перечисленных выше работах метрика считается фиксированной.

В нашей работе мы пользуемся наличием метрики и используем риманово скалярное произведение, которым наделено пространство LDDMM, как ядро классификатора, которым для простоты и наглядности является ядерный дискриминантный анализ Фишера. Точность классификации изображений нормальных мозговых структур и мозговых структур с патологией измеряется как площадь под ROC-кривой (ROC AUC).Используя этот современный подход, в нашей работе мы показываем, как возможно детектировать изменение субкортикальных структур головного мозга под влиянием шизофрении. Эксперименты по классификации форм субкортикальных структур, извлеченных на основе снимков MPT головного мозга из базы данных Schizconnect [7], показали, что левый гиппокамп и правый путамен подвержены наиболее явным изменениям в геометрической форме под влиянием шизофрении.

2 Методы

2.1 Метрика на пространстве диффеоморфизмов

Метод Large Deformation Diffeomorphic Metric Mapping (LDDMM) был впервые представлен в работе [2]. Цель регистрации состоит в том, чтобы найти диффеоморфизм $\phi: \Omega \to \Omega$, где Ω - это пространство изображений. Диффеоморфизм ϕ порождается потоком или полем скоростей v, который зависит от времени и определяется следующим образом:

$$\frac{\partial \phi(t,x)}{\partial t} = v(t,\phi(t,x)),$$

$$\phi(0,x) = id,$$
(1)

где id - это тождественный оператор: $id(x) = x, \forall x \in \Omega$. Это уравнение задает путь на пространстве диффеоморфизмов $\phi_t \colon \Omega \to \Omega, t \in [0, 1]$. Оптимальный диффеоморфизм находится как решение задачи оптимизации в пространстве гладких полей скоростей V на Ω и принимает следующий вид, с краевыми условиями определенными уравнением [];

$$v^* = \underset{v}{\operatorname{argmin}} \left(\int_{0}^{1} \|v_t\|_{L}^{2} dt + \frac{1}{\sigma^2} \|I_0 \circ \phi - I_1\|^{2} \right).$$
(2)

Первое слагаемое в выражении 2 является одной из самых распространенных мер сходства, а требуемая гладкость обеспечивается регуляризацией, определенной нормой поля скоростей с заданной на касательном пространстве римановой метрикой L, входящей в уравнение 2

Риманова метрика L естественным образом определяет геометрическую структуру пространства. Скалярное произведение $||v||_L^2 = \langle Lv, v \rangle$ можно рассматривать как метрику между изображениями, т. е. длиной геодезической на пространстве диффеоморфизмов. Поскольку пространство диффеоморфизмов является группой Ли относительно уравнения [], то риманова метрика является правоинвариантной. В оригинальной работе об LDDMM [2] оператор L определяется как гладкий дифференциальный самосопряженный оператор вида $L = (\alpha \Delta + \beta E)^2$, где E- тождественный оператор, который также называется бигармоническим оператором. Параметры (α, β) соответствуют выпуклости и нормализации (евклидовости) пространства соответственно. Эти параметры существенно влияют на качество регистрации, поэтому используется автоматический метод подбора параметров.

2.2 Модель машинного обучения

Рассмотрим задачу стандартной бинарной классификации: дана выборка $(x_i, y(x_i))_{i=1}^n$, где $y(x_i) \in \{1, -1\}$ - это класс, а классификационная модель определяется функцией \hat{y} , аппроксимирующей истинную y. Одним из стандартных линейных методов анализа статистических данных является линейный дискриминантный анализ Фишера (LDA). Ядерный дискриминантный анализ Фишера (KLDA), введенный в [S], является обобщением классического LDA. Существует несколько подходов для улучшения качества разделимости классов. KLDA изменяет пространство признаков, переводя его в гильбертово пространство, наделенное определенной структурой (RKHS[9]). Это пространство индуцируется ядром k, что соответствует нелинейной функции решения в исходном, или "входном пространстве. Основная идея LDA состоит в том, чтобы найти одномерную проекцию w в пространстве признаков, которая максимизирует межклассовую дисперсию при минимизации внутриклассовой дисперсии. KLDA ищет аналогичную проекцию в новом пространстве признаков, где вычисляются средние (M_z) и ковариационные матрицы (Σ_z) для каждого класса $z \in \{-1, 1\}$. Функция стоимости (K)LDA принимает следующий квадратично-рациональный вид:

$$J(w) = \frac{w^T (M_1 - M_{-1})(M_1 - M_{-1})^T w}{w^T (\Sigma_{-1} + \Sigma_1) w} = \frac{w^T M w}{w^T N w},$$
(3)

где

$$(M_z)_i = \frac{1}{n_z} \sum_{x_\ell: y(x_\ell) = z} k(x_i, x_\ell),$$
$$(\Sigma_z)_{i,j} = \frac{1}{n_z} \sum_{x_\ell: y(x_\ell) = z} k(x_i, x_\ell) k(x_j, x_\ell) - (M_z)_i (M_z)_j.$$

Здесь n_z - количество объектов из класса z в примере.

Явное решение задачи $J(w) \to \min$ получено аналитически и записывается явной формулой $\hat{J} = N^{-1}(M_1 - M_{-1})$. Функция принятия решения для нового наблюдения x основана на прогнозируемом расстоянии до средних обучающей выборки, $w^T(M_z - x)$. В качестве функции потерь для оптимизации параметров ядра используется так называемый Hinge Loss, определенный для KLDA как

$$h(x', z) = \max\{0, 1 - z y(x')\}$$

$$y(x') = \sum_{i=1}^{n} w_i \left(k(x_i, x') - \frac{(M_1)_i + (M_{-1})_i}{2} \right),$$
 (4)

где $z \in \{-1, 1\}$ является истинным значением класса для нового наблюдения x' и $K(x_i, x')$ задает расстояние между x' и наблюдениями из обучающей выборки x_i . Хотя формулировка Hinge Loss в конечном итоге приводит к некоторому локальному оптимуму, но при этом дает некоторую стабильность в обучении, в отличии от альтернативных методов, таких как максимизация правдоподобия.

2.3 Метод оптимизации параметров ядра LDDMM

Для оптимизации параметров ядра регистрации LDDMM в смысле решения задачи классификации изображений в нашей работе выбран автоматический метод. Он заключается в следующем.

Пусть риманова метрика определяет ядро $K_L(\alpha, \beta) = \langle Lv, v \rangle$. На практике параметр β играет незначительную роль и может быть опущен (так как умножение скорости на константу не меняет задачу оптимизации в

LDDMM). Поэтому оптимизация происходит только по параметру α , фиксируя $\beta = 1$ в качестве члена нормализации.

Нахождение оптимального значения параметра метрики происходит с помощью EM-алгоритма. E-шаг заключается в попарной регистрации изображений с каким-то фиксированным значением α и тем самым получаемым ядром $K_L(x_i, x_j)$. Ядро для классификатора определяется как $K(x_i, x_j) =$ $\exp\{-\gamma K_L(x_i, x_j)\}$. Параметр γ оценивается с помощью поиска по сетке, чтобы упростить вычисление, но его также можно оценить по градиентному спуску. Функция потерь определяется [4] при фиксированном значении параметра α . M-шаг состоит в минимизации функции потерь [4] по параметру α . Для реализации градиентного спуска можно выписать явные формулы, которые приведены в оригинальной статье.

$$\frac{dh(x',z)}{d\theta} = \begin{cases} -z\frac{dy(x')}{d\theta}, & \text{if } zy(x') < 1, \\ 0, & \text{otherwise.} \end{cases}$$
(5)

3 Постановка задачи

Цель работы - продемонстрировать эффективность современного метода, описанного выше, для определения на основе МРТ-изображений мозга субкортикальных структур, наиболее существенно различающихся при наличии шизофрении и без патологии.

Логика наших рассуждений такова. Предполагая, что расстояние между нашими объектами задается длиной геодезической, что собственно отражает степень изменения многообразия(субкортикальной структуры), мы ожидаем, что классификатор сможет лучше разделить объекты, принадлежащие различным классам, поскольку явно учитывается соотношение геометрических свойств субкортикальных структур. Говоря о более высоком качестве работы классификатора, мы ожидаем получить не только лучшее качество различения объектов разных классов (более высокое значение площади под ROC-кривой, ROC AUC), но и явным образом визуализировать отличия здоровых пациентов от пациентов с патологией. В нашей работе мы рассматриваем несколько субкортикальных структур, для которых из литературы известно, что они информативны в смысле наличия патологии.

4 Эксперименты

4.1 Данные

Данные были взяты из проекта **17** и представляют собой 3D-изображения субкортикальных структур, выделенных на основе MPT-снимков головного мозга человека. В экспериментах мы использовали правые и левые субкортикальные структуры: гиппокамп, путамен, паллидиум и таламус. Все субкортикальные структуры были выделены с использованием Freesurfer [10]. В данных содержатся снимки 227 пациентов с диагнозом шизофрения и 496 пациентов контрольной группы. Все субкортикальные структуры были предварительно аффинно зарегистрированы на стандартный атлас ENIGMA [11], их бинарная маска строилась при помощи преобразованния мешевой модели данной структуры.

4.2 Алгоритм

В экспериментах участвовали 105 объектов в обучающей выборке и 45 в тестовой, чтобы уменьшить временные затраты на попарные регистрации, количество которых растет как $O(n^2)$, где n размер обучающей выборки. На 105 обучающих объектах проводились все этапы обучения: подсчет ядра (всех попарных расстояний), подбор параметра γ , подсчет результата классификации и нахождение градиента для параметров ядра. Для 45 тестовых объектов были подсчитаны только расстояния до обучающих объектов. Эта схема была проделана для 5 разных разбиений.

4.3 Результаты

Результаты работы алгоритма классификации шизофрении и нормы на основе изображений левых и правых субкортикальных структур головного мозга с использованием оптимизации параметров ядра LDDMM представлены в таблице [].

	Hippocampus	Thalamus	Putamen	Pallidum
left	0.83 ± 0.05	0.64 ± 0.05	0.73 ± 0.10	$0.76 {\pm} 0.06$
right	0.72 ± 0.07	$0.70{\pm}0.08$	0.81 ± 0.10	0.74 ± 0.05

Таблица 1: Метрика ROC AUC для четырех видов субкортикальных структур (левые и правые). Для ROC AUC приведены среднее и стандартное отклонение максимальных значений по 5 реализациям обучающей траектории.

Результаты оптимизации LDDMM-ядра для 5 разных разбиений субкортикальных структур приведены на Рисунке 1. Очевидно, что алгоритм сходится к локальному оптимуму; обучение зависит от способа разбиения и информативности каждой из частей. При этом известно, что вариабельность даже объемов различных структур мозга может отличаться в разных группах.

Тем не менее, полученные данные показывают, что в контексте задачи различения шизофрении и нормы лучшей предсказательной способностью облает левый гиппокамп и правый путамен. При этом вполне информативными с точки зрения предсказательной способности являются и остальные



Рис. 1: Результаты оптимизации LDDMM-ядра для 5 разных разбиений субкортикальных структур (сверху вниз): гиппокампа, паллидума, путамена и таламуса, соответственно левых (в первом столбце) и правых (в правом столбце). Фиолетовая линия соответствует средней линии обучения.

субкортикальные структуры. Полученные нами результаты хорошо соотносятся с современными исследованиями 12, демонстрирующими влиняние

изменений в субкортикальных структурах на социальные и когнитивные способности человека при шизофрении.

5 Заключение

В данной работе мы продемонстрировали, каким образом современный подход к регистрации изображений может быть использован для выявления наиболее существенных различий в субкортикальных структурах мозга в контексте задачи классификации шизофрении и нормы. Мы использовали полученные с помощью LDDMM расстояния и автоматическое выучивание метрики, чтобы построить ядро KLDA-классификатора для различения объектов из групп нормального развития и диагностированной шизофрении. Мы показали, что левый гиппокамп и правый путамен имеют наибольшую предсказательную ценность, а значит являются наиболее значимыми идентификаторами наличия шизофрении.

Наша работа требует продолжения, в первую очередь – верификации полученных результатов на других наборах данных и в других классификационных задачах из данной предметной области. Кроме того, планируется усложнение и модификация самого алгоритма для получения более устойчивых результатов.

Список литературы

- Klein, A., Andersson, J., Ardekani, B.A., Ashburner, J., Avants, B., Chiang, M.C., Christensen, G.E., Collins, D.L., Gee, J., Hellier, P., Song, J.H., Jenkinson, M., Lepage, C., Rueckert, D., Thompson, P., Vercauteren, T., Woods, R.P., Mann, J.J., Parsey, R.V.: Evaluation of 14 nonlinear deformation algorithms applied to human brain mri registration. Neuroimage 46(3) (2009) 786–802
- Beg, M.F., Miller, M.I., Trouvé, A., Younes, L.: Computing large deformation metric mappings via geodesic flows of diffeomorphisms. Int. J. Comput. Vision 61(2) (February 2005) 139–157
- Fletcher, P.T., Venkatasubramanian, S., Joshi, S.C.: The geometric median on riemannian manifolds with application to robust atlas estimation. NeuroImage 45 1 Suppl (2009) S143–52
- Zhang, M., Singh, N., Fletcher, P.T.: Bayesian estimation of regularization and atlas building in diffeomorphic image registration. In Gee, J.C., Joshi, S., Pohl, K.M., Wells, W.M., Zöllei, L., eds.: Information Processing in Medical Imaging, Berlin, Heidelberg, Springer Berlin Heidelberg (2013) 37–48
- Hong, Y., Golland, P., Zhang, M.: Fast geodesic regression for populationbased image analysis. In Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D.L., Duchesne, S., eds.: Medical Image Computing and Computer Assisted Intervention â[^] MICCAI 2017, Cham, Springer International Publishing (2017) 317–325
- Zhang, M., Fletcher, P.T.: Probabilistic principal geodesic analysis. In: Proceedings of the 26th International Conference on Neural Information Processing Systems -Volume 1. NIPS'13, USA, Curran Associates Inc. (2013) 1178–1186

- 7. Wang, L., Alpert, K.I., Calhoun, V.D., Cobia, D.J., Keator, D.B., King, M.D., Kogan, A., Landis, D., Tallis, M., Turner, M.D., Potkin, S.G., Turner, J.A., Ambite, J.L.: Schizconnect: Mediating neuroimaging databases on schizophrenia and related disorders for large-scale integration. NeuroImage **124** (2016) 1155 1167 Sharing the wealth: Brain Imaging Repositories in 2015.
- Mika, S., Ratsch, G., Weston, J., Scholkopf, B., Mullers, K.R.: Fisher discriminant analysis with kernels. In: Neural networks for signal processing IX, 1999. Proceedings of the 1999 IEEE signal processing society workshop., Ieee (1999) 41–48
- 9. Aronszajn, N.: Theory of reproducing kernels. Transactions of the American mathematical society **68**(3) (1950) 337–404
- 10. Fischl, B.: Freesurfer. Neuroimage 62(2) (2012) 774-781
- Roshchupkin*, G.V., Gutman*, B.A., Vernooij, M.W., Jahanshad, N., Martin, N.G., Hofman, A., McMahon, K.L., van der Lee, S.J., van Duijn, C.M., de Zubicaray, G.I., Uitterlinden, A.G., Wright, M.J., Niessen, W.J., Thompson, P.M., Ikram**, M.A., Adams**, H.H.H.: Heritability of the shape of subcortical brain structures in the general population. Nature Communications 7 (2016) 13738
- Koshiyama, D., Fukunaga, M., Okada, N., Yamashita, F., Yamamori, H., Yasuda, Y., Fujimoto, M., Ohi, K., Fujino, H., Watanabe, Y., et al.: Role of subcortical structures on cognitive and social function in schizophrenia. Scientific reports 8(1) (2018) 1183

Центральная предельная теорема для барицентров Васерштейна от гауссовских мер

Алексей Крошнин^{1,2} Александра Суворикова^{1,3}

¹ Институт проблем передачи информации РАН, Москва
 ² Высшая школа экономики, Москва
 ³ Weierstrass Institute for Applied Analysis and Stochastics, Berlin gorianzwei@gmail.com

Аннотация. В работе рассматривается пространство многомерных нормальных распределений, снабженное транспортным расстоянием 2-Васерштейна. Исследуются дифференциальные свойства оптимального транспортного отображения между такими распределениями. Рассматриваются обобщенные средние по Фреше от мер, определенные на основе транспортного расстояния, — барицентры Васерштейна. Для барицентров установлен аналог центральной предельной теоремы, уточняющий известные ранее результаты типа закона больших чисел.

Ключевые слова: барицентр Васерштейна; нормальное распределение; центральная предельная теорема.

1 Введение

Объекты, которые можно рассматривать как вероятностные распределения, возникают в самых разных областях. Зачастую необходимо исследовать их, учитывая геометрию пространства, на котором они определены. Так, один из естественных способов определить "геометрическую" метрику между распределениями — это транспортное расстояние (расстояние Монжа-Канторовича, Васерштейна, Канторовича-Рубинштейна и т.д.) Рассмотрим пространство $\mathcal{P}_2(\mathbb{R}^d)$ вероятностных мер на \mathbb{R}^d , обладающих конечным вторым моментом. *Расстояние 2-Васерштейна* между мерами $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ определяется следующим образом [1,2]:

$$W_2^2(\mu,\nu) := \min_{\gamma \in \Pi(\mu,\nu)} \int \|x - y\|_2^2 \, d\gamma(x,y),$$

где минимум берется по множеству *транспортных планов* из μ в ν , т.е. по всем мерам $\gamma \in \mathcal{P}\left(\mathbb{R}^d \times \mathbb{R}^d\right)$ с маргинальными распределениями μ и ν , соответственно.

Наделив пространство мер метрической структурой, можно определить среднее от распределений (которое мы в данном случае будем называть барицентром), как среднее по Фреше, которое является естественным обобщением линейного среднего на нелинейные метрические пространства. Пусть дана случайная мера $\mu \in \mathcal{P}_2(\mathbb{R}^d)$, имеющая распределение P и конечное мат. ожидание дисперсии. Барицентром Васерштейна распределения P называется мера ν_* , минимизирующая средний квадрат отклонения [3]:

$$\nu_* = \operatorname{bar}(P) := \operatorname{argmin}_{\nu \in \mathcal{P}(\mathbb{R}^d)} \mathbb{E} W_2^2(\mu, \nu).$$

Известно, что для барицентров Васерштейна выполняется закон больших чисел [4,5,6]: если дана i.i.d. последовательность мер μ_1, μ_2, \ldots , то эмпирические барицентры

$$\nu_n = \operatorname{bar}(\mu_1, \dots, \mu_n) := \operatorname*{argmin}_{\nu} \frac{1}{n} \sum_{i=1}^n W_2^2(\mu_i, \nu)$$

сходятся к барицентру распределения $W_2(\nu_n, \nu_*) \to 0$. Тем не менее, в общем случае на данный момент не известно никаких результатов о скорости сходимости барицентров, либо об их предельном распределении.

В данной работе мы рассмотрим частный случай, а именно — нормальные распределения на \mathbb{R}^d (более общо, можно рассматривать так называемые *scatter-location* семейства []), и покажем, что для эмпирических барицентров нормальных мер имеет место центральная предельная теорема.

Для гауссовских мер $\mu = \mathcal{N}(r, Q)$ и $\nu = \mathcal{N}(m, S)$ расстояние 2-Васерштейна можно выразить в явном виде [7,2]:

$$W_2^2(\mu,\nu) = \|r - m\|_2^2 + \operatorname{tr} S + \operatorname{tr} Q - 2\operatorname{tr} \left(S^{1/2}QS^{1/2}\right)^{1/2}$$

Более того, если Q и S невырожденные, то существует оптимальное транспортное отображение, т.е. оптимальный транспортный план γ^* сосредоточен на графике отображения $T^{\nu}_{\mu} \colon \mathbb{R}^d \to \mathbb{R}^d$, переводящего меру μ в меру ν . Данное отображение тоже представимо в явном виде

$$T^{\nu}_{\mu}(x) := m - r + T^{S}_{Q}x, \quad T^{S}_{Q} := S^{1/2} \left(S^{1/2} Q S^{1/2} \right)^{-1/2} S^{1/2}. \tag{1}$$

Известно, что семейство невырожденных гауссовских мер замкнуто относительно взятия барицентра, а именно: если $\mathcal{N}(m,S) \sim \mathbf{P}$, то $\operatorname{bar}(\mathbf{P}) = \mathcal{N}(r,Q)$, где $r = \mathbb{E} m$ и Q — единственное положительно определенное решение уравнения

$$Q = \mathbb{E}\left(S^{1/2}QS^{1/2}\right)^{1/2}$$

или, что эквивалентно,

$$\mathbb{E}T_O^S = I$$

где *I* — единичная матрица [3].

Таким образом, вопрос исследования барицентров Васерштейна от гауссовских мер сводится к задаче на конечномерном пространстве средних и ковариационных матриц $\mathbb{R}^d \times \text{Sym}_+(d)$. Так как средние входят в выражение для барицентра линейно, то без потери общности можно ограничиться рассмотрением центрированных распределений. Соответственно, в дальнейшем мы будем отождествлять центрированную гауссовскую меру с ее ковариационной матрицей и говорить о распределении и барицентрах на множестве положительно определенных матриц размера $d \times d$.

Работа построена следующим образом: в Разделе 2 показывается, что отображение $Q \mapsto T_Q^S$ дифференцируемо, и исследуются его свойства. В Разделе 3 непосредственно доказывается ЦПТ для эмпирических барицентров.

1.1 Используемые обозначения

Через $\|\cdot\|_F$ будем обозначать норму Фробениуса (Гильберта-Шмидта) матрицы:

$$||A||_{F}^{2} := \operatorname{tr} A^{\top} A$$

Соответствующее скалярное произведение матриц обозначим как $\langle \cdot, \cdot \rangle$.

 $\lambda_{\max}(A)$ и $\lambda_{\min}(A)$ означают, соответственно, максимальное и минимальное собственные значения матрицы A.

Множество всех симметричных $d \times d$ матриц обозначим через Sym(d), симметричных положительно определенных матриц — через $Sym_+(d)$.

2 Дифференцирование оптимального транспортного отображения

В Лемме 1 показано, что матрица оптимального транспортного отображения T_O^S , определенная в (1), дифференцируема по отношению к Q:

$$T^S_{Q+X} = T^S_Q + DT(X;Q,S) + o\bigl(\|X\|\bigr), \quad X \to 0,$$

где $DT(\cdot; Q, S)$: Sym $(d) \to$ Sym(d) — самосопряженный отрицательно определенный оператор. Его свойства исследуются в Лемме 2.

Лемма 1. Для любых матриц $Q, S \in \text{Sym}_+(d)$ выполняется

$$T_{Q'}^{S} = T_{Q}^{S} + DT(Q' - Q; Q, S) + o(||Q' - Q||), \quad Q' \to Q,$$

где

$$DT(X;Q,S) := -S^{1/2}O^{\top}\Lambda^{-1/2}\delta\Lambda^{-1/2}OS^{1/2},$$
(2)

 $O^{\top} \Lambda O$ — спектральное разложение $S^{1/2} Q S^{1/2}$:

$$O^{\top} \Lambda O = S^{1/2} Q S^{1/2}, \quad O^{\top} O = O O^{\top} = I, \quad \Lambda = \operatorname{diag}(\lambda);$$

u

$$\delta = (\delta_{ij})_{i,j=1}^d, \quad \delta_{ij} := \frac{\Delta_{ij}}{\sqrt{\lambda_i} + \sqrt{\lambda_j}}, \quad O^{\top} \Delta O = S^{1/2} X S^{1/2}.$$

Доказательство. Доказательство основывается на дифференцировании члена $(S^{1/2}QS^{1/2})^{-1/2}$, так как

$$DT(X;Q,S) = S^{1/2} \left[d_Q \left(S^{1/2} Q S^{1/2} \right)^{-1/2} \right] (X) S^{1/2}.$$

Рассмотрим спектральное разложение $S^{1/2}QS^{1/2}$:

$$S^{1/2}QS^{1/2} = O^{\top}\Lambda O,$$
 (3)

где $\Lambda=\mathrm{diag}(\lambda),$ а O-ортогональная матрица. Зафиксируем произвольный достаточно маленький $X\in \mathrm{Sym}(d)$ (такой, что $Q+X\succ 0)$ и определим $\Delta:=OS^{1/2}XS^{1/2}O^{\top}$ (Δ не обязательно диагональная). Матрицы T_Q^S и T_{Q+X}^S можно переписать в следующем виде:

$$T_Q^S = S^{1/2} O^{\top} \Lambda^{-1/2} O S^{1/2}, \quad T_{Q+X}^S = S^{1/2} O^{\top} \left(\Lambda + \Delta\right)^{-1/2} O S^{1/2}.$$
(4)

Рассмотрим разложение Тейлора

$$\left(\Lambda + \Delta\right)^{1/2} = \Lambda^{1/2} + \delta(\Delta; \Lambda) + o(\|\Delta\|), \tag{5}$$

с линейным оператором δ : Sym $(d) \rightarrow$ Sym(d). В дальнейшем мы опустим зависимость от Λ и Δ , и будем использовать просто δ . Возведя выражение выше в квадрат, имеем

$$\Lambda + \Delta = \Lambda + \Lambda^{1/2} \delta + \delta \Lambda^{1/2} + o(\|\Delta\|).$$

Таким образом, получаем поэлементное выражение для $\delta = (\delta_{ij})$:

$$\delta_{ij} = \frac{\Delta_{ij}}{\sqrt{\lambda_i} + \sqrt{\lambda_j}}.\tag{6}$$

Теперь обратим (5) и применим разложение в ряд Неймана:

$$(\Lambda + \Delta)^{-1/2} = \left(\Lambda^{1/4} \left(I + \Lambda^{-1/4} \delta \Lambda^{-1/4} + o(\|\Delta\|) \right) \Lambda^{1/4} \right)^{-1} =$$

= $\Lambda^{-1/4} \left(I - \Lambda^{-1/4} \delta \Lambda^{-1/4} + o(\|\Delta\|) \right) \Lambda^{-1/4} =$
= $\Lambda^{-1/2} - \Lambda^{-1/2} \delta \Lambda^{-1/2} + o(\|\Delta\|).$

Следовательно, дифференциал $\left[d_Q \left(S^{1/2} Q S^{1/2} \right)^{-1/2} \right] (X)$ имеет вид

$$\left[d_Q \left(S^{1/2} Q S^{1/2}\right)^{-1/2}\right] (X) = O^{\top} \Lambda^{-1/2} \delta \Lambda^{-1/2} O.$$
(7)

Учитывая (4), (5) и (6), получаем, что

$$T_{Q+X}^{S} = T_{Q}^{S} + DT(X;Q,S) + o(||X||),$$

где DT(X; Q, S) определен равенством (2).

Если это не вызывает неоднозначности, мы будем опускать зависимость от Q или от S, Q одновременно и писать просто DT(X;S) или DT(X), соответственно. Следующие две леммы являются техническими, в них исследуются некоторые свойства оператора $DT(\cdot)$, необходимые для получения ЦПТ.

Лемма 2. Оператор $DT(\cdot) = DT(\cdot; Q, S)$, определенный в (2), обладает следующими свойствами

- (I) самосопряженность;
- (II) отрицательная определенность;
- (III) ограничения на собственные числа: для любого $X \in \operatorname{Sym}(d)$

$$-\langle DT(X), X \rangle \leq \frac{\sqrt{\lambda_{\max}(S^{1/2}QS^{1/2})}}{2} \left\| Q^{-1/2}XQ^{-1/2} \right\|_{F}^{2} \leq \frac{\sqrt{\lambda_{\max}(S^{1/2}QS^{1/2})}}{2\lambda_{\min}^{2}(Q)} \left\| X \right\|_{F}^{2}, \\ -\langle DT(X), X \rangle \geq \frac{\sqrt{\lambda_{\min}(S^{1/2}QS^{1/2})}}{2} \left\| Q^{-1/2}XQ^{-1/2} \right\|_{F}^{2} \geq \frac{\sqrt{\lambda_{\max}(S^{1/2}QS^{1/2})}}{2\lambda_{\max}^{2}(Q)} \left\| X \right\|_{F}^{2};$$

- (IV) однородность степени $-\frac{3}{2}$ по отношению к $Q: DT(X; aQ, S) = a^{-3/2}DT(X; Q, S)$ для любого a > 0;
- (V) монотонность по отношению к Q: если $Q \preccurlyeq Q'$, то $DT(\cdot;Q,S) \preccurlyeq DT(\cdot;Q',S)$ в смысле самосопряженных операторов на Sym(d).

Доказательство. Перепишем (2) следующим образом:

$$DT(X;Q,S) = -S^{1/2}O^{\top}\Lambda^{-1/2}\delta^{X}\Lambda^{-1/2}OS^{1/2},$$

где матрицы О и Л определены в (3), и

$$\delta^X = (\delta^X_{ij})^d_{i,j=1}, \quad \delta^X_{ij} = \frac{\Delta^X_{ij}}{\sqrt{\lambda_i} + \sqrt{\lambda_j}}, \quad \Delta^X = OS^{1/2}XS^{1/2}O^\top.$$

(I) Самосопряженность. Рассмотрим скалярное произведение

$$\langle DT(X), Y \rangle = \operatorname{tr} \left(DT(X)Y \right) = -\operatorname{tr} \left(S^{1/2} O^{\top} \Lambda^{-1/2} \delta^X \Lambda^{-1/2} O S^{1/2}Y \right) = \\ = -\operatorname{tr} \left(\Lambda^{-1/2} \delta^X \Lambda^{-1/2} O S^{1/2} Y S^{1/2} O^{\top} \right).$$

Обозначим

$$\Delta^Y := OS^{1/2}YS^{1/2}O^\top.$$

Получаем, что

$$-\operatorname{tr}\left(\Lambda^{-1/2}\delta^{X}\Lambda^{-1/2}OS^{1/2}YS^{1/2}O^{\top}\right) = -\operatorname{tr}\left(\Lambda^{-1/2}\delta^{X}\Lambda^{-1/2}\Delta^{Y}\right) =$$
$$= -\sum_{i,j} \frac{\delta^{X}_{ij}}{\sqrt{\lambda_{i}\lambda_{j}}}\Delta^{Y}_{ij} = -\sum_{i,j} \frac{\Delta^{X}_{ij}\Delta^{Y}_{ij}}{\sqrt{\lambda_{i}\lambda_{j}}(\sqrt{\lambda_{i}} + \sqrt{\lambda_{j}})} =$$
$$= \operatorname{tr}\left(DT(Y)X\right) = \operatorname{tr}\left(XDT(Y)\right) = \langle X, DT(Y) \rangle.$$

Т.е., оператор является самосопряженным.

(II) Отрицательная определенность и (III) собственные числа Обозначив Δ^X через Δ и учитывая полученное выше выражение для скалярного произведения, получаем

$$-\langle DT(X), X \rangle = \sum_{i,j=1}^{d} \frac{\Delta_{ij}^{2}}{\sqrt{\lambda_{i}\lambda_{j}}(\sqrt{\lambda_{i}} + \sqrt{\lambda_{j}})} = \sum_{i,j=1}^{d} \left(\frac{\Delta_{ij}}{\sqrt{\lambda_{i}\lambda_{j}}}\right)^{2} \frac{\sqrt{\lambda_{i}\lambda_{j}}}{\sqrt{\lambda_{i}} + \sqrt{\lambda_{j}}}.$$
(8)

(8) Заметим, что функция $f(\lambda_i,\lambda_j):=\frac{\sqrt{\lambda_i\lambda_j}}{\sqrt{\lambda_i}+\sqrt{\lambda_j}}$ возрастает по обоим аргументам, так что

$$\max_{i,j} f(\lambda_i, \lambda_j) = \frac{\sqrt{\lambda_{\max}}}{2}, \quad \min_{i,j} f(\lambda_i, \lambda_j) = \frac{\sqrt{\lambda_{\min}}}{2}, \tag{9}$$

где $\lambda_{\max} = \lambda_{\max}(\Lambda) = \lambda_{\max} \left(S^{1/2} Q S^{1/2} \right)$ и $\lambda_{\min} = \lambda_{\min}(\Lambda) = \lambda_{\min} \left(S^{1/2} Q S^{1/2} \right)$. Для удобства введем новую переменную

$$\xi := Q^{-1/2} X Q^{-1/2},$$

Ее норма равна

$$\|\xi\|_{F}^{2} = \operatorname{tr}(XQ^{-1}XQ^{-1}) = \operatorname{tr}(\Delta\Lambda^{-1}\Delta\Lambda^{-1}) = \|\Lambda^{-1/2}\Delta\Lambda^{-1/2}\|_{F}^{2}$$

Наконец, из (8) и (9), можно получить следующие оценки на скалярное произведение:

$$-\langle DT(X), X \rangle \le \max_{i,j} f(\lambda_i, \lambda_j) \sum_{i,j=1}^d \left(\frac{\Delta_{ij}}{\sqrt{\lambda_i \lambda_j}}\right)^2 = \frac{\sqrt{\lambda_{\max}}}{2} \|\xi\|_F^2$$

И

$$-\langle DT(X), X \rangle \ge \min_{i,j} f(\lambda_i, \lambda_j) \sum_{i,j=1}^d \left(\frac{\Delta_{ij}}{\sqrt{\lambda_i \lambda_j}}\right)^2 = \frac{\sqrt{\lambda_{\min}}}{2} \left\|\xi\right\|_F^2$$

Учитывая, что

$$\frac{\|X\|_F}{\lambda_{\max}Q} \le \|\xi\|_F \le \frac{\|X\|_F}{\lambda_{\min}Q},$$

отсюда следуют оценки на собственные числа $DT(\cdot)$.

(IV) Однородность и (V) монотонность Однородность немедленно вытекает их явного выражения (2). Теперь перейдем к доказательству монотонности. Имеют место следующие равенства

$$\langle DT(X), X \rangle = \operatorname{tr} \left(S^{1/2} O^{\top} \Lambda^{-1/2} \delta \Lambda^{-1/2} O S^{1/2}, X \right) = = \left\langle O^{\top} \Lambda^{-1/2} \delta \Lambda^{-1/2} O, S^{1/2} X S^{1/2} \right\rangle = = \left\langle \left[d_Q \left(S^{1/2} Q S^{1/2} \right)^{-1/2} \right] (X), S^{1/2} X S^{1/2} \right\rangle = = \left\langle \left[d_M M^{-1/2} \right] \left(S^{1/2} X S^{1/2} \right), S^{1/2} X S^{1/2} \right\rangle,$$

где $[d_Q \cdot](X)$ определен в (7), а $M := S^{1/2}QS^{1/2}$. Достаточно показать, что при фиксированном X дифференциал $[d_M M^{-1/2}](\cdot)$ монотонен по M. Заметим, что обратный оператор $[d_M M^{-1/2}]^{-1}(\cdot)$ в точке M равен дифференциалу обратного отображения $[d_P P^{-2}](\cdot)$ в точке $P = M^{-1/2}$:

$$\left[d_M M^{-1/2} \Big|_M \right]^{-1} (\cdot) = \left[d_P P^{-2} \right] \Big|_{P=M^{-1/2}} (\cdot).$$

В свою очередь, легко проверить, что

$$\left[d_P P^{-2}\right](X) = -P^{-1} \left(P^{-1} X + X P^{-1}\right) P^{-1}.$$

Данный оператор является отрицательно определенным и

$$\langle -P^{-1}(P^{-1}X + XP^{-1})P^{-1}, X \rangle = -2 \operatorname{tr} P^{-2}XP^{-1}X.$$

Рассмотрим произвольные $M_1 \succcurlyeq M_0 \succ 0$ (следовательно, $M_1^{1/2} \succcurlyeq M_0^{1/2}$) и $P_i := M_i^{-1/2}, i = 0, 1$. Тогда для любого заданного $X \in \operatorname{Sym}(d)$ выполняется

$$-\operatorname{tr} P_1^{-2} X P_1^{-1} X = -\operatorname{tr} M_1 X M_1^{1/2} X \le -\operatorname{tr} M_0 X M_0^{1/2} X = -\operatorname{tr} P_0^{-2} X P_0^{-1} X,$$

т.е. $\left[d_P P^{-2}\right]\Big|_{P_1}(\cdot) \preccurlyeq \left[d_P P^{-2}\right]\Big|_{P_0}(\cdot)$, и следовательно для дифференциала $M \mapsto M^{-1/2}$ выполнено обратное неравенство:

$$\left[d_M M^{-1/2}\right]\Big|_{M_0}(\cdot) \preccurlyeq \left[d_M M^{-1/2}\right]\Big|_{M_1}(\cdot),$$

что влечет монотонность $DT(\cdot; Q)$.

Лемма 3. Для любых $Q_0, Q_1, S \in \operatorname{Sym}_+(d)$ рассмотрим

$$Q_t := (1-t)Q_0 + tQ_1, \quad Q' := Q_0^{-1/2}Q_1Q_0^{-1/2}.$$

Имеют место следующие неравенства:

$$\begin{split} \frac{2}{\lambda_{\min}(Q') + \sqrt{\lambda_{\min}(Q')}} DT\left(\cdot; Q_0, S\right) \preccurlyeq & \int_0^1 DT(\cdot; Q_t, S) \, dt \preccurlyeq \\ \preccurlyeq \frac{2}{\lambda_{\max}(Q') + \sqrt{\lambda_{\max}(Q')}} DT\left(\cdot; Q_0, S\right). \end{split}$$

Доказательство. Заметим, что

$$Q_t = Q_0^{1/2} ((1-t)I + tQ') Q_0^{1/2}.$$

Из однородности и монотонности $DT(\cdot)$ (см. Лемма 2) следует, что

$$DT(\cdot; Q_t, S) \preccurlyeq DT\left(\cdot; ((1-t) + t\,\lambda_{\max}(Q'))Q_0, S\right) =$$
$$= \left((1-t) + t\,\lambda_{\max}(Q')\right)^{-3/2} DT(\cdot; Q_0, S)$$

И

$$DT(\cdot; Q_t, S) \succcurlyeq DT\left(\cdot; ((1-t) + t\lambda_{\min}(Q'))Q_0, S\right) =$$
$$= \left((1-t) + t\lambda_{\min}(Q')\right)^{-3/2} DT(\cdot; Q_0, S)$$

Таким образом,

$$\int_0^1 DT(\cdot; Q_t, S) dt \preccurlyeq DT(\cdot; Q_0, S) \int_0^1 \left((1-t) + t \lambda_{\max}(Q') \right)^{-3/2} dt =$$
$$= \frac{2}{\lambda_{\max}(Q') + \sqrt{\lambda_{\max}(Q')}} DT(\cdot; Q_0, S)$$

и, соответственно,

$$\int_0^1 DT(\cdot; Q_t, S) \, dt \succcurlyeq \frac{2}{\lambda_{\min}(Q') + \sqrt{\lambda_{\min}(Q')}} DT\left(\cdot; Q_0, S\right).$$

3 ЦПТ для барицентров Васерштейна

Теперь мы можем доказать центральную предельную теорему для эмпирических барицентров Q_n .

Теорема 4 (Центральная предельная теорема для эмпирических барицентров). Пусть дано распределение P на $\text{Sym}_+(d)$, у которого существует барицентр Q_* . Рассмотрим i.i.d. последовательность случайных матриц $S_1, S_2, \dots \sim P$ и эмпирические барицентры $Q_n := \text{bar}(S_1, \dots, S_n)$. Q_n имеют асимптотически нормальное распределение:

$$\sqrt{n}(Q_n - Q_*) \rightharpoonup \mathcal{N}(0, \Xi),$$

где Ξ : Sym $(d) \to$ Sym(d) - следующий линейный оператор:

$$\Xi := F^{-1} \circ \operatorname{Var} \left(T^S_{Q_*} \right) \circ F^{-1}, \quad F(\cdot) := - \operatorname{\mathbb{E}} DT(\cdot; S, Q_*).$$

Доказательство. Введем обозначения

$$T_i := T_{Q_*}^{S_i}, \quad T_i^n := T_{Q_n}^{S_i}.$$

Согласно Лемме 1

$$T_i^n = T_i + \int_0^1 DT(Q_n - Q_*; Q_t, S_i) dt =$$

= $T_i + DT(Q_n - Q_*; Q_*, S_i) + \alpha (Q_n - Q_*; Q_n, S_i)$ (10)

где $Q_t = (1-t)Q_* + tQ_n$ и $\alpha(X;Q_n,S) = o(\|DT(X;Q_*,S)\|)$ при $Q_n \to Q_*$ равномерно по S и X в силу Леммы 3. Заметим, что $\frac{1}{n} \sum T_i^n = I$, так как Q_n — барицентр матриц S_1, \ldots, S_n . Усредняя (10) по i, получаем

$$I = \overline{T}_n - F_n(Q_n - Q_*) + \alpha_n (Q_n - Q_*), \qquad (11)$$

$$\overline{T}_n := \frac{1}{n} \sum_{i=1}^n T_i, \quad F_n(\cdot) := -\frac{1}{n} \sum_{i=1}^n DT(\cdot; Q_*, S_i), \quad \alpha_n(\cdot) := \frac{1}{n} \sum_{i=1}^n \alpha(\cdot; Q_n, S_i).$$

Отметим, что $F_n(\cdot)$ является выборочным аналогом оператора $F(\cdot) := -\mathbb{E} DT(\cdot; Q_*, S)$. Несложно видеть, что $F(\cdot)$ корректно определен; действительно, в силу Леммы 2 он является самосопряженным, положительно определенным и ограниченным:

$$\left\|F\right\|_{op} \leq \mathbb{E}\left\|DT(\cdot;Q_*,S)\right\|_{op} \leq \mathbb{E}\left\frac{\sqrt{\lambda_{\max}(S^{1/2}Q_*S^{1/2})}}{2\,{\lambda_{\min}}^2(Q_*)} < \infty.$$

В [4] показано, что $W_2^2(\mu_n, \mu_*) \to 0$ п.н., что в нашем случае эквивалентно сходимости $Q_n \to Q_*$. Так как $F_n(\cdot) \to F(\cdot)$ по закону больших чисел, и $\alpha_n(X) = o(||F_n(X)||)$, из (11) немедленно следует, что

$$Q_n = Q_* + F^{-1} \left(\overline{T}_n - I \right) + o \left(\left\| \overline{T}_n - I \right\| \right);$$

 $F^{-1}(\cdot)$ определен корректно в силу положительной определенности $F(\cdot)$. Наконец, применяя ЦПТ для \overline{T}_n , получаем утверждение теоремы.

Список литературы

- 1. C. Villani. Optimal Transport, Old and New. Springer-Verlag, Berlin-Heidelberg, 2009.
- F. Santambrogio. Optimal Transport for Applied Mathematicians. Birkhäuser, Basel, 2015.
- M. Agueh and G. Carlier. Barycenters in the wasserstein space. SIAM Journal on Mathematical Analysis, 43(2):904-924, 2011.
- T. Le Gouic and J.-M. Loubes. Existence and consistency of wasserstein barycenters. ArXiv preprint, 2015.
- Alexey Kroshnin. Fréchet barycenters in the monge-kantorovich spaces. Journal of Convex Analysis, 25(4), 2018.
- J. Bigot and T. Klein. Consistent estimation of a population barycenter in the wasserstein space. ArXiv preprint, 2015.
- Asuka Takatsu. Wasserstein geometry of gaussian measures. Osaka Journal of Mathematics, 48(4):1005-1026, 2011.

где

Notes on the Lagrangian structure of cold dark matter halos and the adhesion approximation in cosmology

Stéphane Colombi

Institut d'Astrophysique de Paris, UMR 7095 CNRS

Andrei Sobolevski*

A.A. Kharkevich Institute for Information Transmission Problems (Dated: September 24, 2018)

Abstract

We show that the adhesion approximation of collisionless cold dark matter (CDM) flows, when considered in the space of initial coordinates of fluid particles (the Lagrangian hydrodynamic representation) rather than in the usual Eulerian representation in the physical space, can be used to understand the internal caustic structure of multi-stream regions in 2D and 3D flows.

^{*} sobolevski@iitp.ru

I. INTRODUCTION

According to observations, the large-scale structure of the Universe, as traced by luminous matter, is formed by a web of relatively thin walls, filaments, and clusters. In the cold dark matter (CDM) model, these formations are described as superpositions of multiple streams of dark matter in which particle velocities have negligible dispersion with respect to a bulk hydrodynamic value. A self-consistent gravitational field, created by all streams collectively, only moderately affects the longitudinal motion of matter in a vicinity of such a multi-stream object. However this field strongly confines this motion depthwise, creating inside a wall or a filament a complex internal structure of intertwined CDM flows.

The Zel'dovich approximation and the adhesion approximation are two related approximations providing analytic, explicitly tractable yet qualitatively meaningful models of this dynamics. Under the *Zel'dovich approximation* [1, 2], the flow of dark matter is supposed to be single-stream, at least at early times, and its particles follow ballistic trajectories unperturbed by gravitational forces. This approximation is known to be exact in 1D before the flow ceases to be single-stream (see, e.g., [3, Appendix A]). In 2D or 3D this approximation still gives a surprisingly good description of multi-stream regions, because in a CDM flow they arise from quasi-1D collapse.

However the Zel'dovich approximation fails to capture dynamics inside multi-stream regions, where gravitational interaction leads to repeated "folding" of individual streams. This mechanism limits the widening of these regions, which would go indefinitely in the purely ballistic flow, and creates formations with complicated caustic structure that are extended longitudinally but restricted in depth.

In 3D these objects form a spatial web that can be divided into components with morphology of walls, filaments, and halos. In 1D the internal structure of halos was well understood both analytically and numerically as early as in the 1980s [4, 5]. In 2D the caustic structure was also confirmed by N-body simulations very early [6], but its sufficiently detailed and reliable direct numerical construction was made possible in 2D and 3D only recently thanks to Lagrangian adaptive tesselation methods [7, 8]. The simulations based on this approach reveal a complex internal structure of superimposed flows, whose analytical understanding remains an open problem.

The adhesion approximation [9, 10], in which a model of structure formation is added

to the ballistic dynamics, provides a next degree of accuracy with respect to the Zel'dovich approximation. Here walls, filaments, and clusters are represented with infinitely thin pieces of 2D manifolds (sheets), 1D manifolds (curves), and zero-dimensional point-like nodes. The adhesion approximation thus overcomes the main drawback of the Zel'dovich approximation, its inability to describe formation of multi-stream structures, but does so rather drastically: all information on the internal structure of multi-stream regions is obliterated. The purpose of this study is to show that the adhesion approximation is nevertheless capable of providing some qualitative information on the internal structure of the multi-stream web in an analytically tractable way.

This report is organized as follows. In Section ...

In the remaining part of the introduction we briefly recall the basic formulas of the Zel'dovich and adhesion approximations. In suitable time and coordinate variables the former takes the form of the following hydrodynamic equation of ballistic motion:

$$\frac{\partial \boldsymbol{u}}{\partial t} + (\boldsymbol{u} \cdot \nabla) \, \boldsymbol{u} = 0, \tag{1}$$

where $\boldsymbol{u}(\boldsymbol{r},t)$ is the velocity field of CDM elements. This equation is supplemented by the potentiality condition:

$$\nabla \times \boldsymbol{u} = 0, \tag{2}$$

which follows from the Jeans linear theory, where the growing mode of initial perturbations is characterized by curl-free velocity. Equation (2) ensures existence of a velocity potential function $\phi(\mathbf{r}, t)$ such that $\mathbf{u} = \nabla \phi$. Equation (1) then implies the Hamilton–Jacobi equation for the velocity potential:

$$\frac{\partial \phi}{\partial t} + \frac{1}{2} |\nabla \phi|^2 = 0.$$
(3)

Both equations (1) and (3) cease to have single-valued solutions after the time of the first caustic formation. This mathematical difficulty, as well as the limited physical validity of the Zel'dovich approximation in description of multi-stream structures, are overcome in the *adhesion approximation* introduced in [9, 10]. Here equation (1) is regularized by introducing a small viscous term:

$$\frac{\partial \boldsymbol{u}}{\partial t} + (\boldsymbol{u} \cdot \nabla) \, \boldsymbol{u} = \nu \nabla^2 \boldsymbol{u} \tag{4}$$

and taking the limit of vanishing viscosity $\nu \to +0$. For the velocity potential the corresponding equation reads

$$\frac{\partial \phi}{\partial t} + \frac{1}{2} |\nabla \phi|^2 = \nu \nabla^2 \phi, \tag{5}$$

whose solution in the vanishing viscosity limit is given by [11]

$$\phi(\mathbf{r},t) = \min_{\mathbf{q}} \left(\phi_0(\mathbf{q}) + \frac{|\mathbf{r} - \mathbf{q}|^2}{2t} \right).$$
(6)

with $\phi_0(q) = \phi(q, t = 0)$.

The velocity potential function (6) is continuous but only piecewise differentiable: at locations (\mathbf{r}, t) where the min_q in (6) is attained for several values of \mathbf{q} , the function ϕ is formed by several smooth branches intersecting along sheets (in 3D) or curves (in 2D) where the values of ϕ coincide but those of $\nabla \phi$ do not match. This implies that the corresponding velocity field $\mathbf{u} = \nabla \phi$ is discontinuous along a network of manifolds of codimension one (infinitely thin 2D "walls" in 3D space, or 1D curves in 2D space), which meet one another along manifolds of yet smaller dimension (1D "filaments" that form boundaries of 2D sheets in 3D space) all the way down to zero-dimensional nodes corresponding to "halos." In what follows, we will refer to all such manifolds where the velocity is discontinuous as *shocks*. Trajectories of fluid particles in the smooth part of the flow terminate at shocks, suggesting that the particles adhere to singularities of the flow, hence the name of the *adhesion* approximation.

We stress that the adhesion approximation is purely phenomenological: the true CDM dynamics is not collisional, whereas the regularizing term in equation (4) would normally arise from some collisional transport mechanism. Still, numerical studies show that the network of infinitely thin singularities in the solution (6) to adhesion model provides a qualitatively realistic "skeleton" of the web of multi-stream regions observed in "real" N-body or Lagrangian tesselation simulations.

II. THE VLASOV–POISSON EQUATIONS AND INITIAL CONDITIONS

THE EXPOSITION IN THIS SECTION CORRESPONDS TO AN EARLIER VERSION OF THE TEXT AND IS OUT OF SYNC WITH THE REST OF THE REPORT.

A suitable mathematical description of multi-streaming in a cold dark matter flow is provided by a kinetic description in the Hamiltonian phase space $(\mathbf{r}, \mathbf{p}), \mathbf{p} = m\mathbf{u}$, where individual streams passing through the same space location \mathbf{r} with different velocities \mathbf{u} can be considered separately. In this section we recall the Vlasov–Poisson equations that provide this description. We also show how the adhesion model can be derived from the kinetic representation.

The Vlasov–Poisson system of equations for a nonrelativistic self-gravitating collisionless fluid that evolves in a self-consistent gravitation field has the form

$$\frac{\partial f}{\partial t} + \frac{1}{m} \boldsymbol{p} \cdot \nabla_{\boldsymbol{r}} f - \nabla_{\boldsymbol{r}} g \cdot \nabla_{\boldsymbol{p}} f = 0, \tag{7}$$

$$\nabla_{\boldsymbol{r}}^2 g = 4\pi G m \int f \,\mathrm{d}\boldsymbol{p}.\tag{8}$$

Here $f(\mathbf{r}, \mathbf{p}, t)$ represents the phase-space density at position $\mathbf{r} = (r_1, r_2, r_3)$, momentum $\mathbf{p} = (p_1, p_2, p_3)$, and time t, $g(\mathbf{r}, t)$ is the gravitational potential, G the gravitational constant, and m the particle mass. We use here notation $\nabla_{\mathbf{r}} = (\partial/\partial r_1, \partial/\partial r_2, \partial/\partial r_3)$ and $\nabla_{\mathbf{p}} = (\partial/\partial p_1, \partial/\partial p_2, \partial/\partial p_3)$ instead of just $\nabla = \nabla_{\mathbf{r}}$, as in the previous section, to distinguish between differentiation with respect to components of two different vector arguments \mathbf{r} and \mathbf{p} .

Now dynamics of fluid particles is determined by the Hamiltonian function of the form

$$H(\boldsymbol{r},\boldsymbol{p},t) = \frac{1}{2m}|\boldsymbol{p}|^2 + g(\boldsymbol{r},t).$$
(9)

In these notes we focus on the *cold* case, in which the phase-space distribution function f is initially concentrated on a 3-dimensional hypersurface in the 6-dimensional phase space:

$$f(\boldsymbol{r}, \boldsymbol{p}, 0) = \frac{1}{m} \rho_0(\boldsymbol{r}) \,\delta_{\mathrm{D}}(\boldsymbol{p} - m\boldsymbol{v}_0(\boldsymbol{r})).$$
(10)

Here $\rho_0(\mathbf{r})$ is the initial mass density field, $\mathbf{v}_0(\mathbf{r})$ the initial velocity field, and $\delta_{\rm D}(\cdot)$ is the Dirac delta function of a 3-dimensional vector argument. Note than the hypersurface $\mathbf{p} = m\mathbf{v}_0(\mathbf{r})$ has one-to-one projection on the coordinate space $(\mathbf{r}, \mathbf{p} = 0)$.

In the commonly accepted picture of large scale structure formation, the initial density field $\rho_0(\mathbf{r})$ is taken close to constant, and the initial velocity field corresponds to the growing mode of the linear perturbation theory, which is potential:

$$\rho_0(\mathbf{r}) \simeq \text{const}, \quad m\mathbf{v}_0(\mathbf{r}) = \nabla_{\mathbf{r}}\phi_0(\mathbf{r}).$$
(11)

To avoid boundary effects we will assume periodic boundary conditions, i.e., suppose that $\rho_0(x + iL, y + jL, z + kL) = \rho_0(x, y, z)$ and $\boldsymbol{v}_0(x + iL, y + jL, z + kL) = \boldsymbol{v}_0(x, y, z)$ for all integer i, j, k, where L is the spatial period. Note that under these conditions $\phi_0(\boldsymbol{r})$ is not necessarily periodic in \boldsymbol{r} ; only its gradient is.

The solution of the Vlasov–Poisson system (7), (8) with initial conditions (10), (11) can be constructed by following trajectories of individual particles. Denote the initial coordinate of a dark matter particle at t = 0 by $\boldsymbol{q} = (q_1, q_2, q_3)$. Then the phase-space trajectory $(\boldsymbol{r}(\boldsymbol{q}, t), \boldsymbol{p}(\boldsymbol{q}, t))$ of this particle solves Hamilton's canonical equations

$$\dot{\boldsymbol{r}} = \frac{1}{m} \boldsymbol{p} = \nabla_{\boldsymbol{p}} H(\boldsymbol{r}, \boldsymbol{p}, t), \qquad (12)$$

$$\dot{\boldsymbol{p}} = -\nabla_{\boldsymbol{r}}g = -\nabla_{\boldsymbol{r}}H(\boldsymbol{r},\boldsymbol{p},t), \qquad (13)$$

where the dot denotes the time derivative taken along the trajectory, i.e., with q fixed, and the initial conditions are given by

$$\boldsymbol{r}(\boldsymbol{q},0) = \boldsymbol{q}, \quad \boldsymbol{p}(\boldsymbol{q},0) = \nabla_{\boldsymbol{r}}\phi_0(\boldsymbol{q}),$$
 (14)

The kinetic equation (7) means that the phase-space density f is constant along each phasespace trajectory. In particular, at time t the solution is concentrated on a 3-dimensional phase space sheet formed by positions $(\mathbf{r}(\mathbf{q}, t), \mathbf{p}(\mathbf{q}, t))$ for all values of the Lagrangian variable \mathbf{q} .

The initial conditions (11) imply that for any components q_i, q_j of the Lagrangian coordinate vector, the following quantity vanishes identically:

$$\sum_{\ell} \left(\frac{\partial r_{\ell}}{\partial q_i} \frac{\partial p_{\ell}}{\partial q_j} - \frac{\partial p_{\ell}}{\partial q_i} \frac{\partial r_{\ell}}{\partial q_j} \right) = 0 \tag{15}$$

Indeed, it follows from (14) that $\partial r_{\ell}/\partial q_k = \delta_{\ell k}$ and thus the sum reduces to

$$\frac{\partial p_i}{\partial q_j} - \frac{\partial p_j}{\partial q_i} = \frac{\partial^2 \phi_0}{\partial q_i \partial q_j} - \frac{\partial^2 \phi_0}{\partial q_j \partial q_i} = 0.$$

In the course of evolution defined by Hamilton's equations (12), (13), the 3-dimensional phase space sheet representing the solution will generally cease to have a one-to-one projection to the coordinate space ($\mathbf{r}, \mathbf{p} = 0$) and will fold in a complicated way, developing a multi-stream flow. However the identity (15) will be preserved for all times, as can be seen by differentiating its left-hand side with respect to t and using Hamilton's equations (12), (13).

Recall that a 3-dimensional manifold in the 6-dimensional phase space given in parametric form by $(\mathbf{r}(\mathbf{q}), \mathbf{p}(\mathbf{q}))$ is called a *Lagrangian manifold* iff relations (15) hold for every q_i, q_j (or, equivalently, if the differential form $d\mathbf{r} \wedge d\mathbf{p}$ vanishes on all pairs of vectors tangent to the manifold). We see that under the initial conditions (11) cold dark matter is described by an evolving Lagrangian manifold in phase space.

Observe now that the scalar field ϕ_0 is closely related to Poincaré invariants, i.e., contour integrals of the form

$$I = \oint \boldsymbol{p} \cdot \mathrm{d}\boldsymbol{r} \tag{16}$$

over closed curves in phase space. Indeed, at time t = 0 we have $\mathbf{p} \cdot d\mathbf{r} = \nabla_{\mathbf{q}} \phi_0(\mathbf{q}) \cdot d\mathbf{q} = d\phi_0$, and $\oint d\phi_0 = 0$ for any closed contour. Since Poincaré invariants are preserved by the flow, the integral $\oint \mathbf{p} \cdot d\mathbf{r}$ along a closed contour whose points follow phase space trajectories (12), (13) must vanish identically, suggesting that a potential ϕ still should be defined consistently on the evolving Lagrangian manifold.

This is indeed the case. In fact the potential ϕ can be defined by the initial condition $\phi(\mathbf{q}, 0) = \phi_0(\mathbf{q})$ and the equation

$$\dot{\phi} = \boldsymbol{p} \cdot \dot{\boldsymbol{r}} - H = L(\boldsymbol{r}, \dot{\boldsymbol{r}}, t), \tag{17}$$

along trajectories of Hamilton's equations (12), (13), so that

$$\phi(\boldsymbol{q},t) = \phi_0(\boldsymbol{q}) + \int_0^t L(\boldsymbol{r}(\boldsymbol{q},s), \dot{\boldsymbol{r}}(\boldsymbol{q},s), s) \,\mathrm{d}s.$$
(18)

In other words, the function ϕ is the *mechanical action* associated to the particle that starts from \boldsymbol{q} at t = 0. Here the Lagrangian function $L(\boldsymbol{r}, \boldsymbol{v}, t)$ is the Legendre transform of the Hamiltonian function (9) with respect to the momentum variable:

$$L(\boldsymbol{r},\boldsymbol{v},t) = \max_{\boldsymbol{p}} \left(\boldsymbol{p} \cdot \boldsymbol{v} - H(\boldsymbol{r},\boldsymbol{p},t) \right) = \frac{m}{2} |\boldsymbol{v}|^2 - \phi(\boldsymbol{r},t).$$
(19)

As $d\phi = \mathbf{p} \cdot d\mathbf{r} - H dt$, for any fixed t the consistency condition $\oint d\phi = \oint \mathbf{p} \cdot d\mathbf{r} = 0$ holds. In particular,

$$\boldsymbol{p} = \nabla_{\boldsymbol{r}} \phi \tag{20}$$

whenever the Jacobian matrix $\partial \boldsymbol{r}/\partial \boldsymbol{q}$ is nondegenerate, so that $\nabla_{\boldsymbol{r}}$ is well defined.

For a smooth initial condition of the form (10), (11) the multi-streaming regime, in which several different values of \boldsymbol{q} are mapped into the same point \boldsymbol{r} , develops only after a positive time t^* . For times $0 \leq t < t^*$, the mechanical action ϕ as well as its gradient \boldsymbol{p} can be considered as functions of the \boldsymbol{r} variable. The full time derivative of ϕ along a trajectory then reads

$$\frac{\mathrm{d}\phi(\boldsymbol{q},t)}{\mathrm{d}t} = \frac{\partial\phi(\boldsymbol{r},t)}{\partial t} + \nabla_{\boldsymbol{r}}\phi(\boldsymbol{r},t)\cdot\dot{\boldsymbol{r}} = \boldsymbol{p}\cdot\dot{\boldsymbol{r}} - \boldsymbol{H},\tag{21}$$

where we used (17). Taking into account (20) we now get the Hamilton–Jacobi equation

$$\frac{\partial \phi(\boldsymbol{r},t)}{\partial t} + H(\boldsymbol{r}, \nabla_{\boldsymbol{r}} \phi(\boldsymbol{r},t), t) = 0, \qquad (22)$$

considered with the initial condition $\phi(\mathbf{r}, 0) = \phi_0(\mathbf{r})$. Formula (18) derived above gives the solution to this initial-value problem as long as the equation $\mathbf{r}(\mathbf{q}, t) = \mathbf{r}$ can be uniquely resolved for \mathbf{q} for any \mathbf{r} , i.e., before the onset of multi-streaming.

In the Zel'dovich approximation, the gravitational potential in the Hamiltonian and Lagrangian functions is neglected, so that $H(\mathbf{r}, \mathbf{p}, t) = |\mathbf{p}|^2/2m$ and $L(\mathbf{r}, \mathbf{v}, t) = m|\mathbf{v}|^2/2$. Then the particle trajectories have the form

$$\boldsymbol{r}(\boldsymbol{q},t) = \boldsymbol{q} + t\boldsymbol{v}_{i}(\boldsymbol{q}), \qquad (23)$$

and the integral in (18) can be calculated explicitly:

$$\phi(\mathbf{r},t) = \phi_0(\mathbf{q}) + \frac{|\mathbf{r} - \mathbf{q}|^2}{2t}$$
(24)

where $\boldsymbol{r}, \boldsymbol{q}$, and t are related by (23).

Even with the simplest dynamics (23), the relation between \boldsymbol{q} and \boldsymbol{r} will cease to be oneto-one after particle trajectories cross. After this, a single-valued solution for $\phi(\boldsymbol{r},t)$ can be constructed for all \boldsymbol{r} by the following selection rule based on the least-action principle (cf (6)):

$$\phi(\mathbf{r},t) = \min_{\mathbf{q}} \left(\phi_0(\mathbf{q}) + \frac{|\mathbf{r} - \mathbf{q}|^2}{2t} \right).$$
(25)

We rewrite this formula to make its relation to the Legendre transform more explicit:

$$\phi(\mathbf{r},t) = \frac{|\mathbf{r}|^2}{2t} - \frac{1}{t} \max_{\mathbf{q}} \left[\mathbf{r} \cdot \mathbf{q} - \left(\frac{|\mathbf{q}|^2}{2} + t\phi_0(\mathbf{q}) \right) \right].$$
(26)

It is now easy to see that the solution $\phi(\mathbf{r}, t)$ is fully determined by the convex hull of the function $\frac{1}{2}|\mathbf{q}|^2 + t\phi_0(\mathbf{q})$. Thus, if ϕ_0 is periodic itself, then for large t the solution is mainly determined by the lowest minimum of ϕ_0 over a period.

III. THE EULERIAN AND LAGRANGIAN REPRESENTATIONS OF THE LARGE-SCALE STRUCTURE

A. Overview

Recall that in fluid dynamics the representation of the flow in physical coordinates r is called *Eulerian*, while its representation parametrized by coordinates q of initial locations of particles is called *Lagrangian*. The Lagrangian representation amounts to following trajectories of particles labelled by their initial locations q, whereas the Eulerian representation amounts to observing particles passing through a fixed physical space location r.

Consider a slightly modified form of the solution (6) to the adhesion model:

$$\phi(\mathbf{r},t) = \frac{1}{t} \min_{\mathbf{q}} \left(t\phi_0(\mathbf{q}) + \frac{|\mathbf{r} - \mathbf{q}|^2}{2} \right), \tag{27}$$

We assume that the initial velocity potential $\phi_0(\mathbf{q})$ has a periodic gradient, denote the spatial period of the problem by L and let $\Phi_0 = \max \phi_0(\mathbf{q}) - \min \phi_0(\mathbf{q})$ characterize the typical scale of the initial condition. At large times $t \gg T_0 = L^2/\Phi_0$ the solution at most locations \mathbf{r} will be determined by an O(1/t) neighborhood of the global minimum \mathbf{q}_0 of $\phi_0(\mathbf{q})$ over the spatial period, and will be characterized by the scale $\Phi_t = O(L^2/t)$. Indeed, up to terms quadratic in 1/t, the value of \mathbf{q} providing minimum to the expression in (27) is given by

$$\boldsymbol{q} = \boldsymbol{q}_0 + \frac{1}{t} H_0^{-1} (\boldsymbol{r} - \boldsymbol{q}_0) + o(1/t), \qquad (28)$$

where H_0 is the Hessian matrix of ϕ_0 at q_0 , so that $t\nabla\phi_0(q) = tH_0(q-q_0) + \cdots$. Essentially this means that the physical (Eulerian) space is divided into "domains of influence" of global minima $q_0 + kL$ of the initial potential ϕ_0 , located at different spatial periods determined by values of $\mathbf{k} = (k_1, k_2, k_3)$ (or $\mathbf{k} = (k_1, k_2)$ in the 2D case).

These domains will be separated with shocks of codimension 1, defined as locations where $\min_{\boldsymbol{q}}$ is achieved at two different values of \boldsymbol{q} . These shocks will meet one another along shocks of smaller dimensions, corresponding to more than two simultaneously minimizing values of \boldsymbol{q} , all the way down to zero-dimensional shock nodes. Fluid particles with initial values of \boldsymbol{q} outside these O(1/t) neighborhoods will not contribute to the left-hand side of (27), which means that they will be absorbed into shocks.

By contrast, in the Lagrangian representation these small neighborhoods will only correspond to negligible volume, concentrated in each period around points of the form $q_0 + kL$. For $t \gg T_0$ most of the remaining Lagrangian space will be occupied by (pre)images of Eulerian shocks of different dimensions, organized in a mosaic of domains which is in duality to the mosaic of "domains of influence" in the Eulerian space. Note that this duality is similar to the duality between the Voronoi diagram and the Delaunay tesselation generated by the same set of points.

Our central observation in this report is that numerical simulations of late stages of evolution of multi-stream flows by the Lagrangian tesselation method form in the Lagrangian space a similar mosaic of domains, which correspond to collapsed formations (walls, filaments, and halos) in the Eulerian space. The structural similarity between the Lagrangian mosaic structures observed in the adhesion approximation and in direct simulations of selfgravitating CDM flows allows to employ the adhesion approximation in order to qualitatively model the stream folding in CDM dynamics.

This approximation by itself does not provide further information about the inner organization of these formations: indeed, in real, or Eulerian, space all shocks in the adhesion approximation are infinitely thin and lack any structure. However, as we will see from numerical data [YET TO BE PROVIDED], the inner organization of Lagrangian preimages of collapsed structures in numerical simulations shows a significant degree of regularity, and it is tempting to expect that the adhesion approximation could be supplemented by another analytic or semi-analytic model that describes this regular organization.

B. Initial conditions

We now provide an example of a 2D spatially periodic flow in which the Lagrangian mosaic of preimages of shocks can be explored analytically in the adhesion approximation and compared to the results of direct numerical simulations by the Lagrangian tesselation method.

Take a parameter $\pi/2 < \alpha < \pi$ and denote $a = (\sin \alpha)^{-1}$. In the graphs and numerical calculations below we set $\alpha = 2\pi/3$, so that $\cos \alpha = -\frac{1}{2}$, $\sin \alpha = \sqrt{3}/2$, and $a = 2/\sqrt{3}$, but using the abstract parameters α and a makes some formulas below more understandable. Consider the initial velocity potential

$$\phi_0(q_1, q_2) = -3 + \frac{1}{2} \exp 2a(\cos(q_1 + \alpha) + \cos(q_2 - \alpha) - 2\cos\alpha) + \frac{1}{2} \exp 2a(\cos(q_1 - \alpha) + \cos(q_2 + \alpha) - 2\cos\alpha) + \exp a(\cos q_1 + \cos q_2 - 2) + \exp(1 - \cos(q_1 + q_2)).$$
(29)

Below we will need explicit formulas for the gradient and the Hessian of this function, which are given in footnote [12]. Inside the period this potential has a global minimum $\phi(0,0) = 0$ at the origin and two symmetric peaks located near points $(-\alpha, \alpha)$ and $(\alpha, -\alpha)$, whose height is approximately given by

$$\phi_0(\mp\alpha,\pm\alpha) = -2 + \frac{1}{2} \exp 4a(1-\cos\alpha) + \frac{1}{2} \exp 4a(\cos 2\alpha - \cos \alpha) + \exp 2a(\cos \alpha - 1).$$
(30)
189

Indeed, although at points $(\mp \alpha, \pm \alpha) = (\mp 2\pi/3, \pm 2\pi/3)$ the gradient of ϕ_0 is given by $\partial \phi_0 / \partial q_1 = -\partial \phi_0 / \partial q_2 = \mp (1 - \exp(-2\sqrt{3}))$ and does not vanish, its Hessian is approximately given by $(-2/\sqrt{3}) \exp(4\sqrt{3}) \simeq -1178$ times the identity matrix (see eq. (34) below), so the true maximum is very close to $(\mp 2\pi/3, \pm 2\pi/3)$.

Note that height of the peaks may be significant: e.g. for $\alpha = 2\pi/3$ and $a = 2/\sqrt{3}$ the value of (30) is close to 509. This makes the time scale $T_0 = L^2/\Phi_0$ small in natural units (e.g. $T_0 \simeq 0.078$ for $\alpha = 2\pi/3$).

C. Formation of shocks in the adhesion approximation

Recall that particle dynamics in the adhesion approximation, as well as in the Zel'dovich approximation, is ballistic: before shell crossing it is given simply by

$$\boldsymbol{r}(\boldsymbol{q},t) = \boldsymbol{q} + t\nabla\phi_0(\boldsymbol{q}),\tag{31}$$

where $\boldsymbol{u}_0(\boldsymbol{q}) = \nabla \phi_0(\boldsymbol{q})$ is the initial velocity field. The Jacobian of this mapping is given by

$$\frac{\partial r_i}{\partial q_j} = \delta_{ij} + t[H_0(\boldsymbol{q})]_{ij},\tag{32}$$

where

$$[H_0(\boldsymbol{q})]_{ij} = \frac{\partial^2 \phi_0(\boldsymbol{q})}{\partial q_i \partial q_j}$$
(33)

is the Hessian matrix of the initial potential ϕ_0 . Note that this matrix is symmetric and therefore it has two real eigenvalues $\alpha(\mathbf{q}) \leq \beta(\mathbf{q})$. For small times the Lagrangian mapping (31) is one to one, and its Jacobian is close to unity. However if $\alpha(\mathbf{q}) < 0$ the trajectory given by (31) will intersect with infinitesimally close trajectories at time $t^*(\mathbf{q}) = 1/\alpha(\mathbf{q})$, and the mapping will cease to be invertible. Singularities thus formed emerge in the Zel'dovich approximation at points $\mathbf{r}(\mathbf{q}, t^*(\mathbf{q}))$ defined by values of \mathbf{q} where the eigenvalue $\alpha(\mathbf{q})$ has negative local minima. These spacetime locations approximate those where shocks emerge in full self-gravitating dynamics.

At each space period the initial potential (29) has five negative local minima of $\alpha(\mathbf{q})$. Two of them correspond to locations of peaks near $(\mp \alpha, \pm \alpha)$, and the other three to saddle points $(-\pi, \pi)$, $(\pm \pi, 0)$ and $(0, \pm \pi)$. We now list the eigenvalues of H_0 at these points for $\alpha = 2\pi/3$. At $(q_1, q_2) = (\mp 2\pi/3, \pm 2\pi/3)$ the Hessian is given by

$$\frac{\partial^2 \phi_0}{\partial q_1{}^2} = \frac{\partial^2 \phi_0}{\partial q_2{}^2} = -\frac{2}{\sqrt{3}} \exp(4\sqrt{3}) + 3 + \frac{1}{\sqrt{3}} + (1 + \frac{1}{\sqrt{3}}) \exp(-2\sqrt{3}),$$

$$\frac{\partial^2 \phi_0}{\partial q_1} \partial q_2 = -(1 + \exp(-2\sqrt{3})),$$

(34)

the eigenvalues are

$$\alpha_{\text{peak}} = -\frac{2}{\sqrt{3}} \exp(4\sqrt{3}) + \frac{1}{\sqrt{3}} (\exp(-2\sqrt{3}) + 1) + 2 < 0,$$

$$\beta_{\text{peak}} = -\frac{2}{\sqrt{3}} \exp(4\sqrt{3}) + \frac{1}{\sqrt{3}} (\exp(-2\sqrt{3}) + 1) + 4 + 2\exp(-2\sqrt{3}) < 0,$$
(35)

and the collapse time is

$$t_{\rm peak}^* = 1/\alpha_{\rm peak} \approx 8.504 \cdot 10^{-4}.$$
 (36)

Note that at true locations of peaks, which are close to the points $(\pm 2\pi/3, \pm 2\pi/3)$ but do not quite coincide with them, these expressions approximate the exact values within a few tenths of percent. Observe also that as α_{peak} and β_{peak} are rather close, the collapse at peak locations takes place in an almost, but note quite, isotropic way. This relative degeneracy is a shortcoming of the present example, but it does not seem to make much harm numerically.

We now turn to the saddle point $(q_1, q_2) = (\mp \pi, \pm \pi)$. Here the Hessian is given by

$$\frac{\partial^2 \phi_0}{\partial q_1^2} = \frac{\partial^2 \phi_0}{\partial q_2^2} = (4 - \frac{2}{\sqrt{3}}) \exp(8/\sqrt{3}) + \frac{2}{\sqrt{3}} \exp(-8/\sqrt{3}) + 1,$$

$$\frac{\partial^2 \phi_0}{\partial q_1 \partial q_2} = -4 \exp(8/\sqrt{3}) + 1,$$
(37)

the eigenvalues are

$$\alpha_{\text{saddle I}} = -\frac{2}{\sqrt{3}} \exp(8/\sqrt{3}) + 2 + \frac{2}{\sqrt{3}} \exp(-8/\sqrt{3}) < 0,$$

$$\beta_{\text{saddle I}} = (8 - \frac{2}{\sqrt{3}}) \exp(8/\sqrt{3}) + \frac{2}{\sqrt{3}} \exp(-8\sqrt{3}) > 0,$$

(38)

and the collapse time is

$$t_{\text{saddle I}}^* = 1/\alpha(\mp\pi, \pm\pi) \approx 8.692 \cdot 10^{-3}.$$
 (39)

Finally turn to two symmetric saddle points $(q_1, q_2) = (\pm \pi, 0)$ and $(q_1, q_2) = (0, \pm \pi)$. Here the Hessian is given by

$$\frac{\partial^2 \phi_0}{\partial q_1^2} = (4 - \frac{2}{\sqrt{3}}) \exp(4/\sqrt{3}) + \frac{2}{\sqrt{3}} \exp(-4/\sqrt{3}) - \exp(2),$$

$$\frac{\partial^2 \phi_0}{\partial q_2^2} = (4 + \frac{2}{\sqrt{3}}) \exp(4/\sqrt{3}) - \frac{2}{\sqrt{3}} \exp(-4/\sqrt{3}) - \exp(2)$$

$$\frac{\partial^2 \phi_0}{\partial q_1 \partial q_2} = 4 \exp(4/\sqrt{3}) - \exp(2),$$

(40)

the eigenvalues are

$$\alpha_{\text{saddle II}} = 4 \exp(4/\sqrt{3}) - \exp(2) - \sqrt{(4 \exp(4/\sqrt{3}) - \exp(2))^2 + \left(\frac{4}{\sqrt{3}} \sinh \frac{4}{\sqrt{3}}\right)^2} < 0,$$

$$\beta_{\text{saddle II}} = 4 \exp(4/\sqrt{3}) - \exp(2) + \sqrt{(4 \exp(4/\sqrt{3}) - \exp(2))^2 + \left(\frac{4}{\sqrt{3}} \sinh \frac{4}{\sqrt{3}}\right)^2} > 0,$$

(41)

and the collapse time is

$$t_{\text{saddle II}}^* = 1/\alpha(\pm\pi, 0) \approx 0.511.$$
 (42)

In particular,

$$\frac{t_{\text{saddle I}}^*}{t_{\text{peak}}^*} \approx 10, \qquad \frac{t_{\text{saddle II}}^*}{t_{\text{saddle I}}^*} \approx 60.$$
(43)

D. Adhesion approximation at large times: contribution of global minima of the potential

Recall that according to the formula (27), for $t \gg T_0 = L^2/\Phi_0$ the solution $\phi(r_1, r_2, t)$ will be determined by global minima of the initial velocity potential ϕ_0 . In this subsection we characterize the contribution of these minima.

Near the global minimum at (0,0) the quadratic part of the Taylor expansion of ϕ_0 has the form

$$\bar{\phi}_0(q_1, q_2) = \frac{4-b}{2}(q_1 - q_2)^2 + \frac{1-b}{2}(q_1 + q_2)^2, \tag{44}$$

where

$$b = \frac{a}{2} \left(2\cos\alpha + 1 \right) \tag{45}$$

(in particular, b = 0 for $\alpha = 2\pi/3$).

Substituting the initial condition (44) into (27) and rearranging terms, we get the exact formula

$$\bar{\phi}(r_1, r_2, t) = \frac{1}{t} \min_{q_1, q_2} \left[\frac{1 + 2(4 - b)t}{4} \left(q_1 - q_2 - \frac{r_1 - r_2}{1 + 2(4 - b)t} \right)^2 + \frac{1 + 2(1 - b)t}{4} \left(q_1 + q_2 - \frac{r_1 + r_2}{1 + 2(1 - b)t} \right)^2 + \frac{r_1^2}{2} + \frac{r_2^2}{2} - \frac{(r_1 - r_2)^2}{4(1 + 2(4 - b)t)} - \frac{(r_1 + r_2)^2}{4(1 + 2(1 - b)t)} \right].$$
(46)

It is clear that min in (q_1, q_2) is attained at

$$\begin{pmatrix} q_1 - q_2 \\ q_1 + q_2 \end{pmatrix} = \begin{pmatrix} \frac{1}{1+2(4-b)t} & 0 \\ 0 & \frac{1}{1+2(1-b)t} \end{pmatrix} \begin{pmatrix} r_1 - r_2 \\ r_1 + r_2 \end{pmatrix}$$
(47)

or

$$\begin{pmatrix} q_1 \\ q_2 \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{pmatrix} \begin{pmatrix} \frac{1}{1+2(4-b)t} & 0 \\ 0 & \frac{1}{1+2(1-b)t} \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} r_1 \\ r_2 \end{pmatrix}.$$
(48)

Observe how this expression agrees with (28), in which the value of (q_1, q_2) minimizing the right-hand side of (27) for finite (r_1, r_2) is shown to lie in an O(1/t) neighborhood of the origin.

For the minimal value in (46) the remaining terms in the right-hand side give an expression

$$\bar{\phi}(r_1, r_2, t) = \frac{4-b}{1+2(4-b)t} \frac{(r_1 - r_2)^2}{2} + \frac{1-b}{1+2(1-b)t} \frac{(r_1 + r_2)^2}{2}.$$
(49)

Equivalently we can write

$$t\bar{\phi}(r_1, r_2, t) = \frac{r_1^2 + r_2^2}{2} - \frac{(r_1 - r_2)^2}{4(1 + 2(4 - b)t)} - \frac{(r_1 + r_2)^2}{4(1 + 2(1 - b)t)}.$$
(50)

Note that at large times this solution decays as $(r_1^2 + r_2^2)/2t$, which stands in agreement with the O(1/t) scaling for Φ_t derived in the overview above. We also observe that the isotropy of the term $|\boldsymbol{r} - \boldsymbol{q}|^2/2$ asymptotically prevails in (27) over the anisotropy of the term $t\bar{\phi}_0(q_1, q_2)$.

Now consider an initial condition represented by the piecewise quadratic periodic function

$$\hat{\phi}_0(q_1, q_2) = \min_{k_1, k_2 \text{ integer}} \left[\frac{4-b}{2} (q_1 - q_2 + 2\pi (k_1 - k_2))^2 + \frac{1-b}{2} (q_1 + q_2 + 2\pi (k_1 + k_2))^2 \right], \quad (51)$$

formed by taking a pointwise minimum of periodic translates of the function $\bar{\phi}_0$. Minimization in (k_1, k_2) in (51) commutes with minimization in (q_1, q_2) in (27), so the solution $\hat{\phi}(r_1, r_2, t)$ provided by (27) for the initial data (51) is a pointwise minimum of periodic translates of (49):

$$\hat{\phi}(r_1, r_2, t) = \min_{k_1, k_2 \text{ integer}} \bar{\phi}(r_1 + 2\pi k_1, r_2 + 2\pi k_2, t).$$
(52)

The scale of this initial condition is equal to $\hat{\Phi}_0 = \max \hat{\phi}_0 = (2\pi)^2 25/32$, so the corresponding time scale is $\hat{T}_0 = (2\pi)^2/\hat{\Phi}_0 = 1.28$. Note that it is larger by an order of magnitude than the time scale $T_0 \simeq 0.078$ defined above for ϕ_0 . However it is the time scale \hat{T}_0 that is relevant to judge if $\hat{\phi}(r_1, r_2, t)$ is a good approximation for $\phi(r_1, r_2, t)$.



FIG. 1. Left: periodicity domain in coordinates (r_1, r_2) and $(\xi = r_1 - r_2, \eta = r_1 + r_2)$. Right: tesselation of the Eulerian plane by translates of the hexagon \hat{H}_t for t = 1/4, $\alpha = 2\pi/3$, b = 0(the boundary of $\hat{H}_{1/4}$ and its translates shown in dashed lines). Note that the value t = 1/4 does not satisfy the condition $t \gg \hat{T}_0 = 1.28$ that ensures validity of $\hat{\phi}$ as approximation to ϕ , but it is chosen here to make the figure more readable. The filled area shows \hat{H}_0 for comparison with $\hat{H}_{1/4}$.

In what follows it is convenient to use the coordinates $\xi = r_1 - r_2$, $\eta = r_1 + r_2$, in which the spatial periodicity domain in the Eulerian space takes the form of a diamond $|\xi - \eta| \leq 2\pi$, $|\xi + \eta| \leq 2\pi$ (Fig. 1, left). The domain in which the periodic solution $\hat{\phi}(r_1, r_2, t)$ coincides with the paraboloid $\bar{\phi}(r_1, r_2, t)$ (i.e., where min in (52) is achieved for $k_1 = k_2 = 0$) is a hexagon \hat{H}_t whose boundaries are determined by the equalities

$$\bar{\phi}(r_1, r_2, t) = \bar{\phi}(r_1 \pm 2\pi, r_2, t), \quad \bar{\phi}(r_1, r_2, t) = \bar{\phi}(r_1, r_2 \pm 2\pi, t), \quad \bar{\phi}(r_1, r_2, t) = \bar{\phi}(r_1 \pm 2\pi, r_2 \pm 2\pi, t).$$
(53)

It is easy to derive from here that \hat{H}_t may be represented as the set of solutions to the system of inequalities

$$\left|\xi \pm \eta \frac{(1-b)(1+2(4-b)t)}{(4-b)(1+2(1-b)t)}\right| \le \pi \frac{5-2b+4(4-b)(1-b)t}{(4-b)(1+2(1-b)t)}, \qquad |\eta| \le 2\pi, \tag{54}$$

see Fig. 1, right (drawn for a relatively small value of time t = 1/4). In particular, for $\eta = \pm 2\pi$ we get

$$|\xi| \le 2\pi \frac{3}{2(4-b)(1+2(1-b)t)}$$
(55)

194



FIG. 2. Eulerian representation of solution for initial data (29) at t = 1/16 (left) and t = 1/4 (right).

and for $\eta = 0$ we get

$$|\xi| \le 2\pi \left(1 - \frac{3}{2(4-b)(1+2(1-b)t)} \right)$$
(56)

For future references we record here coordinates of one of the vertices of hexagon \hat{H}_t ,

$$V_1: \quad \hat{r}_1(t) = -\hat{r}_2(t) = \frac{\hat{\xi}(t)}{2} = \pi \frac{5 - 2b + 4(4 - b)(1 - b)t}{2(4 - b)(1 + 2(1 - b)t)}, \quad \hat{\eta}(t) = 0.$$
(57)

The other vertices in counterclockwise order are, in (r_1, r_2) or (ξ, η) coordinates, given by the formulas

$$V_{2}: (-\hat{r}_{1}(t) + 2\pi, -\hat{r}_{2}(t)), (2\pi - \hat{\xi}(t), 2\pi),$$

$$V_{3}: (\hat{r}_{1}(t), \hat{r}_{2}(t) + 2\pi), (\hat{\xi}(t) - 2\pi, 2\pi),$$

$$V_{4}: (-\hat{r}_{1}(t), -\hat{r}_{2}(t)), (-\hat{\xi}(t), 0),$$

$$V_{5}: (\hat{r}_{1}(t) - 2\pi, \hat{r}_{2}(t)), (\hat{\xi}(t) - 2\pi, -2\pi),$$

$$V_{6}: (-\hat{r}_{1}(t), -\hat{r}_{2}(t) - 2\pi), (2\pi - \hat{\xi}(t), -2\pi).$$
(58)

Periodic translates $\hat{H}_t + (2\pi k_1, 2\pi k_2)$ of this hexagon form a tesselation of the Eulerian plane (Fig. 1, right) [13].

Fig. 2 shows the Eulerian structure of solution $\phi(\xi, \eta, t)$ for t = 1/16 and t = 1/4, computed using the software [14].

We now turn to the Lagrangian picture for the piecewise quadratic solution $\hat{\phi}$. The inverse Lagrangian map from (r_1, r_2) to (q_1, q_2) takes a particularly simple form (47) when



FIG. 3. Lagrangian tesselation shown for the same value of time t = 1/4 and of the parameters $\alpha = 2\pi/3$, b = 0 as in Fig. 1. The filled hexagons are Lagrangian preimages of $\hat{H}_{1/4}$ and its translates, the dashed triangles are Lagrangian preimages of their vertices. Note that the filled hexagons are significantly exaggerated in this figure with respect to the situation when $\hat{\phi}$ becomes a good approximation to the solution ϕ of the adhesion model.

considered in the "rotated" plane $(q_1 - q_2, q_1 + q_2)$, where it reduces just to rescaling the two axes by different time-dependent factors $(1 + 2(4 - b)t)^{-1}$ and $(1 + 2(1 - b)t)^{-1}$. Thus the Lagrangian preimage of the hexagon \hat{H}_t is also a hexagon whose size vanishes as O(1/t)for large times. Moreover, Lagrangian preimages of periodic translates of \hat{H}_t are periodic translates of the Lagrangian preimage of \hat{H}_t , and the rest of the Lagrangian plane is divided into preimages of vertices and sides of these hexagons, see Fig. 3.

E. Adhesion model at finite times

Note that for large times $t \gg T_0$ the solution $\hat{\phi}(r_1, r_2, t)$ is close to the exact solution $\phi(r_1, r_2, t)$ determined by (27) and the initial data (29).

Note that in terms of the solution $\phi(r_1, r_2, t)$ to the original problem with smooth initial data ϕ_0 , this means that shock nodes that appear at the two peaks of the potential ϕ_0 approach one another and asymptotically converge to the same location as $t \to \infty$. In

terms of the full self-gravitating solution with multi-streaming, we expect that the two halos corresponding to these shock nodes approach one another and asymptotically merge at large times.

- [1] Y. B. Zel'dovich, Astron. & Astrophys. 5, 84 (1970).
- [2] S. F. Shandarin and Y. B. Zeldovich, Reviews of Modern Physics 61, 185 (1989).
- [3] Y. Brenier, U. Frisch, M. Hénon, G. Loeper, S. Matarrese, R. Mohayaee, and A. Sobolevskii, Mon. Not. R. Astron. Soc. 346, 501 (2003), astro-ph/0304214.
- [4] A. G. Doroshkevich, E. V. Kotok, A. N. Poliudov, S. F. Shandarin, I. S. Sigov, and I. D. Novikov, Mon. Not. R. Aston. Soc. 192, 321 (1980).
- [5] A. V. Gurevich and K. P. Zybin, Physics-Uspekhi 38, 687 (1995).
- [6] A. L. Melott and S. F. Shandarin, Astrophys. J. 343, 26 (1989).
- [7] O. Hahn, T. Abel, and R. Kaehler, Mon. Not. R. Astron. Soc. 434, 1171 (2013), arXiv:1210.6652.
- [8] T. Sousbie and S. Colombi, Journal of Computational Physics **321**, 644 (2016).
- [9] S. N. Gurbatov, A. I. Saichev, and S. F. Shandarin, Soviet Physics Doklady 30, 921 (1985).
- [10] S. N. Gurbatov, A. I. Saichev, and S. F. Shandarin, Mon. Not. R. Astron. Soc. 236, 385 (1989).
- [11] We briefly recall here how (6) is derived, following [15]. The Cole–Hopf transformation $\phi(\mathbf{r},t) = -2\nu \ln \psi(\mathbf{r},t)$ sends equation (5) into a heat equation $\partial \psi / \partial t = \nu \nabla^2 \psi$ for the function ψ . It can be solved explicitly in terms of heat kernel. Taking the logarithm of the solution and using Laplace's asymptotic formula, we arrive at Hopf's formula (6), where the quadratic term comes from the exponent of the heat kernel.
- [12] The gradient of the function (29) has the form

$$\frac{\partial \phi_0}{\partial q_1} = -a \sin(q_1 + \alpha) \exp 2a(\cos(q_1 + \alpha) + \cos(q_2 - \alpha) - 2\cos\alpha) - a \sin(q_1 - \alpha) \exp 2a(\cos(q_1 - \alpha) + \cos(q_2 + \alpha) - 2\cos\alpha) - a \sin q_1 \exp a(\cos q_1 + \cos q_2 - 2) + \sin(q_1 + q_2) \exp(1 - \cos(q_1 + q_2)),$$
(59)
$$\frac{\partial \phi_0}{\partial q_2} = -a \sin(q_2 - \alpha) \exp 2a(\cos(q_1 + \alpha) + \cos(q_2 - \alpha) - 2\cos\alpha) - a \sin(q_2 + \alpha) \exp 2a(\cos(q_1 - \alpha) + \cos(q_2 + \alpha) - 2\cos\alpha) - a \sin q_2 \exp a(\cos q_1 + \cos q_2 - 2) + \sin(q_1 + q_2) \exp(1 - \cos(q_1 + q_2))$$
(60)

and the Hessian is given by

$$\frac{\partial^2 \phi_0}{\partial q_1^2} = (2a^2 \sin^2(q_1 + \alpha) - a\cos(q_1 + \alpha)) \exp 2a(\cos(q_1 + \alpha) + \cos(q_2 - \alpha) - 2\cos\alpha) + (2a^2 \sin^2(q_1 - \alpha) - a\cos(q_1 - \alpha)) \exp 2a(\cos(q_1 - \alpha) + \cos(q_2 + \alpha) - 2\cos\alpha) + (a^2 \sin^2 q_1 - a\cos q_1) \exp a(\cos q_1 + \cos q_2 - 2) + (\sin^2(q_1 + q_2) + \cos(q_1 + q_2)) \exp(1 - \cos(q_1 + q_2)),$$

(61)

$$\frac{\partial^2 \phi_0}{\partial q_2^2} = (2a^2 \sin^2(q_2 - \alpha) - a\cos(q_2 - \alpha)) \exp 2a(\cos(q_1 + \alpha) + \cos(q_2 - \alpha) - 2\cos\alpha) + (2a^2 \sin^2(q_2 + \alpha) - a\cos(q_2 + \alpha)) \exp 2a(\cos(q_1 - \alpha) + \cos(q_2 + \alpha) - 2\cos\alpha) + (a^2 \sin^2 q_2 - a\cos q_2) \exp a(\cos q_1 + \cos q_2 - 2) + (\sin^2(q_1 + q_2) + \cos(q_1 + q_2)) \exp(1 - \cos(q_1 + q_2)),$$
(62)

$$\frac{\partial^2 \phi_0}{\partial q_1 \partial q_2} = 2a^2 \sin(q_1 + \alpha) \sin(q_2 - \alpha) \exp 2a(\cos(q_1 + \alpha) + \cos(q_2 - \alpha) - 2\cos\alpha) + 2a^2 \sin(q_1 - \alpha) \sin(q_2 + \alpha) \exp 2a(\cos(q_1 - \alpha) + \cos(q_2 + \alpha) - 2\cos\alpha) + a^2 \sin q_1 \sin q_2 \exp a(\cos q_1 + \cos q_2 - 2) + (\sin^2(q_1 + q_2) + \cos(q_1 + q_2)) \exp(1 - \cos(q_1 + q_2)).$$
(63)

- [13] Note that $\hat{r}_1(t), -\hat{r}_2(t) \to \pi$, $\hat{\xi}(t) \to 2\pi$ as $t \to \infty$, so the hexagon \hat{H}_t converges to the periodicity square at an O(1/t) rate. As noted above, this follows from the fact that $\hat{\phi}(r_1, r_2, t)$ becomes isotropic as $t \to \infty$.
- [14] J. Hidding, "jhidding/adhesion-example 1.0.0," (2017).
- [15] E. Hopf, Comm. Pure Appl. Math. 3, 201 (1950).

Исследование кодов с локальным восстановлением для случая нескольких восстанавливающих множеств и их обобщений

Камилла Назирханова^{1,2,3}, Станислав Круглик^{3,1}, Алексей Фролов^{3,1}

¹ Институт Проблем Передачи Информации РАН, Москва, Россия, ² Московский физико-технический институт, Москва, Россия

 3 Сколковский институт науки и технологий, Москва, Россия

kamilla.nazirkhanova@phystech.edu, stanislav.kruglik@skoltech.ru,

al.frolov@skoltech.ru

Аннотация В работе рассмотрена задача локального восстановления линейным кодом. Приведены новые верхняя и нижняя границы на кодовое расстояние.

Ключевые слова: коды с локальным восстановлением, кодовое расстояние, границы

1 Введение

Исследование кодов со свойствами локальности, когда каждый символ кодового слова является функцией небольшого числа других кодовых символов, является одной из наиболее бурно развивающихся областей теории кодирования. Первоначальный интерес к данной области был мотивирован наличием очевидных приложений к системам распределенного хранения информации. Позже область возможного применения таких кодов была расширена на системы баз данных, которые имеют тесную связь с так называемым сетевым кодированием.

Важно понимать, что постоянные или временные отключения серверов в системе хранения информации требуют новых схем кодирования, позволяющих быстро и эффективно восстановить недоступные данные. Важной характеристикой таких систем являются такие используемые системные ресурсы, как: количество узлов, к которым происходит обращение при восстановлении, вычислительные ресурсы, определяемые сложностью операции восстановления и некоторые другие. Решения, основанные на использовании кодов Рида-Соломона, были применены в системах хранения информации таких крупных зарубежных компаний, как Facebook и Google. Подход с их использованием был стандартизирован как часть хорошо известной технологии RAID 6 распределенного хранения информации. Однако, ввиду необходимости осуществления запросов информации с большого числа узлов при проведении операции восстановления, данный подход априори не удовлетворяет всем условиям поставленной задачи. Особенно актуальной является задача восстановления информации при отказе небольшого числа узлов, так как данный сценарий является наиболее вероятным. При этом важно не только обеспечить свойство локальности, но и добиться большого кодового расстояния, чтобы в случае большого числа стираний перейти от локального исправления стираний к глобальному, используя различия в кодовом расстоянии.

Задача конструирования кодов, оптимизирующих данную задачу для случая одного стирания, получила широкое распространение в литературе под названием кодов с локальным восстановлением. Известны точные верхние границы данных кодов, а также семейства так называемых оптимальных кодов [1,7,8,9]. В данной работе рассмотрена аналогичная задача, но для случая, в котором восстанавливающие множества могут пересекаться. Предложены новые верхняя и нижняя границы, полученные на основе широко известного метода укорочений, обобщенных весов Хэмминга и каскадного кода.

2 Коды с локальным восстановлением

Обозначим через F_q поле с элементами q. Пусть $[n] = \{1, 2, ..., n\}$. Код $C \subset F_q^n$ имеет локальность r, если каждый символ кодового слова $c \in C$ можно восстановить из подмножества r других символов c [1]. Другими словами, это означает, что при $c \in C, i \in [n]$, существует подмножество координат $R_i \subset [n] \setminus i, |R_i| \leq r$, такое что ограничение C на координаты из R_i позволяет найти значение C_i . Подмножество R_i называется восстанавливающим множеством для символа c_i .

Рассмотрим обобщение этого определения. Код C имеет t не пересекающиеся восстанавливающие множества, если для каждого $i \in [n]$ существуют t попарно не пересекающиеся подмножества $R_i^1, \ldots, R_I^t \subset [n] \setminus i$, что для всех $j = 1, \ldots, t$ и каждой пары символов $a, a' \in F_q, a \neq a'$

$$C(i,a)_{R_i^j} \cap C(i,a')_{R_i^j} = \emptyset.$$
(1)

В дальнейшем мы будем называть такие коды (r, t)-ЛВ. Далее приведем некоторые известные границы для (r, t)-ЛВ кодов.

Первая граница для (r, t)-ЛВ кодов была предложена в работе [2], [3] и имеет вид:

$$d \le n-k+2 - \lfloor \frac{t(k-1)+1}{t(r-1)+1} \rfloor.$$

Улучшение этой границы было получено в [4]:

$$d \le n - \sum_{i=0}^{t} \lfloor \frac{k-1}{r^i} \rfloor.$$

Алфавито-зависимая граница была предложена в статье [5] и имеет форму :

$$d \le \min_{\substack{1 \le x \le * \frac{k-1}{(r-1)t+1}; 1 \le y_j \le t; j \in [x] \\ A < k; x, y_j \in Z^+}} d_{l-opt}^q [n-B, k-A].$$

где $A=\sum_{j=1}^x(r-1)y_j+x,$ $B=\sum_{j=1}^xry_j+x$
и d^q_{l-opt} определяет максимально возможное кодовое расстояние кода на
д $F_q.$

Следующее улучшение границы на кодовое расстояние было получено в [6]:

$$k \le \min_{sr+1 \le n-d} 1 + (r-1)s + k^*(q, n-1-sr, d)$$
$$d \le \min_{1+(r-1)s \le k} d^*(n-1-sr, k-1-(r-1)s)$$

В [11] была представлена рекурсивная конструкция (r, t)-ЛВ кода с параметрами: $n = \binom{r+t}{t}$, $R = \frac{r}{r+t}$ и d = t+1. Эти коды для удобства далее будем называть WZL кодами. WZL код определяется проверочной матрицей, построение которой произведем далее. Определим матрицу H(m, t) кода t-ЛВ с локальностью m-t следующим образом. Каждая строка H(m, t) связана с (t-1)-подмножеством [m], отсортированным в лексикографическом порядке, а каждый столбец - с t-подмножеством [m], также отсортированным в лексикографическом порядке. В этом случае (i, j)-й элемент H(m, t) равен 1, если $E_i \subseteq F_j$, где $E_i - (t-1)$ -подмножество [m], связанное с i-ой строкой, и F_j является t-подмножеством [m], связанным с j-ым столбцом, в противном случае (i, j)-й элемент H(m, t) равен 0. Следует отметить, что H(m, t) имеет C_m^{t-1} строк и C_m^t столбцов, а также блочную структуру следующего вида. Для m > t > 1:

$$H(m,t) = \begin{pmatrix} H(m-1,t-1) & 0\\ I_{C_{t-1}^{m-1}} & H(m-1,t) \end{pmatrix}$$

И для m = t $H(m, t) = H(m, 1)^{\tau} = (1, ..., 1)^{\tau}$ При этом WZL код имеет скорость $R = \frac{r}{r+t}$.

Свойство доступности уменьшает кодовую скорость, поэтому расширим определение, данное для (r, t)-ЛВ кода, добавив свойство, позволяющее восстанавливающим множествам пересекаться не более чем в x позициях. Такие коды будем называть (r, t, x)-ЛВ. Дадим более формальное определение таким кодам.

Код *C* является (r, t, x)-ЛВ, если для каждого $i \in [n]$ существует t подмножеств $R_i^1, ..., R_i^t \subset [n] \setminus i$, для которых справедливо

1. для каждой пары $l, l' \in [t], l \neq l'$

$$|R_i^l \cap R_i^{l'}| \le x;$$

2. для всех $j=1,\ldots,t$ и каждой пары символов $a,a'\in F_q, a\neq a'$

$$C(i,a)_{R_i^j} \cap C(i,a')_{R_i^j} = \emptyset.$$

Данное обобщение было получено в работе [12], это свойство не влияет на возможность восстановления символов при стираниях, но увеличивает максимально достижимую кодовую скорость, верхняя граница на которую представлена ниже:

$$R(C) \le R^*(r, t, x) = 1 - f(r, t, x),$$

где

$$f(r,t,x) = \sum_{j=1,j=1}^{t} \mod 2 {t \choose j} \frac{1}{\overline{N}(r,j,x)+1}$$
$$- \sum_{j=1,j=0 \mod 2}^{t} {t \choose j} \frac{1}{\underline{N}(r,j,x)+1}$$

Расширенная конструкция (r, t, x)-ЛВ кода [12] использует проверочную матрицу (\tilde{r}, \tilde{t}) -WZL кода, обозначенную как \mathbf{H}_{WZL} .

Матрица **H** расширенной конструкци
и(r,t,x)-ЛВ кода получается следующим образом

$$\mathbf{H} = \mathbf{H}_{\mathrm{WZL}} \otimes \underbrace{[11 \dots 1]}_{x+1},$$

где \otimes означает произведение Кронекера.

Полученный код обладает параметрами: $n = (x+1) {\tilde{t} + \tilde{t} \choose \tilde{t}}, d = 2, r = (\tilde{r}+1)(x+1) - 1, t = \tilde{t}$

3 Обобщенные веса Хэмминга

Для каждого линейного кода C носитель $\chi(C)$ может быть определен следующим образом

$$\chi(C) = \{i : \exists (c_1, c_2, ..., c_n) \in C, c_i \neq 0\}$$

Для каждого $1 \leq i \leq k$
i-ый обобщенный вес Хэмминга d_i линейного код
аC

$$d_i = \min\{|\chi(D)| : D \subset C, \dim(D) = i\}$$

Нетрудно заметить, что в частности d_1 равен расстоянию Хэмминга линейного кода и для *i*-ого обобщенного веса Хэмминга [n, k]-линейного кода справедлива обобщенная граница Синглтона для всех $1 \le i \le k$ [14]:

$$d_i \le n - k + i$$

Иерархией весов кода C называется упорядоченный набор $\{d_i | 1 \le i \le k\}$. Очевидно, что для [n,k]-кода $1 \le d_1 < d_2 < \ldots < d_k = n$ [15]

4 Коды в ранговой метрике

Линеаризованный многочлен f(x) над F_{q^m} степени $q \ell$

$$f(x) = \sum_{i=0}^{\ell} a_i x^{[i]}$$

где $a_i \in F_{q^m}$, $i = 0, ..., \ell$, $a_\ell \neq 0$ and $x^{[i]} = x^{q^i}$.

Теперь покажем, как получить кодовое слово $[n_G, k_G]$ кода Габидулина [16]. Выберем произвольный линеаризованный многочлен f(x) над F_{q^m} таким образом, чтобы его степень, равная q, была меньше или равна k-1. Коэффициенты этого многочлена будут являться k информационными символами. Тогда

$$c_G = (f(\alpha_1), f(\alpha_2), \dots, f(\alpha_{n_G})),$$

где элементы $\alpha_1, \alpha_2, \ldots, \alpha_{n_G} \in F_{q^m}$ и являются линейно независимыми векторами длины m над F_q . Положим $m \ge n$, это условие необходимо для существования n линейно независимых векторов.

Отметим свойство линеаризованных многочленов

$$f(a\beta + b\gamma) = af(\beta) + bf(\gamma),$$

где $a, b \in F_q$ и $\beta, \gamma \in F_{q^m}$.

5 Верхняя граница на кодовое расстояние

Для вывода верхней границы на кодовое расстояние (r, t, x)-ЛВ кодов воспользуемся идеями из [11] и [17] и применим метод укорочения кода для определения иерархии весов. Для этого нам понадобится следующая лемма.

Лемма 5.1. Пусть С линейный (r, t, x)-ЛВ код над F_q с параметрами n, k, d и порождающей матрицей G. Проведем процедуру укорочения по аналогии с [11]. Тогда на s-ом шаге получим линейный код C_s над $_q$ длины не более n - 1 - sr, размерности k - 1 - (r - 1)s и i-ым обобщенным весом Хэмминга $d_i(C_s) \ge d_i(C)$ для $1 \le i \le k - r$

Доказательство непременно следует из описания процедуры укорочения, описанной в [11]. Тогда докажем теорему. **Теорема 5.2.** Пусть C - линейный (r, t, x)-ЛВ код над F_q с параметрами $n, k, d u d_i, ede 1 \le i \le k - r, morda$

$$d_i \le \min_{1 \le s \le \lfloor \frac{k-1-i}{r-1} \rfloor} \{d_i^*(q, n-1-sr, k-1-(r-1)s)\}$$

где $d_i^*(q,n,k)$ верхняя граница на і-ый обобщенный вес Хэмминга любого линейного кода.

Эта граница совпадает с верхней границей [11] в случае i = 1.

Теорема 5.3. Пусть C - линейный (r,t,x)-ЛВ код над F_q с параметрами n, k, d u d_i, где $1 \le i \le k-r$ тогда

$$d_i(C) \le n - (k - i) - \lfloor \frac{k - 1 - i}{r - 1} \rfloor$$

Учитывая обобщенную границу Синглтона

$$d_i(C) \le n - k + i$$

Таким образом,

$$d_i(C) \le \min_{1\le s\le \lfloor \frac{k-1-i}{r-1}\rfloor} (n-1-sr-k+1+(r-1)s+i)$$

$$= \min_{1\le s\le \lfloor \frac{k-1-i}{r-1}\rfloor} (n-k-s+i)$$

$$= n - (k-i) - \lfloor \frac{k-1-i}{r-1} \rfloor.$$

В предыдущей теореме мы определили границу для иерархии весов (r, t, x)-ЛВ кода. Используя связь обобщенных весов Хэмминга с кодовым расстоянием линейного кода (см. [17]) мы можем ограничить минимальное расстояние (r, t, x)-ЛВ кода в соответствии со следующей леммой.

Лемма 5.4. Пусть C - линейный (r,t,x)-ЛВ код над F_q с параметрами $n, k \ u \ d,$ тогда

$$d \leq \min_{1 \leq i \leq k-r} \frac{q^i - q^{i-1}}{q^i - 1} (n - (k - i) - \lfloor \frac{k - 1 - i}{r - 1} \rfloor)$$

Результат в асимптотическом случае имеет форму

$$R \le \frac{r-1}{r}(1-\delta) + o(1)$$

Заметим, что факт пересечения восстанавливающих множеств в x позициях не использовался в доказательстве, поэтому данная граница справедлива также для случая простых (r, t)-ЛВ кодов. Численные результаты показывают, что в данном случае полученная граница улучшает границу из [11].

6 Нижняя граница на кодовое расстояние

Рассмотрим каскадный код $C_G \Diamond C_I$. Кодирование информационных символов будет проходить в два этапа. Во-первых, k информационных символов над F_{2^m} кодируются кодом Габидулина C_G . Далее кодовое слова кода Габидулина длины n_G разбивается на локальные группы, каждая из которых впоследствии кодируется при помощи $[n_I, k_I]$ двоичного (r, t, x)-ЛВ кода C_I с проверочной матрицей \mathbf{H}_I , которая имеет вид $\mathbf{H} = \mathbf{H}_{\text{WZL}} \otimes \underbrace{[11 \dots 1]}_{x+1}$

Заметим, что проверки кода C_I являются точками интерполяции многочлена f(x) в точках F_{2^m} , которые являются линейно зависимыми $\alpha_1, \alpha_2 \ldots, \alpha_{n_G}$. Для осуществления процедуры декодирования C_G необходимо интерполировать f(x). Для этого необходимо найти k точек, которые соответствуют линейно независимым элементам из F_{2^m} . Не теряя общности, примем $k_I | n_G$. Обозначим позиции стираний как E_I , $|E_I| = e_I$.

Известно, что код с минимальным расстоянием d способен исправлять не более, чем d-1 стирание. Для этого найдем наихудшую комбинацию из d-1 стирания и покажем, что в этом случае найдутся необходимые точки в F_{2^m} для корректной интерполяции. Наихудшим случаем (т.е. тот, при котором минимальным числом стираний максимально уменьшается ранг) является тот, при котором стирания покрывают кодовые блоки C_I целиком. Тогда число блоков, в которых не произошли стирания, $\lfloor \frac{n-d+1}{n_I} \rfloor$. Мы можем использовать информационные символы этих блоков. В случае, когда n_I не делит n-d+1, мы имеем один частично стертый блок. Тогда можно использовать $k_I - e_I$ информационный символ из него, где e_I - число стираний, произошедших в данном блоке.

В асимптотическом случае граница имеет вид

$$R \ge \frac{r + (t - 1)x}{r + t + (t - 1)x} (1 - \delta) - o(1).$$

Сравнение верхней и нижней границы на скорость (r, t, x)-ЛВ кода можно увидеть на следующих графиках и таблице

7 Заключение

В данной работе исследуются параметры кодов с локальным восстановлением с доступностью, восстанавливающие множества которого могут пересекаться. Для них получены верхняя и нижняя граница на кодовое расстояние.

(r,t)	x = 0 (WZL)	$R^*_{(\mathrm{lin})}(r,t)$	x = 1	$R^*(r, t, x = 1)$
(3, 2)	0.6000	0.6000	0.6667	0.6667
(5, 2)	0.7143	0.7143	0.7500	0.7667
(7, 2)	0.7778	0.7778	0.8000	0.8214
(3, 3)	0.5000	0.5548	0.6250	0.6500
(5, 3)	0.6250	0.6705	0.7000	0.7375
(7, 3)	0.7000	0.7373	0.7500	0.7938



Сравнение границ при разных r и x

Благодарности

Исследование выполнено при финансовой поддержке РФФИ в рамках научных проектов 16-01-00716, 18-07-01427, 18-37-0045.

Список литературы

- Goplan, P. and Huang, C. and Simitci, H. and Yekhanin, S.: On the Locality of Codeword Symbols. IEEE Trans. Inf. Theory, Vol. 57, no. 11, pp. 6925–6934, 2012.
- A. Wang and Z. Zhang: Repair locality with multiple erasure tolerance. IEEE Trans. Inf. Theory, vol. 60, no. 11, pp. 6979–6987, 2014.
- 3. A. S. Rawat, D. S. Papailiopoulos, A. G. Dimakis, and S. Vishwanath: Locality and availability in distributed storage. Proceedings IEEE International Symposium on Information Theory (ISIT), pp. 681–685, 2014.
- 4. I. Tamo, A. Barg, and A. Frolov : Bounds on the parameters of locally recoverable codes. IEEE Trans. Inf. Theory, vol. 62, no. 6, pp. 3070–3083, 2016.
- P. Huang, E. Yaakobi, H. Uchikawa, and P. H. Siege: Linear locallyrepairable codes with availability. Proceedings IEEE International Symposium on Information Theory, pp. 1871–1875, 2016.
- S. Kruglik and A. Frolov: Bounds and Constructions of Codes withAll-Symbol Locality and Availability. Proceedings IEEE International Symposium on Information Theory, pp. 1023–1027, 2017.
- N. Silberstein, A. S. Rawat, O. O. Koyluoglu, and S. Vishwanath Optimal locally repairable codes via rank-metric codes. Proceedings 2013 IEEE International Symposium on Information Theory, pp. 1819–1823, 2013.
- N. Silberstein, A. S. Rawat, and S. Vishwanath Error-correcting regenerating and locally repairable codes via rank-metric codes. IEEE Trans. Inform. Theory, vol. 61, no. 11, pp. 5765–5778, 2015.

- 9. Itzhak Tamo and Alexander Barg A family of optimal locally recoverable codes. IEEE Trans. Inform. Theory, vol. 60, no. 8, pp. 4661–4676, 2014.
- V.R.Cadambe and A.Mazumdar Bounds on the size of locally recoverable codes. IEEE Trans. Inform. Theory, vol. 61, no. 11, pp. 5787–5794, 2015.
- A. Wang, Z.Zhang and M.Liu Achieving arbitrary locality and availability in binary codes. Proceedings 2015 IEEE International Symposium on Information Theory, pp. 1866–1870, 2015.
- S. Kruglik, M. Didina, V. Potapova and A. Frolov On One Generalization of LRC Codes with Availability. Proceedings 2017 IEEE Information Theory Workshop, pp. 1–5, 2017
- 13. A. Wang and Z. Zhang An integer programming based bound for locally repairbale codes.
- V. K. Wei Generalized Hamming weights for linear codes. Proceedings 1991 IEEE Transactions on Information Theory, vol. 37, no. 5, pp. 1412–1418, 1991
- J. Hao, S.-T. Xia, B. Chen and On the Weight Hierarchy of Locally Repairable Codes. 2017 IEEE Information Theory Workshop (ITW), pp 1–5, 2017.
- Ernst Gabidulin, A brief survey of metrics in coding theory. Mathematics of Distances and Applications, pp. 66–68, 2012.
- A. S. Khaled Abdel-Ghaffar and J. H. Weber Bounds for cooperative locality using generalized hamming weights. 2017 IEEE International Symposium on Information Theory, pp. 699–703, 2017

Исследование методов декодирования на основе глубинных нейронных сетей

Данил Шамсимухаметов^{1,2}, Кирилл Андреев^{3,2}, Алексей Фролов^{3,1}

¹ Институт Проблем Передачи Информации РАН, Москва, Россия, ² Московский физико-технический институт, Москва, Россия

³ Сколковский институт науки и технологий, Москва, Россия

danshr@frtk.edu, k.andreev@skoltech.ru, al.frolov@skoltech.ru

Аннотация Рассматривается задача применения глубоких нейронных сетей для декодирования коротких линейных кодов. Исследуется обобщающая способность нейронных сетей на коротких кодах. Показывается, что при использовании нейронных сетей общего назначения вероятность битовой ошибки, соответствующая методу максимального правдоподобия, достижима только, когда обучающая выборка включает в себя все слова. Проводится эксперимент и показывается, что такая сеть не способна к обобщению: она не способна декодировать кодовые слова, не показанные ей в процессе обучения. Таким образом, остро встает проблема уменьшения объема обучающей выборки, ведь число слов кода экспоненциально.

1 Введение

Интерес к применению нейронных сетей в задаче декодирования существовал с 1980-х годов [9,10,11]. Однако, эти ранние работы не оказали существенного влияния на развитие их применения из-за ограничений сетей, которые были доступны в то время. В последние годы идея применения нейронных сетей обрела новую жизнь. Сейчас потенциал вычислительных машин достаточно велик, широта и спектр сфер применения нейронных сетей поражает воображение. Уже существует множество мощных архитектур и эффективных алгоритмов их реализации. Словом, нейронные сети сейчас очень популярны и их потенциал несравнимо велик, однако задача их применения к декодированию по прежнему актуальна. В последнее время глубокие нейронные сети в декодировании изучались в [6,7,4].

Нахмани [4], [5] предложил модель глубинного обучения, которая построена на декодере, основанном на алгоритме распространения доверия (ВР) для LDPC и имеет хорошую обобщающую способность при малом количестве слов, включаемых в обучающую выборку. Однако недостатком его конструкции является то, что для сохранения симметрии системы, чтобы точно имитировать структуру передачи сообщений ВР разнообразие архитектур нейронной сети ограничено. В частности, связи между нейронами повторяют структуру графа Таннера, лежащего в основе ВР, а также активации в нейронах. Данное обстоятельство существенно ограничивает весь спектр возможностей использования богатства доступных различных архитектур нейронных сетей, и исключает возможное применение мощных архитектур, появившихся в последние годы [13].

Основной проблемой применения нейронных сетей является проблема плохого обобщения на кодовых словах, не вошедших в обучающую выборку. Основанное на канальном декодировании глубокое обучение сталкивается с "проклятием размерности"[3]: для короткого кода длины N = 100 и скорости r = 0.5, существуют 2⁵⁰ разных кодовых слов, которых слишком много, чтобы обучить каждое из них. Единственный путь, при котором нейронная сеть может быть обучена для используемых на практике длин блоков, если она изучает какой-то алгоритм декодирования, который может вывести полную кодовую книгу при обучении на малой доле кодовых слов. Единственная возможность - обучиться на малом числе кодовых слов. Остается вопрос, насколько хороша обобщающая способность.

Стоит подчеркнуть, что эта работа основана на кодах с очень короткими длинами блоков, $N \leq 64$, что позволяет сравнивать с декодированием по методу максимального правдоподобия (MAP), напомним, что при таком декодировании на каждом его шаге необходимо сравнение со всеми кодовыми словами, число которых для больших N экспоненциально велико.

В данной работе мы исследуем зависимость вероятности битовой ошибки (BER) от доли общего числа кодовых слов, вошедших в обучающую выборку. Не оставим без внимания и возможность протестировать метод синдромного декодирования линейного кода. Его можно распространить и на коды с большой длиной блока. Важным результатом является то, что при таком подходе, обучаясь лишь на нулевом кодовом слове, нейронная сеть может обобщать результаты и на другие слова, что следует из особенности линейных кодов.

Метод синдромного декодирования хорошо описан в литературе по алгебраическому декодированию (см., Например, [12] [раздел 3.2]). Этот подход отделяет оценку канального шума от оценки переданного кодового слова. В контексте глубинного обучения, интересующего нас в данной работе, его потенциал заключается в уменьшении размера обучающей выборки.

В данной работе представлены модели нейронных сетей, как для получения зависимости BER от числа кодовых слов, так и для подхода, основанного на синдромном декодировании.

2 Глубинное обучение в задаче канального декодирования

Теория глубинного обучения всесторонне описана в [2]. Тем не менее, для полноты мы кратко объясним основные идеи и концепции, чтобы ввести нейронные сети для канального кодирования/декодирования и его терминологию.

Нейронная сеть состоит из многих связанных нейронов, нейроны образуют слои. В таком нейроне все взвешенные входы складываются, возмож-

но добавление смещения, и результат распространяется через нелинейную активационную функцию. Функцией активации могут служить такие, как например сигмовидная функция или ReLU, которые соответственно определяются как:

$$g_{\text{sigmoid}}(z) = \frac{1}{1 + e^{-z}}, \qquad g_{\text{relu}}(z) = \max(0, z).$$
 (1)

Если нейроны располагаются в слоях без обратной связи, мы говорим о прямой нейронной сети, поскольку информация проходит через сеть слева направо без обратной связи (см. Рис. 1). Каждый уровень *i* с входами n_i и m_i выполняет отображение $\mathbf{f}^{(i)} : \mathbb{R}^{n_i} \to \mathbb{R}^{m_i}$ с весами и смещениями нейронов в качестве параметров. Обозначая \boldsymbol{v} как вход и \boldsymbol{w} как выход нейронной сети, отображение ввода-вывода определяется цепочкой функций, зависящей от набора параметров $\boldsymbol{\theta}$ как:

$$\boldsymbol{w} = \mathbf{f}\left(\boldsymbol{v}; \boldsymbol{\theta}\right) = \mathbf{f}^{(L-1)}\left(\mathbf{f}^{(L-2)}\left(\dots\left(\mathbf{f}^{(0)}\left(\boldsymbol{v}, \theta_{0}\right), \dots\right), \theta_{L-2}\right), \theta_{L-1}\right)$$

где L задает количество слоев и называется глубиной нейронной сети. В [1] показано, что такая нейронная сеть с L = 2 и нелинейными активационными функциями теоретически может аппроксимировать любую непрерывную функцию на ограниченной области сколь угодно близко, если число нейронов велико достаточно.

Для того чтобы найти оптимальные веса нейронной сети необходим обучающий набор известных сопоставлений ввода-вывода и определенная функция потерь. С помощью методов оптимизации градиентного спуска и метода обратного распространения ошибок [14] можно найти веса нейронной сети, которые минимизируют функцию потерь над обучающим множеством. Итоговая цель всего обучения – получать на выходе правильные значения для соответствующих неизвестных во время обучения сети входов. Это называется обобщением. Для оценки способности обобщения нужно определить набор данных, который не использовался для обучения, так называемой тестовой выборке.

Опишем модель передачи кодового слова по зашумленному каналу. Вектор $u \in \{0,1\}^k$ кодируется в вектор $v \in \{0,1\}^N$, умножением на порождающую матрицу G линейного кода.

Если бы при передаче по каналу связи не было опасности возникновения ошибок, то каждое из 2^k сообщений можно было бы передавать посредством вектора u = (u1, u2, ..., uk) дли- ны k. Однако ради создания нужного нам расстояния между кодовыми векторами вектор u должен быть подвергнут некоторому преобразованию. Оно состоит в том, что вектор u задаёт линейную комбинацию строк матрицы G

v = uG

Далее нам нужно преобразовать биты в соответствующие аналоговые сигналы. Для это мы используем модуляцию BPSK (Binary Phase Shift Keying) – бинарной фазовой манипуляцией, позволяющей закодировать 1 бит информации за сигнал. Вектор $v \in \{0,1\}^N$ модулируется в вектор $y \in \{-1,1\}^N$ по правилу:

 $y = 1 - 2v, \quad 0 \to 1, \quad 1 \to -1$

После этого сигнал передается в канале с аддитивным Гауссовым шумом (AWGN). Мы выбрали именно AWGN канал, потому что его легко смоделировать: $y \to \hat{y} = y + noise$, где $noise = (n_1, ..., n_N)$ – вектор длины N со случайными компонентами, выбранными из нормального распределения $n \sim N(0, \sigma^2)$.

В машинном обучении формирование обучающей выборки зачастую является очень сложной и дорогостоящей задачей. Но для канального кодирования использование нейронной сети является исключением из правила, потому что мы имеем дело с AWGN, который, как мы напомним, легко моделировать. Поэтому мы можем генерировать столько обучающих образцов, сколько нам необходимо. Более того нам не нужно как-то хитро подбирать метки, на которых происходит обучение, потому что нам известны и переданные кодовые слова и принятый согнал слова с шумом. Именно для простоты используется BPSK модуляция и канал AWGN. Эта гибкость может быть особым преимуществом обучения нейронной сети.

3 Реализация используемых нейронных сетей

В данном разделе представлены две архитектуры нейронных сетей, которые используются для исследований в этой работе. Принципиально они отличаются как по тому, что подается на их вход, так и по количеству входных данных.

3.1 NN-декодер и обучения на большом количестве кодовых слов

Описанный здесь нейронный декодер мы будем использовать для того, чтобы понять хорошо ли сеть может обучаться в зависимости от числа слов, включенных в обучающую выборку. Чтобы сохранить небольшой набор данных для обучения, можно расширить NN дополнительными слоями для модуляции и добавления шума. Эти дополнительные слои не имеют обучаемых параметров. Такой подход позволяет генерировать различные реализации шума непосредственно в ходе обучения, вместо того, чтобы создавать и хранить, достаточно много зашумленных версий одного и того же кодового слова. Таким образом, обучающий набор \mathcal{X} состоит из всех возможных кодовых слов $\mathbf{x}_i \in \mathbb{F}_2^N$ с $\mathbb{F}_2 \in \{0, 1\}$ (метки являются соответствующими информационными битами) и задается параметрами $\mathcal{X} = \{\mathbf{x}_0, \ldots, \mathbf{x}_{2^{k-1}}\}$.

Как рекомендуется в [2], каждый скрытый слой использует функцию активации ReLU, потому что она нелинейна и в то же время её производная вычисляется достаточно просто. Поскольку выходной уровень представляет ет информационные биты, сигмоидальная функция заставляет выходные значения $\sigma(x), x \in \mathfrak{R}$ находиться в диапазоне от 0 до 1, что можно интерпретировать как вероятность передачи «1». Если вероятность близка к бит метки, потеря должна быть увеличена незначительно, тогда как большие ошибки должны привести к очень большой потере. Примерами таких функций потерь являются метод наименьших квадратов (MSE) и бинарная кросс-энтропия (BCE), определенные соответственно как:

$$\mathcal{L}_{\rm MSE} = \frac{1}{k} \sum_{i} \left(b_i - \hat{b}_i \right)^2 \tag{2}$$

$$\mathcal{L}_{BCE} = -\frac{1}{k} \sum_{i} \left[b_i \ln\left(\hat{b}_i\right) + (1 - b_i) \ln\left(1 - \hat{b}_i\right) \right]$$
(3)

где $b_i \in \{0,1\}$ – *i*-ый информационный бит (метка) и $\hat{b}_i \in (0,1)$ мягкая выходная оценка нейронной сети.

Свободно доступные библиотеки обучения с открытым исходным кодом, такие как Theano (https://github.com/Theano/Theano), помогают внедрять и обучать сложные NN модели. Мы используем Keras (https://github.com/kerasteam/keras) как удобный высокоуровневый абстракционный интерфейс для Theano. Это позволяет быстро и свободно задать архитектуру сети, состоящую из наиболее часто используемых типов слоев, с очень абстрактной точки зрения на языке программирования Python.



Рис. 1: Схема декодирования на большом количестве кодовых слов.

Будем исследовать полярный код длины N = 16 и кодовой скоростью r = 0.5. Порождающая матрица полярного кода для кодового блока длины $N = 2^n$ задается как:

$$\mathbf{G}_N = \mathbf{F}^{\otimes n}, \qquad \mathbf{F} = \begin{bmatrix} 1 & 0\\ 1 & 1 \end{bmatrix}$$
(4)

где через $\mathbf{F}^{\otimes n}$ обозначена *n*-ая степень Кронекера матрицы **F**. Теперь кодовые слова можно получить, умножив исходное сообщение на порождающую матрицу $\mathbf{x} = \mathbf{u} \mathbf{G}_N$, где **u** содержит *k* информационных битов и N - k неизменяемых "замороженных" позиций, более подробно можно посмотреть в [8].

Архитектура нейронной сети

Мы уже описали не обучаемые слои, которые предшествуют основной части – нейронному декодеру (NND). Перейдем к описанию последнего. Введем обозначение 128-64-32, в котором описывается конструкция NND, использующая три скрытых слоя с числом узлов соответственно 128, 64 и 32. Отметим, что существуют и другие параметры проектирования с незначительным воздействием на производительность. Вопрос оптимизации гиперпараметров нейронной сети нас интересует гораздо меньше. Приведем лишь те, которые дали положительные результаты.

Будем обучать нейронную сеть для двадцати значений SNR (отношение сигнал-шум) из диапазона от 0 до 5 dB и сравнивать со значениями MAP. Рассматриваются проверки 20000 примеров для каждого SNR. На всем диапазоне SNR используются одни и те же веса.

Мы обучаем наш декодер в так называемых «эпохах». В каждую эпоху градиент функции потерь вычисляется по всему обучающему набору \mathcal{X} с использованием метода оптимизации стохастического градиентного спуска Adam [15]. Поскольку на каждом новом уровне шума в нашей архитектуре генерируются новые шумы, нейронный декодер никогда не будет видеть одни и те же входные данные дважды. По этой причине, хотя обучающая выборка имеет ограниченный размер кодовых слов 2^k , мы можем тренироваться по существенно неограниченному набору обучения, просто увеличивая количество эпох $M_{\rm ep}$, посколько реализация шума генерируется внутри нейронной сети.

3.2 Синдромный NN-декодер, обучение на единственном кодовом слове

Рассмотрим код BCH(63,45). Модель декодера представлена на рисунке 2. От принятого сигнала мы берем жестко знак sign. После этого мы производим демодуляцию с принятием жесткого решения bin – отображение из алфавита $\{\pm 1\}$ в $\{0,1\}$, которое перевод все значения по правилу $-1 \rightarrow 1, 1 \rightarrow 0$. Полученное слово мы умножаем на проверочную матрицу *H* по модулю 2 и получаем синдром. Который и подается на вход нейронной сети. Выходом NN является оценка вектора ошибок, которая не зависит от кодового слова. Умножив его на знаки выходного канального значения, мы получим оценку переданного входного слова.

Архитектура нейронной сети

Нейронная сеть состоит из 15 полносвязных скрытых слоев, на вход каждого из которых поступает объединенные выход предыдущего слоя и синдром жесткого решения. Каждый из скрытых слоев содержит 10N узлов



Рис. 2: Схема нейронного декодера.

и использует функцию активации ReLU. Выходной слой содержит N узлов и применятся с функцией активации гиперболический тангенс, выходными значениями которого являются значения от -1 до 1. Для её реализации применяется библиотека TensorFlow от Google, и в каждую эпоху градиент функции потерь вычисляется с использованием метода оптимизации Adam, о котором уже было сказано выше. Функцией потерь служит:

$$L = \frac{1}{N} \sum_{i=1}^{N} H_{CE}(z_i, \hat{z}_i)$$
(5)

Здесь H_{CE} – функция бинарной кросс-энтропии. z_i – знак і-ой компоненты битов жестких выходных значений, \hat{z}_i – і компонента кодового слова.



Рис. 3: Многослойная нейронная сеть.

Мы будем обучать нейронную сеть на одном нулевом кодовом слове, и задача нашей нейронной сети научиться строить оценку вектора ошибок.

4 Численные результаты

Обучение на большом числе слов

Мы проводим исследование обобщающей способности нейронного декодера. Для начала поймем может ли нейронная сеть, обучаясь на всём наборе кодовых слов, произвести результат близкий к результатам алгоритма максимального правдоподобия и какое количество эпох для этого необходимо. При $M_{ep} = 2^{18}$ для каждого SNR вероятность битовой ошибки на выходе NND совпадает с MAP.



Рис. 4: Сравнение вероятности битовой ошибки для различных размеров обучающей выборки.

Сформируем обучающую выборку, выбрав случайно из всего набора 75% слов. Число эпох $M_{ep} = 2^{18}$. Тестировать нейронный декодер будем уже на всех кодовых словах. Результат представлен на Рис. 4. Видно, что значения BER-кривой значительно ухудшились.

Тот же опыт проделаем, обучая сеть на половине и далее на четверти случайно выбранных слов. Результат представлен на рис. 4.

Таким образом, даже для очень короткого линейного кода (16,8) нейронная сеть не показала хорошей обобщающей способности и можно сделать вывод, что такой подход применения нейронного декодера на кодах с большей длиной блока крайне не эффективен.

Обучение на одном слове

Исследуем теперь метод синдромного декодирования. Рассмотрим линейный код BCH(63,45). Будем подавать на вход в канал каждый раз нулевое слово. Нейронная сеть будет обучаться, принимая на вход только синдром.



Рис. 5: Сравнение вероятности битовой ошибки для декодированного и не декодированного кодового слова.

Обучим нейронный декодер на значениях E_b/N_0 от 4 до 10 и сравним с BER-кривой не декодированного слова. Результат см. Рис.5. Видно, что на значениях SNR, которые декодер видел во время обучения, результат выхода нейронной сети выигрывает у не декодированного значения. Это означает, что подход, при котором нейронной сети не нужно учиться на большом числе кодовых слов, чтобы обобщить результат имеет место быть и нейронная сеть, учась только на одном слове, может обобщить результат.

5 Заключение

В данной работе мы исследовали обобщающую способность нейронных декодеров. Проблема экспоненциальной зависимости числа кодовых слов от длины кода не позволяет методам нейронного декодирования обучаться на всех словах для кодов большой длины. Мы изучили насколько хорошо обучается нейронный декодер, тренируясь лишь на части кодовых слов и пришли к выводу, что даже для короткого кода не можем рассчитывать на высокую обобщающую способность.

Логичным продолжением является исследование методов, при которых не нужно знать все кодовые слова, чтобы обучить нейронную сеть находить ошибки. Одним из таких методов является метод синдромного декодирования.

Мы исследовали этот метод на коде (63,45) и пришли к выводу, что нейронная сеть может обучаться и выдавать улучшенный результат.

Стоит отметить, что видоизменив архитектуру нейронной сети возможно можно добиться более значительного проигрыша ошибок.

Мы изучим и другие методы декодирования, на которых можно обучить нейронную сеть и в дальнейшем попытаемся улучшить результат, полученный синдромным декодером.

Благодарности

Исследование выполнено в ИППИ РАН при финансовой поддержке РН Φ в рамках научного проекта No 14-50-00150.

Список литературы

- 1. K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," Neural Networks, vol. 2, no. 5, pp. 359–366, 1989.
- 2. I. Goodfellow, Y. Bengio, and A. Courville, "Deep Learning," 2016, book in preparation for MIT Press. [Online]. Available: http://www.deeplearningbook.org
- X.-A. Wang and S. B. Wicker, "An artificial neural net Viterbi decoder," *IEEE Trans. Commun.*, vol. 44, no. 2, pp. 165–171, Feb. 1996.
 T. Richardson and R. Urbanke, "The capacity of low-density parity-check codes under message-passing decoding," *IEEE Trans. Inf. Theory*, vol. 47, pp. 599–618, Feb. 2001.
- E. Nachmani, E. Marciano, L. Lugosch, Loren, W.J. Gross, D. Burshtein and Y. Be'ery, "Deep learning methods for improved decoding of linear codes," arXiv:1706.07043, 2017
- E. Nachmani, Y. Bachar, E. Marciano, D. Burshtein and Y. Be'ery "Near Maximum Likelihood Decoding with Deep Learning," *Int. Zurich Seminar on Inf. and Comm.*, 2018
- T. J. O'Shea and J. Hoydis, "An introduction to machine learning communications systems," arXiv:1702.00832, 2017.
- T. Gruber, S. Cammerer, J. Hoydis, and S. t. Brink, "On deep learning-based channel decoding," 51st Annual Conference on Inf. Sciences and Systems (CISS), 2017.
- E. Arikan, "Channel polarization: A method for constructing capacityachieving codes for symmetric binary-input memoryless channels," *IEEE Trans. Inform. Theory*, vol. 55, no. 7, pp. 3051–3073, 2009.
- L. G. Tallini and P. Cull, "Neural nets for decoding error-correcting codes," Proc. IEEE Tech. Applicat. Conf. and Workshops Northcon95, pp. 89–94, Oct. 1995.

- J.-L. Wu, Y.-H. Tseng, and Y.-M. Huang, "Neural network decoders for linear block codes," *Int. Journ. of Computational Engineering Science*, vol. 3, no. 3, pp. 235–255, 2002.
- 11. J. Bruck and M. Blaum, "Neural networks, error-correcting codes, and polynomials over the binary n-cube." *IEEE Trans. Inf. Theory* vol. 35(5), pp. 976–987, 1989.
- 12. S. Lin and D. J. Costello. "Error control coding," 2nd edition, Prentice Hall, 2004.
- 13. I. Goodfellow, Y. Bengio, and A. Courville, "Deep learning." MIT press, 2016.
- D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Parallel distributed processing: Explorations in the microstructure of cognition, vol. 1." Cambridge, MA, USA: MIT Press, 1986, pp. 318–362.
- D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," CoRR, 2014. [Online]. Available: http://arxiv.org/abs/1412.6980

Сверхнадёжная связь с низкой задержкой в сетях Wi-Fi на основе IEEE 802.11ax*

Е.С. Авдотьин, Д.В. Банков, Е.М. Хоров

avdotin.es@phystech.edu, bankov@iitp.ru, khorov@iitp.ru

ИППИ РАН, НИУ ВШЭ, МФТИ

Аннотация Беспроводные сети нового поколения должны поддерживать сверхнадёжную связь с низкой задержкой, предъявляющую строгие требования к задержке при передаче данных и доле потерянных пакетов. Для выполнения данных требований разрабатываются новые решения, позволяющие передавать данные быстрее и надёжнее. Одним из таких решений является дополнение к стандарту Wi-Fi, IEEE 802.11ax, которое дополняет функциональность Wi-Fi технологией OFDMA, которую можно использовать для разделения частотных ресурсов между пользователями сети и их приоритизации, в частности, для того, чтобы резервировать канальные ресурсы для срочных передач. В данной работе предлагаются и исследуются некоторые методы планирования канальных ресурсов между станциями при использовании OFDMA в Wi-Fi, позволяющие выполнять требования на задержку не более 1 мс с надёжностью 99.999%, При помощи имитационного моделирования показано, что предложенные методы позволяют снизить задержку при доставке данных при незначительном уменьшении пропускной способности, доступной для обслуживания несрочного трафика.

1 Введение

Сверхнадёжная связь с низкой задержкой (англ. Ultra-Reliable Low-Latency Communications, URLLC) является важной составляющей развивающихся беспроводных сетей пятого поколения, цель которых — обеспечение передачи данных с высокой надежностью (вероятность потери пакета не более 10^{-5}) и низкой задержкой (менее одной миллисекунды) [1,2]. Столь высокие требования требуют разработки новых режимов работы для технологии беспроводных сетей.

Помимо сотовых сетей, URLLC также очень важна для сетей Wi-Fi, ставшей основной технологией для локальных беспроводных сетей во многих сценариях, связанных с индустриальным и игровым применением, а также с дистанционным управлением и многим другим.

 ^{*} Исследование выполнено в НИУ ВШЭ за счет гранта Российского научного фонда (проект №18-19-00580)

Обеспечить требуемый уровень надежности и задержки в сетях Wi-Fi представляется непростой задачей, поскольку станции Wi-Fi используют метод случайного доступа к каналу, что влечет за собой коллизии и, как следствие, рост задержки и количества потерянных пакетов.

Чтобы удовлетворить высокие и постоянно растущие требования к задержке и пропускной способности в сетях Wi-Fi, комитет IEEE 802 разрабатывает IEEE 802.11ax — новое дополнение к стандарту IEEE 802.11. Оно включает в себя различные способы повышения эффективности Wi-Fi в известных сценариях. В частности, в 802.11ах впервые для сетей Wi-Fi вводится множественный доступ с ортогональным частотным разделением (OFDMA), который позволяет точке доступа выделять частотные ресурсы нескольким устройствам, позволяя им передавать или принимать данные одновременно. Стандарт предоставляет возможность станциям передавать детерминированным образом в выделенных ресурсных блоках, а также описывает метод случайного доступа, аналогичный методу АЛОХА с несколькими каналами. Следует отметить, что стандарт не описывает алгоритм назначения ресурсов для детерминированного и случайного доступа в OFDMA, а лишь предоставляет набор инструментов. Данный инструментарий можно использовать для того, чтобы резервировать ресурсы для потоков URLLC, обеспечивая им доставку данных с высокой надёжностью и с низкой задержкой.

Резервирование ресурсов для потоков URLLC снижает объём ресурсов, доступных для передачи не-URLLC данных. Поэтому возникает задача удовлетворения требований трафика URLLC при том, чтобы для обычного трафика выделялся максимальный объём ресурсов. Эта задача рассматривается в данной работе, где предлагаются и исследуются некоторые подходы к выделению ресурсов станциям URLLC в сети Wi-Fi.

Эффективность рассматриваемых методов оценивается при помощи имитационного моделирования сети стандарта 802.11ах. Предлагаемые подходы сравниваются в терминах доли канальных ресурсов, доступных для передачи не-URLLC данных и доли пакетов URLLC, переданных с задержкой менее 1 мс.

Дальнейшее изложение построено следующим образом. В Разделе 2 описан используемый в сетях 802.11ах метод множественного доступа с частотным разделением. Далее в Разделе 3 описывается метод случайного доступа стандарта 802.11ах. В Разделе 4 приведён обзор литературы, содержащей решения для обслуживания трафика с низкой задержкой и высокой надёжностью. Раздел 5 содержит формальную постановку задачи. В Разделе 6 описаны рассмотренные алгоритмы планирования ресурсов. В Разделе 7 приведены численные результаты. Раздел 8 содержит заключение.

26	26	26	26	26	26	26	26	26	26	26	26	26	26	26	26	26	26
52 52		26	52		52		52		5	52		52		52			
106			26	106			106		26	106							
242					242												
484 тонов																	

Рис. 1. Расположение РБ в канале 40 МГц

2 Множественный доступ с частотным разделением в сетях 802.11ax

Главной особенностью стандарта IEEE 802.11ах и одним из способов улучшить эффективность сетей Wi-Fi является применение технологии множественного доступа с ортогональным частотным разделением OFDMA, которая расширяет стандартный CSMA/CA путём введения возможности разделения ресурсов канала в частотной области.

В OFDMA канал может быть разделён на несколько наборов OFDMAподнесущих(тонов). Группа тонов, отведённых одной станции, называется ресурсным блоком (PБ). В стандарте IEEE 802.11ах каждый из PБ может состоять из 26, 52, 106, 242, 484, 996 или 2 × 996 тонов. Набор доступных PБ зависит от ширины канала: так, например, в канале 40 МГц, станция может использовать PБ размером до 484 тонов (см. рис. 1). Существуют правила деления PБ на более мелкие по размеру. Их общая тенденция гласит: широкие PБ могут быть разделены на примерно вдвое более узкие PБ независимо от других. Как исключения, некоторые PБ, в частности, 996 и 242-тоновые, могут быть разбиты минимум в три PБ. Данное правило деления канала возникает из-за необходимости иметь для каждого PБ служебные тоны, находящиеся с ним в непосредственной близости в частотной области.

В данной работе рассматривается передача данных при помощи OFDMA в восходящем потоке (см. рис. 2). Передача начинается с того, что точка доступа рассылает триггер-кадр (ТК), при помощи которого осуществляет синхронизацию станций и предоставляет им информацию о выделенных ресурсах и параметры передачи (СКК, время передачи, содержание преамбулы кадра PHY и др.). Спустя короткий межкадровый интервал SIFS (англ. Short Inter-Frame Space) станции передают кадры с данными. Затем точка доступа передаёт кадр блочного подтверждения (БП), подтверждая успешность переданных пакетов.



Рис. 2. Восходящие потоки данных в 802.11ах

3 Особенности случайного доступа в сетях 802.11ax

Стандарт позволяет выделять РБ либо адресно, указывая, какой именно станции назначается данный РБ, либо выделять «общие» РБ для передачи в случайном доступе.

Случайный доступ в 802.11ах действует по принципу, аналогичному слотированному многоканальному протоколу Aloha. Стоит отметить, что данным стандартом не исключаются методы случайного доступа, описанные в старых стандартах Wi-Fi.

Для выбора времени и РБ для передачи станциями используется процедура отсрочки OFDMA Back-off (OBO), состоящая в следующем. Станция выбирает случайное значение из равномерного распределения на отрезке [0, OCW], где OCW — конкурентное окно OFDMA. Если текущее значение счётчика отсрочки OBO меньше числа РБ, выделенных в тригтеркадре для случайного доступа, то станция случайным образом выбирает РБ из числа выделенных для случайного доступа и передаёт в нем кадр. Иначе, станция уменьшает OBO на число РБ, выделенных для случайного доступа и ожидает следующего триггер-кадра, содержащего информацию о РБ для случайного доступа.

В случае коллизии значение OCW станции удваивается, пока оно не достигнет заданного предела OCW_{MAX} . В случае успешной передачи OCWстанции присваивается минимальное значение OCW_{MIN} . Значения параметров OCW_{MIN} и OCW_{MAX} задаются точкой доступа в биконах и в управляющих кадрах типа Probe Response Frame.

4 Обзор литературы

Задача распределения ресурсов при необходимости обеспечения малой задержки часто рассматривается в контексте сетей LTE и сетей пятого поколения. Так, в работе [3] рассматривается задача планирования радиоресурсов при обслуживании трафика Тактильного Интернета, чувствительного к задержке и надёжности передачи данных. Авторами рассматриваются функции полезности двух широко известных планировщиков: Max Rate и Proportional Fair, и на их основе разрабатываются планировщики для трафика URLLC, функции полезности которых учитывают не только объём переданных данных, но и то, могут ли данные быть доставлены в пределах ограничения по времени. Авторами показано, что в данном сценарии задача планирования ресурсов может быть сведена к задаче линейной целочисленной оптимизации, которую возможно эффективно, за полиномиальное время, решать методом динамического программирования.

Задача управления канальным ресурсом при обслуживании трафика, чувствительного к задержкам, рассматривается также в [4]. В данной работе предложен алгоритм управления канальным ресурсом иерархического типа, состоящий из двух компонент: алгоритма приоритизации пользователей и планировщика, и описаны общие принципы построения такого алгоритма. Данный алгоритм выделяет пользователей, у которых доля потерянных пакетов (потери могут быть вызваны тем, что пакеты не были доставлены вовремя) менее заданного порога, и выделяет их в приоритизованную группу, которая получает канальные ресурсы в первую очередь. Применение предложенного алгоритма приоритизации пользователей позволяет избежать деградации производительности сети при росте нагрузки и увеличить долю пользователей, чьи требования к надёжности доставки данных и к задержке удовлетворены.

В работе [5] оценивается эффективность некоторых решений для URLLC на высоких уровнях стека протоколов. Показано, что алгоритм планирования Proportional Fair для сетей с низкой задержкой, предложенный в [3] увеличивает пропускную способность сети на 40% по сравнению с существующими решениями. В работе также выяснено, что выполнение требований URLLC существенно зависит от модели трафика. Так, авторами предложена модель двунаправленного трафика и показано, что формирование трафика согласно данной модели вместе с упомянутым выше планировщиком позволяет увеличить пропускную способность почти в два раза.

Исследование сетей IEEE 802.11ах представлено в [6,7]. В работах уделяется внимание особенностям OFDMA в сетях Wi-Fi, и представлена методология построения планировщиков для сетей Wi-Fi. В данных работах рассматривается задача планирования ресурсов для передачи пользовательских данных в восходящем потоке и разрабатывается алгоритм планирования, направленный на минимизацию задержки. Следует отметить, что в данных работах не рассматривается случайный доступ в OFDMA.

Особенность работ [3–7] и их отличие от текущей работы состоит в том, что планирование ресурсов проводится для конечных устройств, запросивших ресурсы у базовой станции, и базовая станция на основании имеющейся у неё информации может решать задачу оптимизации. В настоящей работе, наоборот, точка доступа не знает заранее, каким станциям нужно выделить ресурсы для передачи данных, а у каких станций данных нет, что усложняет планирование ресурсов и вынуждает устройства использовать методы случайного доступа для передачи данных.

Наряду с настоящей работой, вопрос передачи данных с низкой задержкой в сетях Wi-Fi рассматривается в [8]. Авторами предлагается введение категории доступа с наивысшим приоритетом для трафика URLLC и использовать дополнительный радиоинтерфейс, предоставляющий возможность для передачи сигналов занятости канала в специальном узкополосном канале. Так, после получения пакета URLLC на передачу станция передаёт в данном канале сигнал занятости. Получив сигнал занятости, станции, передающие обычные (не URLLC) пакеты, должны сразу освободить канал и дать возможность передать приоритетный пакет. В то же время, станции, передающие пакеты URLLC соревнуются между собой за доступ к каналу согласно методу случайного доступа. В работе показано, что в случае низкой нагрузки на сеть использование данного метода приоритетного доступа позволяет значительно снизить задержку при доставке данных URLLC при небольшом уменьшении пропускной способности обычных станций.

Авторами работы [9] разработан протокол канального уровня узкополосной системы Wi-Fi, работающей в сценарии Интернета Вещей, причём параметры физического уровня соответствуют стандарту 802.11ах. Отличие предлагаемой системы от стандартных сетей 802.11ах состоит в том, что устройства при передаче прослушивают только узкую полосу, что позволяет им снизить энергопотребление. Точка доступа, в свою очередь, распределяет станции по узким (шириной 5 МГц) частотным подканалам. По результатам имитационного моделирования в сценарии множества точек доступа предложены некоторые способы улучшения работы системы. В предположении, что все станции формируют трафик по одинаковому принципу, установлено, что для снижения вероятности коллизии требуется равномерное распределение трафика по свободным РБ. Данный результат пересекается с тем, что получен в текущей работе.

5 Постановка задачи

В работе рассматривается сеть Wi-Fi, работающая согласно стандарту IEEE 802.11ах. В сети имеется точка доступа, к которой подключено N станций, передающих кадры URLLC точке доступа. После передачи кадра станция генерирует новый кадр через время, распределённое экспоненциально с параметром λ .

Данные передаются только при помощи OFDMA, т.е. в сети не используется случайный доступ предыдущих версий Wi-Fi. Точка доступа выделяет станциям ресурсы для передачи кадров, о чем сообщает станциям в триггеркадрах. Далее интервал времени от отправки триггер-кадра до окончания кадра подтверждения будем называть *слотом*. В рассматриваемом сценарии все слоты имеют одинаковую длительность.

Считается, что кадры URLLC имеют маленький размер, и даже в самом узком РБ станция может передать кадр за один слот.

Для данного сценария ставится задача разработать алгоритм выделения ресурсов для передачи кадров URLLC, который позволяет с вероятностью 99.999% обеспечить станциям задержку при передаче кадров URLLC менее 1 мс, и при этом доля потребляемых канальных ресурсов должна быть минимальной.

6 Алгоритмы планирования ресурсов для URLLC

В данном разделе описываются алгоритмы назначения ресурсов, направленные на минимизацию задержки при передаче данных URLLC. Поскольку перед точкой доступа стоит задача минимизации задержки для наибольшего количества станций, передающих небольшие кадры данных, разработанные алгоритмы выделяют РБ минимального размера. Представленные далее алгоритмы имеют общие параметры: f — число РБ, используемых для случайного доступа и F_{max} — максимальное число РБ, которое можно выделить в данном канале. Оба алгоритма выделяют часть ресурсов в случайный доступ, и передача в данных РБ осуществляется так, как описано в разделе 3.

Алгоритмы решают задачу планирования ресурсов на границе слотов.

Поскольку ограничение на задержку при доставке пакетов URLLC очень мало по сравнению с длительностью слота, параметры OCW_{MIN} и OCW_{MAX} следует установить равными 0 для того, чтобы станция не ожидала своей очереди на передачу, а сразу передавала либо успешно, либо с коллизией, но таким образом сигнализируя точке доступа, что необходимо увеличить объём ресурсов, выделяемых для трафика URLLC.

6.1 Алгоритм циклического назначения ресурсов

В каждом слоте точка доступа фиксирует наличие коллизионных передач в ресурсных блоках. Каждый раз когда по окончании слота оказывается, что не произошло ни одной коллизии, в следующем слоте точка доступа выделяет только f PB для случайного доступа. В подобной ситуации точка доступа знает, что в сети нет устройств, не передавших срочные данные, и потому выделяет только общий ресурс, в котором можно передавать сразу, когда появится новый пакет данных. В случае, если в случайном доступе происходил коллизия, точке доступа известно, что в сети есть станции, которым нужно передать данные URLLC, но не знает, какие именно станции имеют данные на передачу. В данной ситуации точка доступа должна дать каждой станции возможность передачи. Для этого точка доступа предоставляет $F_{max} - f$ PБ для передачи детерминированным образом, а оставшиеся fРБ выделяет остальным станциям для передачи методом случайного доступа. Станции, передающие детерминированным образом, выбираются одна за другой: сначала передают первые $F_{max} - f$ станций, потом следующие F_{max} – f станций и т.д. При этом станции перечисляются циклически, как показано на рис. 3. Если в данном цикле оказывается, что в слоте не было коллизий (это возможно только в РБ, выделенных под случайный доступ), точка доступа снова выделяет станциям только f PB для случайного доступа.

6.2 Алгоритм группового назначения ресурсов

Как и в случае предыдущего алгоритма, каждый раз, когда в слоте не происходит коллизий, в следующем слоте точка доступа выделяет только f

Hacmoma

\$8\$4	NRA III	RA	RA
0	5	10	
1	6	11	
2	7	0	
3	8	1	
4	9	2	

Время

Рис. 3. Иллюстрация работы алгоритма циклического назначения ресурсов, штриховкой показаны РБ, в которых происходит коллизия.



Рис. 4. Иллюстрация работы алгоритма группового назначения ресурсов.

РБ для случайного доступа. В случае, если происходит коллизия, точка доступа случайным образом разбивает станции на F_{max} групп и выделяет каждой группе один РБ, при этом размер всех групп поддерживается примерно одинаковым, равным $\lfloor \frac{N}{F_{max}} \rfloor$ или $\lceil \frac{N}{F_{max}} \rceil$, где N — число станций. Если в какой-то группе не произопло коллизий, то на следующем шаге все станции из данной группы будут передавать в РБ, выделяных для случайного доступа. Если в группе происходит коллизия, то все станции из данной группы помечаются как те, которым нужно выделить ресурсы в следующем слоте. Все помеченные станции в следующем слоте заново случайным образом разбиваются на $F_{max} - f$ групп, и каждой группе выделяется один РБ. Если в случайном доступе тоже помечаются как те, которым нужно выделить ресурсы в следующем слоте в следиющем доступе происходит коллизия, то все станции, которые передавали в случайном доступе тоже помечаются как те, которым нужно выделить ресурсы. Данная процедура повторяется до тех пор, пока в слоте наблюдаются коллизии. Иллюстрация работы алгоритма приведена на рис. 4.

7 Численные результаты

Далее приведены результаты имитационного моделирования и произведён сравнительный анализ работы предложенных алгоритмов распределения ресурсов. При имитационном моделировании станции передавали данные во временных слотах длительностью 250 мкс, а максимальное число используемых РБ равнялось 18.

Сравнивались предложенные алгоритмы циклического (кривые *cyclic* на графиках), группового назначения ресурсов (кривые *group*), а также алгоритм статического выделения ресурсов (кривые *static*), при котором заданное число РБ выделяется для случайного доступа. Алгоритмы исследовались при различном значении параметра f.



Рис. 5. Зависимость доли пакетов, переданных с задержкой более 1 мс, от числа станций. Интенсивность поступления пакетов $\lambda = 200c^{-1}$

На рис. 5 представлена зависимость доли пакетов P_{late} , переданных с задержкой более 1 мс, от числа станций. Зависимость построена в линейном масштабе для значений P_{late} меньше 10^{-5} и в логарифмическом для значений P_{late} больше 10^{-5} . Из графиков видно, что хуже всего требования URLLC к задержке удовлетворяются при статическом выделении ресурсов, при этом чем меньше РБ используется в случайном доступе, тем больше средняя задержка и P_{late} . Также следует обратить внимание, что P_{late} в алгоритме циклического назвачения ресурсов резко увеличивается, когда



Рис. 6. Зависимость доли пакетов, переданных с задержкой более 1 мс, от интенсивности поступления пакетов

число станций превышает $(F_{max} - f) \cdot 2$. При данном количестве станций возможна ситуация, когда станция слишком долго ожидает очередь на передачу в детерминированном доступе, и при этом из-за коллизий не может передать в случайном доступе. Наилучшие результаты по P_{late} демонстрирует алгоритм группового назначения ресурсов, однако P_{late} для данного алгоритма начинает возрастать при количестве станций, превышающем $(F_{max} - f) \cdot 3$. В данном случае рост связан с тем, что из-за нехватки канальных ресурсов в детерминированном доступе на каждый РБ выделяется более одной станции.

На рис. 7 представлена зависимость доли пропускной способности сети, доступной не-URLLC станциям от числа станций, передающих кадры URLLC для различных алгоритмов выделения ресурсов. При статическом выделении ресурсов данный показатель вычисляется как $\frac{F_{max}-f}{F_{max}}$ и уменьшается с ростом числа РБ, выделяемых станциям URLLC. При нестатическом выделении ресурсов доля пропускной способности, доступной не-URLLC станциям, уменьшается при увеличении количества станций что связано с тем, что больше времени уходит на разрешение коллизий и на работу в «детерминированной» фазе работы алгоритмов, при которой все ресурсы канала выделяются для разрешения конфликтов. Алгоритм группового назначения ресурсов показывает лучший результат, чем алгоритм циклического назначения ресурсов, так как первый позволяет быстрее разрешать коллизии.



Рис. 7. Зависимость доли канальных ресурсов, доступным не URLLC станциям, от числа станций. Интенсивность поступления пакетов $\lambda = 200c^{-1}$

8 Заключение

В данной работе представлено исследование метода обеспечения сверхнадёжной связи с низкой задержкой (URLLC) в сетях Wi-Fi стандарта IEEE 802.11ах. Данный стандарт предоставляет устройствам возможность передавать данные с использованием технологии OFDMA, для чего точка доступа может разделять канал на ресурсные блоки и назначать их станциям. При этом ресурсные блоки можно назначать станциям не только детерминированным образом, когда указывается, какая именно станция должна передать в данном ресурсном блоке, но и случайным, когда станции равновероятно выбирают один ресурсный блок из нескольких выделенных для случайного доступа.

В данной работе была рассмотрена задача планирования ресурсов для того, чтобы обеспечить устройствам быструю передачу данных так, чтобы вероятность доставить данные позднее чем за 1 мс не превышала 10^{-5} . Были предложены и исследованы несколько алгоритмов планирования ресурсов и показано, что при определённых ограничениях на нагрузку сети, предложенные алгоритмы позволяют выполнять требования URLLC и более эффективны, чем базовый случайный доступ стандарта IEEE 802.11ах как по вероятности своевременной доставки данных, так и по доле канальных ресурсов, доступных для обслуживания обычного трафика.

Список литературы

- 1. Framework and overall objectives of the future development of IMT for 2020 and beyond : Recommendation : M.2083 / ITU-R : 2015. September.
- 2. Ultra Relible Low Latency Cellular Networks: Usecases, Challenges and Approaches / Chen He, Abbas Rana, Cheng Peng et al. -2006.
- Khorov Evgeny, Krasilov Artem, Malyshev Aleksei. Radio Resource Scheduling for Low-latency Communications in LTE and beyond // Quality of Service (IWQoS), 2017 IEEE/ACM 25th International Symposium on / IEEE. – 2017. – P. 1–6.
- Khorov Evgeny, Krasilov Artem, Malyshev Aleksei. Reliable low latency communications in LTE networks // Black Sea Conference on Communications and Networking (BlackSeaCom), 2017 IEEE International / IEEE. - 2017. - P. 1-5.
- Khorov Evgeny, Krasilov Artem, Malyshev Aleksei. Radio resource and traffic management for ultra-reliable low latency communications // Wireless Communications and Networking Conference (WCNC), 2018 IEEE / IEEE. – 2018. – P. 1–6.
- IEEE 802.11ax Uplink Scheduler to Minimize Delay: a Classic Problem with New Constraints / Dmitry Bankov, Andrey Didenko, Evgeny Khorov et al. // PIMRC 2017, IEEE / IEEE. – 2017.
- OFDMA Uplink Scheduling in IEEE 802.11ax Networks / Dmitry Bankov, Andrey Didenko, Evgeny Khorov, Andrey Lyakhov // Communications (ICC), 2018 International Conference on / IEEE. – 2018.
- Enabling Low Latency Communications in Wi-Fi Networks / Dmitry Bankov, Evgeny Khorov, Andrey Lyakhov, Mark Sandal // Personal, Indoor, and Mobile Radio Communications (PIMRC), 2018 IEEE 29th Annual International Symposium on / IEEE. - 2018. - P. 1-5.
- MAC layer design and evaluation of a narrowband Wi-Fi system / Yu Wang, Luis Felipe Del Carpio, Dennis Sundman et al. // Personal, Indoor, and Mobile Radio Communications (PIMRC), 2017 IEEE 28th Annual International Symposium on / IEEE. - 2017. - P. 1–6.

Имитационная модель передачи данных в слотах переменной длительности для радиоинтерфейса New Radio в среде NS-3 *

И.С. Герасин, А.Н. Красилов

gerasin.is@phystech.edu, krasilov@iitp.ru

ИППИ РАН

Аннотация В начале 2018 года консорциумом 3GPP была опубликована спецификация нового радиоинтерфейса (англ. New Radio, NR) для сетей пятого поколения. Одним из наиболее важных нововведений по сравнению с радиоинтерфейсом Long-Term Evolution (LTE) является возможность конфигурирации параметров физического уровня, в том числе возможность изменения длительности слота, используемого для передачи данных. В данной работе в среде имитационного моделирования NS-3 разработана модель данного механизма. С помощью модели проведено исследование влияния выбора длительности слота на производительность сети при обслуживании различных типов трафика.

1 Введение

В настоящее время технология Long-Term Evolution (LTE) является одной из основных технологий для обеспечения широкополосного доступа мобильных устройств к сети Интернет. Растущие ожидания пользователей требуют развития технологий беспроводной связи. В начале 2018 года консорциум 3GPP опубликовал спецификацию нового радиоинтерфейса (англ. New Radio, NR) для сетей пятого поколения (5G). Одним из главных отличий радиоинтерфейса NR от существующего LTE является возможность адаптивного выбора длительности слота, в котором осуществляется передача данных, в зависимости от типа обслуживаемого трафика. В частности, при использовании радиоинтерфейса LTE передача данных может осуществляется только в слотах фиксированной длительности, равной 1 мс. Однако для ряда сервисов, которые должны будут предоставляться в сетях 5G (например, сервиса сверхнадежной связи с малой задержкой), требуется обеспечить время доставки данных менее одной миллисекунды, что невозможно достичь посредством LTE. Для решения данной проблемы NR предоставляет возможность осуществлять передачу данных в слотах переменной длительности.

^{*} Исследование выполнено в ИППИ РАН за счет гранта Правительства Российской Федерации (Договор No 14.W03.31.0019)

Так, например, для сервиса сверхнадежной связи с малой задержкой длительность слота может быть уменьшена до сотен микросекунд. Уменьшение длительности слота с одной стороны приводит к уменьшению задержки при передаче данных, однако, с другой стороны, приводит к необходимости осуществлять передачу коротких пакетов, что может существенно снижать спектральную эффективность использования ресурсов радиоканала [1,2]. В связи с этим возникает вопрос, каким образом длительность слота влияет на производительность сети при обслуживании различных типов трафика. Для решения данной задачи в работе построена имитационная модель передачи данных в слотах переменной длительности для радиоинтерфейса NR в среде NS-3. С помощью модели проведено исследование влияния длительности слота на производительность сети при обслуживании трафика сверхнадежной связи с малой задержкой и широкополосного трафика (в частности, веб-трафика).

2 Описание модели

В данной работе используется среда имитационного моделирования NS-3 [3], которая является мощным инструментом, доступным для исследователей и разработчиков новых систем связи. В ней имеется реализованная модель технологии LTE. Так как технологии LTE и NR используют схожие принципы, то реализация механизма управления длительностью слота построена на базе существующей модели LTE. В разделе 2.1 описана модель LTE, реализованная в среде NS-3. В разделе 2.2 описана модификация данной модели для реализации механизма управления длительностью слотов.

2.1 Модель физического уровня технологии LTE в среде NS-3

В технологии LTE временная ось делится на слоты фиксированной длительности равной 1 мс. Передача данных осуществляется при помощи метода ортогонального частотного разделения каналов (англ. Orthogonal Frequency-Division Multiplexing, OFDM). Расстояние между поднесущими фиксировано и равно 15 кГц. Тем самым длительность одного OFDM символа при использовании короткого циклического префикса равна 71.4 мкс. Таким образом, слот состоит из 14 OFDM символов. Согласно стандарту [4], служебная информация в нисходящем канале передается в начале каждого слота и занимает от одного до трех OFDM символов по всей ширине канала. Фактическая длительность сообщается во вспомогательном канале управления физическим уровнем (англ. Physical Control Format Indicator Channel, PCFICH). Служебная информация передается в так называемом физическом канале управления нисходящим каналом (англ. Physical Downlink Control Channel, PDCCH). В PDCCH передаются сообщения, поступающие с уровня управления доступом к среде (англ. Media Access Control, MAC), в которых планировщик указывает какие частотно-временные ресурсы выделены каждому пользователю. В NS-3 передача PCFICH и PDCCH моделируются как передача кадра управления с фиксированной длительностью 3



Рис. 1. Структура слота LTE в NS-3 для нисходящего и восходящего канала

OFDM символа, охватывающего всю доступную ширину канала (см. рис. 1). Оставшиеся 11 OFDM символов используются для передачи пользовательской информации (англ. Physical Downlink Shared Channel, PDSCH). В восходящем канале осуществляется передача пользовательских данных (англ. Physical Uplink Shared Channel, PUSCH), служебной информации (англ. Physical Uplink Control Channel, PUCCH), а так же опорных сигналов (англ. Sounding Reference Signal, SRS), используемых для оценки качества канала. В NS-3 передача PUSCH&PUCCH моделируется как передача кадра длительностью 13 OFDM символов, за которой следует передача SRS длительностью 1 OFDM символ.

2.2 Реализация механизма управления структурой слота для технологии NR в среде NS-3

Технологии NR и LTE используют схожие принципы построения физического уровня. В частности, также как и в LTE в NR используется метод OFDM. Однако в отличие от LTE в NR можно адаптивно изменять расстояние между поднесущими, а следовательно изменять длительность OFDM символа. Кроме того, в зависимости от типа обслуживаемого трафика в NR можно изменять структуру слота (например, изменять число OFDM символов, приходящихся на передачу служебной информации и пользовательских данных). Таким образом, структура слота в NR описывается следующими параметрами: (а) расстояние между поднесущими, (б) длительность одно-


Рис. 2. Иерархия классов, реализующих физический и канальный уровень LTE в среде NS-3

го OFDM символа, (в) число OFDM символов, приходящихся на передачу служебной информации и пользовательских данных в восходящем и нисходящих каналах. Как отмечено в разделе 2.1, для модели LTE в NS-3 используются фиксированные значения этих параметров. Для реализации механизма управления структурой слота, предусмотренного технологией NR, были модифицированы следующие классы модуля LTE (см. рис. 2).

 $1. \ LteUeMac$

Класс, отвечающий за моделирование МАС уровня пользователя. В него добавлен член класса, определяющий размер слота. Также в методе SendRaPreamble() добавлена возможность изменения параметра времени ожидания пользователем ответа от базовой станции на сообщение случайного доступа с учетом заданной длительности слота.

2. LteUePhy

Реализует модель физического (англ. РНҮ) уровня на пользователе. В этом классе были добавлены параметры, определяющие длительность передачи PUSCH&PUCCH и SRS. Модифицирован метод SubframeIndication(),

который управляет длительностью слота восходящего канала с учетом новых параметров физического уровня.

3. LteEnbPhy

Отвечает за моделирование физического уровня базовой станции. Добавлены члены класса, определяющие длительность PCFICH&PDCCH и PDSCH в нисходящем канале. В методах *SendDataChannels()* и *StartSubFrame()* длительность слота нисходящего канала изменяется с учетом новых параметров.

4. LteSpecrtumPhy

Отвечает за моделирование беспроводного канала в сети LTE. Добавлены новые члены класса, определяющие длительность опорных сигналов и служебной информации в восходящем и нисходящем каналах соответственно. Изменены методы класса StartTxDlCtrlFrame() и StartTxUlSrsFrame(), отвечающие за конфигурацию слота восходящего и нисходящего канала с учетом новых параметров физического уровня.

5. LteAmc

Осуществляет выбор сигнально-кодовой конструкции и оценку размера транспортного блока (объема данных, которые могут быть переданы в слоте). Добавлены члены класса, определяющие длительность PUSCH в восходящем канале и PDSCH в нисходящем канале. Согласно [5] реализована функция расчета размера транспортного блока в зависимости от параметров физического уровня. В методах GetDlTbSizeFromMcs() и GetUlTbSizeFromMcs() рассчитывается размер транспортного блока в нисходящем и восходящем каналах соответственно.

6. LteHelper

Конфигурация параметров физического уровня осуществляется в начале эксперимента с помощью вспомогательного класса *LteHelper*.

3 Численные результаты

Имитационное моделирование проводилось в следующем сценарии. Базовая станция находится на высоте 30 м, к которой подключен пользователь, находящийся на высоте 1 м. Расстояние между ними – параметр эксперимента. Ширина канала – 25 ресурсных блоков (5 МГц). Рассматривается частотно-селективная модель затухания радиосигнала Extended Pedestrian A [6]. К базовой станции подключен сервер по проводному соединению с пропускной способностью 10 Гбит/с и задержкой передачи пакетов 0 мс. Рассматривается две возможные конфигурации слота.

1. Конфигурация с длинными слотами (соответствует конфигурации слота в сетях LTE).

Длительность OFDM символа – 71.4 мкс. В нисходящем канале 3 OFDM символа используются для передачи PCFICH&PDCCH, 11 – для PDSCH. В восходящем 13 OFDM символов для передачи PUSCH&PUCCH, 1 – для передачи SRS. Длительность слота 1 мс.



Рис. 3. Время передачи пакета в нисходящем (DL) и восходящем (UL) каналах.

 Конфигурация с короткими слотами. Длительность OFDM символа – 71.4 мкс. В нисходящем канале 1 OFDM символ используются для передачи PCFICH&PDCCH, 1 – для PDSCH, в восходящем 1 OFDM символ используется для передачи PUSCH&PUCCH, 1 – для передачи SRS. Таким образом, длительность слота 142.8 мкс.

3.1 Оценка времени передачи пакета

В первом эксперименте исследовалась зависимость времени передачи пакета на уровне приложений при использовании протокола транспортного уровня User Datagram Protocol (UDP) от расстояния между пользователем и базовой станцией. Пакеты могут отправляться с сервера пользователю (нисходящий трафик, англ. downlink, DL) и от пользователя к серверу (восходящий трафик, англ. uplink, UL). Рассматривается передача потока постоянной интенсивности. Размер пакета на уровне приложений равен 32 байтам. Интервал между пакетами – 100.1 мс. Данное значение длительности интервала было выбрано для того, чтобы избежать синхронизации моментов отправки пакетов с границами слотов.

Результаты моделирования представлены на рис. 3. Видно, что использование коротких слотов позволяет существенно снизить задержку при передаче пакета. Рост задержки при больших расстояниях (>200 м) и использовании коротких слотов обусловлен значительным ухудшением качества канала (т.е. необходимостью совершать дополнительные попытки передачи пакета), а также уменьшением размера транспортного блока. При использовании коротких слотов на расстоянии до 150 м и передачи в нисходящем канале задержка уменьшается в 7 раз по сравнению с длинными слотами и



Рис. 4. Пропускная способность сети в нисходящем (DL) и восходящем (UL) каналах.

не превышает 1 мс. Таким образом, использование коротких слотов в нисходящем канале позволяет обеспечить обслуживание трафика сверхнадежной связи с малой задержкой. В восходящем канале при использовании коротких слотов задержка превышает 1 мс. Это связано с тем, что пользователю перед отправкой пакета необходимо запросить частотно-временные ресурсы у базовой станции и получить «разрешение» (англ. grant) на отправку данных.

3.2 Оценка пропускной способности сети

Во втором эксперименте измерялась пропускная способность соединения между пользователем и базовой станцией. Для этого на пользователе/сервере генерировался насыщенный поток данных с помощью протокола UDP. Размер пакета на уровне приложений равен 1460 байтам.

Результаты моделирования представлены на рис. 4. На графике видно, что использование коротких слотов приводит к значительному падению пропускной способности по сравнению с длинными слотами. Это вызвано следующими причинами. Во-первых, в случае использования коротких слотов уменьшается доля OFDM символов, приходящихся на передачу пользовательских данных (передачу PDSCH/PUSCH). Во-вторых, уменьшается размер транспортного блока, передаваемого в слоте, что, в свою очередь, требует использования более робастных сигнально-кодовых конструкций (CKK) для достижения заданной вероятности успешного декодирования транспортного блока. Например, на расстояниях меньше 10 м между базовой станцией и пользователем базовая станция использует СКК с наибольшей номинальной скоростью. При использовании длинных слотов доля OFDM символов, приходящихся на передачу пользовательских данных (передачу PDSCH), равна 11/14, а при использовании коротки слотов – 1/2. Из рис. 4 видно, что отношение пропускных способностей в нисходящем канале при использовании длинных и коротких слотов на расстояниях меньше 10 м приблизительно равно (11/14)/(1/2) = 11/7. При больших расстояниях это соотношение увеличивается из-за того, что при том же расстоянии (при тех же канальных условиях) в случае коротких слотов необходимо использовать более робастную СКК (т.е. СКК с меньшей номинальной скоростью).

3.3 Оценка скорости загрузки веб-данных

В третьем эксперименте осуществляется передача веб-данных (файла ограниченного размера) с помощью протокола транспортного уровня Transmission Control Protocol (TCP). Размер TCP сегмента MSS = 1460 байт. Начальное значение окна перегрузки 10 MSS. Исследуется зависимость средней скорости передачи данных от размера веб-страницы. Скорость передачи данных – это отношение размера веб-страницы к полному времени ее загрузки.

На рис. 5 приведены результаты, полученные для ширины канала 25 ресурсных блоков (5 МГц) и 100 ресурсных блоков (20 МГц), когда пользователь находится на расстоянии 10 м. Видно, что использование коротких слотов позволяет увеличить скорость загрузки при малом размере страниц по сравнению с длинными слотами. Это происходит из-за зависимости от времени начальной стадии передачи данных через протокол ТСР, когда окно перегрузки далеко от своего оптимального значения. Скорость передачи данных по протоколу ТСР оценивается как отношение размера окна перегрузки к промежутку времени между отправкой пакета и получением подтверждения. При загрузке веб-страницы с малым размером на начальной стадии основное влияние оказывает маленькая задержка передачи пакета, а не максимальная пропускная способность, которая еще не достигнута. При больших размерах веб-страницы за время ее передачи протокол TCP успевает увеличить окно перегрузки так, чтобы достичь значения пропускной способности канала. В частности, при использовании коротких слотов и ширине канала 25 ресурсных блоков скорость загрузки веб-страницы достигает значения пропускной способности канала (см. соответствующие результаты на рис. 4).

4 Заключение

В данной работе была построена имитационная модель передачи данных в слотах переменной длительности для радиоинтерфейса NR в среде имитационного моделирования NS-3. Исследована зависимость задержки при передаче пакетов от структуры слота. Показано, что использование коротких слотов позволяет уменьшить время передачи пакетов до 1 мс, что необходимо для сервисов сверхнадежной связи с малой задержкой. Исследована



Рис. 5. Средняя скорость передачи в зависимости от размера страницы в нисходящем (DL) и восходящем (UL) каналах.

связь между пропускной способностью канала и конфигурацией слота. Показано, что использование длинных слотов эффективней в случаях, когда требуется передавать большие объемы данных. Так же исследовано влияние длительности слотов на скорость обслуживания веб-трафика. Показано, что при малых размерах веб-страниц оказывается выгоднее использовать короткие слоты, а для больших веб-страниц – длинные слоты. В дальнейших работах планируется разработать адаптивный алгоритм выбора параметров слота в зависимости от типа обслуживаемого трафика, его характеристик, а также требований к качеству обслуживания.

Список литературы

- 1. Link adaptation design for ultra-reliable communications / H. Shariatmadari, Z. Li, M. A. Uusitalo et al. // 2016 IEEE International Conference on Communications (ICC). -2016. -May. P. 1-5.
- 2. MAC layer enhancements for ultra-reliable low-latency communications in cellular networks / G. Pocovi, B. Soret, K. I. Pedersen, P. Mogensen // 2017 IEEE International Conference on Communications Workshops (ICC Workshops). 2017. May. P. 1005–1010.
- 3. Network Simulator 3. https://www.nsnam.org/.
- 4. Evolved Universal Terrestrial Radio Access (E-UTRA); Physical channels and modulation : Technical Specification (TS) : 36.211 / 3rd Generation Partnership Project (3GPP) ; Executor: 3GPP : 2017. 03. Version 14.2.0.
- 5. NR; Physical layer procedures for data : Technical Specification (TS) : 38.214 / 3rd Generation Partnership Project (3GPP) ; Executor: 3GPP : 2018.-01.-Version 15.0.0.
- Evolved Universal Terrestrial Radio Access (E-UTRA); Base Station (BS) radio transmission and reception : Technical Specification (TS) : 36.104 / 3rd Generation Partnership Project (3GPP) ; Executor: 3GPP : 2017. -03. - Version 14.3.0.

Планирование ресурсов в сетях IEEE 802.11ax для случая частотно-селективного канала*

Д.В. Банков, С.А. Тутельян, Е.М. Хоров

{bankov, tutelian, khorov}@iitp.ru

ИППИ РАН

Аннотация В данной работе рассмотрен процесс передачи в восходящем канале в новом стандарте IEEE 802.11ах. В нем впервые для сетей 802.11 вводится технология множественного доступа с ортогональным частотным разделением каналов, которая предоставляет возможность одновременных многопользовательских передач. В работе исследована задача распределения ресурсов между пользователями в условиях частотно-селективных замираний. Предложен жадный подход к планированию ресурсов точкой доступа, который был выбран по причине высокой вычислительной сложности задачи. Приведено сравнение данного подхода с планировщиками, не учитывающими частотную селективность, и планировщиками, не допускающими одновременные передачи в канале.

1 Введение

Сети стандарта IEEE 802.11 (Wi-Fi) прочно вошли в жизнь большинства людей. Они удовлетворяют большому количеству различных требований к локальным сетям передачи данных, будь то требование максимально быстрой работы в домашней сети или желание надежной работы сети в общественном месте. 802.11ах является новым дополнением данного стандарта, в котором основной упор сделан на повышение эффективности работы пользователей с сетью. В условиях большого количества точек доступа и еще большего количества подключенных к ним устройств задача обеспечить приемлемую работу с сетью для каждого устройства становится довольно важной и непростой.

Начиная с версии 802.11а, в сетях Wi-Fi предусмотрено использование технологии мультиплексирования с ортогональным частотным разделением (Orthogonal Frequency-Division Multiplexing, OFDM) в диапазоне 5 ГГц, которая представляет из себя способ модуляции сигналов при помощи большого количества поднесущих. Основным нововведением стандарта 802.11ах является применение множественного доступа с ортогональным частотным

^{*} Исследование выполнено в ИППИ РАН за счёт гранта Российского научного фонда (проект №16-19-10687)

разделением (Orthogonal Frequency-Division Multiple Access, OFDMA). Данная технология позволяет разделять по частоте весь доступный канал между различными пользователями, т. е. одновременно могут передавать или принимать несколько устройств.

Производительность OFDMA зависит от способа частотного разделения канала по пользователям (станциям). Один из самых простых вариантов заключается в выделении всего канала только одному пользователю. Такое управление каналом может оказаться неэффективным в случае, когда многим станциям надо передавать небольшие порции данных: выделение всего канала каждой избыточно. Поэтому возникает задача планирования ресурсов — построения «хорошего» частотного разбиения канала по пользователям. В данной работе эта задача рассматривается для случая передач в восходящем канале (т. е. от станций к точке доступа), так как в нисходящем канале данное планирование осуществляется проще. Также передаваемые сигналы подвержены частотно-селективным замираниям, и необходимо учитывать их влияние при планировании ресурсов.

В разделе 2 дано описание основ применения OFDMA в стандарте 802.11ах. В разделе 3 приведен обзор литературы. Раздел 4 содержит постановку задачи планирования. В разделе 5 предложен жадный подход решения поставленной задачи. В разделе 6 представлены результаты численного исследования. Раздел 7 содержит заключение.

2 Передачи с использованием OFDMA

При использовании OFDMA весь канал делится на группы поднесущих (тонов) — ресурсные блоки (РБ). РБ может состоять из следующего количества тонов: 26, 52, 106, 242, 484, 996 и 2×996. Доступные наборы РБ отличаются для каналов разной ширины. Например, структура РБ для канала 40 МГц показана на рис. 1. 484-тоновый РБ занимает весь канал целиком. Данный РБ состоит из двух 242-тоновых РБ. Эти РБ, в свою очередь, состоят из РБ меньшего размера.

Согласно стандарту 802.11ах, управление OFDMA-передачей лежит на точке доступа. Она выбирает, каким пользователям когда и как передавать. Каждой станции для передачи может быть выделено не более одного РБ. При этом РБ может быть любого доступного размера. Количество передаваемых данных зависит от ширины РБ и от используемой сигнально-кодовой конструкции (СКК). Чем шире РБ и чем быстрее СКК, тем больше данных способна передать станция в единицу времени. Конкретные значения скоростей передачи в зависимости от СКК и размера РБ представлены в табл. 1. Стоит отметить, что высокоскоростные СКК 1024-QAM доступны для использования только в широких РБ.



Рис. 1: Структура ресурсных блоков в канале 40 МГц

N⁰	CKK	26 тонов	52 тона	106 тонов	242 тона	484 тона	996 тонов
1	BPSK, $1/2$	0.8	1.7	3.5	8.1	16.3	34
2	QPSK, $1/2$	1.7	3.3	7.1	16.3	32.5	68.1
3	QPSK, 3/4	2.5	5	10.6	24.4	48.8	102.1
4	16-QAM, 1/2	3.3	6.7	14.2	32.5	65	136.1
5	16-QAM, 3/4	5	10	21.3	48.8	97.5	204.2
6	64-QAM, 2/3	6.7	13.3	28.3	65	130	272.2
7	64-QAM, 3/4	7.5	15	34.9	73.1	146.3	306.3
8	64-QAM, 5/6	8.3	16.7	35.4	81.3	162.5	340.3
9	256-QAM, 3/4	10	20	42.5	97.5	195	408.3
10	256-QAM, 5/6	11.1	22.2	47.2	108.3	216.7	453.7
11	1024-QAM, 3/4				121.9	243.8	510.4
12	1024-QAM, $5/6$				135.4	270.8	576.1

Таблица 1: Скорость передачи, Мбит/с

Выбор РБ и СКК для передачи зависит от отношения сигнал/шум. Чем быстрее СКК и чем шире РБ, тем более высокое соотношение сигнал/шум требуется для успешной передачи. Для этого необходимо выделять больше мощности на передачу, однако каждая станция имеет собственное ограничение на максимальную мощность и, следовательно, ограничена в выборе пары РБ и СКК.

Еще одним ограничением [1] на выбор РБ и СКК является то, что станции должны передавать данные с такой мощностью, чтобы уровень приходящего на точку доступа сигнала был приблизительно одинаков для всех одновременно передающих станций. Если отношение сигнал/шум будет сильно различаться для разных РБ одного OFDMA-кадра, то точка доступа может принять данные с ошибкой.

Ситуация, когда точка доступа выделяет ресурсы конкретной станции, называется детерминированным доступом. Также 802.11ах предусматривает возможность стандартного для Wi-Fi случайного доступа к каналу. Однако для упрощения модели в данной работе рассматривается только детерминированный доступ. Общий вид процесса передачи в восходящем канале без случайного доступа показан на рис. 2.



Рис. 2: OFDMA передача

Передачи внутри одного OFDMA-кадра должны начинаться и заканчиваться одновременно. Для этого используются триггер-кадры (ТК), которые отправляются точкой доступа, если у станций есть даные на передачу. В них указываются назначения для станций: длительность передачи, СКК, мощность, РБ для передачи и другие параметры. После этого все станции, получившие назначения РБ, передают данные. Затем точка доступа отправляет кадр подтверждения, в котором указывается, данные каких станций удалось принять.

3 Обзор литературы

Несмотря на то, что окончательный вариант стандарта 802.11ax запланирован для публикации лишь в 2019 году, основные его положения известны и вряд ли изменятся. Поэтому уже проводятся исследования его эффективности. В работе [2] рассматривались различные планировщики для разбиения всего канала между пользователями. Показано, что использование возможности одновременной передачи несколькими пользователями дает уменьшение задержки передачи потоков, а так же ощутимо увеличивает пропускную способность, по сравнению со случаем, когда одновременно передавать может лишь один. Однако данное исследование было проведено без учета частотно-селективных замираний, что исправляет настоящая работа.

В работе [3] рассматриваются многопользовательские передачи в восходящем потоке. Показано, что 802.11ах превосходит 802.11ас по пропускной способности и задержке. При этом делается допущение, что канальные условия для всех станций одинаковы, а планировщик просто разделяет весь канал поровну.

В работе [4] исследована задача распределения РБ для пользователей разных категорий обслуживания (VoIP, видеотрафик), принимая во внимание такие факторы, как заполненность буферов передачи, параметры QoS, индикаторы справедливости обслуживания. Также предложен подход предсказания с помощью известных методов регрессии (например, метод *k*-ближайших соседей) значений некоторых из этих факторов в случае, если их значения редко обновляются.

Случайный доступ в передачах с помощью OFDMA в восходящем канале был изучен в работе [5]. Показано, что необходим баланс в выделении ресурсов для детерминированного и случайного доступа. Уменьшение количества ресурсов для случайного доступа ведет к повышению пропускной способности, но в то же время повышает время доставки отчетов о текущем состоянии буфера, что может приводить к большим задержкам передачи некоторых потоков.

4 Задача планирования

Рассматривается следующая модель сети: дана точка доступа, вокруг нее в некотором круге равномерно распределены N станций. Через случайные промежутки времени у станций появляются потоки данных на передачу. В данной работе не учитывается случайный доступ, поэтому считается, что при возникновении потока, точка доступа сразу узнает об этом. Следовательно, в любой момент времени t точка доступа знает, у каких станций есть данные на передачу. Этим станциям точка доступа должна выделить ресурсы.

Каждой станции из некоторого подмножества станций, у которых есть данные на передачу, назначается один РБ. Эти назначения посылаются в триггер-кадре, который передается на самой медленной СКК, и предполагается, что его принимает каждая станция. Станции, которым выделены ресурсы, передают в соответствующих РБ. После этих передач точка доступа снова планирует ресурсы и генерирует следующий триггер-кадр.

Согласно стандарту, при планировании точка доступа должна назначить РБ станциям, удовлетворяя следующим *ограничениям*:

 Каждой станции может быть выделено не более одного РБ. Ширина РБ может быть любой допустимой (26 тонов, 52 тона и т. д.). – Один и тот же частотный отрезок полосы не может быть выделен двум и более станциям одновременно. Например, для структуры РБ с рис. 1 если одной станции был выделен РБ под номером 1, то другим станциям не могут быть выделены РБ с номерами 19, 27, 31, 33.

Также точка доступа выбирает мощность и СКК, на которых будут передавать станции.

Для решения задачи планирования задается некоторая функция полезности U, значение которой планировщик должен максимизировать. Пусть дано множество станций $S = \{s_i\}_{i=1}^n$, где n — количество станций, у которых есть данные на передачу, а также множество всех РБ для данного канала $R = \{r_j\}_{j=1}^J$. r_j отличаются друг от друга шириной и расположением в полосе. Например, все РБ для канала 40 МГц пронумерованы на рис. 1. Множество назначений РБ станциям — множество пар $X = \{(s_i, r_j)\}$, где (s_i, r_j) означает, что станции s_i был выделен РБ r_j . Исходя из структуры РБ, можно определить операции включения и пересечения различных РБ. $r_i \subset r_j$, если r_j содержит все частоты, которые содержит r_i . Например, на рис. 1 $r_1 \subset r_{19}, r_{24} \subset r_{32}$ и так далее. $r_i \cap r_j = r_i$, если $r_i \subset r_j$; $r_i \cap r_j = r_j$, если $r_j \subset r_i$; $r_i \cap r_j = \emptyset$, если $r_i \not\subset r_j$ и $r_j \not\subset r_i$. Стоит отметить, что не может быть частично перекрывающихся РБ: если 2 разных РБ имеют общие частоты, то один из них полностью содержится в другом. Для упрощения записи $s \in X \iff \exists (s, r) \in X$ и $r \in X \iff \exists (s, r) \in X$.

С помощью данных операций можно определить указанные выше ограничения на X:

$$s_i \neq s_j, \forall s_i, s_j \in X,$$

$$\bigcap_{r \in X} r = \emptyset.$$
(1)

Пусть теперь H — множество всех назначений, удовлетворяющих условиям (1), C — множество доступных СКК. $\lambda(s, r, c)$ — прирост функции полезности в случае, если алгоритм выделил станции s РБ r на СКК $c \in C$. Если станция не поддерживает передачу на данной СКК в данном РБ (например, из-за отсутствия необходимой мощности), то $\lambda = 0$.

В разделе 2 отмечено, что кадры от разных станций должны приходить к точке доступа с приблизительно одинаковой мощностью. Для этого требуем, чтобы станции передавали на одной СКК. В этом случае для любого множества назначений X выравнивание мощностей происходит по станции с наихудшими канальными условиями, что довольно просто. Если выбирать разные СКК для разных станций, то необходима более сложная процедура: она должна учитывать, что станции могут менять не только выходную мощность, но и СКК для выравнивания.

5 Построение назначений

Как уже было отмечено, точка доступа должна максимизировать функцию полезности для построения назначений. Пусть СКК $c \in C$ фиксирована.

Тогда задачу максимизации функции полезности можно записать так:

$$\max_{X \in H} \sum_{(s,r) \in X} \lambda(s,r,c).$$
(2)

Так как в условиях частотной селективности все РБ отличаются друг от друга, то прямой перебор по всем $X \in H$ является вычислительно затратным. Поэтому в данной работе предлагается использование жадного алгоритма для приближенного решения задачи (2).

5.1 Жадный алгоритм

Схема работы жадного алгоритма выглядит следующим образом. Все станции сортируются по убыванию метрик — приращений функции полезности. После этого первой станции в списке ставится в соответствие самый широкий доступный РБ, затем второй станции самый широкий из оставшихся и т. д. Назначения продолжаются до исчерпания РБ или станций в списке. В итоге получается множество назначений при фиксированной СКК *с*. Точка доступа строит такие множества для каждой СКК из *С* и затем выбирает наилучшее. Далее шаги данного алгоритма рассмотрены более подробно.

Сначала для каждой станции вычисляются значения $\lambda(s, r, c)$, для $r \in R_{242}$, где R_{242} — множество 242-тоновых РБ (пирина $\approx 20 \text{ МГ ц}$). Канальные условия в этих РБ могут сильно отличаться из-за частотно-селективных замираний, причем их влияние различно для разных станций. Из данных значений берется максимум, т. е. $\lambda_s = \max_{r \in R_{242}} \lambda(s, r, c)$. Для сравнения были выбраны значения метрик для каналов шириной 20 МГ ц, так как данные РБ довольно крупные, и поэтому значения метрик характеризуют «среднее» качество канала у станции. При этом более крупные РБ не были взяты из-за того, что канал обычно разделяют по частоте несколько станций, и передача в очень широких РБ происходит сравнительно редко.

Станции сортируются по убыванию полученных значений λ_s . Затем первой станции в получившимся списке выделяется самый широкий РБ, в котором она может передавать на СКК *с*. После этого выбирается следующая станция в списке, и ей выделяется наиболее широкий РБ из оставшихся допустимых. Получившиеся назначения *X* для двух станций должны удовлетворять ограничениям (1). Так повторяется для всех станций до конца списка или до исчерпания допустимых РБ, т. е. до случая, когда для следующей станции в списке нельзя назначить какой-либо РБ, чтобы $X \in H$.

Далее для получившегося множества назначений считается значение выражения:

$$\sum_{(s,r)\in X} \lambda(s,r,c). \tag{3}$$

Описанная процедура повторяется для всех $c \in C$. В результате выбирается та СКК c и соответствующее ей множество назначений X, для которых значение выражения (3) максимально.

5.2 Вычисление приоритетов

Для работы описанного жадного алгоритма необходимо вычислять значения λ , которые являются приоритетом: станция с большим значением λ обычно получает больше ресурсов. Рассмотрим несколько подходов к их вычислению.

Наиболее простым способом является максимизация пропускной способности Max Rate (MR). Планировщик стремится передать максимальное количество данных по каналу за промежуток времени t, несмотря на то, какие именно станции будут передавать. В этом случае λ считается следующим образом:

$$\lambda(s, r, c) = rate(r, c),$$

где rate(r, c) — скорость передачи на СКК c в РБ r. Как уже было отмечено, если станция s не может передавать на СКК c в РБ r, то $\lambda = 0$. Поэтому для простоты записи данный случай не включается в выражения для λ .

Следующим рассмотрим подход Proportional Fair (PF). В этом случае планировщик учитывает не только количество данных, которое может передать станция, но еще и то количество данных, которое станция уже успела передать к этому моменту, так, что станция, которая мало успела передать имеет высокий приоритет. Таким образом, получается справедливое распределение ресурсов, и

$$\lambda(s, r, c) = \frac{rate(r, c)}{Q(s)},$$

где $Q(s) = \frac{A(s)}{TT(s)}$ — усредненная скорость передачи данных, A(s) — количество уже переданных данных, TT(s) — время, когда у станции были данные на передачу.

Последним рассмотренным планировщиком является Shortest Remaining Processing Time (SRPT). Этот планировщик стремится минимизировать время, за которое поток будет полностью передан. В данной работе рассмотрена следующая версия данного подхода. Сортировка для жадного алгоритма производится по значениям

$$\lambda(s) = \frac{D(s)}{rate(r_{entire}, c_{best})}.$$

где D(s) — количество данных потока, оставшихся для передачи, r_{entire} — РБ, занимающий весь канал, c_{best} — самая быстрая СКК, с которой станция *s* может передавать во всем канале. Данное значение $\lambda(s)$ является временем передачи потока, если этой станции выделить весь канал. Чем меньше это время, тем быстрее станция может передать (при выделении всего канала) оставшуюся часть своего потока, и тем выше у нее приоритет, т. е. сортировка происходит по возрастанию.

Значения $\lambda(s)$ не зависят от выбранной СКК, поэтому порядок сортировки для всех СКК является одинаковым. При этом вместо выражения (3)

после построения множества назначений для СКК \boldsymbol{c} вычисляется следующее:

$$\sum_{(s,r)\in X} \frac{D(s) - \min\{D(s), \tau_{max} \times rate(r,c)\}}{rate(r_{entire}, c_{best})} + \sum_{(s,r)\notin X} \frac{D(s)}{rate(r_{entire}, c_{best})}, \quad (4)$$

где $\tau_{max} = 5484$ мкс — максимальная длительность передачи, согласно стандарту. $\tau_{max} \times rate(r, c)$ определяет максимальное количество данных, которые передаст станция, если ей назначить РБ r и СКК c. Следовательно, числитель левой суммы является количеством оставшихся данных потока после выполнения передачи, т. е. новым значением D(s). Таким образом, выражение (4) определяет суммарное время передачи оставшихся потоков всех станций, если бы им выделяли весь канал. Выбирается множество назначений для той СКК, при которой значение выражения минимально.

5.3 Адаптация информации о состоянии канала

Частотная селективность означает, что замирания дают различный эффект для разных частот сигнала, т. е. разные ресурсные блоки внутри одного канала могут испытывать различное усиление/ослабление мощности. В данной работе используется следующая модель потерь мощности сигнала (конкретные числовые выражения находятся в разделе 6):

$$PL(s,r) = PL(s_{dist}) + P(r_{size}) + Fading(s,r) + const,$$
(5)

где $PL(s_{dist})$ — потери, зависящие от расстояния от станции до точки доступа, $P(r_{size})$ — учет ширины РБ (чем меньше РБ, тем больше плотность мощности), Fading(s,r) — изменения мощности, вызванные замираниями. Точка доступа знает значения PL(s,r) для каждой станции и каждого РБ и определяет, может ли станция передать на данной СКК в данном РБ, на основании чего строятся назначения. Однако значения Fading(s,r), которые знает точка доступа, могут отличатся от актуальных.

Пусть F(s,r) — оценка значения Fading(s,r), имеющаяся у точки доступа. Пусть информация о влиянии замираний на каждую станцию, а именно значения F(s,r), обновляется на точке доступа с некоторым периодом T_{fd} . Тогда спустя некоторое время после последнего обновления оценка F(s,r)может оказаться неактуальной, и тогда точка доступа выделит станции такие РБ и СКК, что данные не будут приняты из-за ухудшившихся канальных условий, и станции придётся сделать новую попытку передачи данных.

Для уменьшения количества повторных передач предлагается алгоритм, который улучшает оценку F(s, r), исходя из результатов прошлых передач. Если станция *s* неуспешно передала в РБ *r* некоторое количество раз подряд, то это означает, что канальные условия ухудшились по сравнению с моментом обновления, и поэтому потери мощности для этой станции в данном РБ увеличиваются. Если же произошло несколько успешных передач подряд, то потери мощности уменьшаются. Далее алгоритм описан более подробно. Пусть $F_{\{T_{fd}\}}(s,r)$ — значения Fading(s,r) в момент последнего обновления. Для каждой станции для каждого РБ $r \in R$ заводятся счетчики успешных и неуспешных передач $count_{success}(s,r)$ и $count_{fail}(s,r)$. Также заводятся переменные add(s,r), которые показывают, как нужно изменить (уменьшить или увеличить и насколько) значения $F_{\{T_{fd}\}}(s,r)$. Увеличение (уменьшение) данного значения означает увеличение (уменьшение) потерь PL(s,r).

В начале работы системы все указанные переменные равны нулю. Ниже показан псевдокод алгоритма изменения переменной F(s, r) в случае успешной/неуспешной передачи станцией *s* в РБ *r*:

Параметры:

```
e<sub>down</sub> — количество повторных неуспешных попыток передач для
 увеличения переменной F(s,r);
e<sub>mid</sub> и e<sub>up</sub> — количество повторных успешных передач для
 уменьшения переменной F(s,r);
step — шаг уменьшения/увеличения F(s,r);
Пусть жадный алгоритм назначил станции s передачу в РБ r.
if Передача неуспешна then
    count_{fail}(s,r) \leftarrow count_{fail}(s,r) + 1;
    count_{success}(s,r) \leftarrow 0;
   if count_{fail}(s, r) \ge e_{down} then
       count_{fail}(s,r) \leftarrow 0;
       add(s,r) \leftarrow add(s,r) + 1;
   end
else
    count_{success}(s,r) \leftarrow count_{success}(s,r) + 1;
    count_{fail}(s,r) \leftarrow 0;
   if add(s,r) > 0 then
     e \leftarrow e_{mid};
   else
    e \leftarrow e_{up};
   end
   if count_{success}(s, r) \ge e then
       count_{success}(s,r) \leftarrow 0;
       add(s,r) \leftarrow add(s,r) - 1;
   end
end
F(s,r) = F_{\{T_{fd}\}}(s,r) + add(s,r) \times step
```

По полученным значениям F(s,r) вычисляются PL(s,r), и, исходя из этих обновленных значений, строятся назначения в следующем цикле. Два различных параметра для уменьшения переменной F(s,r) нужны, чтобы станция быстрее (при $e_{mid} < e_{up}$) уменьшала данные значения, если до этого они были повышены, что позволяет быстро «исправить» оценку, когда качество канала переоценено. Далее выполняются попытки передачи и вычисляются новые значения $count_{success}(s,r)$, $count_{fail}(s,r)$ и add(s,r) по описанному выше алгоритму. Эти переменные обнуляются при получении актуальных значений (обновлении) $Fading_{T_{td}}(s,r)$.

6 Результаты моделирования

Общая модель системы была описана в разделе 4. Рассматривается канал на частоте 5 ГГц шириной 40 МГц. Структура РБ показана на рис. 1. Размеры потока берутся из усеченного логнормального распределения с минимальным, средним и максимальным значениями 1 КБ, 1 МБ и 10 МБ соответственно. Время ожидания нового потока берется из усеченного экспоненциального распределения с минимальным, средним и максимальным значениями 1 с, 3 с и 6 с соответственно.

Для вычисления $PL(s_{dist})$ из формулы (5) была использована следующая модель затуханий сигнала [6]:

 $PL(s_{dist}) = 40.05 + 20lg(f_c/2.4) + 20lg(min(s_{dist}, 5)) + \mathbb{I}(s_{dist} > 5) \times 35lg(s_{dist}/5),$

где f_c — центральная частота в ГГц, \mathbb{I} — индикаторная функция, s_{dist} — расстояние от станции до точки доступа. Данная формула показывает потери мощности с расстоянием, выраженные в децибелах.

Размеры РБ учитываются так: $P(r_{size}) = 10lg(r_{size}/18)$. Здесь r_{size} — количество 26-тоновых РБ, из которых состоит РБ r, согласно схеме 1. Например, 52-тоновый РБ состоит из 2-х 26-тоновых РБ, 242-тоновый РБ — из 9. 18 — максимальная ширина РБ. Значения Fading(s,r) рассчитываются по модели TGac D NLOS [7], рекомендованной в [6].

Для проведения экспериментов был реализован описанный в разделе 5 жадный алгоритм для метрик MR, PF и SRPT (обозначаются на графиках как ах-MR, ах-PF и ах-SRPT соответственно). Также были рассмотрены 6 планировщиков из работы [2] для проверки их эффективности в условиях замираний: 3 планировщика (обозначения на графиках: old ах-MR, old ах-PF и old ах-SRPT), разделяющих по частоте канал, но считающих, что канальные условия в каждом PБ одинаковы (считающих, что канал плоский), а также 3 планировщика, не разделяющих по частоте канал (обозначения на графиках: MR, PF, SRPT).

В первой серии экспериментов станции распределены в круге радиуса R = 5 метров. Также было рассмотрено 2 варианта обновления информации о замираниях: в любой момент времени известна актуальная информация $(T_{fd} = 0 \text{ c})$, или данные обновляются с периодом 0.5 с $(T_{fd} = 0.5 \text{ c})$.

Были построены графики зависимости пропускной способности и среднего времени передачи потока от количества станций. Они показаны на рис. 3 и 4. На таком расстоянии все станции почти всегда способны передавать на самой быстрой СКК в любом РБ, и поэтому разделение по частоте канала между несколькими станциями не увеличивает общую пропускную способность по сравнению с ситуацией, когда канал не разделяется, что и показано на рис. 3. Однако по среднему времени передачи потока алгоритмы SRPT превосходят алгоритмы MR и PF.

При $T_{fd} = 0.5$ с заметно небольшое увеличение времени передачи для всех алгоритмов, связанное с неактуальностью данных о замираниях. Потери в пропускной способности для этого случая являются незначительными.



Рис. 3: R = 5 м, $T_{fd} = 0$ с



Рис. 4: R = 5 м, $T_{fd} = 0.5$ с

Во второй серии экспериментов R = 40 м. Результаты показаны на рис. 5 и 6. По этим графикам видно, что предложенные в данной работе алгоритмы имеют значительное превосходство по пропускной способности и по среднему времени передачи потока. При $T_{fd} = 0.5$ с все алгоритмы теряют в эффективности, однако наибольшие потери проявляются у алгоритмов семейства PF. PF справедливо дает долю канала станциям с плохими канальными условиями. Он продолжает это делать даже при постоянных ошибках передач, связанных с неактуальностью данных о канале, что ведет к сильному падению производительности.



Рис. 5: R = 40 м, $T_{fd} = 0$ с



Рис. 6: R = 40 м, $T_{fd} = 0.5$ с

Также была проведена проверка эффективности работы алгоритма изменения информации о канальных условиях (см. раздел 5.3). Результаты представлены на рис. 7, 8. При R = 5 м заметно небольшое уменьшение времени передачи при использовании адаптации. Если R = 40 м, то адаптация дает значительное преимущество. Особенно это заметно для алгоритма PF: нет такого резкого ухудшения, как в случае, когда адаптация отсутствует.



Рис. 7: R = 5 м, $T_{fd} = 0.5$ с



Рис. 8: R = 40 м, $T_{fd} = 0.5$ с

7 Заключение

В данной работе были рассмотрены передачи с использованием OFDMA в восходящем канале в условиях частотно-селективных замираний. Предложен жадный подход назначения PE для трех различных метрик: MR, PF и SRPT. Было показано, что данный подход работает лучше алгоритмов, не разделяющих по частоте канал, и алгоритмов, не учитывающих частотнуюселективность, в терминах пропускной способности и средней длительности передачи потока.

Также в работе был введен механизм адаптивного изменения информации о канальных условиях для случая, когда данная информация обновляется с некоторым периодом. Показано, что такое адаптивное изменение также увеличивает пропускную способность и уменьшает время передачи потоков за счёт того, что точка доступа при планировании канальных ресурсов располагает более актуальными данными о качестве канала для станций.

Список литературы

- 1. Oteri K, Yang R. Power Control for Multi-User Transmission in 802.11 ax // IEEE, doc. IEEE.
- 2. OFDMA Uplink Scheduling in IEEE 802.11ax Networks / Dmitry Bankov, Andrey Didenko, Evgeny Khorov, Andrey Lyakhov // Communications (ICC), 2018 International Conference on / IEEE. – 2018.
- 3. Sharon Oran, Alpert Yaron. Scheduling Strategies and Throughput Optimization for the Uplink for IEEE 802.11 ax and IEEE 802.11 ac Based Networks // Wireless Sensor Network. -2017.-Vol. 9, no. 08. P. 250.
- 4. Resource Allocation in 802.11 ax Networks / Mukesh Taneja, Bibek Sahu, Ramachandra Murthy et al.
 -2018.
- Naik Gaurang, Bhattarai Sudeep, Park Jung-Min Jerry. Performance Analysis of Uplink Multi-User OFDMA in IEEE 802.11 ax // Communications (ICC), 2018 IEEE International Conference on, IEEE. – 2018.
- 6. ax High Efficiency WLAN Study Group et al. 11-14-0980-14 TGax simulation scenarios.- 2016.
- 7. Committee TGac Channel Models Special et al. TGac channel model addendum // IEEE 802.11-09/0308r0.- 2009.

Анализ эффективности сетей LTE-LAA при использовании резервирующего сигнала ограниченной длительности*

В.Д. Ждановский, В.А. Логинов, А.И. Ляхов

zhdanovskiy.vd@phystech.edu, loginov@iitp.ru, lyakhov@iitp.ru

ИППИ РАН

Аннотация В последнее время становится актуальной проблема повышения производительности сетей LTE, в связи с чем консорциумом 3GPP была предложена новая технология LTE-LAA, использующая нелицензируемый спектр для передачи данных. Для доступа к каналу в LTE-LAA используется механизм LBT, основанный на CSMA/CA. Так как начало передачи данных в LTE-LAA возможно лишь на границе слотов в лицензируемой полосе, во многих работах предлагается резервировать среду специальным сигналом вплоть до границы ближайшего слота, когда передача данных будет возможна. В настоящей работе приводится несколько модификаций механизма LBT, использующих резервирующий сигнал ограниченной длительности и призванных обеспечить справедливое совместное существование сетей LTE-LAA и Wi-Fi, а также приводится анализ эффективности предложенных решений в зависимости от различных параметров.

1 Введение

В ближайшем будущем аналитики прогнозируют значительный рост объема данных, передаваемых в сотовых сетях [1]. Чтобы удовлетворить возросшие требования к производительности сетей LTE (англ. Long Term Evolution), консорциум 3GPP, занимающийся их стандартизацией, разработал новую технологию под названием LTE-LAA (англ. Licensed-Assisted Access).

Для увеличения пропускной способности в сетях LTE-LAA планируется передавать данные в нисходящем канале не только в лицензируемых полосах частот, доступных лишь определенному оператору, но и в нелицензируемых полосах частот, уже используемых другими технологиями беспроводной связи. В частности, LTE-LAA будет использовать диапазон 5 ГГц, в котором работают сети Wi-Fi стандарта IEEE 802.11a/n/ac.

 ^{*} Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 18-07-01223 а

В связи с этим механизм доступа к каналу в LTE-LAA должен учитывать, что среда может оказаться занятой другим устройством. Более того, в некоторых странах существуют законодательные ограничения на использование диапазона 5 ГГц, требующие обязательное применение метода CSMA/CA (англ. Carrier Sense Multiple Access with Collision Avoidance). В LTE-LAA этот метод реализован в виде механизма LBT (англ. Listen Before Talk).

При использовании LBT базовая станция LTE-LAA (eNodeB, eNB) перед передачей должна выполнить процедуру отсрочки. В связи с протокольными ограничениями eNB может начать передачу только на границах временных слотов в лицензируемой полосе, т.е. в моменты времени, идущие с периодом T = 500 мкс. Однако спецификация не дает однозначных указаний, что делать базовой станции после завершения процедуры отсрочки, но до начала ближайшего слота. Если ничего не предпринимать, то до этого момента среду может занять передача Wi-Fi, и производительность сети LTE-LAA может сильно упасть при большом количестве станций Wi-Fi, работающих рядом.

Одним из возможных решений данной проблемы является посылка резервирующего сигнала вплоть до начала ближайшего слота, чтобы не допустить занятия среды другим устройством. Однако это решение имеет множество недостатков. Например, максимальное время непрерывного использования канала ограничено законодательно в ряде стран. Использование резервирующего сигнала в этом случае приводит к тому, что для передачи полезных данных остается меньше времени. Другая проблема заключается в том, что резервирование среды без ее использования на интервалы времени, достигающие 500 мкс, является очень неэффективным использованием частотной полосы.

В связи с этим в настоящей работе предлагается несколько модификаций данного решения, позволяющих снизить длительность резервирующего сигнала, а также анализируется их эффективность при совместном существовании сетей LTE-LAA и Wi-Fi в зависимости от различных параметров.

Дальнейшее изложение построено следующим образом. В разделе 2 приведено подробное описание методов доступа к среде, используемых в Wi-Fi и LTE-LAA. Раздел 3 содержит обзор литературы. В разделе 4 с помощью имитационного моделирования показана эффективность предложенных реализаций LBT в различных ситуациях. Раздел 5 содержит заключение.

2 Методы доступа к среде в сетях Wi-Fi и LTE-LAA

Для доступа к каналу в сети Wi-Fi используется метод EDCA (англ. Enhanced Distributed Channel Access), основанный на CSMA/CA.

При использовании EDCA перед каждой попыткой передачи необходимо выполнить процедуру отсрочки. Перед началом данной процедуры станция слушает канал и ждет, пока среда не будет свободной в течение интервала AIFS (англ. Arbitration Inter-Frame Space). После этого станция инициализирует счетчик отсрочки случайным целым значением, равномерно распределенным на интервале [0, W-1], где W — текущее значение конкурентного окна (англ. Contention Window, CW). По прошествии очередного слота длительностью 9 мкс данный счетчик уменьшается на единицу, если среда была свободна. В противном случае счетчик приостанавливается до того момента, пока канал снова не будет доступен в течение AIFS.

Как только значение счетчика отсрочки достигнет нуля, станция может начать передачу. После приема пакета получатель отправляет кадрподтверждение, посылаемый через интервал SIFS (англ. Short Inter-Frame Space). Если передающая станция по истечении интервала ACK Timeout не получает кадр-подтверждение, то происходит повторная передача. SIFS имеет меньшую длительность, чем AIFS, что позволяет избегать захвата канала другой станцией до получения подтверждения отправителем.

Величина W перед первой попыткой передачи равна W_{min} . С каждой неудачной попыткой передачи W увеличивается вдвое, но не более W_{max} . Увеличение конкурентного окна позволяет уменьшить вероятность повторных коллизий.

При использовании EDCA константы AIFS, W_{min} и W_{max} зависят от типа передаваемого трафика, что дает возможность станции, передающей данные с более высоким приоритетом, захватывать канал быстрее.

В LTE-LAA для доступа к каналу используется механизм LBT, схожий с EDCA вплоть до значений служебных интервалов, но имеющий несколько существенных отличий.

Во-первых, как уже было сказано в разделе 1, во избежание занятия канала станциями Wi-Fi в работах [2-6] базовой станции LTE-LAA предлагается занимать нелицензируемую полосу в момент окончания процедуры отсрочки резервирующим сигналом вплоть до начала ближайшего слота, когда базовая станция сможет начать передачу данных.

Заметим, что использование резервирующего сигнала в отсутствие станций Wi-Fi немного понижает пропускную способность LTE-LAA, что связано с тем, что часть максимально разрешенного времени непрерывной передачи данных тратится на передачу резервирующего сигнала.

Во-вторых, в Wi-Fi подтверждение об успешном приеме пакета через SIFS по нелицензированному каналу. В LTE-LAA, с другой стороны, подтверждение передается с задержкой 5 мс по каналу, находящемуся в лицензированной полосе.

В данной работе предлагается несколько реализаций механизма LBT, позволяющих ограничить максимальную длительность резервирующего сигнала T. Если интервал между моментом завершения процедуры отсрочки и началом следующего слота в лицензированной полосе меньше T, то базовая станция, как и прежде, посылает резервирующий сигнал до границы слотов. В противном случае eNB дожидается, пока до начала ближайшего слота не останется время T, после чего посылается резервирующий сигнал, если среда была свободна в последний интервал времени $T_d = 43$ мкс, аналогичный интервалу AIFS в Wi-Fi. Следует заметить, что при разработке спецификации LTE-LAA решение, что делать базовой станции в случае, если между моментами завершения отсрочки и посылки резервирующего сигнала нелицензируемая полоса оказалась занята передачей Wi-Fi, было оставлено на усмотрение разработчиков оборудования. В данной работе предлагается два возможных варианта действия: первое решение предполагает повторение процедуры отсрочки LBT без повышения конкурентного окна. Второе решение не предполагает повторения отсрочки. В частности, при его работе eNB начинает передачу резервирующего сигнала с последующей передачи данных в ближайший момент времени, когда одновременно выполнены два условия:

- 1. В течение T_d перед этим моментом среда свободна.
- 2. Время до ближайшей границы слота в лицензируемой полосе меньше Т.

Если в текущем слоте это невозможно, базовая станция должна ждать такого момента в последующих слотах.

3 Обзор литературы

Совместное существование LTE-LAA и Wi-Fi уже достаточно широко исследовано в литературе [2-7], однако в большинстве работ рассматривается только реализация LTE-LAA с резервирующим сигналом неограниченной длительности.

Авторы [2] впервые использовали среду имитационного моделирования ns-3 для исследования данной проблемы. В работе заключается, что при совместном существовании с Wi-Fi технология LTE-LAA показывает наибольшую эффективность в сценарии с ненасыщенным трафиком. Авторы также делают вывод, что факторы, воздействующие на занятость канала (например, управляющие сигналы LTE-LAA) оказывают более сильное влияние на производительность обоих сетей, чем изменение параметров в алгоритмах отсрочки.

Аналитическая модель метода доступа LTE-LAA разработана в [3]. Отмечается, что при использовании насыщенного трафика пропускная способность сети Wi-Fi сильно падает, в то время как LTE-LAA сохраняет свою производительность. Стоит отметить, что авторы предполагают использование постоянного конкурентного окна в процедуре отсрочки LBT.

Авторы [4] исследуют совместное существование сетей LTE-LAA и Wi-Fi на примере реальных устройств в зависимости от ширины канала и центральной частоты LTE-LAA, использования MIMO в Wi-Fi, а также от расстояния между сетями и наличия препятствий. Однако в работе рассматривается лишь сценарий с одной станцией Wi-Fi. Кроме того, в работе [4] рассматривается реализация LTE-LAA без процедуры отсрочки.

В [5] изучается влияние выбора значений параметров LBT в LTE-LAA на производительность сети Wi-Fi. Авторы рассматривают две возможных схемы: с конкурентным окном постоянного и переменного размера. Для второй схемы исследуются два случая: стандартный, соответствующий спецификации LTE-LAA, и альтернативный, в котором устройство наблюдает за



Рис. 1. Схема эксперимента

каналом и вычисляет значение конкурентного окна W на основе статистики свободных слотов в течение определенного периода времени. Авторы указывают, что стандартная схема с переменным конкурентным окном является наиболее подходящей для взаимного существования LTE-LAA и Wi-Fi. Следует отметить, что в [5] не исследуется зависимость пропускных способностей обоих сетей в зависимости от количества Wi-Fi станций.

В работе [6] рассматривается реализация LBT с конкурентным окном фиксированного размера. Авторами предложен алгоритм динамического выбора размера конкурентного окна и длительности передачи LTE-LAA. Показано, что предложенный алгоритм настройки параметров позволяет увеличить суммарную пропускную способность системы из сетей LTE-LAA и Wi-Fi по сравнению с использованием фиксированного значения конкурентного окна, при этом не причиняя вреда производительности сети Wi-Fi. К сожалению, авторами не было приведено сравнение работы предложенного алгоритма с реализацией LBT с двоичным экспоненциальным конкурентным окном, которое по умолчанию используется в спецификации LTE-LAA.

Авторы [7], в отличие от [2-6], также исследуют реализации LBT, когда резервирующий сигнал не используется и когда он ограничен тремя OFDMсимволами, указывая на предпочтительность последнего варианта. Однако они исследуют лишь одно ограничение на максимальную длительность резервирующего сигнала и рассматривают только схему с повторением отсрочки.

В литературе еще не была исследована эффективность LTE-LAA при использовании различных ограничений на длительность резервирующего сигнала, а также вариантов действия в случае, если Wi-Fi начал передачу в промежутке между окончанием процедуры отсрочки и ближайшим слотом в лицензируемой полосе, что указывает на актуальность данной работы.

4 Численные результаты

Исследование, проведенное в данной работе, было выполнено в среде имитационного моделирования ns-3. Для этого в исходный код авторов [2] была добавлена возможность работы LTE-LAA с резервирующим сигналом огра-



Рис. 2. Зависимость пропускных способностей от числа станций в случае использования схемы с повторением отсрочки

ниченной длительности с двумя возможными модификациями, описанными в разделе 2.

Рассматривался сценарий, в котором оценивалась производительность сети, состоящей из совместно работающих одной базовой станции и одного пользовательского устройства LTE-LAA, одной точка и доступа (англ. access point, AP) и N станций (STA) Wi-Fi. В обоих сетях передавался насыщенный трафик: в сети LTE-LAA нисходящий, в сети Wi-Fi восходящий. Расстояние между базовой станцией/точкой доступа и конечными устройствами равнялось $d_1 = 10$ м, сами же сети были расположены друг от друга на расстоянии $d_2 = 10$ м, см. рис. 1. Для обоих сетей использовалось ограничение на максимальную длительность занятия среды, равное 10 мс.

В первом эксперименте оценивались пропускные способностей обоих сетей в зависимости от N. Из результатов моделирования, показанных на рис. 2-3, следует, что пропускная способность LTE-LAA при увеличении N стремится к некоторой ненулевой константе, что объясняется способностью разрешать часть возникающих коллизий благодаря выбору робастной низкоскоростной сигнально-кодовой конструкции, позволяющей повысить вероятность успешной передачи данных. Исключением является только случай при использовании механизма с повторением отсрочки при полном отсутствии резервирующего сигнала (T = 0), где пропускная способность стремится к



Рис. 3. Зависимость пропускных способностей от числа станций в случае использования схемы без повторения отсрочки

нулю. Также применение схемы без повторения отсрочки позволяет повысить производительность LTE-LAA по сравнению со схемой с повторением отсрочки. Можно заметить, что уменьшение максимальной длительности резервирующего сигнала T ведет к падению пропускной способности LTE-LAA, поскольку увеличивается вероятность того, что какая-либо станция Wi-Fi начнет передачу в промежутке между концом процедуры отсрочки LTE-LAA и началом ближайшего слота в лицензируемой полосе. Стоит отметить, что T = 0 означает отсутствие резервирующего сигнала, а случай T = 1000 мкс аналогичен ситуации, когда длительность резервирующего сигнала.

Во втором эксперименте оценивался прирост пропускной способности изза использования LTE-LAA по отношению к сценарию, в котором есть N+1устройств Wi-Fi и нет устройств LTE-LAA. Результаты показывают, см. рис. 4-5, что добиться справедливого эффективного совместного существования (т.е. когда имеется положительный прирост пропускной способности как для LTE-LAA, так и для Wi-Fi) можно, выбирая различные T и схемы повторения отсрочки в зависимости от заданного числа N станций Wi-Fi. Например, при N = 3 справедливое существование достигается при использовании схемы без повторения отсрочки и T = 500мкс, в то время как при N = 9 лучше использовать схему с повторением отсрочки и T = 250мкс.



Рис. 4. Зависимость прироста пропускной способности от числа станций в случае использования схемы с повторением отсрочки



Рис. 5. Зависимость прироста пропускной способности от числа станций в случае использования схемы без повторения отсрочки

На рис. 6 показана зависимость прироста пропускной способности от максимальной длительности резервирующего сигнала T при фиксирован-



Рис. 6. Зависимость прироста пропускной способности от максимальной длительности резервирующего сигнала при фиксированном числе N станций Wi-Fi

ном количестве станций. В целях наглядности значения T, при которых прирост пропускной способности отрицательный хотя бы для одного типа станций, не рассматриваются. Результаты показывают, что модификация LBT без повторения отсрочки дает выигрыш в пропускной способности для LTE-LAA на 25-50% больше по сравнению со схемой с повторением отсрочки при малом числе станций (N = 5), однако последняя обеспечивает справедливое эффективное совместное существование в более широком диапазоне изменения T при большом числе станций, что позволяет использовать ее в случае, когда количество Wi-Fi станций точно не известно.

5 Заключение

В данной работе предложены модификации механизма LBT в LTE-LAA, позволяющие обеспечить справедливое совместное существование LTE-LAA и Wi-Fi. В частности, исследовано влияние максимальной длительности резервирующего сигнала на производительность обоих сетей, а также рассмотрены две схемы повторения отсрочки в механизме LBT с последующим анализом их эффективности. С помощью имитационного моделирования было показано, что предложенные модификации LBT позволяют добиться справедливого совместного существования сетей LTE-LAA и Wi-Fi.

Список литературы

- Cisco Visual Networking Index. Global Mobile Data Traffic Forecast Update, 2015– 2020 White Paper. – 2016.
- 2. Simulating LTE and Wi-Fi Coexistence in Unlicensed Spectrum with ns-3 / Lorenza Giupponi, Biljana Bojovic, Marco Miozzo et al. -2016.
- Chen Cheng, Ratasuk Rapeepat, Ghosh Amitava. Downlink performance analysis of LTE and WiFi coexistence in unlicensed bands with a simple listen-beforetalk scheme // Vehicular Technology Conference (VTC Spring), 2015 IEEE 81st / IEEE. - 2015. - P. 1–5.

- 4. Coexistence of Wi-Fi and LAA-LTE: Experimental evaluation, analysis and insights / Y. Jian, C. F. Shih, B. Krishnaswamy, R. Sivakumar // 2015 IEEE International Conference on Communication Workshop (ICCW). $-2015.-June.-P.\ 2325-2331.$
- Wi-Fi-LAA coexistence: Design and evaluation of Listen Before Talk for LAA / A. V. Kini, L. Canonne-Velasquez, M. Hosseinian et al. // 2016 Annual Conference on Information Science and Systems (CISS). – 2016. – March. – P. 157–162.
- Kim Chung K, Yang Chan S, Kang Chung G. Adaptive listen-before-talk (LBT) scheme for LTE and Wi-Fi systems coexisting in unlicensed band // Consumer Communications & Networking Conference (CCNC), 2016 13th IEEE Annual / IEEE. – 2016. – P. 589–594.
- 7. Fairness vs. Performance in Rel-13 LTE Licensed Assisted Access / Michal Cierny, Timo Nihtila, Toni Huovinen et al. // IEEE Communications Magazine.-2017.- Vol. 55, no. 12.-P. 133–139.

Аналитическая модель метода доступа к нелицензируемой полосе в сетях 5G *

П.Н. Куцевол, В.А. Логинов, А.И. Ляхов

kutsevol.pn@phystech.edu, loginov@iitp.ru, lyakhov@iitp.ru

ИППИ РАН

Аннотация В связи с ростом объема информации, передаваемой в сотовых сетях, растут требования к их пропускной способности. Поэтому планируется начать использовать не только лицензируемые полосы частот, но и нелицензируемые, которые в данный момент используются другими технологиями, например, Wi-Fi. Для этого консорциумом 3GPP была разработана новая технология LTE-LAA (англ. License-assisted access), о которой пойдет речь в данной работе. Согласно спецификации, базовая станция LTE-LAA получает доступ к каналу с помощью метода CSMA/CA, но может начать передачу данных только одновременно с началом слота в лицензируемой полосе. В сетях 4G возможна передача станций LTE-LAA с периодичностью 1 мс или 0,5 мс. В отличие от 4G, в разрабатываемых сегодня сетях 5G планируется уменьшить период T моментов, в которые базовая станция 5G может начать передачу данных в нелицензируемой полосе после выполнения процедуры отсрочки. В данной работе представлена аналитическая модель, позволяющая оценить пропускную способность сетей 5G в зависимости от периода T и числа станций Wi-Fi, работающих в той же полосе.

1 Введение

С каждым годом объем данных, передаваемых в сотовых сетях, растет, а вместе с ним растут и требования к пропускной способности системы. На данный момент разработана технология LTE-LAA, которая позволяет сетям LTE использовать для передачи данных нелицензируемые полосы, что позволяет увеличить пропускную способность сотовых сетей. Стоит отметить, что эти частоты уже используются другими технологиями беспроводной связи, например, Wi-Fi, поэтому необходимо рассматривать их взаимодействие со станциями LTE-LAA.

Согласно спецификации, базовые станции LTE-LAA используют механизм доступа к среде, основанный на методе множественного доступа с прослушиванием несущей и избеганием коллизий (англ. Carrier Sense Multiple

^{*} Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 18-07-01223 а

Ассезя With Collision Avoidance, CSMA/CA). Отметим, что механизм доступа к среде в Wi-Fi тоже основан на CSMA/CA. В частности, перед каждой попыткой передачи станция Wi-Fi равновероятно выбирает случайное целое число из интервала [0, W-1], где W — текущее значение конкурентного окна. Станция последовательно уменьшает счетчик отсрочки на единицу каждый раз, когда канал остается пустым в течение $\sigma = 9$ мкс. В противном случае, если канал остается занятым, то счетчик приостанавливается до того момента, пока канал станет снова свободным в течение заданного в стандарте интервала AIFS. Когда счетчик доходит до нуля, станция Wi-Fi начинает попытку передачи данных. Если она получает подтверждение об успешной передаче, значение конкурентного окна W устанавливается равным минимальному значению W_{min} . Иначе, если подтверждение не было получено, она выполняет процедуру отсрочки с удвоенным значение W, но не больше W_{max} .

Станции LTE-LAA так же, как и станции Wi-Fi, перед передачей выполняют процедуру отсрочки, однако начать фактическую передачу данных они могут лишь на границе слотов лицензируемой полосы из-за протокольных ограничений, связанных с необходимостью синхронизации с передачей контрольной информации в лицензируемой полосе.

Спецификация LTE-LAA не дает четких указаний по поводу того, что станция LTE-LAA должна делать в течение интервала между концом процедуры отсрочки и началом слота в лицензируемой полосе. В большинстве работ, например, [1–6], предлагается в данном промежутке отправлять резервирующий (шумовой) сигнал для того чтобы не допустить занятие канала другими станциями, в частности, Wi-Fi. Кроме этого варианта, согласно спецификации LTE-LAA станция может не передавать резервирующий сигнал, а начать передачу на границе слотов в лицензируемой полосе при условии, что канал никто не занял. Заметим, что во втором случае станция Wi-Fi может начать передачу в течение этого интервала. В данной работе будем называть такой интервал «опасным», так как передача станции LTE-LAA может прерваться.

Одна из целей при разработке механизма LTE-LAA заключалась в обеспечении в нелицензируемых частотах справедливого совместного существования с уже работающими там устройствами, например, Wi-Fi. Под справедливым совместным существованием в данном случае имеются ввиду следующее: пропускная способность станций Wi-Fi не должна уменьшаться при замене в системе только из станций Wi-Fi некоторых устройств на станции LTE-LAA; пропускная способность добавленных станций LTE-LAA, напротив, должна возрасти по сравнению с замененными устройствами Wi-Fi.

Достижение справедливого совместного существования без использования резервирующего сигнала может оказаться неосуществимым в сетях 4G LTE, в которых возможна передача станций LTE-LAA с периодичностью 1 мс или 0,5 мс. В разрабатываемых сегодня сетях 5G планируется уменьшить период T моментов, в которые базовая станция 5G может начать передачу данных в нелицензируемой полосе.

Основная задача данной статьи — исследовать взаимодействие станций Wi-Fi и LTE-LAA для случаев использования и отсутствия резервирующего сигнала при разном периоде T.

Дальнейшее изложение построено следующим образом. Раздел 2 содержит обзор литературы. В разделе 3 приведена аналитическая модель для случаев работы системы с и без резервирующего сигнала. Раздел 4 содержит численные результаты и анализ справедливости распределения частотных ресурсов между устройствами разных технологий. В разделе 5 приводится заключение.

2 Обзор литературы

Эффективность механизма доступа к каналу в сетях LTE-LAA исследуется во многих работах, например, в [1–6].

В [1] рассмотрена работа LTE-LAA в двух случаях: если LTE-LAA использует резервирующий сигнал и если LTE-LAA начинает передачу сразу после выполнения процедуры отсрочки. С помощью имитационного моделирования было показано, что во втором случае пропускная способность незначительно увеличивается по сравнению с первым.

В [2] с помощью среды имитационного моделирования ns-3 исследовалась система из двух станций Wi-Fi в ненасыщенном случае, а затем одну из них заменяли на станцию LTE-LAA, использующую резервирующий сигнал. При замене обнаружилось падение производительности станции Wi-Fi и значительный рост производительности станции LTE-LAA, что является несправедливым разделением частотных ресурсов.

В [3] представлена математическая модель для оценки пропускной способности системы станций LTE-LAA при использовании резервирующего сигнала. К сожалению, авторы не рассматривали взаимодействие LTE-LAA со станциями Wi-Fi.

В [4] была представлена аналитическая модель для оценки пропускной способности системы станций Wi-Fi и LTE-LAA с учетом влияния на них параметров механизма определения занятости среды. Похожая модель с резервирующим сигналом рассмотрена в [5], где авторы исследовали производительность сетей LTE-LAA и Wi-Fi, находящихся в режиме насыщения, при различных значениях W_{min} и W_{max} .

Анализ работы LTE-LAA с и без использования резервирующего сигнала представлен в [6]. Авторы пришли к выводу о том, что в случае ненасыщенного трафика оптимальным вариантом для «честного» распределения частотных ресурсов является ограничение резервирующего сигнала до трех OFDM символов (≈ 215 мкс).

Как уже было сказано, во всех работах за исключением [6], посвященных изучению совместного существования сетей LTE-LAA и Wi-Fi, исследуется только реализация LTE-LAA с резервирующим сигналом, в то время как анализу производительности LTE-LAA при отсутствии резервирующего сигнала не было уделено достаточного внимания. В [6] же приводятся результаты имитационного моделирования сети в случае отсутствия резервирующего сигнала для сетей LTE-LAA без варьирования длительности слотов в лицензируемом спектре. В данной работе будет впервые с помощью аналитического моделирования проанализирована эффективность работы сетей LTE-LAA и Wi-Fi при различном периоде *T* моментов времени, когда станция LTE-LAA может начать передачу в нелицензируемом спектре.

3 Аналитические модели

В данной работе будем рассматривать систему из Nстанций Wi-Fi и одной станции LTE-LAA.

Будем считать, что станции работают в режиме насыщения, т. е. в любой момент времени у каждой из них есть пакеты для передачи. Также все станции находятся в области радиовидимости друг друга. Длительность кадров Wi-Fi достаточно велика по сравнению с размером слота в лицензируемой полосе, т. е. если при использовании резервирующего сигнала станция Wi-Fi начала передачу одновременно со станцией LTE-LAA, то закончит она после начала фактической передачи данных станции LTE-LAA. Иными словами, невозможна ситуация, в которой станция Wi-Fi успеет закончить передачу за время, которое станция LTE-LAA отправляла резервирующий сигнал, и не повредить данные станции LTE-LAA. В случае отсутствия резервирующего сигнала, если станция Wi-Fi начнет передачу в «опасном» интервале, то станция LTE-LAA не сможет начать передачу данных на ближайшей границе слотов лицензируемого спектра из-за занятости канала.

3.1 Случай использования резервирующего сигнала

При использовании резервирующего сигнала счетчики отсрочки всех станций работают синхронно, поэтому можно воспользоваться подходом, описанным в [7]. Аналогично [7], будем использовать понятие виртуального слота — интервала между двумя последовательными отсчетами счетчика отсрочки. Для вычисления пропускных способностей станций Wi-Fi S_W и станции LTE-LAA S_L необходимо оценить вероятность p_s^W успешного слота станции Wi-Fi, вероятность p_s^L успешного слота станции LTE-LAA. Кроме того, нужно оценить среднюю длительность слота T_{slot} , которая в случае синхронного отсчета отсрочки одинакова для обоих типов станций. Тогда:

$$S_L = \frac{p_s^{(L)} d_L}{T_{slot}},$$
$$S_W = \frac{p_s^{(W)} d_W}{T}$$

где d_L и d_W — средний размер пакетов LTE-LAA и Wi-Fi, соответственно. Аналогично [7], введем τ_W (τ_L) — вероятность выбора станцией Wi-Fi (LTE-LAA) данного слота для передачи.

 T_{slot}
Если рассматриваемая станция Wi-Fi выбрала данный слот для передачи, то коллизия произойдет с вероятностью ρ_c^W , если еще хотя бы одна станция выберет этот слот:

$$\rho_c^W = 1 - (1 - \tau_L)(1 - \tau_W)^{N-1}.$$
(1)

Соотношение для вероятности ρ_c^L коллизии станции LTE-LAA можно записать аналогично:

$$\rho_c^L = 1 - (1 - \tau_W)^N.$$
(2)

Вероятности τ_W и τ_L можно найти, подсчитав сколько в среднем виртуальных слотов приходится на одну попытку передачи:

$$\tau_W = \frac{1}{1 + \frac{W_{min} - 1}{2} + \rho_c^W \cdot \frac{W_{min}}{2} \sum_{i=0}^{m-1} (2\rho_c^W)^i},$$
(3)

$$\tau_L = \frac{1}{1 + \frac{W_{min} - 1}{2} + \rho_c^L \cdot \frac{W_{min}}{2} \sum_{i=0}^{m-1} (2\rho_c^L)^i},$$
(4)

где

$$m = \log_2(\frac{W_{max}}{W_{min}}).$$

Решая совместно систему уравнений (1)-(4), найдем искомые вероятности τ_W и τ_L .

Найдем вероятность p_e того, что выбранный виртуальный слот окажется пустым. Виртуальный слот пустой, если в нем не передает ни станция LTE-LAA, ни любая из N станций Wi-Fi:

$$p_e = (1 - \tau_L)(1 - \tau_W)^N.$$

Виртуальный слот является успешным для какой-либо станции Wi-Fi, если она выбрала данный слот, а остальные станции — нет:

$$p_s^W = N\tau_W (1 - \tau_L) (1 - \tau_W)^{N-1}$$

Соотношение для станции LTE-LAA можно записать аналогично:

$$p_s^L = \tau_L (1 - \tau_W)^N.$$

Оценим среднюю длительность виртуального слота. Виртуальный слот может быть пустым, если ни одна из станций не выбрала данный слот для передачи, успешным для станции LTE-LAA или Wi-Fi или коллизионным, если его выбрало несколько станций одновременно, причем коллизия может произойти с или без участия станции LTE-LAA. Найдем вероятность коллизии p_c^W станций Wi-Fi без участия станции LTE-LAA:

$$p_c^W = (1 - \tau_L)(1 - (1 - \tau_W)^N - N\tau_W(1 - \tau_W)^{N-1}).$$

Тогда среднюю длительность виртуального слота можно найти следующим образом:

$$T_{slot} = p_e \sigma + p_s^W T_s^W + p_s^L T_s^L + T_c^W p_c^W + (1 - p_s^W - p_s^L - p_c^W - p_e) T_c^L,$$

где T^W_s — длительность успешного слота станции Wi-Fi, T^W_c — длительность слота в случае коллизии станции Wi-Fi без участия станции LTE-LAA, T^L_s — длительность успешного слота станции LTE-LAA, T^L_c — длительность слота в случае коллизии станции LTE-LAA и Wi-Fi.

3.2 Случай отсутствия резервирующего сигнала

В данном случае подход, описанный в [7], оказывается неприменим, так как в течение «опасного» интервала работают только счетчики отсрочки станций Wi-Fi. Для анализа пропускной способности сетей в данном случае воспользуемся подходом, описанным в [8]. Вероятность коллизии LTE-LAA и Wi-Fi будем считать пренебрежимо малой, так совпадение границы слотов в лицензируемом спектре LTE с границей виртуальных слотов Wi-Fi маловероятно. В данном случае, чтобы найти пропускные способности, оценим E_W и E_L — средние длительности интервала между последовательными окончаниями успешных передач станций Wi-Fi и LTE-LAA, соответственно. Тогда пропускные способности можно выразить следующим образом:

$$S_L = \frac{d_L}{E_L},$$
$$S_W = \frac{d_W}{E_W}$$

Отметим, что в [8] используется другое понятие виртуального слота. В [8] виртуальный слот — это интервал между двумя последовательными отсчетами счетчика отсрочки выбранной станции при условии, что эта станция не передает в этом промежутке.

Если рассматриваемая станция Wi-Fi выбрала данный слот для передачи, то коллизия произойдет с вероятностью ρ_c^W , если еще хотя бы одна станция выберет этот слот:

$$\rho_c^W = 1 - (1 - \tau_W)^{N-1}.$$
(5)

После завершения процедуры отсрочки станция LTE-LAA ожидает конца слота в лицензируемом спектре, т.е. возникает «опасный» интервал, в течение которого станция LTE-LAA не борется за доступ к каналу. Станция Wi-Fi может начать передачу в «опасном» интервале с вероятностью ρ_c^L , что означает неудачную попытку передачи станции LTE-LAA. Заметим, что согласно спецификации LTE-LAA, в таком случае станция LTE-LAA должна заново начать процедуру отсрочки, не меняя значения конкурентного окна.

Пусть a_i^W, a_i^L — длительности (i + 1)-й попытки, включающие длительность передачи, процедуры отсрочки, а также «опасного» интервала для

станции LTE-LAA. Тогда среднюю длительность интервала между последовательными окончаниями успешных передач можно найти следующим образом:

$$E = a_0 + \sum_{i=1}^{\infty} a_i (\rho_c)^i.$$
 (6)

Сначала оценим пропускную способность станции LTE-LAA.

Для того чтобы найти вероятность ρ_c^L , воспользуемся результатами модели, описанной в [7]. Как показано в [7], для стационарных вероятностей $b_{i,k}$ нахождения станции Wi-Fi в состоянии с $W = 2^i W_{min}$ и текущим счетчиком отсрочки k верны соотношения (7)-(8):

$$b_{0,0} = \frac{2(1 - 2\rho_c^W)(1 - \rho_c^W)}{(1 - 2\rho_c^W)(W_0 + 1) + pW_0(1 - (2\rho_c^W)^m)},\tag{7}$$

$$b_{i,k} = \frac{W_0 2^i - k}{W_0 2^i} (\rho_c^W)^i b_{0,0}.$$
(8)

Пусть T — длина слота LTE-LAA в лицензируемой полосе. Будем считать, что длительность «опасного» интервала имеет равномерное распределение на интервале [0, T).

Передача станции LTE-LAA будет прервана, если текущее значение счетчика отсрочки хотя бы одной станции Wi-Fi меньше длительности «опасного» интервала станции LTE-LAA, выраженного в целом числе слотов σ :

$$\rho_c^L = \sum_{f=1}^P \frac{\sigma}{T} \left(1 - (1 - \sum_{i=0}^m \sum_{k=0}^{\min(f-1,W_02^i - 1)} b_{i,k})^N \right) + \frac{T - \sigma P}{T} \left(1 - (1 - \sum_{i=0}^m \sum_{k=0}^{\min(P,W_02^i - 1)} b_{i,k})^N \right),$$

где $P = \lfloor \frac{T}{\sigma} \rfloor$.

Среднюю длительность промежутка l_c^L между началом «опасного» интервала и прерыванием станции LTE-LAA можно вычислить, усредняя ее по длине «опасного» интервала и порядковому номеру виртуального слота, в который начнет передачу станция Wi-Fi. Учитывая, что $(1-(1-\tau_W)^N)(1-\tau_W)^{iN}$ — вероятность того, что станция Wi-Fi начнет передачу в *i*-й слот, получаем:

$$l_c^L = \sum_{f=1}^P \frac{\sigma}{T} \frac{\sum_{i=0}^f i\sigma(1 - (1 - \tau_W)^N)(1 - \tau_W)^{iN}}{\sum_{i=0}^f (1 - (1 - \tau_W)^N)(1 - \tau_W)^{iN}}.$$

Среднюю длительность «опасного» интервала l_s^L , завершающегося успешной передачей станции LTE-LAA, можно вычислить следующим образом:

$$l_s^L = \sum_{f=1}^P \frac{\sigma}{T} \frac{\sum_{i=0}^f i\sigma(1-\tau_W)^{(i+1)N}}{\sum_{i=0}^f (1-\tau_W)^{(i+1)N}}.$$

Также необходимо оценить среднюю длительность виртуального слота t_{slot}^L станции LTE-LAA. Виртуальный слот может быть пустым, успешным для одной из станций Wi-Fi или неудачным:

$$t_{slot}^{L} = (1 - \tau_{W})^{N} \sigma + N \tau_{W} (1 - \tau_{W})^{N-1} T_{s}^{W} + (1 - (1 - \tau_{W})^{N} - N \tau_{W} (1 - \tau_{W})^{N-1}) T_{c}^{W}.$$

Заметим, что с вероятностью ρ_c^L «опасный» интервал заканчивается успешным или коллизионным слотом станции Wi-Fi, а с вероятностью $(1-\rho_c^L)$ — успешной передачей станции LTE-LAA. Тогда для длительности попытки получаем:

$$\begin{aligned} a_i^L &= \frac{W_0 - 1}{2} t_{slot}^L + \rho_c^L \Big[\frac{N \tau_W (1 - \tau_W)^{N-1}}{1 - (1 - \tau_W)^N} T_s^W + \\ &+ \frac{1 - N \tau_W (1 - \tau_W)^{N-1} - (1 - \tau_W)^N}{1 - (1 - \tau_W)^N} T_c^W + l_c^L \Big] + (1 - \rho_c^L) (T_s^L + l_s^L). \end{aligned}$$

Так как коллизии станции LTE-LAA отсутствуют, конкурентное окно станции LTE-LAA является постоянным, поэтому τ_L можно найти следующим образом:

$$\tau_L = \frac{2}{W_0 + 1}$$

Теперь оценим пропускную способность S_W станций Wi-Fi.

Найдем вероятность τ_W выбора станцией Wi-Fi данного слота для передачи. Как и в [8], τ_W можно найти как:

$$\tau_W = \frac{f_W}{f_W + w_W},\tag{9}$$

где f_W — среднее количество попыток передачи, w_W — среднее количество слотов отсрочки в расчете на одну передачу.

$$f_W = 1 + \sum_{i=1}^{\infty} (\rho_c^W)^i,$$
(10)

$$w_W = \frac{W_0 - 1}{2} + \sum_{i=1}^{m-1} \frac{W_0 2^i - 1}{2} (\rho_c^W)^i + \frac{W_0 2^m - 1}{2} \frac{(\rho_c^W)^m}{1 - \rho_c^W}.$$
 (11)

Решая совместно (5), (9)-(11), найдем τ_W .

Оценим среднюю длительность t^W_{slot} виртуального слота станции Wi-Fi. Виртуальный слот станции Wi-Fi может быть пустым, занятым передачей или коллизией другой станции Wi-Fi, или занятым передачей станции LTE-LAA. Найдем вероятность \tilde{p} того, что граница слотов в лицензируемой полосе окажется внутри данного виртуального слота станции Wi-Fi, при условии что ни одна из станций Wi-Fi не выбрала данный виртуальный слот для передачи и станция LTE-LAA к этому моменту уже закончила процедуру отсрочки. Проведем усреднение по всем возможным длинам «опасного» интервала и получим:

$$\tilde{p} = \sum_{f=1}^{P} \frac{1}{P} \tau_L (1 - \tau_W)^{Nf},$$

В результате среднюю длительность t_{slot}^W виртуального слота станции Wi-Fi можно найти следующим образом:

$$t_{slot}^{W} = (1 - \tau_{W})^{N-1} (1 - \tilde{p}) \sigma + (1 - \tau_{W})^{N-1} \tilde{p} T_{s}^{L} + (N - 1) \tau_{W} (1 - \tau_{W})^{N-2} T_{s}^{W} + (1 - (N - 1) \tau_{W} (1 - \tau_{W})^{N-2} - (1 - \tau_{W})^{N-1}) T_{c}^{W},$$

Тогда длительность a_i^W попытки передачи станции Wi-Fi можно найти следующим образом:

$$a_i^W = \frac{W_i - 1}{2} t_{slot}^W + (1 - \tau_W)^{N-1} T_s^W + (1 - (1 - \tau_W)^{N-1}) T_c^W.$$

Подставив найденные значения a_i^L и a_i^L в (6), найдем искомые пропускные способности S_W и S_L .

4 Численные результаты

Как уже было сказано, одной из задач работы является проверка справедливости распределения ресурсов между станциями LTE-LAA и Wi-Fi при разных параметрах сети. Для этого рассматривается сценарий с 1 станцией LTE-LAA и N станциями Wi-Fi с и без использования резервирующего сигнала с разной длительностью T слота LTE-LAA в лицензируемом спектре. Для численного анализа выбраны значения конкурентных окон $W_{min} = 16$, $W_{max} = 1024$, соответствующие категории трафика best effort в стандарте Wi-Fi и спецификации LTE-LAA. Для данного типа трафика в спецификации LTE-LAA определена максимально допустимая длительность непрерывной передачи, равная 10 мс, если гарантировано отсутствие устройств, работающих в той же полосе по другой технологии, или 8 мс в остальных случаях.

Максимальная длительность непрерывной передачи станции Wi-Fi в случае передачи одиночного кадра в стандарте IEEE 802.11n ограничена 10 мс, а в стандарте IEEE 802.11ac — 5,484 мс. В случае последовательной передачи



Рис. 1. Пропускная способность станции LTE-LAA и станции Wi-Fi в расчете на одну станцию для случаев с и без использования резервирующего сигнала для длительности слота лицензируемого спектра, равной а) 1 мс; б) 0,5 мс

нескольких кадров вводится дополнительное ограничение на длительность передачи TxOP limit, которое для категории трафика best effort по умолчанию равно 2,5 мс. Также считаем, что и станция LTE-LAA, и станции Wi-Fi используют канал шириной 20 МГц и самую скоростную сигнально-кодовую конструкцию.

Как уже было показано в [1], при использовании резервирующего сигнала пропускная способность практически не зависит от периода T моментов, когда станция LTE-LAA может начать передачу данных. Вследствие этого, при численном моделировании реализации LTE-LAA с резервирующим сигналом предполагалось, что T = 1мс.

На рис. 1 представлены пропускные способности LTE-LAA и Wi-Fi в расчете на одну станцию в случае длительности передачи 10 мс для обоих типов станций. Как нетрудно заметить, в случае использования резервирующего сигнала пропускные способности станций LTE-LAA и Wi-Fi приблизительно равны. Без использования резервирующего сигнала при существующих в сетях 4G длительностях слота 1 мс и 0,5 мс значительно растет пропускная способность станции Wi-Fi, а пропускная способность станции LTE-LAA стремится к нулю. Это связано с большим числом прерываний станции LTE-LAA в «опасном» интервале станциями Wi-Fi.

Для оценки «честности» и эффективности распределения частотных ресурсов, рассмотрим вспомогательный сценарий, в котором взаимодействуют N+1 станций Wi-Fi.

Пусть S_W^1 — пропускная способность одной станции Wi-Fi в сценарии N + 1 станций Wi-Fi, а S_W^2 и S_L^2 - пропускные способности одной станции Wi-Fi и станции LTE-LAA в сценарии N станций Wi-Fi и 1 станция LTE-LAA, соответственно.



Рис. 2. Приращение пропускной способности а) станции Wi-Fi; б) станции LTE-LAA при замене в системе из N + 1 станций Wi-Fi одной станции на станцию LTE-LAA в случае длительности передачи 10 мс для обоих типов станций

Будем рассматривать 2 показателя: G_W - приращение пропускной способности станции Wi-Fi:

$$G_W = \frac{S_W^2 - S_W^1}{S_W^1},$$

и G_L - приращение пропускной способности станции LTE-LAA:

$$G_L = \frac{S_L^2 - S_W^1}{S_W^1}.$$

Очевидно, что технология LTE-LAA будет справедлива по отношению к станциям Wi-Fi и превосходить технологию Wi-Fi только в том случае, если приращение обоих пропускных способностей будет положительным.

Как видно из рис. 2, при одинаковой максимальной длительности передачи, равной 10 мс, реализация LTE-LAA с резервирующим сигналом (р. с.) оказывается неэффективной. Это связано с тем, что резервирующий сигнал занимает часть времени передачи станции LTE-LAA. Реализация LTE-LAA без резервирующего сигнала является справедливой по отношению к станция Wi-Fi, но станции LTE-LAA получают выигрыш лишь при маленькой (T < 100 мкс) длительности слота LTE-LAA в нелицензируемом спектре, причем при большом количестве станций Wi-Fi (N > 15). Станции Wi-Fi получают преимущество, потому что они могут прервать передачу станции LTE-LAA, причем преимущество тем больше, чем больше длительность слота LTE-LAA в лицензируемой полосе. G_L становится положительным при большом количестве станций, так как коллизии LTE-LAA маловероятны, а для станций Wi-Fi вероятность коллизий сильно растет с ростом числа станций.

На рис. 3 длительность передачи станции LTE-LAA равна 8 мс, а станции Wi-Fi-5мс. В случае использования резервирующего сигнала пропускная способность станции Wi-Fi сильно падает, хотя для станции LTE-LAA на-



Рис. 3. Приращение пропускной способности а) станции Wi-Fi; б) станции LTE-LAA при замене в системе из N + 1 станций Wi-Fi одной станции на станцию LTE-LAA в случае длительности передачи 8 мс для станции LTE-LAA и 5 мс для станции Wi-Fi

блюдается значительный рост пропускной способности из-за большей длительности кадра. В отсутствие резервирующего сигнала распределение можно назвать справедливым для обоих типов станций в случае маленьких значений Т (< 100 мкс). При этом при больших Т происходит слишком много прерываний передачи станции LTE-LAA станциями Wi-Fi, вследствие чего падает пропускная способность первой.

В случае длительности передачи Wi-Fi, равной 2,5 мс, см. рис. 4, результаты для LTE-LAA схожи с предыдущим случаем, но «честное» распределение может быть достигнуто при больших значениях T (< 250 мкс). Отличие для станций Wi-Fi заключается в том, что из-за маленькой длительности кадра Wi-Fi вариант без резервирующего сигнала может оказаться несправедливым. Заметим, что в предыдущих случаях (рис. 2-3) при отсутствии резервирующего сигнала станции Wi-Fi всегда имели положительное значение прироста пропускной способности G_W .

5 Заключение

В данной работе представлена аналитическая модель, позволяющая оценить пропускную способность сетей LTE-LAA и Wi-Fi при совместном существовании в нелицензируемом диапазоне частот. Были рассмотрены различные варианты реализации LTE-LAA, в частности, с и без использования резервирующего сигнала. Был проведен анализ эффективности системы при различном периоде моментов времени, когда станция LTE-LAA может начинать передачу данных в нелицензируемом диапазоне. Проанализирована справедливость распределения частотных ресурсов между станциями разных типов при различных значениях параметров. Согласно полученным результатам справедливое и эффективное совместное существование сетей Wi-Fi и



Рис. 4. Приращение пропускной способности
а) станции Wi-Fi; б) станции LTE-LAA при замене в системе из
 N+1 станций Wi-Fi одной станции на станцию LTE-LAA в случае длительности передачи
 8 мс для станции LTE-LAA и 2,5 мс для станции Wi-Fi

LTE-LAA без использования резервирующего сигнала возможно при уменьшении периода моментов, в которые базовая станция может начать передачу данных в нелицензируемой полосе.

Список литературы

- 1. Enabling frequency reuse for licensed-assisted access with listen-before-talk in unlicensed bands / Hua Wang, Markku Kuusela, Claudio Rosa, Antti Sorri // Vehicular Technology Conference (VTC Spring), 2016 IEEE 83rd / IEEE. -2016. P. 1-5.
- 2. Simulating LTE and Wi-Fi coexistence in unlicensed spectrum with NS-3 / Lorenza Giupponi, Thomas Henderson, Biljana Bojovic, Marco Miozzo // arXiv preprint arXiv:1604.06826.-2016.
- Performance Analysis of LTE-LAA Network / Jaehong Yi, Weiping Sun, Seungkeun Park, Sunghyun Choi // IEEE Communications Letters. – 2018. – Vol. 22, no. 6. – P. 1236–1239.
- Analytical Modeling of Wi-Fi and LTE-LAA Coexistence: Throughput and Impact of Energy Detection Threshold / Morteza Mehrnoush, Vanlin Sathya, Sumit Roy, Monisha Ghosh // arXiv preprint arXiv:1803.02444. — 2018.
- 5. Song Yujae, Sung Ki Won, Han Youngnam. Co
existence of Wi-Fi and cellular with listen-before-talk in unlicensed spectrum
 // IEEE Communications Letters. 2016. Vol. 20, no. 1. P. 161–164.
- Fairness vs. Performance in Rel-13 LTE Licensed Assisted Access / Michal Cierny, Timo Nihtila, Toni Huovinen et al. // IEEE Communications Magazine. – 2017. – Vol. 55, no. 12. – P. 133–139.
- 7. Bianchi Giuseppe. Performance analysis of the IEEE 802.11 distributed coordination function // IEEE Journal on selected areas in communications. 2000. Vol. 18, no. 3. P. 535–547.
- Ляхов АИ, Пустогаров ИА, Гудилов АС. Проблема неравномерного распределения пропускной способности канала в сетях IEEE 802.11 // Информационные процессы. 2008. Vol. 8, no. 3. Р. 149–167.

Анализ характеристик видеопотоков, генерируемых видеохостингом YouTube*

К.Б. Рагимова, В.А. Логинов

ragimova.kb@phystech.edu, loginov@iitp.ru

ИППИ РАН

Аннотация За последние годы доля видеопотоков в общем объеме трафика увеличилась, превысив 70%, и продолжает расти. Значительная часть этого трафика генерируется популярными видеохостингами, такими как YouTube, NetFlix и др., использующими технологию адаптивной загрузки видеопотока по протоколу HTTP. В данной работе были исследованы характеристики видеопотоков, генерируемых крупнейшим в мире видеохостинговым сервисом YouTube. Было собрано большое количество данных о видеороликах, загруженных на YouTube, проведен анализ накопленных метаданных. Кроме того, было проведено исследование трафика, генерируемого при просмотре видеороликов на веб-сайте YouTube при использовании ПК и мобильных устройств.

1 Введение

По прогнозам компании Cisco [1] на долю видеотрафика в 2021 г. будет приходиться 79% всего интернет-трафика, в то время как в 2016 г. этот показатель составлял 61%. При этом большая часть этого трафика генерируется при использовании видеохостинговых сервисов, например, YouTube, Vimeo, NetFlix и др.

На сегодняшний день YouTube — это крупнейшая видеохостинговая компания, предоставляющая пользователям возможность загружать, просматривать, оценивать, комментировать видеоролики. Благодаря простоте и удобству использования YouTube стал самым популярным видеохостингом и вторым веб-сайтом в мире по количеству посетителей [2]. По данным YouTube [3] аудитория веб-сайта насчитывает больше одного миллиарда пользователей — это почти треть всех пользователей сети Интернет. Согласно статистике, ежедневно пользователи совершают миллиарды просмотров и сотни миллионов часов воспроизведения. При этом на мобильные устройства приходится больше половины всех просмотров. Поэтому исследование YouTube является важной задачей.

^{*} Исследование выполнено в ИППИ РАН за счет гранта Правительства Российской Федерации (Договор No 14.W03.31.0019).

Как и любой высокотехнологичный сервис, с каждым годом YouTube развивается и меняется. В связи с этим результаты многих работ, исследующих YouTube, могут быть уже неактуальны. В данной работе нам предстоит актуализировать накопленные о YouTube знания и расширить их, проанализировав характеристики видеопотоков, генерируемых популярным видеохостингом.

Хранение и обработка видеороликов при использовании видеохостинга YouTube устроены следующим образом. После загрузки видеоролика на видеохостинг YouTube ему присваивается уникальный идентификатор, называемый *Id*. Далее при помощи различных алгоритмов кодируются аудиои видеодорожки. После этого аудио- и видеодорожки инкапсулируются в файл контейнера. Существует несколько типов контейнеров, используемых в YouTube, а комбинация, состоящая из типа контейнера, кодека, а также разрешения и количества кадров в секунду для видеодорожки и битрейта аудиодорожки имеет уникальный параметр *itag*, не зависящий от видеоролика.

Дальнейшее изложение построено следующим образом. Раздел 2 содержит обзор литературы, посвященной исследованию различных аспектов работы видеохостинга YouTube. В разделе 3 рассматриваются характеристики аудио- и видеодорожек, доступных для различных видеороликов. В разделе 4 описывается поведение устройств при просмотре YouTube. В разделе 5 приводится заключение.

2 Обзор литературы

Анализу трафика, генерируемого при просмотре видеороликов на вебсайте YouTube при доступе с персональных компьютеров (ПК) или мобильных устройств, посвящено множество работ [4–9]. Например, авторы [4] описали основные свойства трафика YouTube при использовании проводных сетей и показали, что загрузка видеоролика при просмотре на веб-сайте YouTube состоит из двух фаз. Первая фаза загрузки видеоролика заканчивается после наполнения буфера. После этого этапа наступает вторая фаза: плеер YouTube ограничивает скорость передачи и серверы YouTube отправляют блоки с данными размером в 64 килобайта для поддержания уровня данных в буфере на приблизительно одном уровне.

В статье [5] трафик YouTube также исследовался при использовании проводных сетей, и было обнаружено, что в течение первой фазы YouTube загружает 40 секунд видеоролика, используя всю доступную пользователю пропускную способность сети. Авторы показали, что во время второй фазы скорость загрузки пропорциональна битрейту видеоролика с коэффициентом 1,25. Кроме того, было показано, что при просмотре видеоролика на веб-сайте YouTube с ПК для 92% видеороликов по умолчанию выбирается разрешение 360р и контейнер Adobe Flash Video (FLV) с битрейтом от 100 кбит/с до примерно 1 Мбит/с (itag = 34). В [6] авторы рассматривают доступ к YouTube с мобильных устройств при использовании Wi-Fi и исследуют характеристики генерируемого трафика. Они получили, что при загрузке видеоролик разделяется на сегменты. Каждый из них запрашивается в отдельном TCP-соединении при помощи протокола HTTP, в заголовке запроса которого указывается запрашиваемая часть (сегмент) видеоролика. В работе [6] показывается, что самые популярные форматы используют разрешение 360р с контейнерами MP4 (itag = 18) и FLV (itag = 34). Также было обнаружено, что высокий процент пользователей прерывает воспроизведение видеоролика до его окончания и что большинство пользователей не переключает разрешение, придерживаясь формата по умолчанию.

В работе [7] анализируется трафик, генерируемый при просмотре YouTube при использовании мобильной сети третьего поколения (3G). Основное отличие от результатов, полученных в [4–6], заключается в том, что преобладающими форматами видеороликов при использовании сети 3G являются itag = 36 и itag = 17. Они используют контейнер 3GP с более низким битрейтом по сравнению с 34-м и 18-м. Кроме того, во время начальной фазы YouTube в среднем загружает 34 секунды видеоролика, а не 40 секунд. Помимо того, если для передачи видеопотока YouTube при использовании проводных сетей коэффициент во второй фазе равняется 1,25, то на мобильных устройствах YouTube он равен 2.

С 2013 года YouTube поддерживает протокол динамической адаптивной потоковой передачи HTTP (Dynamic Adaptive Streaming over HTTP, DASH), который был впервые предложен в статье [8]. Позднее этот протокол был стандартизирован в рамках технологии MPEG-DASH. MPEG-DASH предусматривает разбиение видеопотока на последовательность сегментов. Каждый сегмент кодируется в нескольких битрейтах, отличающихся разрешением, количеством кадров в секунду и другими параметрами. При этом DASHплееру становятся доступными альтернативные сегменты, выровненные на одной временной шкале. По мере проигрывания, плеер автоматически выбирает следующий сегмент для скачивания и воспроизведения из доступных альтернатив. Например, DASH-плеер может переключаться на более низкое разрешение видеоролика, если в буфере имеется мало данных. Затем, как только буфер наполнится, разрешение видеоролика может быть увеличено.

Одной из первых работ, посвященных анализу реализации MPEG-DASH в YouTube, была работа [9]. В ней было исследован алгоритм выбора качества сегментов, используемый в DASH-плеере YouTube, при неизменной и меняющейся во времени пропускной способности канала передачи данных. Было показано, что используемый алгоритм позволяет обеспечить непрерывно воспроизведение видеоролика без прерываний. Однако, при изменениях пропускной способности канала, характерных для беспроводных сетей, исследуемый алгоритм может вести себя слишком консервативно, не повышая качество загружаемого видеоролика в течение долгого времени после повышения пропускной способности канала. В отличие от [4–9], в данной работе будет исследован не алгоритм выбора качества загружаемых сегментов, реализованный в DASH-плеере YouTube, а характеристики доступных аудио- и видеодорожек. Кроме того, будет исследовано, какие подмножества из множества всех доступных дорожек используются при просмотре видеороликов на YouTube с помощью различных устройств, операционных систем и приложений.

3 Анализ характеристик аудио- и видеодорожек, доступных для различных видеороликов

Для исследования характеристик видеопотоков, было необходимо собрать базу данных о видеороликах, доступных на видеохостинге YouTube. Каждый видеоролик, загруженный на YouTube, имеет уникальный идентификатор *Id*, к которому привязаны все метаданные видеоролика. Ключевыми задачами для сбора большого объема метаданных являются, во-первых, создание списка уникальных идентификаторов, и, во-вторых, разработка программы, автоматически собирающей метаданные видеоролика по заданному идентификатору.

Для анализа характеристик видеопотоков требовалась обширная база данных, поэтому было необходимо сгенерировать большое количество уникальных идентификаторов. За основу был взят метод, описанный в работе [10], использующий программный интерфейс YouTube Data API [11]. Этот интерфейс позволяет получить ответ на запрос, аналогичный поисковому запросу на веб-сайте YouTube. Однако с помощью этого запроса можно создать только несколько сотен уникальных идентификаторов, потому что YouTube Data API имеет программные ограничения и не предназначен для возврата большого объема данных. Тем не менее, есть возможность использовать YouTube Data API рекурсивно. Например, вначале взять небольшой набор идентификаторов, состоящий из 10 случайных Id. С помощью YouTube Data API найти идентификаторы 25 «рекомендованных» видеороликов. Таким образом, если повторить последнюю операцию рекурсивно дважды, количество новых идентификаторов увеличится до 10.25.25 = 6250. А значит, таким методом можно составить достаточно большой список идентификаторов для будущей базы данных.

Для генерации начального набора идентификаторов в данной работе был использован метод поиска по ключевым словам. Для каждого ключевого слова YouTube Data API возвращал фиксированное количество идентификаторов. На следующих шагах по каждому *Id* производился поиск «рекомендованных» видео. После проверки полученных идентификаторов было обнаружено несколько проблем. Первая проблема заключалась в том, что окончательный набор идентификаторов, полученный при помощи такого метода, содержал множество повторяющихся *Id*, так как множества видеороликов, «рекомендованных» после просмотра, могут пересекаться для различных видеороликов. Поэтому после выполнения каждой операции дубликаты должны были быть удалены. Следующая проблема возникла из-за



Рис. 1: Варианты доступных дорожек при просмотре видеоролика.

того, что YouTube Data API не учитывает страну, из которой делается запрос, поэтому видеоролики недоступные для проигрывания на территории Российской Федерации были также удалены. Таким образом, был получен окончательный список идентификаторов для составления базы данных.

Далее при помощи утилиты youtube-dl [12] для каждого *Id* был получен список доступных дорожек и информация о них. В частности, полученная база данных хранит следующую информацию: длительность видеоролика, количество просмотров, форматы, кодеки, которыми закодированы дорожки, их размеры, контейнеры и максимальное значение битрейта их сегментов, для видеодорожек также указаны параметры разрешения и количество кадров в секунду, а для аудиодорожек — приблизительный средний битрейт. Оказалось, что каждый видеоролик хранится в множестве форматов, как показано на рис. 1, причем их количество может варьироваться в зависимости от видеоролика от 8 до 40 дорожек.

Все доступные дорожки можно разделить на две части: мультиплексированные и немультиплексированные. В первой части аудио- и видеодорожки объединены в один файл, во второй — они хранятся по отдельности. Мультиплексированные дорожки хранятся в контейнерах MP4, WebM или 3GP. Немультиплексированные аудио- и видеодорожки хранятся в контейнерах MP4 и WebM. При этом в контейнере WebM аудиодорожки могут быть закодированы кодеками Opus или Vorbis. Кроме того, видеодорожки могут различаться количеством кадров в секунду (Frames per Second, FPS): для MP4 встречаются дорожки со стандартным количеством кадров в секунду (Standard frame rate, SFR), равным 25 или 30 FPS, и высоким количеством



Рис. 2: Распределение количества просмотров видеороликов в собранной базе данных.

кадров секунду (High frame rate, HFR) – 50 или 60 FPS; для WebM также встречаются дорожки с высоким динамическим диапазоном и высоким количеством кадров в секунду (HDR + HFR).

Полученная база данных содержит информацию о 33526 видеороликах. Исследуемые видеоролики имеют различную популярность, определяемую количеством просмотров. Рис. 2 показывает, что были охвачены как и очень популярные видеоролики с миллиардами просмотров, так и менее популярные видеоролики.

Стоит отметить, что в собранной базе данных только 178 видеороликов имеют дорожки с поддержкой технологии HDR (меньше 1%). Это связано с тем, что, во-первых, поддержка HDR была добавлена в YouTube относительно недавно (ноябрь, 2016 [13]), и, во-вторых, технологию HDR поддерживают не все устройства. Также 3882 видеоролика имеют разрешение с высоким количеством кадров в секунду (около 11,6%). Следует отметить, что в отличие от результатов, полученных в [5], на сегодняшний день YouTube отказался от использования контейнера FLV.

На рис. 3 показано распределение битрейтов видеодорожек в зависимости от разрешения, вертикальными пунктирными линиями изображены 95-процентили распределения. Согласно собранным данным, битрейты видеодорожек при различных разрешениях существенно пересекаются, так как битрейт видеодорожки зависит от содержания видеоролика.

На рис. 4 показаны распределения характеристик аудиодорожек для различных контейнеров и кодеков. Как можно видеть, для MP4 распределение представляет из себя практически вертикальные прямые, что означает, что в данном случае осуществлялось кодирование с константным битрейтом. С



Рис. 3: Средний битрейт видеодорожек при фиксированных разрешениях.



Рис. 4: Характеристики аудиодорожек для различных контейнеров и кодеков.

другой стороны, для кодеков Opus и Vorbis использовалось кодирование с переменным битрейтом.

На Рис. 5 изображено распределение отношения среднего битрейта видеодорожки в контейнере MP4 к среднему битрейту видеодорожки в контейнере WebM для фиксированного разрешения. Согласно полученным результатам, преимущество для какого-либо из кодеков практически отсутствует для высоких разрешений (360р и более). Для низких разрешений (144р и 240р) более эффективным является кодек vp9, используемый в контейнере WebM. При его использовании более 70% видеодорожек имеют меньший битрейт.

На рис. 6 изображена функция распределения для отношения среднего битрейта мультиплексированных дорожек к сумме средних битрейтов немультиплексированных дорожек с аналогичными параметрами, т.е. контейнером, разрешением для видеодорожек и звуковым кодеком аудиодоро-



Рис. 5: Отношение параметров видеодорожек в MP4 к соответствующим параметрам видеодорожек в WebM при фиксированном разрешении.



Рис. 6: Отношение среднего битрейта мультиплексированных дорожек к сумме средних битрейтов аудиодорожки и видеодорожки с соответствующими параметрами.

жек. Из полученных результатов можно сделать вывод, что практически во всех видео у мультиплексированных дорожек битрейт выше. По всей видимости, это можно объяснить тем, что в мультиплексированных дорожках при кодировании используются более старые версии кодеков, обладающие меньшей эффективностью сжатия аудио- и/или видеоданных.

На рис. 7 показано распределение отношения среднего битрейта видеодорожек с высоким количеством кадров в секунду (HFR) или с высоким динамическим диапазоном и высоким количеством кадров в секунду (HFR + HDR) к среднему битрейту видеодорожек со стандартным количеством кад-



Рис. 7: Отношение среднего битрейта видеодорожек с HFR и HDR к среднему битрейту видеодорожек с SFR.

ров в секунду (SFR). На графике распределениия можно заметить, что при использовании HFR (без HDR) битрейт увеличивается примерно в 1,5-1,7 раза. При этом, при использовании HDR разброс увеличивается до 1,0-3.7. Это может быть объяснено особенностями кодирования при применении технологии HDR.

Кроме того, при анализе видеодорожек был замечен следующий эффект. В некоторых случаях для одного и того же видеоролика перечень доступных дорожек со временем отличался. При этом, количество дорожек оставалось одинаковым, но для немультиплексированных ауди- и видеодорожек с соответствующими параметрами средние битрейты различались. В то же время, средние битрейты мультиплексированных дорожек оставались неизменными. Возможная причина заключается в том, что информация о дорожках зависит от сервера, на который YouTube переадресует пользователя при загрузке видеоролика. Поэтому, если в определенный момент видеоролик был перекодирован, то из-за необходимости синхронизации серверов хранения видеороликов, часть пользователей еще в течение некоторого времени будет проигрывать видеоролик с предыдущим вариантом кодирования.

4 Анализ поведения устройств при просмотре YouTube

При использовании DASH выбор контейнера и кодека происходит один раз в самом начале загрузки видеоролика. Это связано с тем, что технология DASH не гарантирует одинаковое выравнивание сегментов в разных контейнерах/кодеках, поэтому переключение с одного контейнера на другой в течение проигрывания видеоролика незаметно для пользователя сильно затруднительно или невозможно.

Мультиплексирование	1	1	1	1	1	1	I	I	+	+	+	+	+	+	1	1	I	I	+	+	+	+	
Видеодорожка	WebM	WebM	MP4	WebM	WebM	WebM	WebM	MP4	MP4	MP4	MP4	MP4	MP4	MP4	WebM	MP4	WebM	WebM	MP4	MP4	MP4	MP4	WebM
Аудиодорожка	WebM	MP4	MP4	WebM	WebM	WebM	WebM	MP4	MP4	MP4	MP4	MP4	MP4	MP4	WebM	WebM	WebM	WebM	MP4	MP4	MP4	MP4	WebM
Приложение	Google Chrome 67.0.3396.99	Microsoft Edge 42.17134.1.0	Internet Explorer 11.165.17134.0	Mozilla Firefox 61.0.1	Google Chrome 67.0.3396.99	Mozilla Firefox 57.0.4	Google Chrome 67.0.3396.79	Safari 9.1.2	Google Chrome 54.0.204505792	YouTube 13.27.8	Safari 11.1.2	Google Chrome 54.0.204505792	YouTube 13.27.8	Safari 11.1.1	Google Chrome 67.0.3396.87	YouTube 13.24.59	Google Chrome 67.0.3396.87	YouTube 13.25.56	Google Chrome 66.0.3359.158	YouTube 13.20.54	Google Chrome 66.0.3359.158	YouTube 13.20.54	Google Chrome 67.0.3396.87
OC	Windows 10	Windows 10	Windows 10	Windows 10	Ubuntu 17.04	Ubuntu 17.04	MAC OS X 10.11.6	MAC OS X 10.11.6	IOS 11.4	IOS 11.4	IOS 11.4	IOS 11.3.1	IOS 11.3.1	IOS 11.3.1	Android 6.0.1	Android 6.0.1	Android 6.0.1	Android 6.0.1	Android 7.0.0	Android 7.0.0	Android 6.0.1	Android 6.0.1	Android 8.0.0
Устройство	PC	PC	PC	PC	PC	PC	PC	PC	IPhone 5S	IPhone 5S	IPhone 5S	IPhone 6	IPhone 6	IPhone 6	Xiaomi Redmi Note 3	Xiaomi Redmi Note 3	Asus Zenfone 2	Asus Zenfone 2	Samsung Galaxy A3	Samsung Galaxy A3	Samsung Galaxy S5	Samsung Galaxy S5	Huawei P10 Plus

Таблица 1: Исследование YouTube на ШК и мобильных платформах.

В YouTube используются зашифрованные HTTPS-соединения для передачи видеороликов. Для анализа HTTPS-трафика YouTube на ПК и мобильных платформах в данной работе был использован прокси-сервер mitmproxy [14]. Mitmproxy реализует атаку «человек посередине» на используемый в HTTPS протокол шифрования, позволяя тем самым дешифровать и прочитать HTTPSзапросы. Таким образом, с его помощью можно получить подробную информацию о том, сегменты каких аудио- и видеодорожек скачивает DASH-плеер.

В данной работе был произведен анализ выбираемых аудио- и видеодорожек при использовании различных популярных браузеров, устройств и операционных систем. Согласно статистике [15], более 90% пользователей ПК используют исследуемые в данной работе браузеры. Аналогичный показатель для мобильных платформ составляет 72%.

Из результатов, полученных в ходе экспериментов над мобильными устройствами и ПК, следует, что выбор аудио- и видеодорожки зависит от устройства и приложения, см. таблицу 1. Для ПК почти все браузеры на самых распространенных операционных системах используют контейнер WebM и проигрывают немультиплексированные дорожки.

Однако для мобильных устройств ситуация отличается. В ходе экспериментов видеоролики проигрывались при помощи мобильных браузеров и официального приложения YouTube. Стоит отметить, что устройства производителей Apple и Samsung используют при проигрывании мультиплексированные дорожки с контейнером MP4, а устройства производителей Xiaomi, Asus и Huawei — немультиплексированные. При этом во время тестирования устройств удалось встретить разные комбинации при выборе контейнеров для аудио- и видеодорожек. Кроме того, было обнаружено, что выбор аудиодорожки может зависеть от используемого приложения, а выбор видеодорожки не зависит: при фиксированном устройстве для различных приложений выбор видеодорожки оказался одинаковым для всех протестированных устройств. Следовательно, зная производителя устройства, можно определить выбор контейнера видеодорожек.

5 Заключение

В данной работе были проанализированы характеристики видеопотоков, генерируемых самым популярным видеохостингом в мире — YouTube. Была собрана база данных о видеороликах, доступных на YouTube, содержащая информацию о более, чем 33 тыс. видеороликах; проанализированы различные характеристики доступных аудио- и видеодорожек. Кроме того, было исследовано поведение различных устройств и операционных систем при просмотре видеороликов на веб-сайте YouTube при помощи различных браузеров и приложений.

Список литературы

1. Cisco. — Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2016–2021 White Paper, 2017.

- 2. Alexa. -- youtube.com Traffic Statistics, 2018. -- July 24.
- 3. YouTube for Press. Access mode: https://www.youtube.com/intl/en-GB/yt/about/press/.
- Alcock Shane, Nelson Richard. Application flow control in YouTube video streams // ACM SIGCOMM Computer Communication Review. - 2011. -Vol. 41, no. 2. - P. 24-30.
- Analysis and modelling of YouTube traffic / Pablo Ameigeiras, Juan J Ramos-Munoz, Jorge Navarro-Ortiz, Juan M Lopez-Soler // Transactions on Emerging Telecommunications Technologies. - 2012. - Vol. 23, no. 4. - P. 360-377.
- 6. Youtube everywhere: Impact of device and infrastructure synergies on user experience / Alessandro Finamore, Marco Mellia, Maurizio M Munafò et al. // Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference / ACM. - 2011. - P. 345-360.
- Characteristics of mobile youtube traffic / Juan J Ramos-Muñoz, Jonathan Prados-Garzon, Pablo Ameigeiras et al. // IEEE Wireless Communications. 2014. Vol. 21, no. 1. P. 18–25.
- Timmerer Christian, Müller Christopher. HTTP streaming of MPEG media // Streaming Day. - 2010.
- YouTube's DASH implementation analysis / Javier Añorga, Saioa Arrizabalaga, Beatriz Sedano et al. // 19th International Conference on Circuits, Systems, Communications and Computers (CSCC). – 2015. – P. 61–66.
- Malik Haroon, Tian Zifeng. A Framework for collecting YouTube meta-data // Procedia Computer Science. – 2017. – Vol. 113. – P. 194–201.
- 11. YouTube API. Access mode: https://developers.google.com/youtube/.
- 12. youtube-dl utility. Access mode: https://rg3.github.io/youtube-dl/.
- 13. YouTube Official Blog. Access mode: https://youtube.googleblog.com.
- Cortesi Aldo. mitmproxy v4.0. 2018. May 15. Access mode: https://mitmproxy.org/.
- 15. StatCounter Global Stats. Browser Market Share Worldwide, 2018. June.

Реализация эффекта захвата канала в программно-конфигурируемом радио *

Кирилл Глинский, Алексей Куреев, Илья Левицкий kglinsk@yandex.ru, kureev@iitp.ru, levitsky@iitp.ru

Институт Проблем Передачи Информации РАН

Аннотация Одной из главных проблем современных беспроводных сетей стандарта 802.11 является интерференция. Поскольку очень часто устройства находятся вне зоны радиовидимости друга, а также используют механизм случайного доступа для передачи данных, нередко возникает ситуация, когда передача нескольких устройств на одной частоте пересекается во времени. В данном случае, согласно действующему стандарту, приемник не может принять ни один из кадров успешно. Тем не менее, если существенно более слабый сигнал поступил раньше, то существует возможность произвести пересинхронизацию и принять более сильный сигнал при достаточной разности мощностей между ними. Данный эффект называется эффектом захвата канала, наличие которого в беспроводных устройствах способно значительно изменить производительность сети. В данной работе для экспериментального изучения эффекта используется программно-конфигурируемое радио NI USRP 2944 RIO, на котором в рамках исследования был реализован эффект захвата канала.

Ключевые слова: Эффект захвата канала, USRP, скрытые станции, IEEE 802.11

1 Введение

В настоящее время технология беспроводных сетей стандарта 802.11 широко распространена[1], и прогнозируется дальнейший рост как количества подключенных устройств, так и объема передаваемых по сети данных. Таким образом, актуальными становятся исследования механизмов, позволяющих повысить пропускную способность и снизить задержки передачи данных в плотных сетях, а также снизить интерференцию со стороны скрытых станций. Одним из таким механизмов является эффект захвата канала, который при корректном использовании способен существенно повысить пропускную способность сети в некоторых сценариях [2]. Данный эффект заключается в возможности декодировать более сильный кадр из пересекающихся при

^{*} Исследование выполнено в ИППИ РАН за счет гранта Российского научного фонда (проект №16-19-10687)

передаче в одном диапазоне частот. Анализ эффективности использования эффекта захвата канала проводился в достаточно большом количестве работ. Например, в работе [3] авторы показывают, что эффект захвата канала позволяет повысить пропускную способность на границе локальной сети, в то время как в работе [4] было показано, что использование эффекта захвата канала может привести к неравномерному распределению пропускной способности внутри локальной сети. Тем не менее, большинство работ являются основываются на теоретическом исследовании эффекта захвата без детального экспериментального исследования. Одной из первых экспериментальных работ, посвященных исследованию эффекта захвата канала является работа [5], в которой построена достаточно громоздкая экспериментальная установка только для исследования одной модели сетевой карты, что исключает массовое исследование эффекта захвата канала. В работе [2] авторы впервые построили удобную в использовании экспериментальную установку, позволяющую исследовать абсолютно любую сетевую карту на наличие эффекта захвата канала без внесения изменений в драйвера

Тем не менее, конкретная реализация эффекта захвата канала, основанная на приципе, описаннов в [6], остается скрытой для исследователей и существует возможности изменять параметры принимающего устройства для изменения принципов приема кадра. В данной работе реализован эффект захвата канала в программно-конфигурируемом радио, что позволяет осуществлять гибкую настройку параметров приемника для дальнейшего его исследования.

2 Описание эффекта захвата канала в сетях 802.11

Согласно стандарту 802.11 после корректного приема преамбулы и поля L-SIG приемник определяет длительность кадра и производит прием и декодирование сигнала в течение всей продолжительности кадра. Поэтому в случае, когда происходит происходит наложение как минимум двух кадров во времени и наиболее мощный кадр принимается первым, приемник способен его успешно декодировать. Данная ситуация изображена на рис. ??. Для такого случая не нужно вносить дополнительные изменения в работу приемника.

Совсем иначе реагирует приемник на пересечение кадров, пример которых изображен на рис. 1 и рис. 2. В отсутстие эффекта захвата, оба пакета будут считаться недействительными. Однако, если эффект захвата канала реализован, то возможно принять более сильный кадр, если разница мощностей между ними ΔW_{RX} превосходит необходимое для используемой сигнально-кодовой конструкции соотношение сигнал-шум. Определив, что уровень принимаемого сигнала вырос не менее чем на пороговую разность мощностей $\Delta W_{Threshold}$, приемник прекращает принимать первый кадр и инициализирует процедуру поиска новой преамбулы. При этом, в обоих возможных случаях, изображенных на рис. 1 и рис. 2 сильный кадр прини-

мается успешно при разности мощностей, достаточной для декодирования данной сигнально-кодовой конструкции (СКК).



Рис. 1. Сценарий приема 2. Сильный кадр идет после слабого.



Рис. 2. Сценарий приема 3. Сильный кадр внутри слабого.

Данный эффект способен оказывать как положительное влияние на пропускную способность сети в сценариях с многочисленными скрытыми станциями и плотной сетью подключений, что было проверено в исследовании[2], так и потенциально нарушать справедливость механизма распределения ресурсов между участниками сети, предоставляя наиболее мощному источнику преимущество при приеме[7].

3 Имплементация алгоритма в USRP

3.1 Формулировка требований к алгоритму

К алгоритму, реализованному в USRP для наблюдения эффекта захвата канала, предъявляются следующие требования

- 1. Обладать обратной совместимостью с оборудованием не поддерживающим данный эффект, и соответствовать основным требованиям различных редакций стандарта 802.11
- 2. Иметь возможность регулировать предельные параметры сигналов, при которых начинает происходить пересинхронизация приемника.

- 3. Обеспечивать корректное начало приема более сильного сигнала вместе с освобождением всех данных слабого пакета из памяти и буферов обмена, разделяя обработку данных принятых из двух пакетов вне зависимости от сдвига по времени двух пакетов между собой.
- 4. Минимизировать количество ложных срабатываний эффекта при различных сценариях использования устройства.

В качестве основного признака, на основании которого можно предполагать коллизию двух пакетов с различными параметрами, было выбрано изменение значение мощности входящего сигнала во время приема пакета на величину выше определенного порога. Данный признак был выбран в силу следующих предположений

- 1. Мощность сигнала от одной станции слабо меняется в течении передачи одного кадра.
- Характеристики канала также не изменяются за время порядка длительности одного пакета.
- 3. Детектирование эффекта должно быть произведено и обработано за минимально возможное время.

Условие срабатывания эффекта по превышению порога мощности удовлетворяет поставленным условиям, так как вычисление мощности входного радиосигнала выполняется за 64 отсчета АЦП с частотой 80 MSPS, а изменение мощности приемника за время передачи кадра предполагается малым.

3.2 Реализация эффекта

После обнаружения скачка входной мощности логическая защелка устанавливается в новое состояние, не допуская дальнейшие избыточные попытки пересинхронизации до прихода сигнала сброса защелки. Сигналом сброса выступает падение входной мощности на величину выше порога, сигнализирующая о завершении кадра и готовности к принятию следующих. Данный алгоритм отличается от предлагаемого в патенте[6], в котором при наложении двух кадров возможно несколько пересинхронизаций из=за дискретности и задержек отсчета мощности, что может привести к невозможности принять кадр.

После срабатывания данного эффекта происходит переключение машины состояний, реализованной внутри FPGA USRP в состояние поиска синхронизации (детектирование преамбулы). Одновременно происходит перенастройка модуля CFO (Central Frequency Offset), устанавливающий величину сдвига частоты относительно эталонной.

На следующем этапе происходит обработка IQ, состоящая из преобразования Фурье, который сбрасывается в исходное состояние на время от детектирования подъема мощности до сигнала о начале пакета, оценки канала и коррекции фазы (перерасчитываемой при подъеме мощности) и выделении пилотных поднесущих сигнала. В дальнейшем поднесущие поступают в модуль обработки битов, состоящий из сериализатора битов, демаппера, декодера Витерби, которые приводятся к исходному состоянию. Также сбрасываются очереди данных и счетчики на все время действия защелки 1. Помимо этого все биты пришедшие во время защелки, обозначаются как не валидные и не передаются на дальнейшую обработку. Еще одна очистка происходит на этапе перевода массива битов в байты PSDU, при которой биты и индекс записываемого бита в массиве перезаписываются нулями.

Изменения были внесены и в машину состояний, отвечающую за управление приемом пакетов. Добавленные функции позволяют переключиться в режим окончания приема и корректно его завершить из любого исходного состояния машины.

Модули, обрабатывающие пакеты на МАС уровне также были переработаны. Был реализован сброс дизассемблера пакета, расчета контрольной суммы, а также машины состояний МАС уровня.

Вышеописанные изменения позволяют обработать и принять данные более сильного пакета до завершения приема и сигнала о принятии последнего сэмпла, после чего происходит сброс защелки (вследствие падения мощности радиосигнала) и станция становится готова к приему следующего пакета.

Схема реализованной установки представлена на рис. 3



Рис. 3. Блок-схема РНҮ уровня установки

4 Экспериментальная проверка эффекта

Для экспериментальной проверки работоспособности установки использовался сценарий эффекта захвата с двумя станциями-передатчиками, отправляющими пакеты с заданными разностью мощностей сигналов при приеме ΔW_{RX} и сдвигом по времени Δt .

Для проверки работоспособности реализованного эффекта захвата канала были произведены 2 эксперимента

- Измерение отношения принятых кадров к количеству отправленных (англ.: Frame Reception Ratio, FRR) в зависимости от временного сдвига между двумя пакетами. В эксперименте длина пакета составила 1000 байт, а время передачи кадра - 1430 мкс при СКК 0.
- 2. Измерение FRR для различных СКК и варьируемой разности мощности между пакетами.



Рис. 4. Зависимость FRR от сдвига между пакетами.

Зависимость FRR от временного сдвига изображена на рис. 4. Полученные результаты показывают, что данная установка позволяет обеспечить близкий к 1 FRR для второго пакета при пересечении пакетов на более чем 12 мкс, то есть после синхронизации на первый пакет, в отличие от стандартного режима работы, при котором FRR равен 0. Отметим, что при смещении кадров на 4 мкс FRR обращается в 0 для стандартного режима вследствие того, что приемник успевает начать синхронизоваться на первый кадр, что приводит к некорректному приему. Для реализованной установки второй кадр не принимается только при сдвиге от 6 до 12 мкс, что связано



с проведенной не до конца синхронизацией на первый кадр, за счет чего пересинхронизация не срабатывает.

Рис. 5. Зависимость FRR от соотношения ΔW_{RX} .

Рис. 5 отображает зависимость между FRR и разностью мощностей между пакетами для различных CKK при фиксированном пороге детектирования второго пакета, равному ΔW_{detect} =8 dB. Видно, что с увеличением номера CKK необходимая для корректного приема разность мощностей возрастает. Поскольку после пересинхронизации первый кадр продолжает передаваться, то соотношение сигнал/шум в канале определяется разностью мощности двух кадров ΔW_{RX} . Таким образом, на возможность принятия кадра влияет как порог детектирования второго кадра, равный 8 дБ в эксперименте, так и минимальное значение SNR для CKK. Из графика видно, что при CKK выше второго определяющую роль играет значение SNR.

5 Заключение

Сети стандарта 802.11 стандарта 802.11 подвержены влиянию эффекта захвата канала, способному существено улучшать параметры сети в условиях высокой плотности соединений. Задачей данной работы было создать установку, способную производить прием пакетов с учетом эффекта захвата канала. В рамках данной работы был реализован программно-аппаратный комплекс на базе USRP NI 2944 RIO для исследования эффекта захвата канала с возможностью настройки параметров детектирования эффекта. Созданный программно-аппаратный комплекс был протестирован в сценарии со скрытыми станциями и его работоспособность была экспериментально подтверждена.

Список литературы

- 1. Cisco. Cisco visual networking index: Global mobile data traffic forecast update, 2016 2021, 2016.
- Evgeny Khorov, Aleksei Kureev, Andrey Lyakhov, and Ilya Levitsky. Testbed to study the capture effect: Can we rely on this effect in modern wi-fi networks. In proc. of IEEE International Black Sea Conference on Communications and Networking, Batumi, Georgia, 2018, 2017.
- 3. Paul Patras, Hanghang Qi, and David Malone. Exploiting the capture effect to improve WLAN throughput. In 2012 IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM). IEEE, jun 2012.
- Andrey Lyakhov and Vladimir Vishnevsky. Wi-Fi hot spots. In Personal, Indoor and Mobile Radio Communications, 2007. PIMRC 2007. IEEE 18th International Symposium on, pages 1-5. IEEE, 2007.
- 5. Jeongkeun Lee, Wonho Kim, Sung-Ju Lee, Daehyung Jo, Jiho Ryu, Taekyoung Kwon, and Yanghee Choi. An experimental study on the capture effect in 802.11a networks. In Proceedings of the the second ACM international workshop on Wireless network testbeds, experimental evaluation and characterization. ACM Press, 2007.
- 6. Jan Boer. Wireless lan with enhanced capture provision, 1997.
- Yigal Bejerano, Hyoung-Gyu Choi, and Seung-Jae Han. Fairness analysis of physical layer capture effects in IEEE 802.11 networks. In 2015 13th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt). IEEE, may 2015.

Экспериментальное исследование неортогонального множественного доступа в Wi-Fi устройствах *

Алексей Куреев, Илья Левицкий, Евгений Хоров

Институт Проблем Передачи Информации РАН {kureev, levitsky, khorov}@iitp.ru

Аннотация Heoptoronaльный множественный доступ (англ.: Non-Orthogonal Multiple Access, NOMA) является на данный момент одной из наиболее перспективных технологий для повышения производительности сети в гетерогенных сценариях с различными типами трафика и условиями в канале. Хотя исследованию NOMA было посвящено большое количество работ, только в нескольких работах было произведено экспериментальное исследование данной технологии и, преимущественно, в сотовых сетях. В данной работе представлена экспериментальная установка, позволяющая исследовать NOMA в сетях Wi-Fi. Поскольку технология Wi-Fi гораздо проще и дешевле, чем сотовые сети, данная установка может быть использована для исследования NOMA в реальных условиях.

Ключевые слова: Неортогональный множественный доступ, IEEE 802.11, USRP

1 Введение

Неортогональный множественный доступ (англ.: Non-Orthogonal Multiple Access, NOMA) является на данный момент одной из наиболее перспективных технологий для повышения производительности сети в гетерогенных сценариях с различными типами трафика и условиями в канале. Большое количество теоретических работ демонстрировало эффективность использования NOMA в беспроводных сетях, исследовало механизмы распределения мощности передачи между пользователями и предлагало различные сигнально-кодовые конструкции, которые могли бы быть эффективнее классических (например, в [1]. В то же время, прототипирование NOMA устройств является весьма затруднительным из-за вычислительной сложности, что приводит к необходимости использовать программируемые логические интегральные схемы (ПЛИС). Более того, решение, работающее в теории, может оказаться совсем неэффективным на практике. Не смотря на

^{*} Исследование выполнено в ИППИ РАН за счет гранта Правительства Российской Федерации (Договор No 14.W03.31.0019)

трудности, прототипирование NOMA является необходимым для определения тех решений, которые окажутся наиболее эффективными на практике. К сожалению, существует всего лишь несколько работ, посвященных экспериментальному исследованию NOMA, и большая часть относится к исследованию NOMA в сотовых сетях и опубликованы не позднее 2017 года. Например, в работе [2] исследуется NOMA на основе сетей четвертого поколения с помощью программно определяемой радиосистемы USRP B210. Авторы используют USRP B210 для передачи нисходящего трафика в частотных каналах шириной 5 и 10 МГц. Ширина канала является такой узкой ввиду того, что данная установка требовала выполнять большое количество вычислений в режиме реального времени, которые исполнялись в процессоре. Поэтому авторам не удалось достичь значительного повышения пропускной способности в сравнении со стандартной технологией. Экспериментальное исследование передачи восходящего трафика было представлено в работе [3]. Авторы предлагают новые методы оценки каналы и синхронизации для NOMA приемника и также используют вышеупомянутые USRP В210. С помощью разработанной установки, авторы исследуют эффективность NOMA при использовании турбо кодов и множественного доступа с частотным разделением каналов с одной несущей частотой.

В данной работе представлена экспериментальная установка, позволяющая исследовать поведение NOMA в сетях Wi-Fi. В меру знаний авторов, данная установка является первой, в которой реализован NOMA на основе технологии Wi-Fi.

2 Описание установки

В разработанной установке используется программно определяемая радиосистема USRP-2944R [4], позволяющая принимать и передавать сигнал шириной до 160 МГц, используемой в станадрте 802.11ас. USRP-2944 имеет ПЛИС, внутри, благодаря которому вся цифровая обработка сигналов может происходить внутри USRP, не загружая центральный процессор.

Для реализации NOMA в USRP используется программа 802.11 Application Framework, содержащая упрощенную версию Wi-Fi. Авторами рассматривается сценарий передачи нисходящего трафика с помощью NOMA нескольким пользователям. Для этого должна быть произведена суперпозиция сигналов (англ.: Superposition Coding) на передатчике, а также должна быть реализован метод последовательного подавления помех (англ.: Successive Interference Cancellation, SIC).

2.1 Передатчик

Упрощенная схема передатчика, использующего суперпозицию сигналов изображена на рис. 1. Два независимых приложения генерируют данные на управляющем компьютере и помещают данные в очередь на USRP. ПЛИС получает данные и обрабатывает их в модулях MPDU Generation Block, где



Рис. 1. Схема передатчика

данные приводятся к формату, совместимому с МАС уровнем Wi-Fi. Поскольку время обработки данных может быть разными на верхней и нижней ветвях схемы, была проведена синхронизация после MPDU Generation Block. Далее, в модуле TX IQ Processing Block данные приобретают общую преамбулу с полями для синхронизации и оценки канала согласно стандарту Wi-Fi a затем каждый сэмпл одного кадра складывается с сэмплом другого. Для реализации NOMA каждый сэмпл одного из сигналов перед сложением должен быть домножен но коэффициент масштабирования s, который показывает, во сколько амплитуда одного сигнала меньше амплитуды другого. Таким образом, на выходе из модуля TX IQ Processing Block получается сигнал, являющийся суперпозицией двух — сильного и слабого, амплитуда которого в 1/s раз меньше амплитуды сильного. После сложения происходит быстрое преобразование Фурье, и затем сэмплы поступают в радиомодуль USRP.

2.2 Приемник

На рис. 2 изображена часть приемника, использующего последовательное подавление помех. После синхронизации данные отправляются в модуль RX IQ Processing Block, где происходит обратное преобразование Фурье и оценка канала. Далее, принятые сэмплы помещаются в буфер и их копия обрабатывается в верхних модулях RX Bit Processing и MAC RX Blocks. Если разница в мощностях между сильным и слабым сигналом достаточно велика, то данные модули могут без труда декодировать сигнал и получить данные. Устройство, в котором реализован SIC определяет сначала предназначен ли для него сильный кадр. Если нет, то данное устройство восстанавливает сэмплы сильного кадра и производит вычитание по сэмплам сильного кадра из суммы. Далее, результат подается на нижнюю часть схемы через модули Bit Processing и MAC RX, при успешном прохождении которых можно считать, что данные приняты корректно.



Рис. 2. Схема приемника

Стоит отметить, что стандартные устройства работают по схожим принципам, по которым функционирует верхняя часть схемы, и поэтому любое устройство может принять сильный сигнал, порожденный USRP без изменения в процесс приема.

3 Экспериментальные результаты

Для анализа разработанного механизма был проведен ряд экспериментов, в которых устройство передает пары кадров двум пользователям с помощью NOMA. Мощность передачи сильного сигнала была на 5 дБ выше мощности сигнала, при котором 95% кадров успешно принимались на СКК 3. Варьируя коэффициент масштабирования, была получена зависимость доли принятых кадров (англ.: Frame Reception Ratio, FRR) в зависимости от отношения сигнал/шум. Стоит отметить, что значение коэффициента масштабирования напрямую влияет на отношение сигнал/шум. Результаты эксперимента для различных сочетаний СКК изображены на рис. 3, где первое значение легенды показывает СКК сильного кадра, а второе — СКК слабого.

4 Заключение

В данной работе была разработана экспериментальная установка, позволяющая организовать неортогональный множественный доступ в сетях Wi-Fi. Данная установка позволяет без труда реализовывать и проверять теоретические работы, посвященные NOMA, практике. В отличие от сотовых сетей, сети Wi-Fi гораздо проще и дешевле, поэтому, предложенная установка



Рис. 3. Доля принятых кадров в зависимости от отношения сигнал/шум.

является удобной в использовании для будущих исследований в лабораториях и индустриальных исследовательских центрах. Поскольку технология Wi-Fi основана на методе случайного доступа в нелицензируемом спектре, необходимо по возможности снижать интерференцию и устанавливать надежную синхронизацию между устройствами, столь необходимую для неортогонального множественного доступа. С другой стороны, данная установка позволяет проводить эксперименты с обыкновенными устройствами Wi-Fi, что позволяет исследовать эффективность NOMA в сценариях с плотными сетями и высокой интерференции. Более того, данный подход без труда может быть дополнен такими широко используемыми технологиями как MU-MIMO и OFDMA, которые уже включены в стандарт. Поэтому, данная установка может быть крайне полезна при создании Wi-Fi будущего.

Список литературы

- S. R. Islam, N. Avazov, O. A. Dobre, and K.-S. Kwak, "Power-domain nonorthogonal multiple access (noma) in 5g systems: Potentials and challenges," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 2, pp. 721-742, 2017.
- X. Wei, Z. Geng, H. Liu, K. Zheng, and R. Xu, "A portable sdr non-orthogonal multiple access testbed for 5g networks," in *Vehicular Technology Conference (VTC Spring)*, 2017 IEEE 85th, pp. 1-5, IEEE, 2017.
- S. Abeywickrama, L. Liu, Y. Chi, and C. Yuen, "Over-the-air implementation of uplink noma," arXiv preprint arXiv:1801.05541, 2018.

4. National Instruments, DEVICE SPECIFICATIONS NI USRP-2944R, 2016.

Исследование эффекта захвата канала в сетях Wi-Fi *

Илья Левицкий, Алексей Куреев

Институт Проблем Передачи Информации РАН levitskiy@iitp.ru, kureev@iitp.ru

Аннотация В плотных сетях интерференция является основной проблемой, которая приводит к снижению пропускной способности. Часто считается, что перекрывающиеся передачи кадров приводят к потере обоих кадров. На самом деле иногда по крайней мере один кадр может быть декодирован, если его мощность приема выше, чем у другого кадра на некоторую предопределенную величину. Способность устройства получать более сильный кадр, если он приходит позже слабого, называется эффектом захвата канала. Для различных устройств эффект захвата проявляется по-разному. В частности, мы можем различать эффекты преамбулы и захвата кадра. Основным вкладом в работу является разработанная экспериментальная установка, с помощью которой можно изучать поведение эффекта захвата в обычных устойствах. А также в работе представлена еще одна установка, позволяющая проанализировать эффективность устройств с эффектом захвата в условиях реальной сети. С помощью этих установок показано, что обладают эффектом немногие устройства, но иммено такие устройства позволяют поддерживать пропускную способность сети на высоком уровне в условиях сильной интерференции.

Ключевые слова: Эффект захвата канала, Беспроводные сети, Wi-Fi, USRP, SDR, Интерференция

1 Введение

В радиосвязи эффект захвата – это явление, при котором сильнейший из двух одновременно передаваемых на одной частоте сигналов корректно демодулируется приемником. Способность приемника декодировать самый сильный кадр, даже если он приходит позже, является очень полезной при плотном развертывании сетей с десятками точек доступа Wi-Fi, расположенных в одной области пространства и в одном и том же частотном канале. Известными примерами таких сценариев являются жилые здания с точками доступа в каждой квартире, стадионы или торговые центры с тысячами точек доступа, развернутых для обеспечения широкополосного беспроводного доступа в Интернет для множества пользователей.

^{*} Исследование выполнено в ИППИ РАН за счет гранта Российского научного фонда (проект №16-19-10687)
Известно, что чаще всего пользователи подключены к точке доступа с самым высоким качеством принимаемого сигнала. Следовательно, сигнал от данной точки доступа является наиболее мощным. Более того, если разность мощностей между точкой доступа и интерферирующим устройством достаточно велика, то приёмник имеет возможность успешно принять более мощный кадр. Отметим, что в примере конечного автомата, описывающего приём кадра в стандарте IEEE 802.11 [1, раздел 17.3.12], устройство, после получения преамбулы кадра устройство, принимает кадр до его завершения, не пытаясь получить другую преамбулу. Это значит, что любой кадр, передача которого попадает на передачу уже принимаемого сигнала, неизбежно попадает в коллизию и оба кадра будут потеряны. Эффект захвата предназначен для помощи устройству получать мощные кадры корректно независимо от времени их появления.

Многие широко используемые результаты в математическом моделировании сетей Wi-Fi, такие как [2], были основаны на предположении, что устройства принимают кадры в соответствии со стандартом [1]. В связи с этим возникает вопрос, как производители реализуют механизм приёма кадров в устройствах и наблюдается ли в них эффект захвата.

Данный вопрос представляет особый интерес для разработчиков нового стандарта 802.11 ах [3]. Для повышения производительности плотных сетей стандарт предполагает эффективное использование пользователями частотно-временных ресурсов. Поэтому исследования эффекта захвата канала, предотвращающего значительное снижение пропускной способности в некоторых сценариях, являются особенно важными.

Научная новизна данной работы заключается в разработке метода, позволяющего экспериментально определить наличие эффекта захвата в произвольных сетевых картах, а также в исследовании характера его проявления. Простой способ, позволяющий определить, как именно проявляется эффект захвата в любых сетевых картах, предложен впервые.

Работа организована следующим образом. В разделе 2 определяется эффект захвата и объясняются различия между трактовками понятия «эффект захвата канала», используемыми в других работах. Раздел 3 рассматривает предыдущие исследования эффекта захвата. В разделе 4 подробно описывается экспериментальная установка для обнаружения эффекта захвата, приводится методика использования установки, представляются и анализируются экспериментальные результаты. В разделе 5 показано, как эффект захвата повышает пропускную способность соединений в условиях сильной интерференции. Раздел 6 завершает работу.

2 Об эффекте захвата канала

В различных работах термин «эффект захвата» используется для описания разных явлений. Некоторые авторы [4], [5] используют термин «эффект захвата», на самом деле описывая эффект подавления (англ. starvation effect). Эффект подавления проявляется в сценариях с несколькими потоками, где один поток полностью или частично блокирует другой. По этой причине в беспроводных сетях, в которых устройства используют множественный доступ с контролем несущей и предотвращением коллизий (англ. Carrier Sense Multiple Access With Collision Avoidance, CSMA/CA), приёмники испытывают серьёзные помехи в сценариях со скрытыми станциями.

Рассмотрим пример, приведённый на рис. 1. Пусть станция 1 и станция 3 находятся вне зоны радиовидимости друг друга. Если станции 1 и 3 передают насыщенные потоки станциям 2 и 4 соответственно, то пропускная способность соединения «1 - 2» ухудшается, поскольку станции 1 и 3 считают, что канал не занят, и почти всегда передают. Тогда с точки зрения станции 2 кадры от обоих передающих станций попадают в коллизию, и нет возможности принять ни один из них. В работе [6] показано, что в сетях Wi-Fi пропускная способность соединения «1 - 2» может быть на порядок ниже, чем пропускная способность соединения «3 - 4». Причина такого явления связана с правилами доступа к каналу и не вызвана принципами приёма кадра.



Рис. 1. Сценарий «эффекта подавления»

В данной работе под эффектом захвата понимают следующее: кадр с наибольшей мощностью с точки зрения приёмника может быть правильно декодирован даже при одновременной передаче несколькими станциями. Мощный кадр может начаться во время преамбулы слабого кадра или во время его тела. Для различия этих случаев, в данной работе они определяются как эффект захвата преамбулы и эффект захвата тела кадра, соответственно.

3 Обзор литературы

Влияние эффекта захвата в сети Wi-Fi исследуется 3 способами: с помощью построения аналитических моделей и их анализа, с помощью имитационного

моделирования и проводя эксперименты с реальными устройствами. Однако работ, использующих последний подход исследования, не так много.

В одной из таких работ ([7]) авторы разработали экспериментальную установку с использованием чипсетов Atheros AR5112 и доказали наличие эффекта захвата в этом устройстве. Используемая для изучения установка довольно сложна и состоит из пяти устройств, среди которых один приемник, два передатчика и два сниффера. Снифферы используются для измерения времени передачи кадров от передатчика и от источника помех. В установке также необходимо использовать препятствие между двумя передатчиками для уменьшения их взаимной радиовидимости. Для проведения экспериментов необходимо вносить изменения в драйвера устройства, что исключает массовость эксперимента. Несмотря на то, что в статье представлено исследование с использованием реального устройства, в экспериментах используется только одна модель уже устаревшего чипсета, и неясно, каким образом использовать установку для других чипсетов. Кроме того, наличие препятствия требует довольно сложной корректировки местоположения для достижения требуемой топологии сети.

В экспериментальной установке, разработанной в рамках данной работы, используется только одно программно-конфигурируемое приёмопередающее устройство, которое позволяет исследовать эффект захвата для любого устройства Wi-Fi. Более подробная информация об установке приведена в следующем разделе.

4 Анализ устройств на наличие эффекта захвата канала

4.1 Экспериментальная установка

Для исследования эффекта захвата необходимо создать поток перекрывающихся кадров с заданной мощностью передачи W_1 и W_2 и смещением по времени Δt , относительно начала слабого кадра, как показано на рис. 2. Затем необходимо определить, как такие перекрывающиеся кадры принимаются исследуемыми устройствами.

Для этого была создана экспериментальная установка, состоящая из двухканального передатчика, приемника (исследуемого Wi-Fi устройства), и вспомогательной точки доступа, см. рис. 3. Стоит обратить внимание, что разработанная установка позволяет исследовать несколько устройств Wi-Fi одновременно.

Приёмник В качестве приемника можно использовать, например, ноутбук, для сетевой карты которого необходимо оценить эффект захвата. Чтобы отслеживать пакеты, полученные данным устройством, используется программа Wireshark [8]. Поскольку многие устройства не поддерживают режим отслеживания (англ. monitor mode), передатчик должен генерировать



Рис. 2. Перекрывающиеся кадры

пакеты, которые должны быть обработаны на уровне выше канального, чтобы Wireshark обнаружила приём кадра. Чтобы была возможность легко обрабатывать полученные измерения, в каждом кадре указывается мощность передачи и смещение между перекрывающимися кадрами. Подробнее структура кадров описана ниже.

Вспомогательная точка доступа Поскольку без режима отслеживания устройство не может принимать пакеты, полученные из сети, к которой оно не подключено, используется вспомогательная точка доступа. В начале эксперимента устройство подключается к точке доступа. Во время эксперимента точка доступа продолжает отправлять служебную информацию, необходимую для поддержания соединения. Поэтому устройство остается подключенным к точке доступа.

В то же время передатчик генерирует кадры с адресом transmission address, который соответствует адресу точки доступа. Таким образом, устройство рассматривает эти кадры как отправленные точкой доступа.

Передатчик В качестве передатчика используется универсальная программноопределяемая радиосистема (англ. Universal Software Radio Peripheral, USRP). При подключении к компьютеру, USRP позволяет создавать прототипы систем беспроводной связи с использованием языка программирования LabVIEW.

USRP содержит программируемую логическую интегральную схему (ПЛИС), которая является эффективным инструментом для построения



Рис. 3. Схема экспериментальной установки

цифровых устройств, выполняющих передачу данных на высокой скорости. Поэтому с помощью USRP возможно реализовать функции управления доступом к среде (англ. medium access control, MAC). В частности, в разработанной установке используется NI USRP-2944R [9], которая имеет два отдельных встроенных приёмопередатчика (UBX-160 [10]), работающих в диапазоне от 10 МГц до 6 ГГц с максимальной полосой пропускания 160 МГц.

Для каждого приёмопередатчика имеется возможность варьировать мощность передачи в широком диапазоне. Для 2,4 ГГц аппаратное обеспечение USRP позволяет варьировать уровень мощности в диапазоне от -8 дБм до 20 дБм, а для 5 ГГц - в диапазоне от -20 дБм до 10 дБм. Для увеличения диапазона мощности дополнительно используется аттенюатор на 30 дБ.

Для генерации кадров с заданной мощностью передачи и смещением используется LabVIEW Communications 802.11 Application Framework. С его помощью можно осуществить передачу по стандарту 802.11, которые принимаются обычными устройствами Wi-Fi.

Application Framework был модифицирован в рамках работы, чтобы обеспечить одновременную передачу кадров с двух независимых приемопередатчиков USRP-2944R. Упрощенная схема двухканального передатчика Wi-Fi показана на рис. 4.

Хост-приложение на стороне компьютера генерирует два независимых потока данных для обоих приемопередатчиков и помещает данные в очереди FIFO прямого доступа к памяти (англ. direct memory access, DMA). USRP, который подключен к главному компьютеру, получает данные из очередей DMA FIFO и синхронно обрабатывает два потока данных параллельно. Для



Рис. 4. Схема двухканального передатчика

простоты реализации, в качестве данных используются запросы ARP [11], которые предназначены для определения MAC-адреса, имея IP-адрес другого устройства.

Обработка данных начинается с блока генерации данных протокола MAC (MPDU generation block), который отвечает за формирование кадров согласно стандарту 802.11. В частности, он добавляет MAC заголовок к входящему блоку данных.

Чтобы отличить кадры, принадлежащие различным сериям экспериментов и упростить обработку результатов экспериментов, параметры эксперимента включены в заголовок MAC. А именно, каждый кадр содержит три поля адреса. Первый, адрес получателя, является широковещательным, так как отсылаются ARP-запросы. Второй, адрес передатчика, устанавливается равным адресу AP. Наконец, третий, который является адресом инициатора запроса. В этот адрес и занесена следующая информация. Байты 2-й и 3-й указывают мощность кадров двух передатчиков внутри USRP, 4-й байт используется для различения передатчиков, а 5-й и 6-й байты содержат смещение кадра, выраженное в мкс.

В блоке MAC TX реализованы функции передачи на MAC уровне. Модуль поддерживает отсчет отсрочки, контролирует состояние канала и обеспечивает правильное межкадровое расстояние. Синхронизация обоих передатчиков происходит с частотой 10 МГц. Стандартная процедура по контролю состояния канала, реализованная в 802.11 Application Framework, была отключена с целью обеспечения наилучшей синхронизации между двумя передатчиками. Таким образом, состояние канала не влияет на процесс передачи. Чтобы настроить смещение между кадрами от двух передатчиков, им назначается разное количество и продолжительность слотов отсрочки. На соответствующее время один из потоков будет задержан относительно другого в блоке MAC TX. При таком подходе можно варьировать смещение с шагом 0.1 мкс.

В блоке обработки передаваемых битов (TX bit processing block) кадр MAC кодируется в соответствии с выбранной MCS.

Блок обработки TX IQ преобразовывает закодированные биты в I/Q отсчеты и посылает их в RF модуль, в котором происходит генерация конечного сигнала и его отправка.

4.2 Методология эксперимента

Описание эксперимента В экспериментах кадры передавались с использованием наиболее надежной сигнально-кодовой конструкции (MCS 0) стандартов 11а или 11g с номинальной скоростью передачи данных 6 Мбит/с. При таком MCS длительность передаваемых кадров равна 136 мкс.

Устройства расположены в пределах 1 м от антенн передатчика. Первоначально, мощность передачи W_{TX2} устанавливается максимальной, а W_{TX1} – минимальной, при которой устройство все еще получает большинство пакетов ($\approx 90\%$).

Смещение по времени Δt варьируется диапазоне [0, 200] мкс. Для каждого смещения, передается 1000 пар перекрывающихся кадров.

Затем, в случае обнаружения эффекта захвата на определенном ноутбуке, эксперимент повторяется с различными уровнями мощности, чтобы выяснить свойства чипа и, в частности, как эффект захвата зависит от уровней мощности обоих пакетов.

Для этого проводятся эксперименты при временных сдвигах 1 мкс и 80 мкс, соответствующие преамбуле PHY и телу кадра.

Калибровка установки Стоит отметить, что сигналы от двух антенн передатчика могут испытывать разное затухание PL_1 и PL_2 , и фактическая разница мощности передатчиков $\Delta W_{\rm TX}$ не равна разнице на приемнике ΔW_{RX} . Для тех устройств, которые не поддерживают режим отслеживания, нет возможности непосредственно получить разницу между мощностью приема сильного и слабого кадров. Чтобы преодолеть эту проблему, применяется следующий метод.

Разница мощностей на приёме ΔW_{RX} определяется как

$$W_{RX2} - W_{RX1} = (W_{TX2} - PL_2) - (W_{TX1} - PL_1)$$
(1)
= $\Delta W_{TX} - PL_2 + PL_1.$

где W_{RX1} и W_{RX2} – принимаемые мощности от двух антенн передатчика, W_{TX1} и W_{TX2} – мощности при передаче. Отсюда видно, что если строить графики в зависимости от разницы мощности на приемнике, предполагая $\Delta W_{RX} = \Delta W_{TX}$, то график был бы смещён на величину $PL_2 - PL_1$ по оси абсцисс. Следующий шаг заключается в изменении отображения между антеннами и потоками передачи. Тогда значение $\Delta \widetilde{W}_{RX}$ равно

$$\widetilde{W}_{RX2} - \widetilde{W}_{RX1} = \Delta \widetilde{W}_{TX} - PL_1 + PL_2.$$
⁽²⁾

При такой инверсии отображения потоков на антенны смещение графика изменит знак. Таким образом, проведя два идентичных эксперимента, в результате получаются 2 одинаковых графика, один смещённый относительно другого на величину $2|PL_2 - PL_1|$. Для определения действительных значений ΔW_{RX} достаточно взять полусумму выражений 1 и 2 и получить следующее выражение:

$$\Delta W_{RX} = \frac{\Delta W_{TX} + \Delta \widetilde{W}_{TX}}{2}.$$
(3)

Это означает, что, если провести 2 идентичных эксперимента, за исключением взаимной замены потоков передачи, то можно определить истинную зависимость доли принятых кадров (англ. Frame Reception Ratio, FRR) от ΔW_{RX} .

4.3 Численные результаты

Следующим этапом проводится анализ полученных результатов для двух случаев: когда наблюдается только эффект захвата преамбулы и когда также наблюдается эффект захвата тела кадра.

Рис. 5 показывает FRR для Intel WiFi Link 5300. Данное устройство может переключиться на более сильный кадр только во время первых 2 мкс, что соответствует участку для обнаружения кадра. Это приводит к почти 100% получению сильных кадров и абсолютным потерям слабых. Если более сильный кадр приходит позже, приемник не может переключиться на этот кадр, что приводит к потере обоих кадров. Оба кадра могут быть получены только при не перекрывающихся по времени передачах.

Результаты для чипсета AR9485, который обладает эффектом захвата тела кадра, изображены на рис. 6. Данный чипсет позволяет получать сильный кадр вне зависимости от начала его передачи. Единственным исключением является промежуток $\Delta t \in [3,7]$ мкс, когда AGC начинает обрабатывать первый кадр, игнорируя второй, что приводит к низкой FRR в этой области.

В дальнейшем, для краткости, будем использовать термин «эффект захвата», подразумевая именно эффект захвата тела кадра. В противном случае будем говорить, что эффект захвата отсутствует.

В данном разделе определяется порог проявления эффекта захвата для чипсета AR9485. Рис. 7 иллюстрирует FRR кадров TX2 для $\Delta t = 80$ мкс и различных уровней мощности TX1. Видно, что для того, чтобы устройство переключилось на более сильный кадр, разница в силе сигнала должна быть около 8 дБ. Это значение почти не зависит от абсолютных значений мощностей кадров. Помимо этого также была исследована зависимость величины порога от используемой сигнально-кодовой конструкции. Полученные результаты отражены на рис. 8. Отчетливо видно, что порог, заложенный в чипе, не меняется для различных MCS. При высоких индексах MCS порог смещен, что связано уже не с параметрами эффекта захвата, а с помехоустойчивостью используемого MCS. Значения этих порогов хорошо согласуются с другими проведенными экспериментальными результатами, например [12].

Таблица 1 подводит итоги результатов по обнаружению эффекта захвата в чипах Wi-Fi. Среди протестированных устройств лишь некоторые чипсеты производства Qualcomm имеют такой эффект.

Производитель	Дата релиза	Стандарты	Модель	Наличие
				эффекта
Broadcom	2012	a/b/g/n/ac	BCM4360	нет
Broadcom	2012	b/g/n	BCM43438	нет
Broadcom	2013	b/g/n	BCM4324A	нет
Intel	2008	a/b/g/n	Link 5300	нет
Intel	2009	b/g/n	Wireless-N 1000	нет
Intel	2009	a/b/g/n	Wireless-N 6205	нет
Intel	2009	a/b/g/n	Advanced-N 6250	нет
Intel	2015	a/b/g/n/ac	Wireless-AC 3165	нет
Intel	2014	a/b/g/n/ac	Wireless-AC 7265	нет
Qualcomm	2008	b/g/n	AR9285	нет
Qualcomm	2008	b/g/n	AR5B95	нет
Qualcomm	2008	b/g/n	AR928X	да
Qualcomm	2011	b/g/n	AR9485	да
Qualcomm	2014	a/b/g/n/ac	QCA9377	да
Qualcomm	2014	a/b/g/n/ac	QCNFA34AC	да

Таблица 1. Эффект захвата на теле кадра в различных устройствах



Рис. 5. Эффект захвата преамбулы



Рис. 6. Эффект захвата на всём кадре



Рис. 7. Зависимость FRR сильного кадра от разницы мощностей, $\varDelta t = 80~\mu {\rm s}$



Рис. 8. Зависимость FRR сильного кадра от разницы мощностей, параметр - MCS

5 Анализ эффективности устройств с эффектом захвата в реальных сетях

5.1 Постановка задачи

После обнаружения наличия или отсутствия эффекта захвата в устройствах и его подробного исследования, следующей задачей является анализ эффективности данных устройств в реальных сетях. Схема используемого сценария изображена на рис. 1. Такой сценарий отвечает реальному случаю, например, когда две пары «точка доступа - мобильное устройство» находятся в соседних комнатах. Точки доступа активно отправляют данные своим мобильным устройствам, при этом являясь скрытыми друг от друга.

5.2 Описание проведения эксперимента

Новая схема отличается от предыдущей из раздела 4.1 тем, что теперь необходимо два USRP. Один из них выступает в роли передатчика 1, а второй ведет передачу в соединении «3 – 4». Приёмником 2 выступает тестируемое устройство. Использование именно USRP вместо точек доступа или других компьютеров позволяет гибко управлять мощностью передачи и порогом мощности принимаемого сигнала. Это позволяет поставить необходимый эксперимент в ограниченном пространстве. Для возможности проведения такого эксперимента, используется последняя версия 802.11 Application Framework, в котором корректным образом реализован DCF. Полезная информация в соединении «1 – 2» представляет собой кадры стандарта 802.11а, передающиеся на MSC0. Интерферирующие кадры в соединении «3 – 4» будут посылаться с теми же параметрами. Эксперименты проводились с размерами полезных данных в пакете 36 байт; длительность таких кадров равна 112 мкс.

5.3 Численные результаты

Измерения проводились с сетевой картой (AR9485) с эффектом захвата и с сетевой картой (BCM4324A) без эффекта захвата. На рис. 9 изображена зависимость пропускной способности соединения «1 – 2» от длины интерферирующего кадра в соединении «3 – 4». Графики отражают характерную ситуацию для устройств с наличием (пунктир) и отсутствием эффекта захвата (сплошная линия). Без эффекта захвата пропускная способность соединения «1 – 2» даже при низких мощностях «3 – 4» составляет не более 20% от максимальной, теоретически равной 1,05 Мбит/с.

При наличии эффекта пропускная способность близка к максимальной и снижается только при разнице мощностей близких к порогу приема в 10 дБ. Причина во флуктуации мощности, которые можно пронаблюдать при помощи Wireshark. Статистика мощности приёма при одной серии эксперимента показывает, что значения мощностей колеблются в пределах ±3 дБ.



Рис. 9. Зависимость пропускной способности соединения «1 – 2» от длины интерферирующего пакета, объём пакета 36 байт

Из-за этой флуктуации с уменьшением разницы уменьшается и вероятность соблюдения порога, и, следовательно, срабатывания эффекта захвата.

Согласно построенным графикам, при одних и тех же параметрах эксперимента пропускная способность для устройства с эффектом захвата может превосходить пропускную способность устройства без эффекта в 5 раз при минимальных длинах интерферирующих кадров. При разности по мощности -12, -15 и -18 дБ и длинных интерферирующих кадрах первое устройство максимальную пропускную способность соединения, в то время как второе не способно принимать данные на фоне интерференции.

6 Заключение

В данной работе был изучен захват канала в устройствах Wi-Fi. Для этого была разработана и развернута экспериментальная установка, которая позволяет исследовать эффект захвата в произвольных устройствах без внесения изменений в их драйвера. В отличие от существующих установок, она обеспечивает высокую точность по времени и мощности посылки сигнала. Были введены и исследованы два типа эффекта захвата. Первый – эффект захвата преамбулы, второй – эффект захвата тела кадра. Результаты проведённых экспериментов показывают, что эффект захвата тела кадра наблюдается лишь в нескольких устройствах, в то время как эффект захвата преамбулы актуален для всех протестированных устройств.

Хотя эффект захвата тела кадра и не регламентирован стандартом, он довольно полезен для плотного развертывания со многими интерферирующими устройствами. В плотных сетях каждое устройство получает много чужих кадров низкой мощности, что не позволяет ему получать свои собственные кадры. Однако, как показывают результаты, большинство устройств не поддерживают эффект захвата и получают кадры по стандарту Wi-Fi.

В то время как большинство исследований рассматривают только один тип эффекта захвата или вообще не рассматривают эффект захвата, на самом деле в сети могут быть устройства с различными типами эффектов захвата. Это поднимает вопрос об эффективности и справедливости распределения ресурсов в такой сети. Соответствующие исследования имеют большое значение для разработки будущих технологий, работающих в плотных сетях.

Список литературы

- 1. "802.11-2016 IEEE Standard for Information technology–Telecommunications and information exchange between systems Local and metropolitan area networks-Specific requirements - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications," ANSI/IEEE Std. 802.11, pp. 1-3534, December 2016.
- 2. G. Bianchi, "Performance analysis of the ieee 802.11 distributed coordination function," IEEE Journal on selected areas in communications, vol. 18, no. 3, pp. 535–547, 2000.
- 3. E. Khorov, A. Kiryanov, and A. Lyakhov, "IEEE 802.11 ax: How to Build High Efficiency WLANs," in 2015 International Conference on Engineering and Telecommunication (EnT), pp. 14–19, IEEE, 2015.
- 4. W. Wang, B. Leong, and W. T. Ooi, "Mitigating unfairness due to physical layer capture in practical 802.11 mesh networks," IEEE Transactions on Mobile Computing, vol. 14, no. 1, pp. 99-112, 2015.
- 5. V. Subramanian, K. Ramakrishnan, S. Kalyanaraman, and L. Ji, "Impact of interference and capture effects in 802.11 wireless networks on tcp," in Proceedings of the second international workshop on Wireless traffic measurements and modeling, p. 2, ACM, 2006.
- 6. A. Lyakhov, I. Pustogarov, A. Safonov, and M. Yakimov, "Starvation effect study in IEEE 802.11 mesh networks," in Mobile Adhoc and Sensor Systems, 2009. $MASS'09.\ IEEE\ 6th\ International\ Conference\ on,\ pp.\ 651-656,\ IEEE,\ 2009.$
- 7. J. Lee, W. Kim, S.-J. Lee, D. Jo, J. Ryu, T. Kwon, and Y. Choi, "An experimental study on the capture effect in 802.11 a networks," in Proceedings of the second ACM international workshop on Wireless network testbeds, experimental evaluation and characterization, pp. 19–26, ACM, 2007.
 8. G. Combs et al., "Wireshark," Web page: http://www. wireshark. org/last modified,
- pp. 12-02, 2007.
- 9. National Instruments, Device specifications NI USRP-2944R, 2016.
- 10. P. H. Dr., UBX Daughterboard. National Instruments.
- 11. D. C. Plummer, "RFC 826: An ethernet address resolution protocol," InterNet Network Working Group, 1982.
- 12. G. Pei and T. R. Henderson, "Validation of ofdm error rate model in ns-3," Boeing Research and Technology, 2010.

Исследование кодированной слотированной ALOHA в неасимптотическом режиме

Дарья Устинова^{1,2}, Алексей Фролов^{3,1}, Павел Рыбин^{3,1}

¹ Институт Проблем Передачи Информации РАН, Москва, Россия, ² Московский физико-технический институт, Москва, Россия

³ Сколковский институт науки и технологий, Москва, Россия

ustinova.DT@phystech.edu, al.frolov@skoltech.ru, p.rybin@skoltech.ru

Аннотация Проведено исследование кодированной слотированной АЛОХИ (IRSA) в неасимптотическом режиме. Сравниваются различные распределения числа передач: все пользователи посылают одинаковое число копий: 1,2 или 3, распределение солитона, а также распределение, полученное с помощью метода эволюции плотностей. Показано, что асимптотически оптимальные распределения не являются лучшими для случая фрейма конечной длины. В частности, показано, что распределение солитона проигрывает модели, где каждый пользователь посылает по 3 копии.

Keywords: кодированная слотированная АЛОХА, метод эволюции плотностей, распределение солитона, режим конечной длины.

1 Введение

В данной работе мы рассматриваем проблему неэффективности асимптотических распределений количества попыток для конечного числа слотов IRSA и сравниваем поведение многочленов в неасимптотическом режиме.

В 70-х годах Норман Абрамсон со своими коллегами из университета Гаваи предложил новый способ распределения канала [1]. Система получила название ALOHA. Основная идея заключалась в возможности любому количеству пользователей независимо друг от друга использовать один и тот же канал. Несколько лет спустя была предложена модель слотированной АЛОХИ, т. е. разбитой на временные слоты. При этом каждый пользователь может передавать информацию только в начале данного слота.

Модель contention resolution diversity slotted ALOHA (CRDSA) предложена в [2]. Это довольно простое, но эффективное усовершенствование слотированной ALOHA. Основная идея заключается в том, что вместе с пакетом информации каждый пользователь посылает 2 копии в разные слоты. Таким образом улучшается вероятность успешного приёма. Далее для этой модели будет предоставлен график зависимости пропускной способности⁴

⁴ здесь и далее в этой работе под пропускной способностью (англ. throughput) мы понимаем среднее число успешно переданных пакетов за один слот

от среднего числа пакетов на слот T(G) в случаях, когда все пользователи посылают одинаковое число копий: 1,2 или 3.

В 2011г. Лива в своей статье [3] предложил ещё более усовершенствованную модель - irregular repetition slotted ALOHA (IRSA), которая основывается на оптимизации CRDSA с помощью двудольного графа. Основное отличие данной модели заключается в том, что для каждого пользователя независимо выбирается количество посылаемых им копий в соответствии с заданным распределением вероятности.

Главной целью работы является исследование кодированной слотированной АЛОХИ в неасимптотическом режиме. Основные результаты заключаются в следующем. Показано, что оптимальное распределение вероятности для бесконечной длины не является наилучшим в неасимптотическом режиме. Получено распределение для конечной длины с помощью метода эволюции плотностей, приведено его сравнение с другими распределениями.

2 Предварительные сведения

2.1 Обозначения и модель системы

Рассмотрим систему с К пользователями, в которой передача ведётся фреймами. Фрейм состоит из M слотов. Тогда G = K/M - среднее число пакетов на слот. Среднее число успешно переданных пакетов на слот составляет $T(G) = G\mathbb{P}_{suc} = G(1 - P_{err})$. Будем рассматривать случай $G \leq 1$.

Каждый пользователь передает пакет в нескольких слотах. Пользователи выбирают число передач случайно, в соответствии с заданным распределением. Это распределение одинаково для всех пользователей. Слоты, в которые передаются копии пакета, тоже выбираются случайно. В соответствии с моделью из [3] в каждом пакете хранятся адреса его копий.

2.2 Схема декодирования

Процесс декодирования удобно представить с помощью двудольного графа (см. Рис. 1). Вершинами являются пользователи (В) и слоты (S), а ребро обозначает посылку копии.

На вход декодера поступает матрица смежности для пользователей и слотов, в которой 1 означает посылку пакета, 0 - её отсутствие. Номер слота соответствует номеру столбца, а пакета - номеру строки. Пример матрицы для рисунка 1:

Γ	1	0	1	0]
	1	1	0	0
	0	0	1	1
L	0	0	0	1

Декодирование происходит следующим образом. Ищутся слоты, в которые был послан только один пакет. Если одновременно было послано более

одной копии, то такую информацию нельзя распознать. Далее этот пакет декодируется, при этом удаляются его копии из других слотов. Так делаем до тех пор, пока не останется вершин S степени 1. В конечном итоге нас интересует доля недекодированных пакетов, то есть вероятность ошибки $P_{\rm err}$.



Рис. 1: процесс декодирования

3 Методы выбора оптимального распределения

3.1 Оптимальное распределение в асимптотическом режиме

В статье [4] была доказана оптимальность распределения солитона для бесконечного числа слотов:

$$L^{N}(x) = \sum_{i} L_{i} x^{i} = \frac{\sum_{i=2}^{N+1} \frac{x^{i}}{i(i-1)} - ax^{2}/2}{\sum_{i=2}^{N+1} \frac{1}{i(i-1)} - a/2}$$
(1)

Где N+1- число слотов, коэффициент $a\in[0,1]$ влияет на эффективность $\eta^N{=}\frac{N}{N{+}1}{-}a.$ Для бесконечной длины мы не можем взять a=0,т. к. сходимость будет медленной (количество итераций будет очень большим). Однако

в нашем случае конечного числа слотов можно принебречь этим и принять a = 0.

В этом многочлене смысл коэффициентов при x^i - вероятность того, что данный пользователь пошлёт і копий. Заметим, что $L_0 = L_1 = 0$.

3.2 Метод эволюции плотностей

В данной работе мы будем пользоваться методом эволюции плотностей (density evolution, или DE), который также позволяет найти близкое к оптимальному распределение в асимптотическом режиме.

В статье [4] приведен вывод вероятности стирания на (l+1)-м шаге:

$$x_{l+1} = \lambda (1 - \rho (1 - x_l))$$
(2)

Здесь λ , ρ - распределения вероятностей для пользователей и слотов соответственно, начальное условие $x_0 = \varepsilon$. Основная идея заключается в нахождении максимального ε такого, что $x_{\infty} \longrightarrow 0$.

Более подробное описание метода можно найти в [5].

3.3 Фиксированное число передач для всех пользователей

Также для сравнения будут рассмотрены случаи, когда все пользоатели посылают одинаковое число копий : 1, 2 или 3. Дальнейшее увеличение числа попыток приводит к резкому снижению эффективности, поэтому мы ограничимся только этими значениями.

4 Численные результаты

Мы будем рассматривать случай фиксированного конечного числа слотов M = 100, меняя лишь число пользователей K.

Для исследования эффективности распределений в неасимптотическом режиме были рассмотрены случаи, когда каждый пользователь посылает: 1) 1 копию

2) 2 копии

3) 3 копии (случай с 4 и более копиями был рассмотрен, однако исключен из графиков из-за значительного снижения эффективности)

4) число копий, зависящее от распределения солитона

5) число копий, зависящее от распределения, выведенного с помощью метода эволюции плотностей

При этом критерием оптимизации является увеличение пропускной способности.

Многочлен DE:

$$L^{DE}(x) = 0.5x^2 + 0.24x^3 + 0.16x^7 + 0.02x^{15} + 0.06x^{16} + 0.01x^{54} + 0.01x^{64}$$
(3)

Исследуя полученные зависимости, можно сделать выводы:

1) Наиболее эффективным оказалось распределение, в котором каждый



Рис. 2: Зависимость $log_{10}P_{err}(G)$

пользователь посылает 3 копии

2) Распределение DE практически совпало по эфективности с распределением солитона

3) Распределения, в котором каждый пользователь посылает 2 или 3 копии выигрывают у случая, когда каждый пакет был отправлен ровно один раз.

Также были получены графики зависимости $T(G) = G(1-P_{err})$



Рис. 3: Зависимость T(G)

Здесь также видно, что при 3 попытках код показывает наиболее высокую эффективность.

5 Заключение

В результате моделирования показано, что распределение, в котором каждый пользователь посылает по 3 копии, показывает большую эффективность, чем распределение солитона, а также распределение, полученное с помощью DE на рассмотренном фиксированном количестве слотов.

В дальнейшем работу можно развить на поиск наиболее оптимального распределения для конечной длины, а также рассмотреть случай Т-кратной ALOHA.

Благодарности

Исследование выполнено в ИППИ РАН при финансовой поддержке РНФ в рамках научного проекта No 14-50-00150.

Список литературы

- 1. N. Abramson 1969 annual report THE ALOHA SYSTEM University of Hawaii Honolulu Hawaii January 1970
- E. Casini, R. De Gaudenzi, and O. del Rio Herrero "Contention resolution diversity slotted ALOHA (CRDSA): an enhanced random access scheme for satellite access packet networks" *IEEE Trans. Wireless*, Commun., vol. 6, no. 4, pp. 1408-1419, Apr. 2007.
- G. Liva. "Graph-Based Analysis and Optimization of Contention Resolution Diversity Slotted ALOHA" IEEE Trans. on Commun., vol. 59, no. 2, pp. 477–487, 2011.
- Narayanan, K. R.; Pfister, H. D. Iterative collision resolution for slotted ALOHA: An optimal uncoordinated transmission policy.. Processing (ISTC), 2012 7th International Symposium on, 136,139, 27-31
- 5. T. Richardson and R. Urbanke Modern coding theory Cambridge university press, 2008.

Сверхнадежная связь с низкой задержкой в сетях Wi-Fi на основе IEEE 802.11ba*

М.Л. Сандал, Д.В. Банков, Е.М. Хоров

mark.sandal@gmail.com, {bankov, khorov}@iitp.ru ИППИ РАН, НИУ ВШЭ, МФТИ, École Polytechnique

Аннотация Одной из особенностей беспроводных сетей нового поколения является поддержка сверхнадёжной связи с низкой задержкой (англ. Ultra Reliable Low Latency Communications, URLLC), предъявляющей строгие требования к задержке при передаче данных и доле потерянных пакетов. Практически невозможно удовлетворить данные требования в рамках существующих на данный момент технологий передачи данных. В данной работе предлагается и исследуется подход, основанный на дополнительном радиоинтерфейсе, описанном в дополнении к стандарту IEEE 802.11ba, который обеспечит поддержку URLLC в сетях Wi-Fi. Для исследования эффективности данного подхода построена математическая модель гетерогенной сети, состоящей из устройств, передающих обычные и приоритетные потоки данных. При помоши модели исследована достижимая задержка при доставке данных устройств URLLC и пропускная способность, доступная обычным станциям. Показано, что предложенный подход позволяет на порядки понизить задержку при доставке данных, в то время как пропускная способность уменьшается незначительно.

1 Введение

Сверхнадежная связь с низкой задержкой (англ. Ultra-Reliable Low-Latency Communications, URLLC) является составной частью развивающихся беспроводных сетей пятого поколения, цель которых — обеспечение передачи данных с высокой надежностью (вероятность потери пакета порядка 10^{-5}) и низкой задержкой (менее миллисекунды) [1,2]. Такие высокие требования требуют изменения существующей архитектуры сети и разработки новых режимов работы для технологии беспроводных сетей. Примером инноваций, вызванных URLLC, может служить разработка нового радиоинтерфейса (англ. New Radio air interface) для сотовой сети, что существенно снизило гранулярность передач, поскольку текущая гранулярность планирования LTE в 1 мс слишком велика, чтобы удовлетворить нормам URLLC на задержку [3,4].

^{*} Исследование выполнено в МИЭМ НИУ ВШЭ за счет гранта Российского научного фонда (проект №18-19-00580)

Помимо сотовых сетей, URLLC также очень важна для сетей Wi-Fi во многих сценариях, связанных с индустриальным и игровым применением, а также с дистанционным управлением и многим другим.

Обеспечение требуемого уровня надежности и задержки в сетях Wi-Fi представляет собой непростую задачу, поскольку станции Wi-Fi используют метод случайного доступа к каналу, что влечет за собой коллизии и, как следствие, рост задержки и количества потерянных пакетов. Другой проблемой является тот факт, что станция обязана дождаться момента, когда канал освободится, прежде чем начать передачу, при том что передача кадра может занять значительное время (до 5 мс). Следовательно, при существующих стандартах Wi-Fi удовлетворить требованиям URLLC невозможно.

В данной работе исследуется и оценивается подход к внедрению URLLC в сети Wi-Fi, предложенный на пленарном заседании IEEE 802 в ноябре 2017 года [5]. В основе этого подхода лежит использование вторичного радиоинтерфейса, как, например, в стандарте 802.11ba [6].

Чтобы оценить эффективность метода, разработана математическая модель передачи данных в гетерогенной сети с двумя типами станций: передающими пакеты URLLC и передающими обычные пакеты. Эта модель позволяет вычислить такие важные показатели как задержка пакетов URLLC и пропускная способность обычных станций.

Дальнейшее изложение построено следующим образом. В Разделе 2 описан метод доступа к каналу, используемый в сетях Wi-Fi. Далее в Разделе 3 описана модификация метода доступа, позволяющая предоставлять приоритетный доступ к каналу для потоков URLLC. В Разделе 4 приведён обзор литературы, содержащей решения для обслуживания трафика с низкой задержкой и высокой надёжностью. Раздел 5 содержит описание сценариев моделирования и формальную постановку задачи. В Разделе 6 строится математическая модель процесса передачи данных в заданных сценариях. В Разделе 7 приведены численные результаты. Раздел 8 содержит заключение.

2 Доступ к каналу в сетях Wi-Fi

Основным методом доступа к каналу в современных сетях Wi-Fi является EDCA (англ. Enhanced Distributed Channel Access) — разновидность метода множественного доступа с контролем несущей и избежанием коллизий с двоичной экспоненциальной отсрочкой. Ниже приведено его краткое описание.

Каждая станция имеет очередь пакетов на передачу, куда добавляет новоприбывшие пакеты. Если по прибытии пакета эта очередь пуста, станция проверяет, происходит ли на данный момент в канале какая-либо передача. В случае, если канал свободен, станция тут же начинает передачу новоприбывшего пакета. Если же канал занят, то станция взводит счетчик отсрочки, представляющий собой случайную величину, равномерно распределенную на интервале $[0, W_r - 1]$, где r — число неудавшихся попыток передачи текущего пакета (изначально равный 0), а W_r — конкурентное окно, которое считается следующим образом:

$$W_{r} = \begin{cases} W_{min}, & r = 0\\ \min(2W_{r-1}, W_{max}), & r > 0 \end{cases}$$

Здесь W_{min} и W_{max} означают нижнюю и верхнюю границы величины конкурентного окна соответственно.

Если канал занят, станция замораживает свой счетчик. В случае же, если канал свободен в течение заданного временного интервала AIFS (англ. Arbitration Inter-Frame Space, конфликтный межкадровый интервал), станция вновь активирует свой счетчик и уменьшает его на 1 каждый временной слот σ . Как только счетчик достигает нуля, станция делает попытку передать свой пакет. Получение этого пакета должно быть подтверждено адресатом посредством служебного кадра ACK по истечении интервала SIFS (англ. Short Inter-Frame Space, короткий межкадровый интервал) после получения. При своевременном получении подтверждения станция удаляет текущий пакет из очереди, сбрасывает в нуль счетчик r и переходит к следующему пакету. Если же отправляющая станция в течение ACKTimeout не получает подтверждения, она считает отправленный пакет потерянным, увеличивает свой счетчик r и пробует снова.

ЕDCA различает четыре типа категорий доступа (англ. Access Categories, ACs), каждая из которых имеет собственную очередь и свой счетчик отсрочки. При этом, W_{min} , W_{max} и AIFS для каждой очереди тоже различаются, что позволяет EDCA приоритезировать очереди.

3 Приоритетная схема доступа

Идея предлагаемого метода состоит в добавлении новой категории доступа для трафика URLLC и использовать добавочный радиоинтерфейс для передачи предупредительных служебных кадров. Для этой цели подойдет так называемый пробуждающий радиоинтерфейс (англ. Wake-Up Radio, WUR), работающий в узкой частотной полосе вне основного канала и обеспечивающий медленную, но надежную связь.

Когда очередь URLLC не пуста, станция отправляет сигнал занятости канала (англ. busy tone) в служебном канале. По получении этого сигнала, все станции, передающие обычные (не URLLC) пакеты тут же освобождают канал, чтобы дать URLLC пакетам возможность быть переданными (см. рис.1.1). Далее станция URLLC ждет AIFS и, если основной канал свободен, осуществляет попытку передачи. В противном случае станция соревнуется за канал с остальными станциями URLLC в соответствии с правилами EDCA. Важно заметить, что первая передача осуществляется без взвода счетчика отсрочки, что аналогично передаче EDCA в случае, когда пакет приходит в пустую очередь. Ввиду того, что такая передача осуществляется асинхронно, она происходит с минимально возможной задержкой и с наименьшей вероятностью коллизии (по сравнению с передачей, сделанной по истечении счетчика отсрочки).



Рис. 1. Приоритетный доступ для пакетов URLLC

Пока служебный канал занят, обычные станции считают обычный канал занятым, то есть, их счетчики отсрочки заморожены. Если же обычная станция получает сигнал занятости канала, находясь в состоянии передачи, она считает, что для передаваемого ей кадра произошла коллизия.

4 Обзор литературы

Задача обеспечения высокой надежности и малой задержки уже была рассмотрена в контексте технологии Ethernet, что привело к появлению решения в виде стандартов 802.1 TSN (англ. Time Sensitive Networking — сети, чувствительные к задержке) [7,8]. Эти стандарты описывают методы обеспечения высокой надежности посредством повторения пакетов, контроля пути и резервирования потока, в то время как низкая задержка достигается путем временного планирования и механизма пресечения кадров. Проблема планирования решается на уровне сетевого коммутатора, который должен обслуживать пакеты из разных потоков в таком порядке, который гарантирует им определённую «наихудшую» задержку, что может быть интерпретировано как задача оптимизации [9]. Однако решения, предложенные для Ethernet, не решают проблемы, возникающие в беспроводных сетях, так как современные сети Ethernet состоят в основном из соединений типа «точкаточка» и не подвержены коллизиям кадров [10].

Стандарт IEEE 802.15.4 также был рассмотрен с перспективы беспроводных сетей [11], чувствительных ко времени. В этом стандарте время делится на период конкуренции за канал и период от нее свободный. Конечное устройство может запросить гарантированный временной слот (англ. Guaranteed Time Slot, GTS) от координатора сети и получить временной промежуток, отведенный специально под его передачи, что позволяет гарантировать определённый уровень задержки. Однако, одновременно возможно выделить лишь 7 GTS, и, несмотря на то, что данную схему можно улучшить, позволив нескольким потокам низкой интенсивности разделять один и тот же GTS, задержка ниже, чем 1 мс, таким путем недостижима.

В работе [12] разработана одна из наиболее известных моделей метода случайного доступа, используемого в сетях Wi-Fi. Модель описывает передачу данных в режиме насыщения в сети равноправных устройств. Процесс передачи данных станцией описывается при помощи цепи Маркова с дискретным временем, а пропускная способность сети выражается через вероятность того, что слот будет успешным, коллизионным или пустым соответственно, которая находится через стационарное распределение цепи Маркова. Данная модель является основой для модели, построенной в разделе 6.

Проблема сосуществования устаревших станций и станций, использующих новые возможности физического уровня рассмотрена в [13]. Данная статья рассматривает сценарий, в котором сеть состоит из IEEE 802.11ах и устаревших станций. Все они передают в одном и том же канале и соревнуются друг с другом и с точкой доступа за доступ к каналу. Станции в сетях 802.11ах могут, помимо EDCA, получать доступ к каналу, не соревнуясь, на основе расписаний, составляемых точкой доступа. Авторы предлагают обеспечить честность между станциями, работающими согласно стандарту 802.11ах и работающими согласно более старым версиям стандарта путем использования различных наборов параметров EDCA для разных типов станций. В данной работе используется подход к моделированию, похожий на тот, что был применен в [13], чтобы описать обычные станции, хотя в данной работе не рассматривается проблема честности распределения ресурсов.

5 Постановка задачи

Предложенная схема позволяет значительно уменьшить задержку доставки пакетов URLLC за счет уменьшения пропускной способности для обычных станций. В то же время, станции URLLC все еще должны соревноваться за доступ к каналу между собой, а потому важно определить интенсивность трафика URLLC, при которой требования на задержку и долю потерянных пакетов (англ. Packet Loss Ratio, PLR) для станций URLLC будут выполнены. Поэтому целью данной работы является оценка эффективности предложенной схемы доступа с точки зрения задержки доставки URLLC пакетов и доступной полосы пропускания для обычных станций, а также поиск границы применимости данной схемы.

Для достижения данной цели рассматриваются два сценария: приоритезированный и равноправный. В обоих случаях рассматривается сеть, состоящая из M станций URLLC и N обычных станций. При этом считается, что обычные станции работают в режиме насыщения, то есть у них всегда есть пакеты в очереди. Станции URLLC, напротив, после успешной передачи пакета получают следующий через временной промежуток, подчиняющийся экспоненциальному распределению с параметром λ . Таким образом, вероятность того, что новый пакет придет до того, как пройдет время t, равна:

$$F(t) = 1 - exp(-\lambda t). \tag{1}$$

Считается, что все станции находятся в зоне радиовидимости друг друга (т.е. нет проблемы скрытой станции).

В приоритезированном сценарии устройства используют предложенную схему, т.е. станции URLLC имеют приоритет при доступе к каналу. В равноправном сценарии обычные станции не прослушивают служебный канал и, следовательно, не реагируют на служебные сигналы занятости канала. Таким образом, они соревнуются за канал на равных правах со станциями URLLC.

Для обоих сценариев ставится задача разработки математической модели процесса передачи станций в сети, позволяющей при заданной интенсивности трафика URLLC и при заданном количестве станций обоих типов найти задержку при передаче кадров URLLC и пропускную способность, доступную для обычных станций. При помощи построенной модели требуется определить максимальную нагрузку на сеть, при которой задержка не превышает заданный порог 1 мс. Также требуется сравнить производительность сети в обоих сценариях и таким образом оценить эффективность предложенного подхода.

6 Аналитическая модель

Чтобы численно описать предложенный метод, ниже приводится математическая модель работы сети для обоих вышеупомянутых сценариев. Для того, чтобы описать поведение станций и найти задержку для станций URLLC и полосу пропускания обычных станций, используется подход к моделированию сетей Wi-Fi, похожий на представленный в [12].

6.1 Приоритезированный сценарий, станции URLLC.

Станции URLLC имеют исключительные права доступа к каналу, вследствие чего их поведение не зависит от обычных станций. Выберем случайную станцию URLLC и будем наблюдать ее состояние в начале передач пакетов и в моменты отсчёта её счетчика отсрочки. Интервал между такими событиями будем называть слотом. Когда очередь станции пуста, станция по определению пребывает в свободном (idle) состоянии. Когда станция получает пакет на передачу и при этом канал не занят, происходит спонтанная асинхронная передача, что приводит к переходу станции в состояние ST (определенное таким образом). Если же канал оказывается занятым, станция взводит счетчик отсрочки и начинает его обратный отсчет. Во время этого отсчета ее состояние описывается парой чисел (b, s), где b — текущее значение счётчика отсрочки, а s — текущая стадия отсрочки.

Чтобы описать вероятности перехода из одного состояния в другое, введём обозначения τ_s и τ_n — вероятности того, что в случайно выбранный слот станция совершает соответственно асинхронную и синхронную передачу. Предполагается, что эти вероятности равны для всех станций и не меняются во времени. Таким образом, вероятность τ того, что станция URLLC передает в случайно выбранный слот равна: $\tau = \tau_s + \tau_n$.

С точки зрения рассматриваемой станции, слоты могут быть пустыми, успешными или коллизионными. Пустым слот становится в случае, если ни одна из станций в его начале не начинает передачу, и вероятность этого

$$p_e = (1 - \tau)^{M-1}.$$
 (2)

Слот успешен, если в его начале ровно одна станция начинает передачу. Таким образом, его вероятность:

$$p_s = (M-1)\tau_n (1-\tau_n)^{M-2} + (M-1)\tau_s.$$
(3)

При этом делается предположение, что все асинхронные передачи успешны. Вероятность коллизионного слота считается как

$$p_c = 1 - p_e - p_s.$$
 (4)

Длительности пустого, успешного и коллизионного слота обозначены как σ, T^u_s и $T^u_c,$ соответственно.

Переход станции из свободного состояния (idle) к отсчету счетчика происходит в случае, когда во время успешного или коллизионного слота ей приходит пакет. Вероятность этого события:

$$\tilde{p}_{idl} = p_s F(T_s^u) + p_c F(T_c^u) \tag{5}$$

Похожим образом можно определить вероятность перехода станции из состояния idle в состояние ST:

$$p_{st} = p_e F(\sigma). \tag{6}$$

Пусть теперь станция находится в состоянии (s = i, b = k). Когда b превышает 1, станция уменьшает значение счетчика. Когда b становится нулем, станция пытается осуществить передачу, которая заканчивается коллизией с вероятностью

$$p = 1 - (1 - \tau_n)^{M-1}, \tag{7}$$

где M — число станций URLLC. Значит, с вероятностью 1 - p станция переходит в состояние idle, а с вероятностью p выбирает новую отсрочку, равномерно распределенную среди W_{s+1} возможных вариантов.

Пусть m — максимально возможная стадия счетчика, т.е. такая стадия, что $W_{max} = W_{min} \times 2^m$. Если, имея b = m, станция совершает неуспешную попытку передачи, инкремента b не происходит, оно остается равным m. Если обозначить Pr(state2|state1) — вероятность перехода из состояния state1 в состояние state2, то получатся следующие вероятности перехода марковской цепи:

,

$$\begin{cases}
Pr(i, k|i, k+1) = 1; k \in [0, ..., W_i - 2], i \in [0, ..., m] \\
Pr(i, k|i - 1, 0) = p/W_i; k \in [0, ..., W_i - 1], i \in [1, ..., m] \\
Pr(m, k|m, 0) = p/W_m; k \in [0, ..., W_m - 1] \\
Pr(idle|i, 0) = (1 - p); i \in [0, ..., m] \\
Pr(0, k|Idle) = \tilde{p}_{idl}/W_0; k \in [0, ..., W_0 - 1] \\
Pr(ST|Idle) = p_{st} \\
Pr(Idle|ST) = 1 \\
Pr(Idle|ST) = 1 \\
Pr(Idle|Idle) = 1 - \tilde{p}_{idl} - p_{st}
\end{cases}$$
(8)

Пусть $q_{i,k}$ — стационарное распределение цепи. Заметив, что

$$q_{i,0} = pq_{i-1,0} = p^2 q_{i-2,0} = \dots = p^i q_{0,0}, i \in [1, m-1]$$
(9)

$$q_{m,0} = p^m q_{0,0} / (1-p) \tag{10}$$

и что остальные вероятности могут бы выражены следующим образом:

$$q_{i,k} = \frac{W_i - k}{W_i} \begin{cases} p^i q_{0,0}; 0 \le i \le m - 1, \\ \frac{p^i}{1 - p} q_{0,0}; i = m, \end{cases}$$
(11)

$$q_{Idle} = \frac{q_{0,0}}{\tilde{p}_{idl}},\tag{12}$$

$$q_{ST} = q_{Idle} p_{st}.\tag{13}$$

можно, используя условие нормировки

$$q_{ST} + q_{Idle} + \sum_{i=0}^{m} \sum_{k=0}^{W_i - 1} q_{i,k} = 1,$$
(14)

найти q_{0,0}:

$$q_{0,0} = \left(\frac{1+p_{st}}{\tilde{p}_{idl}} + \frac{1}{2}\left[W_0\left(\sum_{i=0}^{m-1} (2p)^i + \frac{(2p)^m}{1-p}\right) + \frac{1}{1-p}\right]\right)^{-1}$$
(15)

Зная эту величину, можно найти вероятности передачи:

$$\tau_n = \sum_{i=0}^m q_{i,0} = \frac{q_{0,0}}{1-p},\tag{16}$$

$$\tau_s = q_{Idle} \tilde{p}_{st} = \frac{q_{0,0} p_{st}}{\tilde{p}_{idl}},\tag{17}$$

$$\tau = \tau_s + \tau_n = \tau_n \left(1 + \frac{(1-p) p_{st}}{\tilde{p}_{idl}} \right).$$
(18)

Получившуюся систему можно решить численными методами.

Найдём теперь среднюю задержку D при передаче кадров URLLC. Она состоит из двух частей:

$$D = D_1 + D_2. (19)$$

Здесь D_1 — средняя задержка для асинхронной передачи (равная минимальной задержке) и может быть приближенно подсчитана как $D_1 = p_{libre}T_s$, где p_{libre} — вероятность того, что в случайный момент времени канал пуст, т.е.

$$p_{libre} = \frac{p_e \sigma}{p_e \sigma + p_c T_c^u + p_s T s^u}.$$
(20)

Вторая часть $D_2 = (1 - p_{libre}) \times D'$, где D' – средняя задержка синхронной передачи. Определив среднюю длину виртуального слота, в котором рассматриваемая станция не пытается провести передачу, как

$$\hat{\sigma} = p_e \sigma + p_s T_s^u + p_c T_c^u. \tag{21}$$

можно заметить, что D' состоит, во-первых, из времени, которое станции придется ждать освобождения канала, а во-вторых, из коллизий, в которые станция попадает при неуспешных попытках передачи. Вероятность того, что станция попадет в *i* коллизий равна p^i и каждая коллизия длится T_c^u . При этом попытка номер *i* в среднем добавляет к задержке $\frac{\hat{\sigma}(W_i-1)}{2}$ — время, в течение которого станция отсчитывает счетчик. В конце концов, с вероятностью 1 - p станция передает успешно за время T_{succ}^u . Суммарная средняя задержка:

$$D' = \frac{T_s^u}{1 - e^{-\lambda T_s^u}} - \frac{1}{\lambda} + \sum_{i=0}^{\infty} p^i \left(\frac{\hat{\sigma}\left(W_i - 1\right)}{2} + (1 - p) T_s^u\right) + \sum_{i=1}^{\infty} p^i T_c^u = \frac{T_s^u}{1 - e^{-\lambda T_s^u}} - \frac{1}{\lambda} + T_s^u + \frac{T_c^u p - \frac{\hat{\sigma}}{2}}{1 - p} + \frac{W_0 \hat{\sigma}}{2} \left[\frac{1 - (2p)^m}{1 - 2p} + \frac{(2p)^m}{1 - p}\right]$$
(22)

Найдя среднюю задержку, можно приступить к поиску функции pacпределения задержки для станций URLLC.

С вероятностью p_{libre} передача станции будет асинхронной, а значит, будет длиться T_s^u . С вероятностью $1 - p_{libre}$ пакет будет сгенерирован тогда, когда канал занят другой передачей. Станция дождется конца слота и начнет отсчитывать счетчик отсрочки. Функция распределения этого времени ожидания может быть найдена как

$$F_{wait}(t) = \frac{\Pr\left(T_s^u - x < t | x < T_s^u\right)}{\Pr\left(x < T_s^u\right)} = \begin{cases} 0, & t < 0, \\ \frac{e^{\lambda(t - T_s^u)} - e^{-\lambda T_s^u}}{1 - e^{-\lambda T_s^u}}, & 0 \le t \le T_s^u, \\ 1, & t > T_s^u. \end{cases}$$
(23)

Пусть $d_i(e, s, c)$ — вероятность того, что течение *i*-ой передачи станция простаивает *e* пустых (с ее точки зрения), *s* успешных и *c* коллизионных слотов. Тогда

$$d_i(e, s, c) = \frac{1}{W_i} \frac{(e+s+c)!}{e!s!c!} p_e^e p_s^s p_c^c,$$
(24)

если $0 \le e + s + c < W_i - 1$ и 0 в противном случае. Далее, пусть $w_i(e, s, c)$ - вероятность того, что во время попыток передач с 0 по i станция ожидает e, s и c соответственно пустых, успешных и коллизионных слотов. В таком случае

$$w_0(e, s, c) = d_0(e, s, c),$$

$$w_i(e, s, c) = \sum_{e_i, s_i, c_i} d_i(e_i, s_i, c_i) \times w_{i-1}(e - e_i, s - s_i, c - c_i), i > 0,$$
(25)

а функция распределения задержки $F_D(t)$ записывается как

$$F_D(t) = \begin{cases} 0, & t < T_s^u, \\ p_{libre} + (1 - p_{libre}) \times F_{D'}(t - T_s^u), & t > T_s^u, \end{cases}$$
(26)

где $F_{D'}(t)$ — функция распределения времени, проведенного в состоянии синхронной передачи:

$$F_{D'}(t) = \sum_{i=0}^{\infty} p^i (1-p) \sum_{e,s,c} w_i(e,s,c) \times F_{wait}(t - \sigma e - sT_s^u - cT_c^u).$$
(27)

6.2 Приоритезированный сценарий, обычные станции

Аналогично случаю со станциями URLLC, рассмотрим случайную обычную станцию. Обозначим за τ_r вероятность того, что в начале случайного слота обычная станция делает попытку передачи.

Вероятность того, что передача является неуспешной равна:

$$p_r = 1 - (1 - \tau_r)^{N-1} e^{-M\lambda T_s^r}, \qquad (28)$$

где T_s^r — длительность успешной передачи для обычных станций. Здесь учтен тот факт, что, передача обычной станции становится неудачной (коллизионной), если хотя бы одна станция URLLC в это время передает пакет.

Поскольку обычные станции работают в режиме насыщения, τ_r может быть найдена более простым путем, а именно как среднее число попыток передачи на один кадр, деленное на среднее время, необходимое для успешной передачи кадра:

$$\tau_r = \frac{\sum_{i=0}^{\infty} p_i^i}{\sum_{i=0}^{m} (\frac{W_i - 1}{2} + 1)p_i^i + \sum_{i=m+1}^{\infty} (\frac{W_m - 1}{2} + 1)p_i^i} = \frac{2(1 - 2p_r)}{(1 - 2p_r)(W_0 + 1) + p_r W_0 (1 - (2p_r)^m)}.$$
(29)

Решив систему из уравнений 28 и 29, можно получить τ_r .

Чтобы найти пропускную способность обычных станций нужно учесть, что на их передачу в значительной степени влияют станции URLLC: даже если слот обычной станции начался как пустой или успешный, он может быть прерван сигналом занятости канала и таким образом стать коллизионным (с точки зрения рассматриваемой обычной станции).

Пусть p_{empty} — вероятность того, что ни одна из обычных станций не начинает передачу в случайно выбранный слот. Тогда

$$p_{empty} = (1 - \tau_r)^N. \tag{30}$$

Как было упомянуто выше, пустой слот может быть прерван сигналом занятости канала, в результате чего канал будет занят до тех пор, пока станции URLLC не передадут свои кадры. Средняя длина такого слота равна

$$\langle \sigma \rangle = \sigma \left(1 - F(\sigma) \right)^M + D \left(1 - \left(1 - F(\sigma) \right)^M \right), \tag{31}$$

где D — задержка станций URLLC, о которой говорилось ранее. Пусть p_{succ} — вероятность того, что в случайно выбранный слот лишь одна обычная станций совершает попытку передачи:

$$p_{suc} = N\tau_r \left(1 - \tau_r\right)^{N-1}.$$
 (32)

Он закончится успешно лишь в том случае, если ни одна из станций URLLC не начнет передачу до его окончания. Вероятность такого исхода:

$$p_{S,S} = p_{suc} \left(1 - F \left(T_s^r \right) \right)^M.$$
(33)

Длина такого слота — T_s^r .

 T^{r}

Если в течение слота приходит пакет URLLC, потенциально успешный слот завершается коллизией. Вероятность этого равна

$$p_{S,C} = p_{suc} \left(1 - \left(1 - F(T_s^r) \right)^M \right).$$
(34)

Средняя длительность такого слота может быть приближена выражением

$$\langle D_{S,C} \rangle = \frac{\int\limits_{0}^{T_s} M\lambda t e^{-M\lambda t} dt}{1 - e^{-M\lambda T_s^r}} + D = \frac{1}{M\lambda} - \frac{T_s^r e^{-M\lambda T_s^r}}{1 - e^{-M\lambda T_s^r}} + D, \qquad (35)$$

где первое слагаемое — среднее время, прошедшее до того, как сигнал занятости канала прервал передачу кадра.

Коллизии можно разделить на два типа. Обыкновенные коллизии (между пакетами обычных станций) длятся T_c^r и случаются с вероятностью

$$p_{C,R} = (1 - p_{empty} - p_{suc}) \left(1 - F(T_c^r)\right)^M.$$
(36)

Как и успешные передачи, обыкновенные коллизии могут прерваны станциями URLLC, что случается с вероятностью

$$p_{C,BT} = (1 - p_{empty} - p_{suc}) \left(1 - (1 - F(T_c^r))^M \right).$$
(37)

Средняя длина такой коллизии:

$$\langle D_{C_BT} \rangle = \frac{1}{M\lambda} - \frac{T_c^r e^{-M\lambda T_c^r}}{1 - e^{-M\lambda T_c^r}} + D, \qquad (38)$$

что аналогично уравнению (35), только вместо T_s^r используется T_c^r . Таким образом, средняя длина коллизионного слота дается выражением

$$\langle T_{coll} \rangle = p_{S,C} \langle D_{S,C} \rangle + p_{C,BT} \langle D_{C,BT} \rangle + p_{C,R} T_c^r, \tag{39}$$

в то время как пропускная способность приближается формулой

$$S = \frac{p_{S,S}L}{\langle \sigma \rangle p_{empty} + T_s^r p_{S,S} + \langle T_{coll} \rangle},\tag{40}$$

где L — размер кадра. Далее будет показано, что это приближение справедливо, когда плотность потока пакетов URLLC низка.

6.3 Равноправный сценарий

В этом сценарии обычные станции не реагируют на сигналы занятости канала, что приводит к равноправной конкуренции за канал между ними и станциями URLLC.

Как и раньше, p и τ описывают станции URLLC, а $p_r \tau_r$ — обычные. Если выбрать случайную обычную станцию, то вероятность того, что случайный слот будет пустым с ее точки зрения при условия, что сама станция не передает, будет

$$p_{er} = (1 - \tau_r)^{N-1} (1 - \tau)^M.$$
(41)

Аналогичная вероятность с точки зрения случайной станции URLLC:

$$p_{eu} = (1 - \tau_r)^N (1 - \tau)^{M-1}.$$
(42)

Аналогичным образом описываются остальные вероятности, описывающие ситуацию с точки зрения выбранной станции (индекс *s* для успешных слотов, *с* — для коллизий)

$$p_{su} = (M-1)\tau_n (1-\tau_n)^{M-2} (1-\tau_r)^N + (M-1)\tau_s + N\tau_r (1-\tau_r)^{N-1} (1-\tau_n)^{M-1},$$
(43)

$$p_{sr} = (N-1)\tau_r (1-\tau_r)^{N-2} (1-\tau_n)^{M} + \frac{1}{2} \sum_{n=1}^{N-1} \frac{1}{2} \sum_{n=1}^{N$$

$$+ M\tau_n (1 - \tau_n)^{m-1} (1 - \tau_r)^{n-1} + M\tau_s, \tag{44}$$

$$p_{cu} = 1 - p_{su} - p_{eu},\tag{45}$$

$$p_{cr} = 1 - p_{sr} - p_{er}. (46)$$

Коллизии делятся на две группы: с участием обычных станций и без них. С точки зрения станции URLLC, вероятность коллизии без участия обычных станций:

$$p_{cnu} = (1 - \tau_r)^N (1 - (1 - \tau)^{M-1} - (M - 1)\tau_n (1 - \tau_n)^{M-2} - (M - 1)\tau_s),$$
(47)

а вероятность коллизии с участием как минимум одной обычной станции:

$$p_{cru} = 1 - (1 - \tau_r)^N - N\tau_r (1 - \tau_r)^{N-1} (1 - \tau)^{M-1}.$$
(48)

Полная вероятность коллизии вычисляется как $p_{cu} = p_{cnu} + p_{cru}$. Похожим образом можно записать $p_{su} = p_{snu} + p_{sru}$, где p_{snu} и p_{sru} — вероятности успешной передачи для URLLC и обычных станций соответственно. В этих обозначениях нужно переопределить \tilde{p}_{idl} и p_{st} :

$$p_{st} = p_{eu}F(\sigma),\tag{49}$$

$$\tilde{p}_{idl} = p_{snu}F(T_s^u) + p_{sru}F(T_s^r) + p_{cnu}F(T_c^u) + p_{cru}F(T_c^r),$$
(50)

а также вероятность того, что синхронная передача станции URLLC станет коллизией:

$$p = 1 - (1 - \tau_r)^N \left(1 - \tau_n\right)^{M-1}, \tag{51}$$

Та же вероятность для обычных станций:

$$p_r = 1 - (1 - \tau_r)^{N-1} (1 - \tau_n)^M.$$
(52)

Таким образом получается та же цепь Маркова для станций URLLC, что была описана системой (8).

Учитывая сделанные переопределения, уравнения для τ из предыдущего раздела, до сих пор справедливы. Однако, чтобы найти его численное значение, придется решить систему уравнений (16), (29), (51) и (52), так как вероятности попасть в коллизию зависят как от τ_n , так и от τ_r .

Зная эти вероятности, можно найти среднюю задержку для станций URLLC и пропускную способность для обычных методами, аналогичными тем, что были применены в приоритезированном сценарии. Так, пропускная способность будет определена формулой

$$S = \frac{p_{sru}L}{\langle \sigma \rangle p_{empty} + T_c^u p_{su} + T_s^r p_{sr} + T_c^r p_{cr} + T_c^u p_{cu}} = p_{sru}L/\langle \tilde{\sigma} \rangle, \quad (53)$$

где $p_{empty}=(1-\tau)^M(1-\tau_r)^N,\,\langle\tilde{\sigma}\rangle$ — средняя длина слота, а средняя задержка станций URLLC находится как

$$\langle D \rangle = p_{libre} T_s^u + (1 - p_{libre}) [\langle \tilde{\sigma} \rangle \frac{W_0 - 1}{2} + (1 - p) T_s^u + p (T_c^u + \langle \tilde{\sigma} \rangle \frac{W_1 - 1}{2} + (1 - p) T_s^u + p (\dots + p) (1 - p) T_s^u + p (\dots + p) (1 - p) T_s^u + p (\dots + p) (1 - p$$

где $p_{libre} = p_{empty}\sigma/\langle \tilde{\sigma} \rangle$. Последняя формула имеет простую вероятностную интерпретацию: либо пакет передается сразу, асинхронно, либо начинается отсчет счетчика в диапазоне $[0, W_0 - 1]$. Далее с вероятностью (1 - p) пакет успешно передается, а с вероятностью p случается коллизия, взводится счетчик уже в большем диапазоне, и т.д.

6.4 Границы применимости

Как будет показано далее, предложенная модель дает хорошие результаты либо для низкоинтенсивного трафика URLLC, либо для высокоинтенсивного, практически эквивалентного режиму насыщения. Ниже приведён вывод границы, в которых данная модель хорошо описывает реальную сеть.

Когда интенсивность трафика низка, большинство передач URLLC асинхронны, а значит, имеют минимальную задержку. По мере возрастания интенсивности, доля асинхронных передач уменьшается, т.е τ_s уменьшается в то время как τ_n растет. Нижнюю границу λ_{min} можно оценить как интенсивность, при которой они равны:

$$\tau_s(M, \lambda_{min}) = \tau_n(M, \lambda_{min}), \tag{55}$$

и тогда для всех $\lambda < \lambda_{min}$ модель дает хорошую оценку средней задержки.

Если интенсивность продолжит увеличиваться, станции передут полностью в синхронный режим. Верхняя граница λ_{max} может быть оценена как



Рис. 2. Зависимость задержки пакетов URLLC от числа станций URLLC. Интенсивность поступления пакетов: $\lambda=100c^{-1}$

интенсивность, при которой за время передачи пакета в среднем 2 станции URLLC генерируют новый пакет. Учитывая экспоненциальное распределение времени генерации пакета, это дает:

$$M\lambda_{max}T_s^u = 2 \to \lambda_{max} = \frac{2}{MT_s^u} \tag{56}$$

При интенсивности трафика λ_{max} и выше, станции URLLC начинают передавать свои пакеты в синхронном режиме и соревноваться за канал, а за время передачи другие станции URLLC с большой вероятностью генерируют новые кадры и присоединяются к соревнованию за канал.

Уравнения 55 и 56 могут быть использованы для того, чтобы найти подходящее число станций URLLC для данной интенсивности трафика.

7 Численные результаты

Далее приведены результаты имитационного моделирования и их сравнение с результатами, полученными при помощи аналитической модели.

При имитационном моделировании станции передавали данные в канале шириной 20 МГц на сигнально-кодовой конструкции MCS 5, обеспечивающей скорость передачи данных 25 Мбит в секунду. Станции URLLC передавали кадры размером 200 байт, а обычные станции — размером 1 Кбайт.



Рис. 3. Зависимость пропускной способности обычных станций от числа станций URLLC

Станции находились на таком расстоянии от точки доступа, что заданная сигнально-кодовая конструкция обеспечивала надёжную передачу данных, то есть передачи могли быть неуспешными только из-за коллизий кадров от разных станций.

На рис. 2 представлена зависимость средней задержки пакетов URLLC от числа станций URLLC для обоих сценариев.

Как видно, наделение станций URLLC исключительными правами доступа к каналу значительно (более, чем на порядок) уменьшает время доставки при условии, что интенсивность потока пакетов URLLC относительно низка. Так же можно наблюдать резкий рост задержки в равноправном сценарии, начиная с некоторого количества станций. Это можно объяснить тем, что при определенном количестве станций URLLC, начинается постоянная конкуренция за канал даже между ними, поскольку оказывается, что всегда есть станция, желающая передавать. В равноправном сценарии такого скачка нет, так как на протяжении всей кривой станции URLLC передают, преимущественно соревнуясь за канал с обычными станциями, то есть станции *сразу* находятся в условиях конкуренции.

На рис. 3 представлена зависимость пропускной способности обычных станций от числа станций URLLC. В приоритезированном сценарии на определенном этапе станции URLLC занимают своими передачами все время, и потому пропускная способность обычных станций обращается в нуль. В рав-



Рис. 4. Функция распределения задержки URLLC для разного числа станций

ноправном сценарии такого не происходит, пропускная способность плавно убывает с добавлением новых претендентов на канал.

На рис. 4 представлена функция распределения задержки доставки пакетов URLLC для различных чисел станций URLLC. Видно, что большинство станций передает асинхронно, то есть с наименьшей возможной задержкой. С другой стороны, если станция передает синхронно, то время доставки пакета может колебаться в довольно широких пределах и даже выходить за желаемую границу в 1 мс. Из этого можно сделать вывод, что при жестких ограничениях на задержку данный метод стоит использовать только для небольшого числа станций URLLC или же для коротких пакетов.

Анализируя получившиеся результаты, можно заключить, что при малой интенсивности прихода пакетов URLLC предложенный метод очень эффективен: незначительно понижая пропускную способность обычных станций, он дает возможность станциям URLLC передавать гораздо быстрее. Тем не менее, при росте интенсивности URLLC-трафика, применение данной схемы может привести к полному прекращению передач обычных станций, при этом, с некоторого момента выигрыш во времени доставки пакетов URLLC перестает быть настолько значительным. Таким образом, целесообразность применения описанного метода зависит от параметров сети.

Наконец, на рис. 5 приведена зависимость пропускной способности обычных станций от интенсивности трафика URLLC, а также указаны границы сходимости модели.


Рис. 5. Зависимость пропускной способности обычных станций от интенсивности трафика URLLC

8 Заключение

В данной работе представлено исследование метода обеспечения сверхнадёжной связи с низкой задержкой (URLLC) в беспроводных сетях Wi-Fi. Данный метод состоит в использовании дополнительного надёжного и низкоскоростного радиоинтерфейса для того, чтобы уведомлять устройства в сети о необходимости передачи срочных кадров и таким образом организовать приоритетный доступ к каналу для передающих устройств.

Предложенный метод оказался эффективным в условиях низкой интенсивности поступления пакетов URLLC, значительно сокращая их время доставки при малых потерях пропускной способности обычных станций.

Описанная схема позволяет доставлять срочные пакеты с очень малой задержкой (менее 1 мс), что невозможно в стандартных сетях Wi-Fi вследствие того, что, согласно действующим протоколам, устройства не могут передавать данные, если канал уже занят другими станциями; отсутствует механизм прерывания менее приоритетной передачи. Тем не менее, построенные функции распределения задержки доставки показывают, что при большом количестве станций URLLC задержка может колебаться в весьма пироких пределах, что может быть критичным для некоторых приложений.

Разработанная модель также показывает существование максимальной нагрузки, превышение которой ведет к потере возможности обеспечить требуемую малую задержку доставки. В зависимости от параметров конкретной сети, эта задержка может быть найдена с помощью представленной аналитической модели.

Список литературы

- 1. Framework and overall objectives of the future development of IMT for 2020 and beyond : Recommendation : M.2083 / ITU-R : 2015. September.
- 2. Ultra Relible Low Latency Cellular Networks: Usecases, Challenges and Approaches / Chen He, Abbas Rana, Cheng Peng et al. $-\,2006.$
- 3. J. Vihriälä et al. Numerology and frame structure for 5G radio access // 2016 IEEE 27th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC). -2016. Sept. P. 1–5.
- 4. NR; Overall description; Stage-2 : Rep. : TS 38.300 / 3GPP : 2017. December.
- 5. Wi-Fi Time Sensitive Networking. Access mode: https://mentor.ieee.org/802.11/dcn/17/11-17-1734-01-0wng-wtsn.pptx.
- 6. IEEE P802.11baTM Draft Standard for Information Technology— Telecommunications and information exchange between systems— Local and metropolitan area networks— Specific requirements Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications — Amendment 9: Wake-Up Radio Operation, 2018. — March.
- 7. Kehrer Stephan, Kleineberg Oliver, Heffernan Donal. A Comparison of Fault-Tolerance Concepts for IEEE 802.1 Time Sensitive Networks (TSN) // Emerging Technology and Factory Automation (ETFA), 2014 IEEE / IEEE. 2014. P. 1–8.
- 8. Morteza Hashemi Farzaneh, Alois Knoll. Time-Sensetive Networking: an experimental setup. IEEE 2017.
- 9. Scheduling Real-Time Communication in IEEE 802.1 Qbv Time Sensitive Networks / Silviu S Craciunas, Ramon Serna Oliver, Martin Chmelík, Wilfried Steiner // Proceedings of the 24th International Conference on Real-Time Networks and Systems / ACM. - 2016. - P. 183–192.
- 10. IEEE Standard for Ethernet : Rep. / IEEE : 2012. December.
- Koubaa Anis, Alves Mário, Tovar Eduardo. i-GAME: an Implicit GTS Allocation Mechanism in IEEE 802.15.4 for Time-Sensitive Wireless Sensor Networks // Real-Time Systems, 2006. 18th Euromicro Conference on / IEEE. – 2006. – P. 10–pp.
- 12. Bianchi Giuseppe. Performance Analysis of the IEEE 802.11 Distributed Coordination Function // IEEE Journal on selected areas in communications. 2000. Vol. 18, no. 3. P. 535–547.
- Khorov Evgeny, Loginov Vyacheslav, Lyakhov Andrey. Several EDCA Parameter Sets for Improving Channel Access in IEEE 802.11ax Networks // Wireless Communication Systems (ISWCS), 2016 International Symposium on / IEEE. – 2016. – P. 419–423.

Приближенная оценка пропускной способности специального класса векторных дизъюктивных каналов*

Смешко А.А.^{1,2}, Иванов Ф.И.¹

¹ Институт проблем передачи информации им. А.А. Харкевича Российской академии наук, Москва, Россияя

 $^2\,$ Московский физико-технический институт, Москва, Россия

Аннотация В данной работе рассматривается векторный дизъюктивный канал, в котором пользователи передают пакеты, состоящие из L бит. Рассмотрена пропускная способность данного канала и её асимптотическое приближение. Для данной оценки были получены численные результаты при различных параметрах канала, а также при различном числе активных пользователей.

1 Введение

В настоящее время ввиду быстрого развития концепций интернета вещей и умного дома одной из ключевых особенностей беспроводных сетей является резкое увеличение числа пользователей в системе. Следовательно, взаимное влияние пользователей друг на друга стало более ощутимым. В то же время требования, которые предъявляются к вероятности успешной передачи и к скорости передачи данных, постоянно растут.

В связи с вышесказанным, очень большое число пользователей ведет некоординированную передачу в беспроводных сетях. Среди общего числа пользователей следует особо выделить подмножество из U пользователей, которые активны в любой момент времени, и каждый из них ведет передачу коротких пакетов информации на базовую станцию. При этом базовая станция восстанавливает список сообщений без идентификации пользователя, который передал конкретное сообщение. Некоординированный способ передачи, а также короткая длина передаваемых сообщений существенно влияют на методы построения сигнально-кодовых конструкций, которые были бы практически применимыми в данных условиях.

Практически все широко известные сигнально-кодовые конструкции, используемые в классических многопользовательских каналах (например, на основе кодового разделения CDMA), разделения скоростей [1], чередования [2] предполагают координирование передачи между пользователями. Также параметры таких схем зависят от числа пользователей.

^{*} Работа выполнена при поддержке гранта 18-37-00319 мол_а

В данной работе рассматривается специальный класс некоординированного многопользовательского канала – векторного дизъюктивного канала, известного также, как Z-канал. Данный канал был предложен Cohen, Heller и Viterbi в работе [3]. Свойства такого канала исследуются в работах [4]-[9].

Мы рассматриваем специальный класс векторного-дизъюктивного канала, где каждый пользователь передает некоторый двоичный вектор длины $L \geq 1$. Мы получили оценку на пропускную способность данного канала, учитывая то, что векторы длины L могут не быть равновероятными. Полученная величина является функцией параметров канала и числа активных пользователей.

2 Модель канала

В данной главе мы рассмотрим специальный векторный дизъюктивный канал и исследуем аналитически его свойства.

2.1 Описание канала

Рассмотрим векторный дизъюнктивный канал, который состоит из Q > 1 независимых подканалов (частот), которые мы будем называть элементарными подканалами. Предположим, что имеется U > 0 пользователей, которые используют канал для передачи. Пусть Q = LS, т.е. многопользовательский канал разделен на S независимых каналов, каждый из которых состоит из L элементарных подканалов. Будем называть такие "обобщенные" каналы L-каналами или слотами. В произвольный момент времени τ каждый i-ый пользователь выбирает произвольный j-ый L-канал $\left(1 \le j \le \frac{Q}{L}\right)$ и передает в нем блок из L бит. Для каждого слота на выходе получаем дизъюнкцию всей переданной информации (векторы длины L).

Рассмотрим случай передачи двух пользователей *i* и *k*, которые выбрали общий слот для передачи в *j*-момент времени. Предположим, что пользователи *i* и *k* передают двоичные векторы длиной *L*: $\mathbf{X}_{ij} = \left(x_{ij}^{(1)}, x_{ij}^{(2)}, \ldots, x_{ij}^{(L)}\right)$

и $\boldsymbol{X}_{ik} = \left(x_{ik}^{(1)}, x_{ik}^{(2)}, \dots, x_{ik}^{(L)}\right)$ соответственно в одном и том же *L*-канале. Ввиду предположения о том, что канал является дизъюнктивным, то в общем слоте на приемник придет: $\boldsymbol{X}_{ij} \vee \boldsymbol{X}_{kj}$.

Например, пусть L = 3 и два пользователя выбирают один общий слот для передачи. Если первый пользователь передает вектор (0,1,0) и второй – (0,1,1), тогда на выходе будет $(0,1,0) \vee (0,1,1) = (0,1,1)$.

Очевидно, что максимальное количество пользователей, которые могут передавать ортогонально в данных условиях, равно S.

2.2 Модель коллизии

Рассмотрим модель передачи, описанную выше. В такой модели коллизии могут иметь место в том случае, когда два или более пользователей выбирают для передачи один и тот же *L*-канал в один момент времени.

Обозначим через κ случайную величину – кратность коллизии. Данная величина может быть рассмотрена как количество активных пользователей, которые выбрали заданный *j*-ый *L*-канал, в котором ведет передачу рассматриваемый пользователь. Очевидно, что $0 \le \kappa \le U - 1$. Кроме того, легко посчитать вероятность события $\kappa = t$ (построить распределение случайной величины κ):

$$P(\kappa = t) = {\binom{U-1}{t}} \left(\frac{1}{S}\right)^t \left(1 - \frac{1}{S}\right)^{U-t-1}$$

В данной работе мы рассматриваем бесшумный канал и некоординированную передачу, поэтому для заданного пользователя передача остальных U – 1 пользователей будут рассматриваться как шум (источник коллизий).

В некоторых случаях передача других пользователей в том же слоте, где ведет передачу заданный пользователь, не приводит к коллизии. Например, если пользователь передает (1, 1, ..., 1), то такая информация будет принята корректно независимо от того, что передают другие пользователи в том же слоте.

В наиболее общем случае на приемнике не будет наблюдаться коллизия для *i*-ого пользователя, если бинарный вектор, передаваемый пользователем, покрывает все векторы, которые передаются в том же *L*-канале.

3 Анализ свойств канала

Предыдущие рассуждения могут быть обобщены для произвольной кратности коллизии. Для лучшего понимания предложенного метода рассмотрим пример канала, представленны на рис.1:

На рис.1 представлены все возможные переходы между входами и выходами канала. Как можно заметить, для входного вектора длины L и веса wсуществует 2^{L-w} возможных переходов в выходные векторы. Таким образом количество переходов и вероятность перехода из входа в выход зависит только от веса входного вектора и не зависит от его конкретного представителя(носителя). Более того, можно заметить, что вес выходного вектора увеличится по сравнению с весом входного вектора тогда и только тогда, когда при передаче произошла коллизия. Вообще говоря, входы не являются равновероятными, как и выходы. Наш канал с 2^L входами и 2^L выходами можно представить как канал с L + 1 входом и L + 1 выходом, которые соответствуют весам входных и выходных векторов(см. рис. 2).

Распределение на входе канала имеет вид:

$$P_{in}(w) = \binom{L}{w} p^w (1-p)^{L-w}$$

Обозначим через $[L]=\{1,2,\ldots,L\}$ и через $supp({\boldsymbol X})$ носитель вектора ${\boldsymbol X},$ т.е.

$$supp(\mathbf{X}) = \{i : x_i \neq 0\}$$
.



Рис. 1. Возможные переходы для L = 3



Рис. 2. Возможные переходы для L = 3

Очевидно, что входное распределение не зависит от кратности коллизии κ . В тоже время выходное распределение зависит от этой величины. Предположим, что $\kappa = t$. Если входой вектор \boldsymbol{u} имеет вес w < L, то выходной вектор \boldsymbol{v} имеет вес w' > w, если хотя бы один из t пользователей (которые используют тот же слот для передачи, что и выбранный пользователь) передает 1 на любой из позиций из множества $[L] \setminus supp(\boldsymbol{u})$. Вероятность p(t) того, что хотя бы один из t пользователей передает 1 на любой позиции $k \in [L] \setminus supp(\boldsymbol{u})$ равна:

$$p(t) = 1 - (1 - p)^t$$
,

где p(1) = p вероятность передачи 1.

Легко посчитать условные вероятности P(w'|w,t) того, что мы получим выходной вектор веса w' для заданного входного вектора веса w, если произошла коллизия порядка t:

$$P(w'|w,t) = \begin{cases} 0, & w' < w \\ \binom{L-w}{w'-w} p(t)^{w'-w} (1-p(t))^{L-w'}, & w' \ge w \end{cases}.$$

Наконец, распределение выхода канала следующее:

$$P_{out}(w'|t) = \sum_{i=0}^{w'} P(w'|i,t) P_{in}(i) \; .$$

После ряда упрощений и преобразований получим:

$$P_{out}(w'|t) = \binom{L}{w'} (1-p)^{(t+1)(L-w')} (1-(1-p)^{t+1})^{w'}.$$

4 Оценка пропускной способности

Оценим пропускную способность канала, описанного выше. Вначале мы получим оценку на пропускную способность для заданной кратности коллизии $\kappa = t$, а затем рассмотрим математическое ожидание данной величины при всех возможных кратностях коллизии, получив таким образом нижнюю границу на пропускную способность векторного дизъюктивного канала, представленного в данной работе.

Вначале запишем определение пропускной способности канала. Пусть \mathcal{X} и \mathcal{Y} – случайные величины, описывающие вход и выход канала соответственно. Пропускная способность канала \mathcal{C} определяется следующим образом:

$$\mathcal{C} = \sup_{p_{in}(\mathcal{X})} I(\mathcal{X}, \mathcal{Y}),$$

где $I(\mathcal{X}, \mathcal{Y})$ - взаимная информация между \mathcal{X} и \mathcal{Y} . Супремум берется по всем возможным распределениям $p_{in}(\mathcal{X})$ случайной величины \mathcal{X} .

Если кратность коллизии фиксирована $\kappa = t$, то пропускная способность C(t) канала для коллизии кратности t оценивается как:

$$C(t) \ge \sum_{w=0}^{L} P_{in}(w) \sum_{w'=w}^{L} P(w'|w,t) \log_2 P(w'|w,t) - \sum_{w''=0}^{L} P_{out}(w''|t) \log_2 P_{out}(w''|t)$$
(1)

Таким образом, пропускная способность C оценивается снизу математическим ожиданием величины C(t) по всем возможным кратностям коллизии t:

$$\mathcal{C} \ge C^* = \sum_{t=0}^{U-1} P(\kappa = t) \mathcal{C}(t) .$$
⁽²⁾

4.1 Асимптотическая оценка пропускной способности

В данном разделе получим приближенную формулу для оценки снизу пропускной способности, так как формула (2) для C^* сложная и громоздкая

для анализа. Для начала запишем формулу для энтропии биномиального распределения, с помощью которой оценим пропускную способность канала $\mathcal{C}(t)$:

$$-\sum_{i=0}^{n} \binom{n}{i} p^{i} (1-p)^{n-i} \log_{2} \left(\binom{n}{i} p^{i} (1-p)^{n-i} \right) =$$
$$= \frac{1}{2} \log_{2} (2\pi enp(1-p)) + O\left(\frac{1}{n}\right) , \qquad (3)$$

Тогда подставив (3) в формулу (1) для C(t) получим оценку снизу:

$$\begin{aligned} \mathcal{C}(t) \geq &-\sum_{w=0}^{L} \binom{L}{w} p^{w} (1-p)^{L-w} \frac{1}{2} \log_{2}(L-w) + \\ &+ \frac{1}{2} \log_{2} \left(L(1-p) \frac{1-(1-p)^{t+1}}{1-(1-p)^{t}} \right) \;. \end{aligned}$$

Устремим tк бесконечности и так как $\frac{1-(1-p)^{t+1}}{1-(1-p)^t} \to 1$ при $t \to \infty,$ то:

$$\mathcal{C}(\infty) \ge -\sum_{w=0}^{L} {\binom{L}{w}} p^{w} (1-p)^{L-w} \frac{1}{2} \log_2(L-w) + \frac{1}{2} \log_2(L(1-p))$$

Пользуясь тем, что $\mathcal{C}(t) \to 0$ при $t \to \infty$, получим:

$$\sum_{w=0}^{L} {\binom{L}{w}} p^{w} (1-p)^{L-w} \frac{1}{2} \log_2(L-w) = \frac{1}{2} \log_2(L(1-p)) .$$

Тогда оценка будет выглядеть следующим образом:

$$\mathcal{C}(t) \ge \frac{1}{2} \log_2 \left(\frac{1 - (1 - p)^{t+1}}{1 - (1 - p)^t} \right)$$

Следовательно, оценка пропускной способности имеет вид:

$$C^* \ge \frac{1}{2} \sum_{t=0}^{U-1} \binom{U-1}{t} \left(\frac{1}{S}\right)^t \left(1 - \frac{1}{S}\right)^{U-t-1} \log_2\left(\frac{1 - (1-p)^{t+1}}{1 - (1-p)^t}\right) \ .$$

Далее, можем воспользоваться тем, что биномиальное распределение переходит в пуассоновское (при $U \to +\infty$ и $\frac{1}{S} \to 0$ так, что $\frac{U-1}{S} \to \xi$, где ξ конечно). Таким образом, придем к следующей оценке пропускной способности:

$$C^* \ge \frac{e^{-\xi}}{2ln(2)} \sum_{t=0}^{U-1} \frac{\xi^t}{t!} \ln\left(\frac{1-(1-p)^{t+1}}{1-(1-p)^t}\right) \ .$$

Если оценить сверху натуральный логарифм, как $\ln(1+x) \geq \frac{x}{1+x},$ то придем к еще одной оценке:

$$C^* \ge \frac{pe^{-\xi}}{2\ln(2)} \sum_{t=0}^{U-1} \frac{(\xi(1-p))^t}{t!(1-(1-p)^{t+1})} \ .$$

Далее пользуемся оценкой снизу знаменателя и тем, что U очень велико. Тогда пользуясь рядом для экспоненты $\sum_{t=0}^{+\infty} \frac{\lambda^t}{t!} = e^{\lambda}$, получим окончательную оценку:

$$C \ge C^* \ge e^{-\xi} \left(1 - \frac{p}{2\ln(2)}\right) + \frac{e^{-\xi p}p}{2\ln(2)}.$$

5 Численные результаты

Для фиксированного $Q = 2^{15}$ проверим с помощью построения графиков справедливость наших приближений. Выберем наиболее естественную вероятность передачи 1: p(1) = 0.5 и посмотрим на результаты при различных параметрах канала(см. рис. 3)



Рис. 3. Зависимость нижней границы пропускной способности C^* от числа активных пользователей U для p(1) = 0.5 при различных параметрах канала

Как видим, оценка на пропускную способность больше для больших S. Так как нижнюю оценку пропускной способности C^* можно рассматривать как долю информации, которая может быть передана корректно, то имеет смысл ввести "скорость передачи данных" в виде $T = C^*L$. Данная величина определяет долю корректно переданных бит из общего числа L бит, которые пользователь отправляет на приемник через канал. На рис. 4 представлена зависимость между нижней границей на "скорость передачи данных" T и числом активных пользователей U для фиксированного $Q = 2^{15}$ и вероятности передачи 1 в канал p(1) = 0.5 при различных параметрах канала.



Рис. 4. Зависимость скорости передачи данных T от числа активных пользователей U для p(1) = 0.5 при различных параметрах канала

Можно заметить, что для некоторых диапазонов числа пользователей для меньших значений S наблюдаются большие значения T. То есть, чтобы увеличить скорость передачи данных T при относительно небольшом количестве пользователей, имеет смысл увеличить число L, тем самым уменьшая S. Для больших значений числа пользователей, наоборот, имеет смысл увеличивать значение S.

6 Заключение

В данной работе рассматривается векторный дизъюктивный канал с неравновероятными входами, в котором пользователи передают на приемник двоичные векторы длины L. Была получена оценка на пропускную способность такого канала и приближенная формула, хорошо описывающая характер зависимости пропускной способности от числа пользователей и может быть применима при любой вероятности передачи в канал 1 на конкретной позиции – т.е. при любом *p*. Также была рассмотрена "скорость передачи данных" *T* и получены численные результаты при различных параметрах канала.

Список литературы

- B. Rimoldi and R. Urbanke, "A rate-splitting approach to the Gaussian multipleaccess channel," IEEE Transactions on Information Theory, vol. 42, no. 2, pp. 364-375, 1996.
- L. Ping, L. Liu, K. Wu and W. K. Leung, "Interleave division multiple-access," IEEE Transactions on Wireless Communications, vol. 5, no. 4, pp. 938-947, 2006.
- A. R. Cohen, J. A. Heller and A. J. Viterbi, "A new coding technique for asynchronous multiple access communication," IEEE Transactions on Communication Technology, vol. 19, pp. 849-855, October 1971.
- D.S. Osipov, A.A.Frolov and V.V.Zyablov, "Multiple access system for a vector disjunctive channel," Problems of Information Transmission, vol. 48, no. 3, pp. 243-249, 2012.
- A.A.Frolov, V.V.Zyablov, V.R.Sidorenko and R.Fischer, "On a multiple-access in a vector disjunctive channel," Information Theory (ISIT), Proceeding of IEEE International Symposium on Information Theory (ISIT), pp. 211-215, 2013.
- L.Wilhelmsson and K.S.Zigangirov, "On the Asymptotic Capacity of a Multiple-Access Channel," Problems of Information Transmission, vol. 33, no. 1, pp. 9-16, 1997.
- S.C.Chang and J.Wolf, "On the t-User m-Frequency Noiseless Multiple-Access Channel with and without Intensity Information," IEEE Transaction on Information Theory, vol. 27, no. 1, pp. 41-48, Jan. 1981.
- L.A.Bassalygo and M.S.Pinsker, "Evaluation of the Asymptotics of the Summarized Capacity of an m-Frequency t-User Noiseless Multiple-Access Channel," Problems of Information Transmission, vol. 36, no. 2, pp.91-97, 2000.
- A.Han Vinck and K.Keuning, "On the Capacity of the Asynchronous t-User m-Frequency Noiseless Multiple-Access Channel without Intensity Information," IEEE Transactions on Information Theory, vol.42, no. 6, pp. 2235-2238, Nov. 1996.

Моделирование передачи информации о чрезвычайных ситуациях в сетях IEEE 802.11ah при использовании окна ограниченного доступа *

И.А. Наседкин, Р.Р. Юсупов

ИППИ РАН nasedkin.ia@phystech.edu, yusupov@iitp.ru

Аннотация Интернет вещей в настоящий момент является быстроразвивающейся областью науки и техники. Примером сценария Интернета вещей является использование полей датчиков для извещения о чрезвычайных ситуациях. Стандарт IEEE 802.11ah создан для удовлетворения нужд Интернета вещей. В стандарте введен механизм ограничения конкуренции, называемый окном ограниченного доступа. В статье рассматривается использование этого механизма в сценарии передачи сообщений о чрезвычайных ситуациях. Предложен новый метод доступа, являющийся модификацией метода EDCA, при использовании которого станция совершает одну попытку передачи. Построена аналитическая модель этого метода, проведено сравнение со стандартным. Согласно полученным результатам предложенный метод доступа обеспечивает задержку и энергопотребление не выше, чем при использовании EDCA.

Keywords: IEEE 802.11ah, окно ограниченного доступа

1 Введение

Изначально сети Wi-Fi проектировались для обеспечения беспроводного доступа к сети Интернет небольшому числу устройств, передающих большой объем данных. Развитие Интернета вещей привело к возникновению сценариев, когда к одной точке доступа подключается большое число устройств. В 2017 году был опубликован [1] стандарт IEEE 802.11ah, удовлетворяющий нужды Интернета вещей. Сети IEEE 802.11ah поддерживают более 6000 станций, одновременно подключенных к точке доступа.

Одним из сценариев Интернета вещей является использование множества сенсоров для мониторинга. Сенсоры должны отправлять сообщения при возникновении чрезвычайных ситуаций, причем одно событие может активировать передачу данных одновременно с группы сенсоров. Возникает

^{*} Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 18-07-01356 а

необходимость доставить информацию о произошедшем событии с наименьшей задержкой. Нужно защитить станции, которые передают информацию о чрезвычайном событии от интерференции с другими станциями. Для этого может быть использован механизм окна ограниченного доступа (Restricted Access Window – RAW), введенный в стандарте IEEE 802.11ah.

Механизм RAW является механизмом ограничения конкуренции за доступ к каналу, при использовании которого точка доступа назначает временные интервалы группам станций. Внутри RAW станции используют EDCA, который основан на методе доступа CSMA/CA с двоичным экспоненциальным окном. Поскольку в рассматриваемом сценарии станциям необходимо осуществить доставку кадра данных от любой станции из группы, вместо стандартного метода EDCA можно использовать его упрощенную модификацию без повторных попыток передач. В данной статье строится аналитическая модель предложенного метода доступа и проводится сравнение со стандартным.

Дальнейшее изложение построено следующим образом. В разделе 2 приведено краткое описание механизма RAW, подробнее данный механизм рассмотрен в [2]. В разделе 4 выполнен краткий обзор предыдущих работ. В разделе 3 представлена постановка задачи. В разделе 5 кратко описана аналитическая модель для предложенного метода доступа. В разделе 6 приводятся численные результаты.

2 Окно ограниченного доступа

Основным предназначением RAW является уменьшение числа станций, пытающихся одновременно получить доступ к каналу путем разделения доступа к каналу по времени. Для этого точка доступа выбирает подмножество станций, которые будут передавать в RAW. Затем это подмножество разбивается на группы, каждой из которых назначается временной интервал внутри RAW, называемый RAW-слот. Станциям запрещено передавать в чужих RAW-слотах и пересекать границу RAW-слота.

Разделение станций на группы и назначение им RAW-слотов уменьшает конкуренцию за доступ к каналу. Однако, точка доступа обычно не знает заранее, сколько станций имеют кадры для передачи. В таком случае назначение RAW-слотов отдельно каждой станции приводит к излишнему потреблению канальных ресурсов и уменьшению пропускной способности. Эффективнее разбивать станции на группы, содержащие больше одной станции, в то время как только у некоторых из них будут кадры для передачи в начале RAW-слота. В частности, стандарт предполагает, что при разбиении станций на группы точка доступа может учитывать типы станций, ограничения мощности, характер трафика при группировании станций. Оценив число станций, имеющих данные для передачи, и объем необходимого канального ресурса, точка доступа определяет длительность RAW и момент его начала относительно периодически рассылаемых бикон-кадров (beacon). Точка доступа периодически рассылает все параметры RAW в биконкадрах, позволяя станциям определить, к какой группе они принадлежат и когда начнется назначенный RAW-слот. Станциям разумно переходить в спящий режим на все время, кроме назначенных RAW-слотов.

Внутри своих RAW-слотов для доступа к каналу станции используют EDCA (Enchanced Distributed Channel Access). Перед началом передачи станция прослушивает канал и не передает до тех пор, пока канал не освободится. Кроме того, чтобы уменьшить вероятность коллизии сразу после освобождения канала, EDCA использует двоичный экспоненциальный счетчик отсрочки.

В начале RAW-слота станция генерирует случайное целое значение счетчика отсрочки из интервала $[0, CW_0)$, где CW_0 – начальный размер конкурентного окна. Затем станция прослушивает канал. Каждый раз, когда канал свободен на протяжении времени $T_{\rm e}$, станция уменьшает счетчик отсрочки на единицу. Если канал занят, станция замораживает счетчик отсрочки до тех пор, пока канал не будет свободен некоторое время. Это время равно AIFS, если станция получила кадр успешно, либо EIFS, если станция получила кадр успешно, либо EIFS, если станция не смогла декодировать кадр.

Когда счетчик отсрочки достигает нулевого значения, станция совершает попытку передачи. Если передача кадра будет пересекать границу RAWслота, то станция не передает и может переключиться в спящий режим. В ином случае станция передает кадр и ждет ACK от точки доступа. Если ACK получен в течение T_{ACK} , станция считает, что кадр передан успешно. Если ACK получен не был и счетчик повторных попыток передач не равен RL, станция начинает следующую попытку передачи. Перед *i*-ой попыткой передачи, станция генерирует новое целое значение счетчика отсрочки из интервала $[0, CW_{i-1})$, где размер конкурентного окна CW_i определяется следующим образом:

$$CW_{i} = \begin{cases} CW_{0}, & i = 0, \\ \min(2CW_{i-1}, CW_{\max}), & i > 0. \end{cases}$$

В начале RAW-слота станция устанавливает минимальный размер конкурентного окна, сбрасывает значения счетчика повторных попыток передач и регенерирует значение счетчика отсрочки.

3 Постановка задачи

Рассматривается группа из M станций, передающих в одном RAW-слоте длительности $T_{\rm R}$, следующим с периодом $T_{\rm per}$, причем доля канального времени, которую занимает RAW-слот, постоянна, и рассматривается только восходящий трафик. Важна успешная передача хотя бы одного кадра. Примем следующие допущения:

1. Кадр возникает в случайный момент времени в промежутке между RAW

- 2. Длительность RAW мала по сравнению с периодом следования RAW, во время RAW станция не генерирует новых кадров для передачи
- 3. Длительность передачи кадра для всех станций одинакова
- 4. Передача может быть неуспешной только в результате интерференции с другими станциями

Подобный сценарий может быть использован для моделирования передачи данных в сенсорных сетях с большим числом станций, в которых выделена группа сенсоров, предназначенных для информирования о возможных чрезвычайных ситуациях. Для доставки кадра данных, содержащем информацию о чрезвычайной ситуации станции по умолчанию используют стандартный метод доступа EDCA. Однако, поскольку важна успешная доставка кадра данных хотя бы от одной станции, возможно, имеет смысл применять модификацию метода доступа, при использовании которой станции совершают одну попытку передачи.

Необходимо построить аналитическую модель, которая позволит рассчитать вероятность хотя бы одной успешной передачи в RAW-слоте, задержку до первой успешной передачи и среднее энергопотребление в слоте, а также сравнить две политики передачи: с повторными попытками передачи при неуспешной передаче и без повторных попыток.

Построенная таким образом модель позволит определить оптимальные политику передач, длительность RAW-слота и начальный размер конкурентного окна.

4 Предыдущие работы

В статье [3] описана математическая модель механизма доступа Distributed Coordination Function (DCF), которая может быть модифицирована для описания механизма EDCA с одной категорией доступа. Однако, в данной статье рассматривается насыщенный трафик в стационарном режиме, что не подходит для исследуемого сценария. Кроме того, модель из статьи [3] нельзя применять для описания сценариев с использованием окна ограниченного доступа, так как при использовании RAW не выполняется условие стационарности за счет реинициализации счетчика отсрочки в начале RAW-слота и ограниченной длительности RAW-слота.

В статье [4] используется модификация подхода из статьи [3] для анализа описанного ими метода Group Synchronized DCF, основанного на механизме окна ограниченного доступа. Построенная в этой статье модель сверяется с результатами имитационного моделирования, полученными в [5]. Однако, для исследуемого сценария использовать модель из [4] некорректно, так как она основана на модели из [3], неприменимость которой уже обоснована выше.

В статье [6] рассмотрена модель передачи данных в RAW-слоте, но при ее описании использовалось предположение о постоянстве вероятности коллизии на протяжении всего RAW-слота, что возможно только в стационарном режиме. Однако отсутствие стационарности обосновано ранее. В статье [7] рассмотрена модель, которая достаточно точно описывает передачу трафика в RAW-слоте и может быть использована для получения необходимых результатов, однако не подходит для описания системы с RL = 1, так как в ней не учитываются станции, не совершившие успешной передачи, но больше не конкурирующие за доступ к каналу из-за достижения счетчиком попыток повторных передачи значения RL. При RL > 1данное упрощение несильно влияет на точность, но при RL = 1 после каждого коллизионного слота как минимум две станции больше не пытаются получить доступ к каналу, что сильно снижает конкуренцию.

5 Модель

5.1 Вероятность успеха и средняя задержка при отсутствии повторных попыток передачи

При отсутствии повторных передач при неуспешной передаче поведение системы в RAW-слоте определяется начальными значениями счетчиков отсрочки всех станций. Каждая станция независимо от других с равной вероятностью генерирует значение начального счетчика отсрочки в интервале [0, CW_0). Рассмотрим пространство равновероятных исходов, где каждый исход – последовательность начальных значений счетчиков отсрочки всех станций. Тогда общее число исходов:

$$N_{\text{total}} = \left(CW_0\right)^M.$$

Найдем число исходов N(v, k, c) таких, что первая успешная передача произошла в виртуальном слоте с номером k (нумерация с нуля) с c станциями, попавшими в v коллизий до первой успешной передачи:

$$N(v,k,c) = MC_k^v C_{M-1}^c V(v,c) (CW_0 - k - 1)^{M-1-c}$$

Действительно, M способами можно выбрать станцию, которая передает в первом успешном слоте, $(CW_0 - k - 1)^{M-1-c}$ – число способов разместить станции после первой успешной передачи, C_{M-1}^c – число способов выбрать c станций, которые попадут в v коллизий до успешной передачи, которые можно расположить по k слотам до передачи C_k^v способами. V(v,c) – число способов разместить c станций по v слотам так, чтобы в каждом слоте возникла коллизия.

Выражение для V(v, c) было получено в [8]:

$$V(v,c) = \begin{cases} 0, & \text{если } v < 2c, \\ 1, & \text{если } v = 1, c > 2, \\ v^c - \sum_{y=1}^{v-1} C_v^y V(y,c) - \\ - \sum_{u=1}^{v-1} C_v^u \frac{c!}{(c-u)!} \sum_{y=1}^{v-u} C_{v-u}^y V(y,c-u), & \text{иначе.} \end{cases}$$

Используя N(v, k, c) найдем число исходов L(v, k) таких, что первая успешная передача произошла в виртуальном слоте с номером $k \, c \, v$ коллизионными слотами до этого слота:

$$L(v,k) = \sum_{c=2v}^{M-1} N(v,k,c).$$

Учтем, что RAW-слот имеет ограниченную длительность $T_{\rm R}$, пустой виртуальный слот, слоты с коллизией и с успешной передачей – длительности $T_{\rm e}$, $T_{\rm c}$ и $T_{\rm s}$ соответственно. Тогда число исходов N_{succ} таких, что успешная передача произошла в пределах RAW-слота:

$$N_{\text{succ}} = \sum_{\substack{v \le k < CW_0 \\ (k-v)T_e + vT_e + T_s \le T_R}} L(v,k).$$

Вероятность наличия успешной передачи в RAW-слоте:

$$P_{\rm succ} = \frac{N_{\rm succ}}{N_{\rm total}}.$$

Таким образом, среднее число полных RAW-слотов, которое пройдет до RAW-слота с успешной передачей:

$$N_{\rm RAW} = \frac{1 - P_{\rm succ}}{P_{\rm succ}}.$$

Средняя задержка до первой успешной передачи внутри RAW-слота с успешной передачей:

$$T_{\text{delay}} = \frac{1}{N_{succ}} \sum_{\substack{v \le k < CW_0 \\ (k-v)T_e + vT_c + T_s < T_{\text{B}}}} ((k-v)T_e + vT_c)L(v,k).$$

Если $T_{\rm per}$ – период, с которым следуют RAW-слоты, то средняя задержка до первой успешной передачи в предположении, что станция может сгенерировать кадр в любой момент между RAW-слотами:

$$T_{\rm td} = N_{\rm RAW}T_{\rm per} + T_{\rm delay} + \frac{T_{\rm per}}{2}.$$

5.2 Среднее энергопотребление в RAW-слоте при отсутствии повторных попыток передачи

Для расчета среднего энергопотребления в RAW-слоте рассматривается произвольно выбранная фиксировання станция. Рассчитывается среднее число пустых, успешных и коллизионных слотов, произошедших до передачи этой станции, а также вероятность передачи данной станции в RAW-слоте. Для этого вводится марковская цепь с состояниями (t, c, s, v), где t – число прошедших виртуальных слотов, c – число прошедших коллизионных слотов, s – число прошедших слотов с успешной передачей, v – число станций, больше не конкурирующих за доступ к каналу. Тогда $T = cT_{\rm c} + sT_{\rm s} + (t - c - s)T_{\rm e}$ – время, прошедшее с начала RAW-слота. Начальное распределение цепи Маркова:

$$P(0, c, s, v) = \begin{cases} 1, & (0, c, s, v) = (0, 0, 0, 0) \\ 0, & \text{иначе} \end{cases}$$

Если выбранная станция совершает попытку передачи, либо оканчивается RAW-слот, цепь переходит в поглощающее состояние, причем необходимо проверять, уместилась ли передача станции до окончания RAW-слота. Тогда, если $I_{\rm p}$ – индикатор того, что передача данной станции успеет завершиться до окончания RAW-слота, то потребленная выбранной станцией энергия может быть рассчитана следующим образом:

$$Q = (t - c - s)Q_{\text{idle}} + cQ_{\text{c}} + sQ_{\text{s}} + I_{\text{p}}Q_{\text{tx}},$$

где $Q_{\rm idle}, Q_{\rm c}, Q_{\rm s}$ и $Q_{\rm tx}$ – энергии, потребляемые при прослушивании пустого, коллизионного, успешного слота и при передаче кадра соответственно.

Для получения среднего энергопотребления достаточно просуммировать значения потребленной энергии по всем поглощающим состояниям цепи для данной станции.

Вероятность, что станция имеет назначенную передачу в виртуальном слоте t и переходные вероятности цепи Маркова могут быть рассчитаны с использованием подхода из [7].

Данная модель может быть также применена для вычисления вероятности наличия успешной передачи в RAW-слоте и задержки до первой успешной передачи, но имеет большую вычислительную сложность по сравнению с моделью, рассмотренной в разделе 5.1.

5.3 Модель для политики с повторными попытками передачи

Для политики с повторными передачами аналитическая модель подробно описана в [7]. В данной статье рассмотрены два процесса: процесс A и процесс Б.

Процесс А описывает поведение произвольно выбранной станции и является марковской цепью с состояниями (t, c, s, r), где t – число прошедших виртуальных слотов с начала RAW-слота, c и s – число прошедших коллизионных и успешных слотов, r – значение счетчика повторных попыток передачи.

Процесс Б описывает поведение станций в совокупности и является марковской цепью с состояниями (t, c, s), где t, c и s имеют тот же смысл, что и в процессе А. Для расчета вероятности успешной передачи и средней задержки до первой успешной передачи был использован процесс Б с поглощающими состояниями (t, c, 1), которые соответствуют первой успешной передаче в RAWслоте. Для расчета энергопотребления был использован процесс A без изменений.

6 Численные результаты

6.1 Параметры модели

Рассматривались RAW-слоты длительностью $T_{\rm R}$, следующие с периодом $T_{\rm per}$. При расчете параметров считалось, что станции не используют RTS/CTS, так что $T_{\rm s} = T_{\rm c}$. Кроме этого, было введено упрощение: энергопотребление при прослушивании слотов с успешной и коллизионной передачами считалось одинаковым и равным $Q_{\rm busy}$. Энергия, затрачиваемая при передаче, также считалась независимой от успешности передачи и равной $Q_{\rm tx}$. Значения для потребляемой энергии взяты из [9].

В Таблице 1 приведены параметры модели, использованные для получения численных результатов.

$T_{\rm e}$	$52 \mu s$	$T_{\rm s}$	$1064 \mu s$
$T_{\rm c}$	$1064 \mu s$	M	64
RL	7	CW_0	64
CW_{\max}	4096	$Q_{\rm idle}$	$0.2288 \mu J$
$Q_{\rm busy}$	$4.6816 \mu J$	$Q_{\rm tx}$	$5.656 \mu J$
$T_{\rm R}$	$5000 \mu s$	$T_{\rm per}$	$50000 \mu s$
Таблица 1. Параметры модели			

6.2 Численные результаты

Проведем проверку точности построенной модели, а также сделаем анализ полученных результатов. При моделировании значения всех параметров устанавливались в соответствии с Таблицей 1, за исключением варьируемых. Если значения менялись, это указано отдельно.

На рис. 1–3 построены зависимости вероятности наличия успешной передачи $P_{\rm succ}$ в RAW-слоте, средней задержки $T_{\rm td}$ до первой успешной передачи и среднего энергопотребления соответственно для политик без повторных передач (RL = 1) и с повторными передачами (RL = 7) при различном числе передающих станций и при неизменных остальных параметрах. Построенная в разделе 5.1 модель достаточно точно описывает зависимость при RL = 1, как и модель из [7] при RL = 7.

Как видно из рис. 1, при достаточно большом числе станций вероятность наличия успешной передачи в RAW-слоте уменьшается. Это связано с тем, что уменьшается вероятность попадания в один виртуальный слот только одной станции, а также с ограниченным числом передач, которые умещаются в RAW-слот. Кроме того, если передает всего две станции, то вероятность наличия успешной передачи падает, это происходит из-за того, что при попадании сразу двух станций в один виртуальный слот обе станции попадают в коллизию, что при RL = 1 приводит к тому, что в RAW-слоте не будет успешных передач. Если RL > 1, то после неуспешной передачи происходит увеличение размера конкурентого окна, поэтому следующая попытка передачи также может не уместиться в RAW-слот из-за большого числа прошедших пустых слотов.

При малом числе станций при уменьшении числа станций происходит увеличение задержки до первой передачи, так как до первой попытки передачи проходит большое число пустых слотов. Затем, за счет уменьшения вероятности наличия успешной передачи в RAW-слоте при увеличении числа станций увеличивается средняя задержка до первой успешной передачи, что и показано на рис. 2. Кроме того, при большем числе передающих станций произвольная фиксированная станция вынуждена дольше прослушивать канал, в котором происходят передачи других станций, что увеличивает энергопотребление (рис. 3).

На рис. 4–6 построены зависимости вероятности наличия успешной передачи $P_{\rm succ}$ в RAW-слоте, средней задержки $T_{\rm td}$ до первой успешной передачи и среднего энергопотребления соответственно для политик без повторных передач (RL = 1) и с повторными передачами (RL = 7) при различной длительности RAW-слота, причем период следования RAW-слотов также менялся и всегда был равен $T_{\rm per} = 10T_{\rm R}$.

Из рис. 4 заключаем, что при увеличении $T_{\rm R}$ вероятность наличия успешной передачи в RAW-слоте ступенчато увеличивается при увеличении длительности RAW-слота, причем расстояние между скачками равно длительности виртуального слота с передачей. Это связано с тем, что с увеличением длительности в RAW-слот умещается больше виртуальных слотов с передачей, что резко повышает вероятность наличия успешной передачи.

Зависимость средней задержки $T_{\rm td}$ до первой успешной передачи выглядит пилообразно (рис. 5). Постепенное увеличение связано с увеличением периода $T_{\rm per}$ следования RAW-слотов при практически неизменной вероятности наличия успешной передачи, а локальные минимумы объясняются резким увеличением вероятности наличия успешной передачи в отдельных RAW-слотах (рис. 4).

Потребляемая мощность меняется скачкообразно в зависимости от длительности RAW-слота, однако наблюдается тенденция к уменьшению с увеличением $T_{\rm R}$. Длительная тенденция связана с тем, что большее число станций совершает попытку передачи до окончания RAW-слота и переходит в спящий режим до его завершения, причем при RL = 1 уменьшение потребляемой мощности выражено больше, чем при RL = 7, так как если RL > 1, то при неудачной попытке передачи станция не переходит в спящий режим, а продолжает конкуренцию за доступ к каналу. Локальные максимумы вызваны увеличением числа виртуальных слотов с передачей, которые помещаются в RAW-слот, и соответствуют локальным минимумам средней задержки до первой успешной передачи (рис. 5).

В целом, политика передачи без повторных попыток передач имеет большую вероятность наличия успешной передачи в RAW-слоте, меньшую задержку до первой успешной передачи и меньшее энергопотребление. Это связано с тем, что станции, попавшие в коллизию, переходят в спящий режим и перестают конкурировать за доступ к каналу.

7 Заключение

В работе была исследована группа станций, передающих информацию об одном событии в окне ограниченного доступа. Была построена модель, позволяющая найти вероятность наличия хотя бы одной успешной передачи в RAW-слоте, задержку до первой успешной передачи, среднее энергопотребление, а также сравнить две политики передачи: с повторными попытками при неудачной передаче и без.

По результатам можно сделать вывод, что помимо того, что политика с RL = 1 проще в реализации, чем с RL > 1, при ее использовании получается несколько меньшая задержка до первой успешной передачи и меньшее энергопотребление.

Список литературы

- 802.11ah-2016 IEEE Standard for Information technology–Telecommunications and information exchange between systems - Local and metropolitan area networks– Specific requirements - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 2: Sub 1 GHz License Exempt Operation. – 2017.
- A survey on IEEE 802.11 ah: An enabling networking technology for smart cities / Evgeny Khorov, Andrey Lyakhov, Alexander Krotov, Andrey Guschin // Computer Communications. - 2015. - Vol. 58. - P. 53–69.
- 3. Bianchi Giuseppe. Performance analysis of the IEEE 802.11 distributed coordination function // IEEE Journal on selected areas in communications. -2000. Vol. 18, no. 3. P. 535–547.
- Performance analysis of group-synchronized DCF for dense IEEE 802.11 networks / Lei Zheng, Minming Ni, Lin Cai et al. // IEEE Transactions on Wireless Communications. - 2014. - Vol. 13, no. 11. - P. 6180–6192.
- 5. Performance analysis of IoT-enabling IEEE 802.11 ah technology and its RAW mechanism with non-cross slot boundary holding schemes / Ali Hazmi, Behnam Badihi, Anna Larmo et al. // World of Wireless, Mobile and Multimedia Networks (WoWMoM), 2015 IEEE 16th International Symposium on a / IEEE. 2015. P. 1–6.

- Performance evaluation of IEEE 802.11 ah and its restricted access window mechanism / Orod Raeesi, Juho Pirskanen, Ali Hazmi et al. // Communications Workshops (ICC), 2014 IEEE International Conference on / IEEE. – 2014. – P. 460–466.
- Khorov Evgeny, Krotov Alexander, Lyakhov Andrey. Modelling machine type communication in IEEE 802.11 ah networks // Communication Workshop (ICCW), 2015 IEEE International Conference on / IEEE. 2015. P. 1149-1154.
- 8. Study of beaconing in multihop wireless PAN with distributed control / Vladimir M Vishnevsky, Andrey I Lyakhov, Alexander A Safonov et al. // IEEE Transactions on Mobile Computing. 2008. Vol. 7, no. 1. P. 113-126.
- A 4mW-RX 7mW-TX IEEE 802.11 ah fully-integrated RF transceiver / Ao Ba, Kia Salimi, Paul Mateman et al. // Radio Frequency Integrated Circuits Symposium (RFIC), 2017 IEEE / IEEE. – 2017. – P. 232–235.



Рис. 1. Зависимость вероятности наличия успешной передачи *P*_{succ} в RAW-слоте от числа передающих станций.



Рис. 2. Зависимость средней задержки $T_{\rm td}$ до первой успешной передачи от числа передающих станций.



Рис. 3. Зависимость средней потребляемой мощности от числа передающих станций.



Рис. 4. Зависимость вероятности наличия успешной передачи *P*_{succ} в RAW-слоте от длительности RAW-слота.



Рис. 5. Зависимость средней задержки $T_{\rm td}$ до первой успешной передачи от длительности RAW-слота.



Рис. 6. Зависимость средней потребляемой мощности от длительности RAW-слота.

Исследование механизма Target Wake Time в сетях Wi-Fi*

Е.А. Степанова, Д.В. Банков, Е.М. Хоров

{stepanova, bankov, khorov}@iitp.ru

ИППИ РАН, НИУ ВШЭ, МФТИ

Аннотация В современном мире все большую популярность получают такие устройства, как датчики, счетчики и сенсоры, которые особенны тем, что генерируют ненасыщенные потоки данных. Для удобства их использования необходимо, чтобы эти устройства обладали следующими свойствами: передавали данные по беспроводному каналу связи и являлись энергонезависимыми. Для обеспечения энергонезависимости сенсоров можно использовать механизмы экономии энергии, представленные в стандарте Wi-Fi. Одним из таких механизмов является Target Wake Time (TWT), основанный на передаче данных согласно расписанию и переходе в режим пониженного энергопотребления в оставшееся время. Главной проблемой этого механизма является такое явление, как дрейф часов, из-за которого передачи устройств перестают четко подчиняться расписанию. Это может вызывать коллизии передач сенсоров и, как следствие, их повышенное энергопотребление. В этой статье мы исследовали влияние дрейфа часов на эффективность механизма TWT и разработали рекомендации для настройки параметров механизма TWT.

1 Введение

В наши дни в каждом доме появляется все больше устройств, передающих информацию по беспроводному каналу связи, но не подключенных к сети электропитания. Примерами таких устройств являются некоторые современные счетчики (газа, воды или электроэнергии) или датчики (движения или влажности). Одна из главных проблем таких устройств заключается в том, что при постоянной интенсивной работе их батареи быстро разряжаются. Следует отметить, что большая часть энергии такими устройствами тратится именно на прием и передачу информации, а не на ее считывание. Замена батарей финансово не выгодна, а при установке аккумуляторов значительно увеличивается стоимость обслуживания. Кроме того специфика самих датчиков не позволяет делать их зависимыми от сети электропитания, так как они могут быть установлены в подвалах, погребах или на улице.

^{*} Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 18-07-01356 а

Именно поэтому очень важно использовать различные механизмы сохранения энергии для таких устройств.

В настоящее время технология Wi-Fi является одной из наиболее популярных технологий беспроводной передачи данных. По оценкам компании Cisco [1] к 2021 году половину всего трафика данных будут составлять передачи при помощи этой технологии. Именно поэтому в качестве механизма беспроводной передачи информации далее рассматриваем только технологию Wi-Fi. В стандарте Wi-Fi IEEE 802.11 описана возможность перехода станции в спящий режим, который является режимом с пониженным потреблением энергии. Оно снижается за счет того, что станция не передает информацию и не прослушивает канал. Так как датчики генерируют ненасыщенный трафик, одно измерение в час и даже реже, то переход в спящий режим в оставшееся время является хорошим решением проблемы энергосбережения. Однако этот режим не решает проблемы траты энергии на повторные попытки передачи пакета, если произошла коллизия. Она может произойти, если в сети есть большое число датчиков, генерирующих данные в одно и то же время, или станции, работающие в режиме насыщения. Это приводит к тому, что передачи датчиков следует разносить во времени.

В дополнении IEEE 802.11ах к стандарту Wi-Fi представлен механизм TWT (англ. Target Wake Time, назначенное время пробуждения), позволяющий реализовать временное разделение передач от различных датчиков. Суть механизма заключается в том, что станции могут договориться заранее о времени обмена информацией. К примеру, точка доступа будет ждать передачу от датчика каждый день в 9:00. В таком случае на время передачи датчика можно зарезервировать канал, чтобы передача прошла без коллизий. Для того, чтобы настроить механизм TWT, две станции обмениваются специальными кадрами, содержащими параметры механизма, после чего устанавливают TWT SP (англ. TWT Service Period, период обслуживания TWT) — временной интервал, в течение которого станция должна проснуться для приема или передачи информации. Главная проблема этого механизма заключается в так называемом эффекте дрейфа часов, так как часы и у датчиков, и у точки доступа не являются идеальными. Стандарт Wi-Fi разрешает часам устройства по прошествии временного интервала отклоняться не более чем на 100 ppm = 0.01% этого интервала. Таким образом максимальное отклонение от назначенного времени пробуждения может быть равно $\delta T = T * 200 \ ppm$, где T — это интервал между пробуждениями одного датчика. Если T = 1 час, то $\delta T = 3600 * 200 * 10^{-6} = 0.72$ сек. Для успешной передачи датчика следует зарезервировать $2\delta T$ канального времени, что значительно скажется на производительности сети. Уменьшить снижение производительности сети можно путем увеличения числа станций, передающих в течение временного интервала $2\delta T$, но это в свою очередь увеличит вероятность коллизий между передачами сенсоров. Пока коллизий в канале не слишком много, их успешно может разрешить механизм случайного доступа к каналу. В связи с этим возникает задача выбора параметров механизма TWT таким образом, чтобы, доставляя наибольшее число пакетов, сенсоры тратили как можно меньше канального времени и собственной энергии.

Дальнейшее изложение построено следующим образом. В разделе 2 кратко описаны механизм TWT и метод случайного доступа к каналу. Описание объекта исследования и постановка задачи подробно представлены в разделе 3. В разделе 4 приведен обзор работ, посвященных механизму TWT и эффекту дрейфа часов. Результаты исследования эффективности механизма TWT в зависимости от собственных параметров и параметров метода случайного доступа к каналу приведены в разделе 5. Раздел 6 содержит заключение.

2 Методы доступа к каналу и энергосбережения в сетях Wi-Fi

Метод доступа устройств к каналу. В сетях Wi-Fi большинство кадров передаётся согласно методу случайного доступа к каналу EDCA (англ. Enhanced Distributed Channel Access, улучшенный распределённый доступ к каналу), в том числе в течение TWT SP в сценарии, исследуемом в данной работе. EDCA работает следующим образом. Каждая станция имеет очередь кадров на передачу. Когда кадр поступает в пустую очередь, станция прослушивает канал и, если канал свободен, передает этот кадр. Если канал занят, станция взводит счетчик отсрочки, который инициализируется целым числом, выбираемым равновероятно из промежутка от 0 до CW_r , где CW_r – конкурентное окно, а r — счетчик повторных попыток передачи. Изначально rравен нулю, и CW_0 равно параметру CW_{min} (по умолчанию 15). Когда канал занят, счетчик отсрочки замораживается. Если же канал освободился и пробыл свободным в течение интервала AIFS, счетчик отсрочки размораживается. Значение счетчика отсрочки уменьшается на единицу каждый раз, когда канал пробыл свободным в течение промежутка времени σ . Когда значение счетчика отсрочки становится равным нулю, станция передает кадр. Станция, получившая кадр, должна через интервал времен
иSIFSотправить подтверждение АСК станции-отправителю. Получив АСК, станция считает кадр успешно переданным и переходит к следующему кадру в очереди, если таковой имеется. Если же АСК не пришел в течение интервала AckTimeout после отправки кадра, станция считает попытку передачи неуспешной. В этом случае станция увеличивает счетчик попыток передачи и выбирает новое значение счетчика отсрочки в новом окне, которое задается следующей формулой:

$$CW_r = \begin{cases} CW_{min} & r = 0, \\ \min\{2(CW_{r-1} + 1) - 1, CW_{max}\} & r > 0, \end{cases}$$

где CW_{max} — максимальное значение конкурентного окна, по умолчанию равное 1023. Если счетчик повторных попыток передачи достигает ограничения RL (англ. Retry Limit, ограничение на количество повторов передачи), станция отбрасывает кадр.

Чем больше число станций одновременно пытается передать кадры, тем больше вероятность коллизии и тем большее время требуется для того, чтобы их разрешить. Механизм ТWT позволяет любой поддерживающей его станции регулировать количество станций, одновременно получающих доступ к каналу, и таким образом увеличивать вероятность доставки кадров данных и уменьшать затраты энергии и канальных ресурсов сенсорами.

TWT. Теперь опишем подробнее сам механизм TWT. Как уже было сказано ранее, для того чтобы настроить механизм TWT, две станции должны обменяться специальными кадрами. Эти кадры содержат информацию о согласии или отказе в установлении TWT, а при согласии параметры самого механизма, о которых более подробно рассказано далее.

До начала настройки TWT станции разделяются на станцию, запрашивающую TWT, и станцию, отвечающую на TWT. Это происходит автоматически и определяется тем, какая станция первая отправляет другой запрос на установление TWT. От этого разделения зависит настройка механизма. Также она зависит от параметра Flow type и типа установки следующего TWT ("implicit" или "explicit").

Параметр Flow type может принимать только два значения: 0 или 1. Если Flow type = 0, то отвечающая станция может начать передачу в TWT SP только после получения от запрашивающей станции специального менеджмент кадра (PS Poll или триггер кадра в зависимости от того, является ли запрашивающая станция точкой доступа или нет), именно после получения которого можно выполнять передачу кадра данных. Если же Flow type = 1, то нет ограничений на то, когда отвечающая станция должна начинать отправку кадров.

Тип установки TWT implicit означает, что следующий TWT SP строго периодически назначает запрашивающая станция. Тип TWT explicit означает, что следующий TWT SP устанавливает <u>отвечающая</u> станция, причем может назначать как периодически, так и апериодически.

3 Сценарий и постановка задачи

Рассмотрим сеть Wi-Fi, состоящую из точки доступа и некоторого фиксированного количества подключенных к ней сенсоров, использующих для передачи информации точке доступа механизм TWT. Сенсоры считаются подключившимися и установившими TWT, то есть обменявшимися всей вспомогательной информацией и назначившими TWT SP, в нулевой момент времени, то есть перед началом моделирования. Пусть интервалы между последовательно назначенными TWT различных станций равны μ , а временное отклонение часов сенсоров от назначенного точкой доступа TWT распределено нормально с дисперсией σ^2 .

Параметры механизма TWT настроены следующим образом. Так как точка доступа является отвечающей и не генерирует трафик для станций, то значение параметра Flow type не влияет на работоспособность сети. Но если у точки доступа могут появиться данные для передачи сенсору, то следует устанавливать Flow type = 0, так как в этом случае не возникнет ситуации, при которой точка доступа пытается передать данные спящей станции. Используется тип TWT explicit, так как отвечающей является точка доступа, которая налаживает передачу всех сенсоров.

В данном сценарии сенсоры находятся в спящем режиме все время, кроме TWT SP. Каждый из сенсоров просыпается, выполняет попытку передачи точке доступа одного пакета и опять переходит в спящий режим. Данный сценарий предусматривает как то, что все рассматриваемые станции находятся в зоне слышимости друг друга, так и то, что часть из них является скрытой для других.

Также в рассматриваемом сценарии есть эффект захвата канала. Когда несколько станций работают в одном канале и их счетчики отсрочки заканчиваются одновременно, кадры, отправленные этими станциями, вступают в коллизию на получателе (в этом случае отсутствует эффект захвата канала). При наличии эффекта захвата канала, если точка доступа принимает кадр от станции с низкой мощностью передачи и в это же время фиксирует преамбулу передачи от более мощной станции, то точка доступа может завершить принятие кадра от маломощной станции и начать принимать кадр от более мощной станции. Этот эффект может уменьшить число коллизий, так как при отсутствии переключения кадры попали бы в коллизию, в то время как при переключении на прием более мощной передачи один из кадров удается принять.

В контексте данного сценария возникает следующая задача: Исследовать зависимость среднего времени отправки сообщений группой станций и средней энергии, затрачиваемой каждой станцией, от временного шага между TWT различных станций, параметров распределения отклонения часов, параметров стандартного метода доступа к каналу и наличия или отсутствия эффекта захвата канала и скрытых станций.

4 Обзор литературы

В настоящий момент существует большое количество обзорных работ, в которых представлено подробное описание работы механизма TWT. Примерами таких работ являются [2], [3] и [4]. Авторы [5] пишут о своих намерениях в будущем реализовать механизм TWT на платформе имитационного моделирования ns-3 [6]. В работе [7] упоминается, что механизм TWT можно использовать для уменьшения числа коллизий без описания или исследования этого процесса. Работ, подробно исследовавших этот механизм в рамках технологии Wi-Fi пока не опубликовано.

Существует большое число работ, исследующих явление дрифта часов: его оценку и процесс синхронизации, как в работах [8] и [9]. В работе [10] предлагается, чтобы при каждом обмене информацией с сенсорами точка доступа оценивала и отправляла им в кадре ACK значение текущего временного отклонения для подстройки часов. Если инициатором обмена инфор-



Рис. 1. Зависимость среднего времени отправки сообщений группой станций от временного шага между их TWT.

мацией является точка доступа, то она должна в кадре данных отправить текущее значение своих часов. Этом метод неприменим в рассматриваемом нами сценарии, так как для успешной передачи необходимо знать отклонение часов заранее. Авторы работы [11] представили алгоритм, оценивающий скорость дрифта часов. Для этого сенсоры должны отслеживать и накапливать временной сдвиг между запланированными и реальными передачами каждого из своих соседей.

5 Результаты исследования

Для того, чтобы оценить эффективность механизма TWT, был проведен ряд экспериментов, в которых моделировался описанный в разделе 3 сценарий. Моделирование проводилось при помощи комплекса имитационного моделирования ns-3 [6]. Точка доступа и все станции передавали кадры размера 1024 байта на скорости 6 Мбит/сек с использованием самой надежной сигнально-кодовой конструкции MCS0. Все станции размещались случайным образом вокруг точки доступа внутри круга радиуса 5 м или 50 м в зависимости от сценария. Радиус слышимости станций равен 65 м, поэтому некоторые станции являются скрытыми для других при радиусе круга 50 м. Далее сценарий с размещением станций в круге радиуса 5 м будем называть «сценарий без скрытых станций», а другой соответственно «сценарий со скрытыми станциями».

В процессе моделирования описанного выше сценария для каждого из рассматриваемых наборов параметров было проведено 100 запусков имитационной модели.

На рис. 1 представлена зависимость среднего времени отправки сообщений группой станций от временного шага между их ТWТ. При маленьком временном шаге между TWT много станций пытаются одновременно передавать свои кадры, из-за чего возрастает число коллизий и растет среднее время передачи кадров. Из графиков 1 видно, что при увеличении CW_{min} уменьшается среднее время передачи кадров. Это происходит из-за уменьшения вероятности коллизий. Вероятность коллизий при малом временном шаге между TWT также уменьшается при увеличении величины дисперсии, так как уменьшается вероятность одновременного перехода из спящего режима нескольких станций. Это особенно заметно в сценарии без скрытых станций. Еще следует обратить внимание на то, что все зависимости при шаге TWT большем, чем две длительности успешной передачи, становятся линейными и имеют вид $T = Nsta*\mu+\sigma$, где Nsta — общее число сенсоров, μ — временной шаг между их TWT, σ — корень из дисперсии в распределении времении времении часов.

Из этой формулы видно, что время передачи группы станций пропорционально временному шагу между их ТWТ. Именно поэтому для минимизации использования канальных ресурсов следует устанавливать минимально возможный шаг между TWT, при котором зависимость имеет линейный вид. Теперь обратимся к рис. 2, на котором представлена зависимость среднего числа доставленных пакетов от временного шага между их TWT. Из этих графиков можно заметить, что при малом временном шаге между ближайшими TWT различных станций количество потерянных пакетов велико. Поэтому при назначении TWT следует соблюдать баланс между вероятностью доставки пакета и временем его доставки, выбирая не очень большой, но и не очень малый шаг между TWT. Следует заметить, что минимальный временной шаг, при котором гарантируется доставка данных, достигается при максимизации параметра CW_{min} .

Обратим внимание на то, что чем меньше повторных попыток передачи выполняет станция, тем меньше энергии она затрачивает. Это подтверждается рис. 3, на котором представлена зависимость средней энергии, затрачиваемой одной станцией на передачу кадра, от временного шага между TWT. Из этого графика видно, что минимум энергии достигается приблизительно в той же точке, в которой достигается максимальная вероятность переда-



Рис. 2. Зависимость среднего количества переданных сообщений группой станций от временного шага между их TWT.

чи пакета. Различие заключается в том, что для уменьшения потребления энергии требуется также уменьшение числа повторных попыток передач.

Теперь сравним представленные выше графики с эффектом захвата канала и без него. Как было сказано в разделе 3, эффект захвата канала проявляет себя в случае коллизии менее мощной передачи и более мощной. Следовательно этот эффект ярче всего себя проявляет в сценариях с большим числом коллизий, уменьшая количество потребляемых канальных ресурсов. Это особенно заметно на рис. 3, который показывает, что при $CW_{min} = 3$ максимально расходуемая энергия уменьшается в 1.2 раз при наличии эффекта захвата.

Обратим внимание на то, что длительность успешной передачи равна $T_{data} + SIFS + T_{ack} + DIFS = 1574$ мкс в рассмотренной имитационной модели. Начиная с удвоенной длительности успешной передачи количество расходуемой станциями энергии выходит на минимум. Если же рассматри-



Рис. 3. Зависимость средней затрачиваемой энергии одной станции от временного шага между ближайшими TWT станций.

вать интересующие нас сценарии с малым числом коллизий, то в них эффект захвата не дает видимого улучшения.

Из всех представленных выше графиков можно сделать следующий вывод: для уменьшения потребления энергии сенсоров и расхода канальных ресурсов следует устанавливать CW_{min} максимально возможным, а шаг между последовательными TWT задавать равным двум успешным передачам пакета.

6 Заключение

В данной работе исследовался механизм TWT (Target Wake Time) в рамках стандарта Wi-Fi IEEE 802.11ax. Этот механизм предназначен для планирования передач станций через длительные временные промежутки, в течение которых станции находятся в спящем режиме. Главной проблемой этого

механизма является дрейф часов, причем как часов сенсоров, так и часов точки доступа. Из-за этого явления моменты выхода станций из режима сна перестают четко подчиняться расписанию, что приводит к коллизиям, повторным попыткам передачи и увеличению энергопотребления станций. Также в этой статье применительно в механизму TWT исследовались проблема скрытых станций, увеличивающая число коллизий, и явление захвата канала, уменьшающее число коллизий.

Приведенный в разделе 3 сценарий был реализован на платформе имитационного моделирования ns-3. Отдельно были рассмотрены сценарии со скрытыми станциями и без оных, с эффектом захвата канала и без него для различных параметров метода доступа к каналу EDCA, различных значений дисперсии в распределении отклонений часов сенсоров и различных временных интервалов между TWT станций.

В работе было показано, что для наибольшей эффективности работы механизма TWT следует устанавливать наибольший из возможных CW_{min} , а длительность шага между назначаемыми TWT устанавливать в зависимости от выбранного для минимизации показателя эффективности работы сети: для уменьшения количества потребляемой станциями энергии нужно устанавливать как можно больший шаг между их TWT, в то время как для уменьшения среднего времени, затрачиваемого на передачу группы станций, шаг между их TWT следует устанавливать минимально возможным при максимальной вероятности доставки пакета. В качестве некоторого компромисса между этими двумя задачами можно предложить устанавливать шаг между TWT различных станций равным двум успешным передачам кадра данных.

Список литературы

- 1. Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2016-2021 White Access mode: Paper. https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visualnetworking-index-vni/mobile-white-paper-c11-520862.html.
- A survey on IEEE 802.11 ah: An enabling networking technology for smart cities / Evgeny Khorov, Andrey Lyakhov, Alexander Krotov, Andrey Guschin // Computer Communications. - 2015. - Vol. 58. - P. 53-69.
- 3. Park Minyoung. IEEE 802.11 ah: sub-1-GHz license-exempt operation for the internet of things // IEEE Communications Magazine.- 2015.- Vol. 53, no. 9.- P. 145–151.
- 4. Afaqui M Shahwaiz, Villegas EG, Aguilera EL. IEEE 802.11 ax: Challenges and requirements for future high efficiency WiFi // IEEE Wireless Communications. 2016. Vol. 99. P. 2–9.
- 5. Implementation and Validation of an IEEE 802.11 ah Module for ns-3 / Le Tian, Sebastien Deronne, Steven Latré, Jeroen Famaey // Proceedings of the Workshop on ns-3 / ACM. 2016. P. 49–56.
- 6. The ns-3 Network Simulator. Access mode: http://www.nsnam.org/.
- 7. Tian Le, Famaey Jeroen, Latré Steven. Evaluation of the IEEE 802.11 ah restricted access window mechanism for dense IoT networks // World of Wireless, Mobile

and Multimedia Networks (WoWMoM), 2016 IEEE 17th International Symposium on A / IEEE. - 2016. - P. 1–9.

- 8. Römer Kay. Time synchronization in ad hoc networks // Proceedings of the 2nd ACM international symposium on Mobile ad hoc networking & computing / ACM. -2001. P. 173-182.
- 9. Srikanth TK, Toueg Sam. Optimal clock synchronization // Journal of the ACM (JACM).- 1987.- Vol. 34, no. 3. P. 626–645.
- WirelessHART: Applying wireless technology in real-time industrial process control / Jianping Song, Song Han, Al Mok et al. // IEEE real-time and embedded technology and applications symposium / IEEE. – 2008. – P. 377–386.
- Clock drift reduction for relative time slot TDMA-based sensor networks / Randolph Tjoa, Kim Loon Chee, PK Sivaprasad et al. // Personal, Indoor and Mobile Radio Communications, 2004. PIMRC 2004. 15th IEEE International Symposium on / IEEE. – Vol. 2. – 2004. – P. 1042–1047.
Разработка методов построения схем множественного неортогонального доступа на основе кодов с малой плотностью проверок на четность.

Луиза Медова^{1,2}, Павел Рыбин¹, Алексей Фролов³, Антон Глебов³.

Институт Проблем Передачи Информации РАН, Москва, Россия, ² Московский физико-технический институт, Москва, Россия

³ Сколковский институт науки и технологий, Москва, Россия

luiza.medova@phystech.edu, prybin@iitp.ru, al.frolov@skoltech.ru, anton.glebov@skolkovotech.ru.

Аннотация В данной работе исследованы МПП-коды для канала множественного доступа. Описан итеративный алгоритм совместного декодирования и метод построения МПП-кодов, оптимизированных для совместного декодирования. Метод анализа для многопользовательского канала был обобщён на случай произвольного количества пользователей. На основе данного обобщенного анализа была реализованы методы оптимизации конструкции МПП-кода. Проведено имитационное моделирование. Результаты имитационного моделирования позволяют заключить, что оптимизированные для совместного декодирования конструкции МПП-кодов превосходят рассмотренные раннее МПП-коды и находятся ближе к теоретической границе.

Ключевые слова: Неортогональный множественный доступ, совместное итеративное декодирование, МПП-коды, протограф алгоритм распространения доверия.

1 Введение

В настоящее время одной из наиболее важных проблем является проблема повышения эффективности использования доступных ресурсов в системах множественного доступа. Одной из причин, по которой решение проблемы множественного доступа играет все большую роль в процессе проектирования современных систем беспроводной связи, является резкий рост числа устройств, которые используют беспроводные каналы связи для передачи информации. Значительное внимание уделяется развитию методов множественного доступа, в которых пользователи могут использовать одни и те же части частотно-временного ресурса (неортогональный множественный доступ), что позволяет поддерживать множество подключений от пользователей и/или устройств и минимизировать расход энергии и трафика на управление системой связи. Для разделения пользователей используются специализированные сигнально-кодовые конструкции, которые позволят существенно уменьшить вероятность возможных коллизий.

В последнее десятилетие наибольшую популярность приобрели коды малой плотностью проверок на четность (МПП-коды) [3]. Представляет интерес вопрос о том, как последние достижения в разработке и анализе МППкодов могут быть использованы для канала множественного доступа.

В работе [1] были исследованы МПП-коды для многопользовательского канала и было показано, что выбирая соответствующие МПП-коды и применяя совместный итеративный декодер, можно достичь любую точку внешей границы пропускной способности без разделения времени или разделения скорости.

В данной работе мы исследовали МПП-коды для многопользовательского канала, построенные на протографах. Описан алгоритм совместного итеративного декодирования и предложен метод анализа корректирующей способности МПП-кода для канала с произвольным числом пользователей. На основе данного метода был реализован метод оптимизации конструкции МПП-кода и проведено имитационное моделирование.

2 МПП-коды, построенные на протографах.

Коды с малой плотностью проверок (МПП-коды) на четность [3] - это блочные коды, проверочные матрицы которых содержат в основном нули и относительно малое количество единиц.

Существует несколько способов описания МПП-кодов. Распространнёным способом задания МПП-кода является представление кода в виде двудольного графа [8], вершины которого разделены на два множества: множество символьных вершин, соответствующих столбцам проверочной матрицы, и множество проверочных вершин, соответствующих строкам проверочной матрицы. Такой граф принято называть графом Таннера. Ребра в графе Таннера соответствуют ненулевым позициям в проверочной матрице кода.

Матрицы МПП-кода могут быть построены с помощью протографа [10]. Протораф - это граф Таннера с небольшим числом вершин, соединённых небольшим числом рёбер. МПП-код строится из протографа путем копирования протографа T раз и перестановки ребер между соответствующих одному и тому же типу вершинами исходного протографа. В качестве простого примера рассмотрим протограф, изображенный на рис.1. Этот граф состоит из |V| = 4 вершин-символов и |C| = 3 вершин-проверок, соединенных |E| = 8 рёбрами. Каждая вершина-проверка и вершина-символ в протографа. Три копии наложены друг на друга так, что вершины одного типа находятся в непосредственной близости, однако общий граф состоит из трех несвязанных подграфов. На рис.3 ребра графа перегруппированы между вершинами-проверками, принадлежащему одному и тому же типу. После



Рис. 1: Протограф



Рис. 2: Три копии исходного протографа

этой перестановки три подграфа теперь взаимосвязаны. Полученный граф на рис.3 представляет собой граф Таннера МПП-кода (n = 12, k = 3).

3 Модель канала множественного доступа

В качестве модели канала будем рассматривать канал множественного доступа с аддитивным белым гауссовским шумом с двоичным входом (рис. 4), в котором T пользователей с равными мощностями P передают данные:

$$Y = \sum_{t=1}^{T} X_t + Z,$$

где $Z \sim \mathcal{N}(0, \frac{N_0}{2})$ и $\mathsf{E}[|X_i|^2] \leq P.$

4 Алгоритм совместного итеративного декодирования

Рассмотрим систему множественного доступа, в которой T пользователей передают данные в АБГШ-канале. Обозначим через $C^{(t)}$ код t-ого пользователя с длиной n и скоростью R. Пользователей посылают кодовые слова $\mathbf{c}^{(1)}, \mathbf{c}^{(2)}, ..., \mathbf{c}^{(T)}$ с использованием двоичной фазовой модуляции. На вход канала поступают $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, ..., \mathbf{x}^{(T)}$ - последовательности сигналов, соответствующие переданным кодовым словам, $\mathbf{x}^{(i)} \in \{-\sqrt{P}, +\sqrt{P}\}^n$. На приемнике мы получаем (**y**) - поэлементную сумму переданных последовательностей сигналов с аддитивным гауссовским шумом.

380



Рис. 3: Полученный граф



Рис. 4: Канал множественного доступа с аддитивным белым гауссовским шумом с двоичным входом.

Задачей совместного декодирования является восстановление всех переданных кодовых слов на основе принятого вектора (у). Совместный декодер основан на итеративном алгоритме распространения доверия [2].

Декодер может быть представлен в виде графа (фактор-графа), пример которого изображен на рис. 5 для случая T = 3.

На рис. 5 МПП-коды, используемые пользователями, представлены в виде



Рис. 5: Граф совместного декодера для случая T = 3.

соответствующих двудольных графов Таннера, где вершины одной доли соответствуют вершинам-проверкам (обозначены черными квадратами), а вершины другой – вершинам-символам (обозначены красными кругами). В тоже время на рисунке присутствуют вершины третьего вида, так называемые, функциональные узлы (обозначены зеленным треугольниками). Эти вершины соответствуют элементам принятой последовательности (у). На рис. 6 отдельно изображен k—ый функциональный узел и сообщения, посылаемые и получаемые данным узлом на каждой итерации. Соответству-



Рис. 6: Граф совместного декодера для случая T = 3.

ющий алгоритм совместного итеративного декодирования состоит в следующем. Сначала мы инициализируем LLR значения вершин-символов для каждого пользователя нулем, предполагая равную вероятность для значения 1 и -1. Затем выполняем I_O внешних итераций, каждая из которых состоит из следующих шагов:

- декодирование по максимуму правдоподобия каждого функционального узла (т.е. вычисляем вычисляем сообщения для вершин-символов);
- декодирование МПП-кодов пользовалелей и обновление LLR значения для вершин-символов (все МПП-коды могут декодироваться параллельно).

Отметим, что в приведенном алгоритме каждый пользователь использует стандартный алгоритм распространения доверия. Наиболее интересная часть приведенного алгоритма – это декодирование функциональных узлов. Согласно принципу алгоритма обмена сообщениями [6], [4], получим правило для вычисления сообщения (μ), посылаемого в i-ю вершину-символ k-го пользователя (k = 1, ..., T, i = 1, ..., N) из функционального узла F_i .

Сообщения, исходящие из вершин-символов для k—ого пользователя вершинам проверкам при совместном итеративном декодировании переписываются следующим образом:

$$m_{vc,k}^{(l)} = \begin{cases} m_{sv,k}^{(0)}, & \text{если } l = 1, \\ m_{sv,k}^{(l-1)} + \sum_{c' \in C_v \setminus \{c\}} m_{c'v}^{(l-1)}, & \text{если } l > 1, \end{cases}$$
(1)

где $m_{sv,k}^{(l)}$ - сообщение, отправленное функциональным узлом на шаге l пользователю k.

Определим в функциональном узле функцию:

$$\mu_{s \to v}(x_i^{(k)}) = \sum_{\sim x_i^{(k)}} \prod_{j \neq k} \mu_{v \to s}(x_i^{(j)}) \times Pr(y_i | x_i^{(1)}, \dots, x_i^{(T)}),$$

где $x_i^{(k)}$ обозначает i-й переданный бит, y_i - принятый бит, а $\mu_{v \to s}(x_i^{(k)})$ функция вершины-символа пользователя k, определяющее сообщение, отправленное функциональному узлу :

$$m_{sv,k} = \log \frac{\mu_{v \to s}(x_i^{(k)} = 1)}{\mu_{v \to s}(x_i^{(k)} = -1)}$$

Для упрощения дальнейших вычислений положим:

$$m_{vs,k} \equiv X_k, \quad e^{X_k} = \frac{\mu_{k \to F_i}(x_i^{(k)} = 1)}{\mu_{k \to F_i}(x_i^{(k)} = -1)}$$

Вычислим сообщение, посылаемое из функционального узлаi-й вершине-проверки пользователя k :

$$m_{sv,k} = \log \frac{\mu_{s \to v}(x_i^{(k)} = 1|y)}{\mu_{s \to v}(x_i^{(k)} = -1|y)} = \\ = \log \left(\frac{\sum_{\sim x_i^{(k)}} \prod_{j \neq k} \mu_{v \to s}(x_i^{(k)} = 1) Pr(y_i | x_i^{(1)}, ..., x_i^{(k)} = 1, ..., x_i^{(n)})}{\sum_{\sim x_i^{(k)}} \prod_{j \neq k} \mu_{v \to s}(x_i^{(k)} = -1) Pr(y_i | x_i^{(1)}, ..., x_i^{(k)} = -1, ..., x_i^{(n)})} \right)$$

Упростим выражение следующим образом:

$$m_{sv,k} = \log\left(\frac{\sum_{\sim x_i^{(k)}} \prod_{j \neq k} e^{1_{x_j} X_j} p(y_i | x_i^{(1)}, \dots, x_i^{(k)} = 1, \dots, x_i^{(n)})}{\sum_{\sim x_i^{(k)}} \prod_{j \neq k} e^{1_{x_j} X_j} p(y_i | x_i^{(1)}, \dots, x_i^{(k)} = -1, \dots, x_i^{(n)})}\right), \quad (2)$$
rge $1_{x_k} = \begin{cases} 1, & x_i^{(j)} = 1\\ 0, & x_i^{(j)} = -1. \end{cases}$

4.1 Метод построения МПП-кодов

Для построения МПП-кодов будем использовать метод аппроксимации плотности эволюции с использованием меры взаимной информации (PEXIT-

анализ). Аппроксимация плотности эволюции с использованием меры взаимной информации (EXIT-анализ) была впервые представлена в [9]. Основная идея данного анализа заключается в вычислении взаимной информации I(x;m) между переданными сигналами x и LLR в схеме итеративного кодирования. Это способ представления того, как изменяется взаимная информация между выходом декодера и переданными битами.

Однако EXIT-анализ, описанный в [9], не может быть напрямую применен для исследования МПП-кодов, основанных на протографах. В [5] приведены примеры протографов с одинаковыми распределениями степеней вершин, но разными порогами декодирования. Поэтому мы будем использовать его модификацию для МПП-кодов на протографах (PEXIT-анализ) [5]. Описание модфицированного метода для однопользовательского канала приведено в [5], [12].

Модифицицируем данный метод для анализа совместного декодирования. Взаимная информация I_{Ev} между сообщением, исходящим из вершинысимвола, и соответствующей вершиной-проверкой:

$$I_{Ev} = I_{Ev}(I_{Av}, I_{Es}) \tag{3}$$

где I_{Av} - взаимная информация, определенная в разделе [12], а I_{Es} - взаимная информация между вершиной-символом и сообщениями, полученными этой вершиной от функциональных узлов.

Поскольку PEXIT-анализ отслеживает взаимную информацию вдоль ребер фактор-графа, то $I_{Ev}(i, j)$, взаимную информацию между сообщением, посланным вершиной-символом V_j , вершине-проверки C_i и соответствующим битом кодового слова, определяется, как:

$$I_{Ev}(i,j) = J\left(\sqrt{\sum_{s \neq i} [J^{-1}(I_{Av}(s,j))]^2 + [J^{-1}(I_{Es}(j))]^2}\right)$$
(4)

Взаимная информация между вершиной-символом V_j и сообщением, отправленным функциональному узлу, обозначается как $I_{Evs}(j)$ и определяется следующим образом:

$$I_{Evs}(j) = J\left(\sqrt{\sum_{s} [J^{-1}(I_{av}(s,j))]^2}\right)$$
(5)

Взаимная информация I_{Ec} между сообщением, исходящим из вершиныпроверки, и битом кодового слова, соответствующему вершине-символу, остается без изменений и вычисляется каждым пользователем аналогично случаю одного пользователя.

Теперь нам нужно вычислить взаимную информацию I_{Es} . Для того, чтобы получить представление о плотности вероятности (2) для *j*-ого пользователя, сгенерируем выборку исходящих LLR-сообщений используя (2) на основании LLR-сообщений, полученных от всех других пользователей, чьи плотности распределения могут быть аппроксимированы гауссовским распределением $\mathcal{N}(\mu_{Evs}, 2\mu_{Evs})$, где $\mu_{Evs} = \frac{J^{-1}(I_{Evs})}{2}$. Для количественной оценки μ_{Es} и вычисления требуемой взаимной информации как $I_{Es} = J(\mu_{Es})$ мы сошлемся на работу [7], где предложены были следующие три подхода:

- Аппроксимация гауссовским распределением по среднему: среднее значение μ оценивается для выборки, а затем устанавливается $\mu_{Evs} = \mu$ и $\sigma_{Evs}^2 = 2\mu$.
- Аппроксимация гауссовским распределение по моде: мода m и мы определяем $\mu_{Evs} = m$ и $\sigma_{Evs}^2 = 2m$.
- Аппроксимация смесью гауссовских распределений: средние значения $\mu_1, ..., \mu_k =$ и веса $a_1, ..., a_k$ и и веса $a_1, ..., a_k$ оцениваются по выборке $I_{Es} = a_1 J(\mu_1) + ... + a_k J(\mu_k).$

Обоснование этих методов продемонстрировано в [7]. Помимо этого, авторы сравнили эффективность данных приближений. Аппроксимация гауссовским распределением по моде позволяет получить максимальную взаимную информацию на выходе функциональных узлов и МПП-коды, построенные с использованием этой аппроксимации, при совместном декодировании позволяют получить вероятность ошибки на бит меньшую, чем МПП-коды, полученными с использованием двух других аппроксимаций.

Каждый пользователь вычисляет $I_{APP}(j)$, совместную информацию между отношением правдоподобия апостериорной вероятности, вычисленным вершиной-символом v_i и соответствующим битом кодового слова.

$$I_{APP}(j) = J\left(\sqrt{\sum_{s} [J^{-1}(I_{Av}(s,j))]^2 + [J^{-1}(I_{Es}(j))]^2}\right)$$
(6)

Алгоритм итеративного декодирования сходится, если $I_{APP}(j)$ достигает 1, в то время как число итераций стремится к бесконечности.

При фиксированной конструкции кода и фиксированном количестве итераций мы находим порог (наименьшее $\frac{E_b}{N_0}$ при котором $I_{app}(j) \rightarrow 1$ для всех вершин-символов каждого пользователя). Используя жадный алгоритм, мы меняем строчки в проверочной матрице и находим оптимизированную матрицу для наших параметров.

4.2 Результаты моделирования

Мы провели имитационное моделирование для случаев T = 2 и T = 4. Рассмотрим сначала результаты моделирования для случая T = 2 (рис. 7). Для этого случая мы сравнили вероятность ошибки на блок МПП-кода (364, 91) из [11], полученного в результате двукратного повторения каждого бита регулярного (3, 6) МПП-кода, вероятность ошибки на блок МПП-кода (364, 91), построенного с помощью РЕХІТ-анализа, и границу Полянского для случая двух кодов с конечной длиной. Как мы можем видеть на рис.

7 МПП-код, построенный с использованием предложенного PEXIT-метода, превосходит МПП-код из [11] примерно на 0.5 дБ. В тоже время разрыв между границей Полянского и лучшим МПП-кодом составляет порядка 3 дБ. Но стоит отметить, что использованная здесь граница Полянского получена для Гауссовских сигналов, а не для сигналов с двоичной фазовой модуляцией, использованной при моделировании. Поэтому, можно предположить, что разрыв значительно сократиться, если использовать границу Полянского для сигналов с двоичной фазовой модуляцией. Теперь рассмотрим результаты имитационного моделирования для T = 4 (рис. 8). Для этого случая мы построили новый МПП-код (364, 91) с использованием РЕХІТанализа и сравнили его вероятность ошибки на блок с вероятностью ошибки на блок предыдущего МПП-кода из [11] и границы Полянского для случая 4 кодов с конечной длиной. Как можно видеть на рис. 8 МПП-код, построенный с использованием PEXIT-анализа, превосходит МПП-код из [11] на более чем 3 дБ. И опять разрыв между границей Полянского и построенным МПП-кодом немного меньше, чем 3 дБ.



Рис. 7: Результаты имитационного моделирования для T=2 и МПП-кода (364, 91)

5 Заключение

Мы описали итеративный алгоритм совместного декодирования и обобщенный метод построения МПП-кодов, оптимизированных для совместного декодирования. На основе данного обобщенного анализа была реализованы методы оптимизации конструкции МПП-кода. Представленные результаты моделирования показали эффективность предложенного метода.



Рис. 8: Результаты имитационного моделирования для Т=4 и МПП-кода (364, 91)

Благодарности

Исследование выполнено при финансовой поддержке РФФИ в рамках научных проектов No 18-07-01409 A, 18-37-00322, 18-37-00319 и при финансовой поддержке Министерства образования и науки РФ (договор No 03.G25.31.0222) в рамках реализации Постановления Правительства РФ No 218 "О мерах государственной поддержки развития кооперации российских образовательных организаций высшего образования, государственных научных учреждений и организаций, реализующих комплексные проекты по созданию высокотехнологичного производства, в рамках подпрограммы "Институциональное развитие научно-исследовательского сектора" государственной программы Российской Федерации "Развитие науки и технологий" на 2013–2020 годы".

Список литературы

- A. Amraoui, S. Dusad, and R. Urbanke. Achieving general points in the 2-user Gaussian MAC without time-sharing or rate- splitting by means of iterative coding . in Information Theory, 2002. Proceedings. 2002 IEEE International Symposium on, 2002.
- Amin Shokrollahi. LDPC Codes: An Introduction. Coding, Cryptography and Combinatorics, 23:85–110, 2003.
- 3. Robert G. Gallager. Low-Density Parity-Check Codes. 1963.
- Frank R Kschischang, Brendan J Frey, and H-A Loeliger. Factor graphs and the sum-product algorithm. *IEEE Transactions on information theory*, 47(2):498–519, 2001.

- G. Liva and M. Chiani. Protograph ldpc codes design based on exit analysis. In IEEE GLOBECOM 2007 - IEEE Global Telecommunications Conference, pages 3250–3254, Nov 2007.
- Tom Richardson and Ruediger Urbanke. Modern Coding Theory. Cambridge University Press, New York, NY, USA, 2008.
- Iqbal Shahid and Pradeepa Yahampath. Distributed joint source-channel code design for gmac using irregular ldpc codes. *EURASIP Journal on Wireless Communications and Networking*, (1):3, Jan 2014.
- 8. R Tanner. A recursive approach to low complexity codes. *IEEE Transactions on information theory*, 27(5):533–547, 1981.
- S. ten Brink. Designing iterative decoding schemes with the extrinsic information chart. International Journal of Electronics and Communications, 54(6):498–519, 2000.
- J. Thorpe. Low-Density Parity-Check (LDPC) Codes Constructed from Protographs. Interplanetary Network Progress Report, 154:1–7, August 2003.
- Avinash Vem, Krishna R. Narayanan, Jun Cheng, and Jean-Francois Chamberland. A user-independent serial interference cancellation based coding scheme for the unsourced random access gaussian channel. In *Proc. IEEE Information Theory Workshop (ITW)*, pages 1–5, 2017.
- Рыбин П.С. Медова Л.Р. Сравнение методов оценки корректирующей способности МПП-кодов при итеративном декодировании. Сборник трудов школыконференции "Информационные технологии и системы", 2017.

Разработка методов построения схем неортогонального множественного доступа на основе полярных кодов

Глеб Балицкий^{1,2,3}, Андрей Дзись^{1,2,3}, Алексей Фролов^{3,1}

¹ Институт Проблем Передачи Информации РАН, Москва, Россия,

² Московский физико-технический институт, Москва, Россия

 $^{3}\,$ Сколковский институт науки и технологий, Москва, Россия

 $\verb"gleb.balitskiy@phystech.edu", and \verb"rey.dzis@phystech.edu", and \verb"rey.dzis@phystech.edu", and \verb"rey.dzis@phystech.edu", and "rey.dzis@phystech.edu", and "rey$

al.frolov@skoltech.ru

Аннотация В работе рассмотрены способы построения схем неортогонального множественного доступа на основе полярных кодов. Была сформулирована и доказана теорема о максимальной скорости передачи данных, на которой еще сохраняется возможность однозначного декодирования, для любого кода в случае двух пользователей. Также был обобщен метод эволюции плотности и на основе этих результатов построена конструкция, позволяющая эффективно передавать данные и различать пользователей.

Ключевые слова: Построение полярных кодов, каналы с аддитивным белым гауссовским шумом, декодирование методом последовательного исключения (successive cancelation), эволюция плотности, множественный доступ, интернет вещей, межмашинное взаимодействие

1 Введение

Построение схем неортогонального множественного доступа на основе полярных кодов является перспективным методом решения задач, возникающих в сценариях межмашинного взаимодействия. Это можно объяснить двумя фактами:

- Невысокая субквадратичная сложность кодирования $\mathcal{O}(mN \log N)$ и декодирования $\mathcal{O}((2^m 1)N \log N)$, где N длина кода, m количество пользователей[1][5].
- Одна из особенностей сценария межмашинного взаимодействия заключается в использовании коротких сообщений для передачи данных. Именно полярный код является одним из наиболее эффективных кодов в случаях передачи коротких сообщений в однопользовательском канале. Например, он значительно превосходит часто используемые в данный момент МПП-коды и турбо-коды на длинах N < 10³[4]. Также полярный код был использован в контрольном канале беспроводных сетей пятого поколения (5G), где показал свою эффективность.

Конечно, высокая эффективность в случае одного пользователя не гарантирует таковую для двух и более пользователей, но дает хорошую мотивацию для исследования именно полярных кодов в контексте межмашинного взаимодействия. Постановка задачи энергоэффективного неортогонального множественного доступа в случае передачи коротких сообщений рождает две основные проблемы:

- Не любой код, построенный для случая большего числа пользователей, может быть однозначно декодирован даже в каналах без шума. А на коротких длинах сообщений этот эффект вносит ощутимый вклад в вероятность ошибочного декодирования.
- Высокая энергоэффективность подразумевает низкие отношения "сигналшум". Поэтому построенный код должен иметь высокую помехоустойчивость.

В данной работе мы рассмотрели обе проблемы.

2 Полярные коды в случае двух пользователей

В данной работе рассмотрен случай двух пользователей для гауссовского канала множественного доступа с двоичной фазовой манипуляцией и синхронной передачей данных с одинаковой мощностью сигнала. Задачу будем решать в контексте массового межмашинного взаимодействия. Следовательно, нам нужно посторить код небольшой длины, работающий в диапозоне малых значений параметра "сигнал-шум". Рассмотрим процесс передачи. Оба пользователя формируют информационный вектор и кодируют его, используя порождающую матрицу $(u_1^{k_1}, v_1^{k_2}) \rightarrow (x_1^N, y_1^N)$:

$$\begin{cases} x_1^N = u_1^k G_u \\ y_1^N = v_1^k G_v \end{cases}$$
(1)

Далее пара кодовых слов (x_1^N, y_1^N) поступает в канал W, и, соотвественно, формируется выходной вектор z. В случае гауссовского канала:

$$z = x_{1b} + y_{1b} + \mathcal{N}(0, \sigma^2)$$
(2)

где вектора x_{1b}^N и y_{1b}^N – это вектора x_1^N и y_1^N после применения двоичной фазовой манипуляции. Задача заключается в построении такого кода, чтобы стало возможным определение пары информационных векторов $(u_1^{k_1}, v_1^{k_2})$. Для декодирования будем пользоватся алгоритмом последовательного исключения для случая двух пользователей.

2.1 Поляризация двухпользовательского канала множественного доступа

Введем двухпользовательский канал множественного доступа

$$W: \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$$

где $\mathcal{X} \in \{0,1\}$ и $\mathcal{Y} \in \{0,1\}$. W(y|x,y) назовем переходной вероятностью. Преобразование для поляризации двухпользовательского канала строится способом, аналогичным случаю одного пользователя.

– Слияние каналов. Аналогично случаю одного пользователя получим новый канал $W_N: \mathcal{X}^N \times \mathcal{Y}^N \to \mathcal{Z}^N:$

$$W_N(z_1^N | u_1^N, v_1^N) = \prod_{i=1}^{i=N} W(z_i | (uG_N)_i, (vG_N)_i) = W^N(z_1^N | u_1^N G_N, v_1^N G_N)$$

где G_N – порождающая матрица полярного кода.

 Разделение канала. Разобьем получившийся канал W_N на N виртуальных каналов с вероятностями передачи:

$$W_N^{(i)}(z_1^N, u_1^{i-1}, v_1^{i-1} | u_i, v_i) = \sum_{\substack{u_{i+1}^N, v_{i+1}^N \in \mathcal{X}^{N-i}}} \frac{1}{(2^N - 1)^2} W_N(z_1^N | u_1^N, v_1^N)$$
(3)

2.2 Построение полярного кода для случая двухпользовательского канала множественного доступа

Рассмотрим преобразование $(u_1^N, v_1^N) \to (x_1^N, y_1^N)$. Для первого пользователя выберем два непересекающихся множества индексов $\mathcal{A}^{(1)}$ и $\mathcal{A}_f^{(1)}$, относящихся к информационным и замороженным битам. Аналогично выберем для второго $\mathcal{A}^{(2)}$ и $\mathcal{A}_f^{(2)}$. Преобразование 2.10 приобретет вид:

$$\begin{cases} x_1^N = u_{\mathcal{A}_f^{(1)}} G_N(\mathcal{A}_f^{(1)}) \oplus u_{\mathcal{A}^{(1)}} G_N(\mathcal{A}^{(1)}) \\ y_1^N = v_{\mathcal{A}_f^{(2)}} G_N(\mathcal{A}_f^{(2)}) \oplus v_{\mathcal{A}^{(2)}} G_N(\mathcal{A}^{(2)}) \end{cases}$$
(4)

Обозначим полученный код $(N, \mathcal{A}^{(1)}, \mathcal{A}^{(2)}, u_{\mathcal{A}_{f}^{(1)}}, v_{\mathcal{A}_{f}^{(2)}})$. Скоростью кода будем считать $R = \frac{|\mathcal{A}^{(1)}|}{N} + \frac{|\mathcal{A}^{(2)}|}{N}$.

2.3 Проблема однозначного декодирования

Рассмотрим канал без шума. Тогда выходной вектор z:

$$z_1^N = x_1^N + y_1^N$$

Соответственно, вектор $z_i \in \{0, 1, 2\}.$

Определение 2. Будем называть код однозначно декодируемым, если никакие две пары информационных векторов $(u_1^{k_1}, v_1^{k_2})$ не преобразуются в один и тот же выходной вектор z.

один и тот же выходной вектор z. **Утверждение 1.** Если код $(N, \mathcal{A}^{(1)}, \mathcal{A}^{(2)}, u_{\mathcal{A}_{f}^{(1)}}, v_{\mathcal{A}_{f}^{(2)}})$ однозначно декодируемый , тогда $NR \leq \lfloor N \log_2 3 \rfloor$, где $R = \frac{|\mathcal{A}^1|}{N} + \frac{|\mathcal{A}^2|}{N}$. Доказательство. Максимальное количество выходнных векторов 3^N . Количество кодовых слов 2^{RN} . Тогда для одназначной декодируемости необходимо, чтобы количество кодовых слов было не более количества выходных векторов.

$$2^{NR} < 3^{I}$$

тогда

 $NR \leq |N \log_2 3|$

Рассмотрим вопрос однозначного декодирования более детально. Предыдущее утверждение сформулированно на основе того факта, что при построении кода мы сможем "попасть" во все 3^N выходных вектора. Но в действительности это сделать невозможно. И этот результат отражает следующая теорема.

Теорема 1. Если код, построенный на основе преобразования,

$$\begin{cases} x_1^N = u_1^{k_1} G_u \\ y_1^N = v_1^{k_2} G_v \end{cases}$$
$$z_1^N = x_1^N + y_1^N$$

однозначно декодируемый, то его скорость $R = \frac{k_1}{N} + \frac{k_2}{N}$ не больше 1.

Доказательство. Рассмотрим выходой вектор z. Множество индексов, для которых $z_i = 1$, обозначим как C_1 , а множество индексов, для которых $z_i = 0$ или $z_i = 2$, как C_0 . Пусть $|C_1| = c$, тогда $|C_0| = N - c$. Если $i \in C_1$, то возможны две комбинации $(x_i = 0, y_i = 1)$ и $(x_i = 1, y_i = 0)$, если $i \in C_0$, то одна для фиксированного выходного вектора. Зафиксируем множетсва C_0 и C_1 . Тогда количество различимых выходных векторов, удовлетворяющих данному условию 2^{N-c} . Соответственно, если количество информационных пар слов, преобразующихся в выходной вектор данного вида, больше 2^{N-c} , тогда код не может быть отдекодирован однозначно. Опишем множетсво таких информационных векторов $(u_1^{(2)}, v_1^{(2)})$, которые преобразуются в выходные вектора z для заданных C_1 и C_0 :

$$u_1^{k_1} G_u \oplus v_1^{k_2} G_v = z_2 \tag{5}$$

где $z_2(\mathcal{C}_1) = (1, 1, ..1)_1^c$ и $z_2(\mathcal{C}_0) = (0, 0, ..0)_1^{N-c}$. Перепишем систему 5 в другом виде:

$$(u_1^{k_1}v_1^{k_2})G = z_2 (6)$$

где

$$G = \begin{pmatrix} G_u \\ G_v \end{pmatrix}$$

Пусть rg(G) = r. Пусть $k = k_1 + k_2 > N$. Тогда количество решений системы 5 либо 0, либо 2^{k-r} . Кроме того, существует такой z_2 , для которого система 5 совместна и его вес Хемминга не менее r, так как rg(G) = r. Тогда количество различимых пар 2^{N-c} меньше количества решений 2^{k-r} , поэтому код с k > N или R > 1 не может быть отдекодирован однозначно.

Теорема 1 также верна для случая преобразования 4, где в качестве G_N можно брать любую невырожденную матрицу. Так как новое множество векторов z_{2_f} , по сути смещенное на вектор $u_{\mathcal{A}_f^{(1)}}G_N(\mathcal{A}_f^{(1)}) \oplus v_{\mathcal{A}_f^{(2)}}G_N(\mathcal{A}_f^{(2)})$ множество z_2 из доказательства, которое будет образовывать смежный класс, и в нем тоже будет существовать вектор веса не менее r. Итак, мы получили границу на скорость кода для однозначного декодирования. Хотелось бы почеркнуть, что Теорема 1 не дает никаких утверждений про существование однозначно декодируемого кода на скоростях не более единицы.

Вернемся к рассмотрению полярных кодов, где G_N – порождающая матрица полярного кода. Будем рассматривать только однозначно декодируемые коды и попробуем построить код для скоростей $R \leq 1$. Также будем учитывать эмпирический факт, что подавляющая часть "хороших" вирутальных каналов $W_N^{(i)}$ принадлежит множеству, где $i \in \{N/2, N/2+1, ..., N\}$. Рассмотрим рекурсивную схему построения полярного кода. Выделим уровни G_N и $G_{N/2}$ (рисунок 1). Определим преобразование из G_N в $G_{N/2}$ как кода в целом.



Рис. 1: Преобразование G в $G_{N/2}$

Обозначим информационные биты как inf, и замороженные как fro, $(G_{N/2}^{up})_i - i$ - ый бит вверхнего кода на уровне $G_{N/2}$, $(G_{N/2}^{down})_i - i$ - ый бит нижнего кода на уровне $G_{N/2}$, $(G_N)_i - i$ - ый бит кода на уровне G_N . Тогда:

$(G_{N/2}^{up})_i:$	fro	fro	inf	inf	inf
$(G_{N/2}^{down})_i$:	fro	inf	fro	$ \inf $	\inf
$(G_N)_{2i-1}$:	fro	none	\inf	\inf	fro
$(G_N)_{2i}$:	fro	none	fro	$ \inf $	\inf

Утверждение 2. Если оба кода на уровне $G_{N/2}$ однозначно декодируемые, то код G_N тоже однозанчно декодируемый.

Доказательство. Единственная ситуация в данных условиях, когда выходной вектор мог не изменить своего значения для двух разных пар информационных слов на уровне G_N – это одинаковые пары информационных слов для каждого кода на уровне $G_{N/2}$, но это невозможно. Тогда если на уровне $G_{N/2}$ оба кода однозначно декодируемые, то код на уровне G тоже.

Утверждение 3. Код $(N, \mathcal{A}^{(1)}, \mathcal{A}^{(2)}, u_{\mathcal{A}_{f}^{(1)}}, v_{\mathcal{A}_{f}^{(2)}})$, где $\mathcal{A}^{(2)} = \mathcal{A}^{(1)} = (N/2, N/2 + 1, ..N), u_{\mathcal{A}_{f}^{(1)}} = (0, 0, ..0, 1)$ и $u_{\mathcal{A}_{f}^{(2)}} = (0, 0, ..0, 0)$ однозначно декодируемый.

Доказательство. Рассмотрим уровень G_2 для данного кода. Для каждого кода на этом уровне мы получим конфигурацию (2, (2), (2), (1), (0)). Она является однозначно декодируемой, что легко проверить простым перебором. Тогда на основе утверждения 2 мы приходим к выводу, что код $(N, \mathcal{A}^{(1)}, \mathcal{A}^{(2)}, u_{\mathcal{A}_{f}^{(1)}}, v_{\mathcal{A}_{f}^{(2)}})$, где $\mathcal{A}^{(2)} = \mathcal{A}^{(1)} = (N/2, N/2 + 1, ..N), u_{\mathcal{A}_{f}^{(1)}} = (0, 0, ..0, 1)$ и $u_{\mathcal{A}_{f}^{(2)}} = (0, 0, ..0, 0)$ однозначно декодируемый.

2.4 Оценка помехоустойчивости витуальных каналов

Итак, мы получили конструкцию, которая является однозначно декодируемой и содержит большую часть "хороших" каналов. Теперь вернем в рассмотрение нашей задачи гауссовский шум, оценим вероятности ошибки, связанные с его наличием и воспользуемся этими оценками для посроения кода. Для построения оценок обобщим алгоритм эволюции плотности на случай двух пользователей. Начнем с подробного рассмотрения формулы 3. Не сложно показать, что данная формула преобразуется и может быть посчитано рекурсивно:

$$W_{2N}^{(2i-1)}(z_{1}^{2N}, u_{1}^{2i-2}, v_{1}^{2i-2} | u_{2i-1}, v_{2i-1}) = \sum_{u_{2i}, v_{2i}} \frac{1}{4} W_{N}^{(i)}(z_{1}^{N}, u_{1,e}^{2i-2} \oplus u_{1,o}^{2i-2}, v_{1,e}^{2i-2} \oplus v_{1,o}^{2i-2} | u_{2i-1} \oplus u_{2i}, v_{2i-1} \oplus v_{2i}) \times \\ \times W_{N}^{(i)}(z_{N+1}^{2N}, u_{1,e}^{2i-2}, v_{1,e}^{2i-2} | u_{2i}, v_{2i})$$

$$(7)$$

$$W_{2N}^{(2i)}(z_1^{2N}, u_1^{2i-1}, v_1^{2i-1} | u_{2i}, v_{2i}) =$$

$$\frac{1}{4}W_{N}^{(i)}(z_{1}^{N}, u_{1,e}^{2i-2} \oplus u_{1,o}^{2i-2}, v_{1,e}^{2i-2} \oplus v_{1,o}^{2i-2}|u_{2i-1} \oplus u_{2i}, v_{2i-1} \oplus v_{2i}) \times \\ \times W_{N}^{(i)}(z_{N+1}^{2N}, u_{1,e}^{2i-2}, v_{1,e}^{2i-2}|u_{2i}, v_{2i})$$

$$(8)$$

Введем новые обозначения, и сократим количество переменных:

– Для вероятности передачи:

$$W_N^{(i)}(X,Y) = W_N^{(i)}(z_1^N, u_1^{i-1}, v_1^{i-1} | u_i = X, v_i = Y)$$

– Логарифмическое отношение правдоподобия для двух пользователей:

$$L_N^{(i)}(X,Y) = \ln \frac{W_N^{(i)}(X,Y)}{W_N^{(i)}(0,0)}$$

Использование таких переменных позволяет значительно ускорить декодирование до сложности $\mathcal{O}((2^m - 1)N\log_2 N)$, где m - количество пользователей. Начнем построение алгоритма эволюции плотности. Для аппроксимации гауссовским распределением нам нужно в качестве метрики выбрать величину, которая на нулевом шаге имеет таковое.Нам подходит логарифмическое отношение правдоподобия:

$$LLR_{N}^{(i)}(X,Y) = \ln \frac{W_{N}^{(i)}(X,Y)}{W_{N}^{(i)}(0,0)}$$

Метод, как и в случае одного пользователя, будем строить для нулевой пары слов. В случае двух пользователей это значительно ускоряет вычисления. Тогда найдем априорное распределение на нулевом шаге рекурсии:

$$LLR_1^{(1)}(z_i, 1, 0) = LLR_1^{(1)}(z_i, 0, 1) = \ln \frac{W_1^{(1)}(z_i|0, 1)}{W_1^{(1)}(z_i|0, 0)} = -\frac{(2z_i - 2)}{\sigma^2}$$
(9)

$$LLR_{1}^{(1)}(z_{i},1,1) = \ln \frac{W_{1}^{(1)}(z_{i}|1,1)}{W_{1}^{(1)}(z_{i}|0,0)} = -\frac{4z_{i}}{\sigma^{2}}$$
(10)

Найдем математическое ожидание $E[LLR_1^{(1)}]$ и дисперсию $D[LLR_1^{(1)}]$. Так как мы считаем, что передавалось нулевое слово, получим для любого $i \in \{1, 2, 3, ..., N\}$:

$$\begin{split} E[LLR_1^{(1)}(z_i, 1, 0)] &= E[LLR_1^{(1)}(z_i, 0, 1)] = -\frac{2}{\sigma^2} \\ D[LLR_1^{(1)}(z_i, 1, 0)] &= D[LLR_1^{(1)}(z_i, 0, 1)] = \frac{4}{\sigma^2} \\ E[LLR_1^{(1)}(z_i, 1, 1)] &= -\frac{8}{\sigma^2} \\ D[LLR_1^{(1)}(z_i, 1, 1)] &= \frac{16}{\sigma^2} \end{split}$$

Преобразуем формулы 7 и 8:

$$LLR_{2N}^{(2i-1)}(z_1^{2N}, 1, 1) = \ln \frac{2(e^{L_{11}} + e^{2L_{01}})}{1 + 2e^{2L_{11}} + e^{L_{01}}}$$
(11)

$$LLR_{2N}^{(2i-1)}(z_1^{2N}, 1, 0) = LLR_{2N}^{(2i-1)}(z_1^{2N}, 0, 1) = \ln \frac{2e^{L_{01}}(1+e^{L_{11}})}{1+2e^{2L_{11}}+e^{L_{01}}}$$
(12)

$$LLR_{2N}^{(2i)}(z_1^{2N}, 0, 1) = 2L_{01}$$
(13)

$$LLR_{2N}^{(2i)}(z_1^{2N}, 1, 1) = 2L_{11}$$
(14)

Где, в случае нулевой пары слов, выполняется:

$$L_{01} = LLR_N^{(i)}(z_1^N u_1^{i-1} v_1^{i-1} | 0, 1) = LLR_N^{(i)}(z_1^N u_1^{i-1} v_1^{i-1} | 1, 0) =$$

$$LLR_N^{(i)}(z_{N+1}^{2N} u_1^{i-1} v_1^{i-1} | 0, 1) = LLR_N^{(i)}(z_{N+1}^{2N} u_1^{i-1} v_1^{i-1} | 1, 0)$$

$$L_{11} = LLR_N^{(i)}(z_1^N u_1^{i-1} v_1^{i-1} | 1, 1) = LLR_N^{(i)}(z_{N+1}^{2N} u_1^{i-1} v_1^{i-1} | 1, 1)$$

Эти равенства позволяют сократить количество переменных с 6 до 2. Далее, считая что D=2Eи используя гауссовскую аппроксимацию найдем $E(LLR_N^{(i)})$ для всех $i\in\{1,2,3,..,N\}$. Считая распределение $LLR_N^{(i)}$ гауссовским, найдем ошибку. Для этого воспользуемся формулами:

$$\begin{split} E[LLR_{2N}^{(2i-1)}(z_1^{2N},1,1)] &= \int \int \ln \frac{2(e^y + e^{2x})}{1 + 2e^y + e^x} \mathcal{P}_{L_{01}}(x) \mathcal{P}_{L_{11}}(y) dx dy \\ E[LLR_{2N}^{(2i-1)}(z_1^{2N},0,1)] &= E[LLR_{2N}^{(2i-1)}(z_1^{2N},1,0)] = \\ &= \int \int \ln \frac{2(e^y + e^{2x})}{1 + 2e^y + e^x} \mathcal{P}_{L_{01}}(x) \mathcal{P}_{L_{11}}(y) dx dy \\ E[LLR_{2N}^{(2i)}(z_1^{2N},0,1)] &= 2E[L_{01}] \\ E[LLR_{2N}^{(2i)}(z_1^{2N},1,1)] &= 2E[L_{11}] \\ e_i &= 1 - [\int_{-\infty}^0 \mathcal{P}_{LLR_N^{(i)}}(0,1)(x) dx]^2 [\int_{-\infty}^0 \mathcal{P}_{LLR_N^{(i)}(1,1)}(x) dx] \end{split}$$

Где \mathcal{P}_X – плотность распределения случайной величины X. Канал, описанный в начале главы, не является симметричным, поэтому в отличие от случая одного пользователя данный метод нельзя использовать для оценки абсолютного значения ошибки в каждом виртуальном канале, но можно брать за основу оценки надежности каналов относительно друг друга. Итак, мы получили однозначно декодируемую конструкцию, и в то же время нашли способ оценить надежность каналов за приемлимое время. Возьмем конструкцию кода из утверждения 3, и будем выбирать наиболее надежные

каналы только из каналов с индексами $i \in A$. Таким образом, мы построили помехоустойчивый однозначно декодируемый код.

2.5 Численные результаты

В этом разделе представлены результаты имитационного моделирования полученной конструкции, для скоростей R = 0.375, R = 0.25 и длин N = 64, N = 128. Для сравнения мы выбрали конструкцию для двух пользователей, основанной на МПП-кодах.



Рис. 2: Сравнение корректирующей способности полярного кода и МПП-кода:
а) на $N=64,\ R=0.25$ б) на $N=64,\ R=0.375$ в) на
 $N=128,\ R=0.375$ г) на $N=128,\ R=0.25$

Итак, построенный нами полярный код, как минимум, не уступает МППкодам в корректирующей способности. Главные преимущества нашей конструкции: возможность различать двух пользователей, не внося дополнительной информации; высокая скорость кодирования и декодирования. Также можно использовать списочный декодер последовательного исключения, который усложнит декодирование до $\mathcal{O}(LN \log N)$, но позволит добиться значительных улучшений помехоустойчивости.

3 Заключение

В работе исследован случай двух пользователей. Для данного случая сформулирована и доказана теорема о максимальной скорости однозначно декодируемого кода. Получена оценка надежности виртуальных каналов, для построения помехоустойчивого Полярного кода. Предложена конструкция однозначно декодируемого Полярного кода. Проведено исследование корректирующей способности предложенной конструкции методами имитационного моделирования. Из полученных численных результатов можно сделать вывод, что данная конструкция имеет корректирующую способность, не уступающюю современным методам построения схем неортогонального множественного доступа. В дальшейшем мы выделяем два направления исследования:

- Улучшение полученной конструкции. Она скорее всего не является оптимальной, так как, возможно, целесообразней допускать неоднозначную декодируемость кода, и тогда мы получим доступ к большему числу помехоустойчивых виртуальных каналов. Но в этом случае необходимо научиться с высокой точностью оценивать количество неоднозначно декодируемых слов.
- Обобщение полученных результатов на случай числа пользователей более двух.

Благодарности

Исследование выполнено в ИППИ РАН при финансовой поддержке РФФИ в рамках научных проектов № 18-07-01409 А, 18-37-00322.

Список литературы

- E. Arıkan.: Channel Polarization: A Method for Constructing Capacity-Achieving Codes for Symmetric Binary-Input Memoryless Channels. IEEE Trans. Inf. Theory, vol. 55, no. 7, pp. 3051–3073, Jul. 2009.
- P. Trifonov.: Efficient Design and Decoding of Polar Codes. IEEE Trans. Inf. Theory, vol.60, no. 11, pp.3221-3227, Nov. 2012.
- 3. H. D. Pfister.: A Brief Introduction to Polar Codes. Supplemental Material for Advanced Channel Coding
- I.Tal, A. Vardy.: List Decoding of Polar Codes. IEEE Trans. Inf. Theory, vol.1, no. 5, pp. 2213-2226, May. 2015.
- 5. Saygun Onay Successive Cancellation Decoding of Polar Codes for the Two-User Binary-Input MAC. IEEE International Symposium on Information Theory, 2013.
- Petar Popovski, Kasper F. Trillingsgaard, Osvaldo Simeone, Giuseppe Durisi: 5G Wireless Network Slicing for eMBB, URLLC, and mMTC: A Communication-Theoretic View,2018.
- 7. Erdal Arikan Polar Coding for the Slepian-Wolf Problem Based on Monotone Chain Rules. IEEE International Symposium on Information Theory Proceedings, 2012.

- 8. Huazi Zhang, Rong Li, Jian Wang, Shengchen Dai, Gongzheng Zhang, Ying Chen, Hejia Luo, Jun Wang Parity-Check Polar Coding for 5G and Beyond, 2018.
- 9. Abdelaziz Amraoui, Sanket Dusad, Ruediger Urbanke Iterative Coding for Discrete-Time Multiple Access Channel, 2002.
- Г.Балицкий, А.Дзись, А.Фролов Сравнение методов построения полярных кодов для каналов с аддитивным белым гауссовским шумом. Труды 41-ой школы-конференции, стр.1-8, 2017.
- 11. A.Dzis, G.Balitskiy, A.Frolov On Methods for Decoding Non-orthogonal Multiple Access Schemes Based on Polar Codes. IEEE European School of Information Theory, 2018.

Разработка методов декодирования схем неортогонального множественного доступа на основе полярных кодов

Глеб Балицкий^{1,2,3}, Андрей Дзись^{1,2,3}, Алексей Фролов^{3,1}

¹ Институт Проблем Передачи Информации РАН, Москва, Россия,

² Московский физико-технический институт, Москва, Россия

 $^{3}\,$ Сколковский институт науки и технологий, Москва, Россия

 $\verb"gleb.balitskiy@phystech.edu", and \verb"rey.dzis@phystech.edu", and \verb"rey.dzis@phystech.edu", and \verb"rey.dzis@phystech.edu", and "rey.dzis@phystech.edu", and "rey$

al.frolov@skoltech.ru

Аннотация В работе рассмотрена задача декодирования полярных кодов для многопользовательских систем в рамках двухпользовательского случая, для которой характерны низкие отношения сигналшум (SNR), короткие кодовые длины и низкие скорости передачи. Был исследован, реализован и оптимизирован декодер, основанный на расширенном методе последовательного исключения (successive cancellation) для случая двухпользовательского канала множественного доступа. Также была предложена конструкция и реализован итеративный декодер. Было произведено сравнение корректирующий способности двух предложенных методов декодирования для разных отношений сигнал-шум.

Ключевые слова: Декодирование полярных кодов, каналы с аддитивным белым гауссовским шумом, декодирование методом последовательного исключения (successive cancelation), итеративное декодирование, множественный доступ, интернет вещей, межмашинное взаимодействие

1 Введение

Полярный код основан на эффекте поляризации канала. Он является первым известным типом кодовых конструкций с субквадратичной сложностью $\mathcal{O}(N \log N)$ кодирования и декодирования. Доказано [1], что кодовые конструкции полярных кодов достигают пропускной способности в дискретных двоичных симметричных каналах без памяти. Высокая эффективность кодовых систем полярных кодов в условиях низкоскоростного режима передачи данных на коротких длинах [6] привела к использованию полярных кодов в контрольном канале сетей пятого поколения (5G control channel) [8].

В рамках сценария межмашинного взаимодействия необходимо рассмотрение коротких низкоскоростных кодовых конструкций. Постановка задачи межмашинного взаимодействия подразумевает одновременную работу огромного числа автономных устройств (сенсоров, датчиков и т.д), подключенных к одной точке доступа, именно поэтому, ввиду ограниченных спектрально-временных ресурсов, необходимо рассмотрение схем неортогонального множественного доступа. Требования энергоэффективности диктуют рассмотрения маленьких отношений сигнал - шум.

2 Полярные коды

2.1 Эффект поляризации

Поляризация канала заключается в том, что из набора N-независимых копий данного B-DMC канала можем создать второй набор N двоичных каналов $W_N^{(i)}: 1 \le i \le N$ так, что с увеличением N для i, для которых $I(W_N^{(i)})$ стремится к 1, пропускная способность соответствующих каналов достигает I(W). Соответственно, для i, для которых $I(W_N^{(i)})$ стремится к 0, достигает значения 1 - I(W).

Введем некоторые обозначения:

Вектор x, состоящий из элементов x_i , $i \in \{1, 2, ..., N\}$ будем обозначать x_1^N . Поэлементную сумму векторов x_1^N и y_1^N в поле GF(2) будем обозначать: $x_1^N \oplus y_1^N$. Для частей векторов, состоящих из четных или нечетных компонент вектора x_1^N введем обозначения: $x_{1,o}^N = (x_1, x_3, x_5, ..., x_{N-1}); x_{1,e}^N = (x_2, x_4, x_6, ... x_N).$

Пропускная способность через переходную вероятность канала W для дискретных двоичных симметричных каналов без памяти (B-DMC):

$$I(W) \triangleq \sum_{y \in \mathcal{Y}} \sum_{x \in \mathcal{X}} \frac{1}{2} W(y|x) \log \frac{W(y|x)}{\frac{1}{2}W(y|0) + \frac{1}{2}W(y|1)}$$
(1)

Эффект поляризации состоит из объединения и разделения каналов.

Объединение каналов

Данный процесс заключается в комбинировании N независимых копий данного канала $W : \mathcal{X} \to \mathcal{Y}$, где \mathcal{X} и \mathcal{Y} входной и выходной алфавиты соответственно, с последующим получением нового канала, который обозначим:

$$W_N: \mathcal{X}^N \to \mathcal{Y}^N$$

где $N = 2^n, n = 1, 2, 3, ...$

Получение нового канала W_N представимо в рекурсивного процесса. На нулевом шаге имеем одну копию канала W. Откуда переходная вероятность нулевого шага рекурсии равна $W_1 \triangleq W(y|x)$. Первый шаг рекурсии изображен на рис. 1.



Рис. 1: Вид канала W_2

Результатом первого шага является построение канала:

$$W_2: \mathcal{X}^2 \to \mathcal{Y}^2$$

с переходными вероятностями:

$$W_2(y_1, y_2|u_1, u_2) = W(y_1|u_1 \oplus u_2)W(y_2|u_2)$$

Результатом следующего, второго шага рекурсии является канал:

$$W_4: \mathcal{X}^4 \to \mathcal{Y}^4$$

для него переходная вероятность:

 $W_4(y_1, y_2, y_3, y_4 | u_1, u_2, u_3, u_4) = W_2(y_1, y_2 | u_1 \oplus u_2, u_3 \oplus u_4) W_2(y_3, y_4 | u_2, u_4)$



Рис. 2: Вид канала W₄

Как видно из рис. 2, в преобразовании используется перестановка, описываемая матрицой R_4 , которая преобразует вектор $(u_1, u_2, u_3, u_4) \rightarrow (u_1, u_3, u_2, u_4)$. Переходная вероятность на n шаге рекурсии (см. рис. 3), где $n = \log_2 N$:



Рис. 3: Вид канала W_N , составленного из $W_{N/2}$

$$W_N = W_{N/2}(y_1^{N/2} | u_{1,e}^N \oplus u_{1,o}^N) W_{N/2}(y_{N/2}^N | u_{1,e}^N) = W^N(y_1^N | u_1^N G_N)$$
(2)

где G_N - порождающая матрица, общий вид которой:

$$G_N = (I_{N/2} \otimes G_2) R_N (I_2 \otimes G_{N/2})$$

а ${\cal R}_N$ - матрица перестановок, общий вид которой:

$$R_N: (u_1, u_2, u_3, .., u_N) \to (u_1, u_3, .., u_{N-1}, u_2, u_4, .., u_N)$$

Разделение канала

Процедура разделения канала представляет собой разбиение получившегося при объединении канала W_N на N виртуальных двоичных каналов с вероятностями передачи.

$$W_N^{(i)}(y_1^N, u_1^{i-1} | u_i) = \sum_{\substack{u_{i+1}^N \in \mathcal{X}^{N-i}}} \frac{1}{2^{N-1}} W_N(y_1^N | u_1^N)$$
(3)

2.2 Декодирование полярных кодов

Предположим, что у нас есть полярный код с параметрами $(N, K, \mathcal{A}, u_{\mathcal{A}_c})$. Пусть информационный двоичный вектор u_1^N кодируется в x_1^N , который,

передается по каналу W_N . y_1^N - выходной вектор канала W_N . Тогда за-дача декодера заключается в нахождении вектора решений \hat{u}_1^N на основе известного \mathcal{A} , принятого вектора y_1^N , значений "замороженных битов" $u_{\mathcal{A}_c}$. Декодер избегает ошибок на "замороженных битах т.е задача декодера, заключается в нахождении вектора приближения информационных символов.

Для метода последовательного исключения задача декодера заключается в подсчете вектора \hat{u}_1^N по следующей формуле:

$$\hat{u}_i \equiv \begin{cases} u_i, & \text{если } i \in \mathcal{A}_c \\ h_i(y_1^N, \hat{u}_1^{i-1}), & \text{если } i \in \mathcal{A} \end{cases}$$

*h*_i - функция принятия решения, определяемая как:

$$h_{i} \equiv \begin{cases} 0, & \text{если } \frac{W_{N}^{(i)}(y_{1}^{N}, \hat{u}_{1}^{i-1}|0)}{W_{N}^{(i)}(y_{1}^{N}, \hat{u}_{1}^{i-1}|1)} \ge 1\\ 1, & \text{в противном случае} \end{cases}$$
(4)

Утверждение Будем считать, что декодер совершил ошибку на блоке, если вектора $u_1^N \neq \hat{u}_1^N$, что эквивалентно $u_A \neq \hat{u}_A$. Определение: Отношением правдоподобия (Likelihood Ratio, LR) назы-

вается величина:

$$L_N^{(i)}(y_1^N, \hat{u}_1^{i-1}) \equiv \frac{W_N^{(i)}(y_1^N, \hat{u}_1^{i-1} | \hat{u}_i = 0)}{W_N^{(i)}(y_1^N, \hat{u}_1^{i-1} | \hat{u}_i = 1)}$$

Из формул 2, 3, согласно [1], для вероятностей передачи четных и нечетных подканалов можно записать следующие рекурсивные формулы:

$$W_{2N}^{(2i-1)}(y_1^{2N}, \hat{u}_1^{2i-2} | \hat{u}_{2i-1}) = \sum_{\hat{u}_{2i}} W_N^{(i)}(y_1^N, \hat{u}_{1,e}^{(2i-2)} \oplus \hat{u}_{1,o}^{(2i-2)} | \hat{u}_{2i} \oplus \hat{u}_{2i-1}) W_N^{(i)}(y_{N+1}^{2N}, \hat{u}_{1,e}^{(2i-2)} | \hat{u}_{2i}) W_{2N}^{(2i)}(y_1^{2N}, \hat{u}_1^{2i-1} | \hat{u}_{2i}) = W_N^{(i)}(y_1^N, \hat{u}_{1,e}^{(2i-2)} \oplus \hat{u}_{1,o}^{2i-2} | \hat{u}_{2i} \oplus \hat{u}_{2i-1}) W_N^{(i)}(y_{N+1}^{2N}, \hat{u}_{1,e}^{(2i-2)} | \hat{u}_{2i})$$
(6)

Тогда формулы для отношения правдоподобия можно переписать в виде:

$$L_{N}^{(2i-1)}(y_{1}^{N},\hat{u}_{1}^{2i-2}) = \frac{L_{N/2}^{(i)}(y_{1}^{N/2},\hat{u}_{1,o}^{2i-2} \oplus \hat{u}_{1,e}^{2i-2})L_{N/2}^{(i)}(y_{N/2+1}^{N},\hat{u}_{1,e}^{2i-2}) + 1}{L_{N/2}^{(i)}(y_{1}^{N/2},\hat{u}_{1,o}^{2i-2} \oplus \hat{u}_{1,e}^{2i-2}) + L_{N/2}^{(i)}(y_{N/2+1}^{N},\hat{u}_{1,e}^{2i-2})}$$
(7)

И

$$L_{N}^{(2i)}(y_{1}^{N},\hat{u}_{1}^{2i-1}) = [L_{N/2}^{(i)}(y_{1}^{N/2},\hat{u}_{1,o}^{2i-2} \oplus \hat{u}_{1,e}^{2i-2})]^{1-2\hat{u}_{2i-1}}L_{N/2}^{(i)}(y_{N/2+1}^{N},\hat{u}_{1,e}^{2i-2})$$
(8)

Определение: Логарифмическим отношением правдоподобия (LLR) называется отношение:

$$LLR_{N}^{(i)}(y_{1}^{N}, \hat{u}_{1}^{i-1}) = \ln \frac{W_{N}^{(i)}(y_{1}^{N}\hat{u}_{1}^{i-1}|\hat{u}_{i}=0)}{W_{N}^{(i)}(y_{1}^{N}\hat{u}_{1}^{i-1}|\hat{u}_{i}=1)}$$
(9)

Тогда представленные выше формулы 7, 8 перепишутся с учетом 9 в виде:

$$LLR_{N}^{(2i-1)}(y_{1}^{N},\hat{u}_{1}^{2i-2}) = 2\tanh^{-1}(\tanh(\frac{LLR_{e}}{2})\tanh(\frac{LLR_{o}}{2}))$$
(10)

$$LLR_N^{(2i)}(y_1^N, \hat{u}_1^{2i-2}) = LLR_o + (-1)^{\hat{u}_{2i-1}}LLR_e$$
(11)

здесь для простоты введены обозначения:

$$LLR_e = L_{N/2}^{(i)}(y_1^{N/2}, \hat{u}_{1,e}^{(2i-2)} \oplus \hat{u}_{1,o}^{(2i-2)})$$
$$LLR_o = L_{N/2}^{(i)}(y_{N/2}^N, \hat{u}_{1,e}^{(2i-2)})$$

Тогда условие 4 для функции принятия решения примет вид:

$$h_{i} \equiv \begin{cases} 0, & \text{если } LLR_{N}^{(i)}(y_{1}^{N}, \hat{u}_{1}^{i-1}) \geq 0\\ 1, & \text{в противном случае} \end{cases}$$

Из сказанного выше видно, что подсчет $LLR_N^{(i)}$ сводится к подсчету ровно N логарифмических отношений правдоподобия на длине $\frac{N}{2}$, подсчет $LLR_{N/2}^{(i)}$ аналогично сводится к подсчету $\frac{N}{2}$ раз $LLR_{N/4}^{(i)}$. Таким образом, подсчет имеет следующую схему $N \to N/2 \to N/4 \to ... \to 1$, где для N = 1 выражение вычисляется напрямую. Сложность декодирования $\mathcal{O}(NlogN)$.

3 Полярные коды для случая двух пользователей

3.1 Модель канала множественного доступа

Рассмотрим систему из одинаковых пользователей X и Y, передающих данные в канал множественного доступа с гауссовским шумом на одинаковой мощности (рис. 4). Обозначим данные пользователя X за u_1^N , а Y за v_1^N . В канал передаются кодовые слова x_1^N , y_1^N , полученные из данных путем кодирования полярным кодом, которые после двухпозиционной фазовой модуляции, кодовые слова передаются по каналу:

$$x_1^N = u_1^N G_x$$
$$y_1^N = v_1^N G_y$$

где G_x и G_y - порождающие матрицы для соответствующих пользователей.

На выходе канала получаем вектор
$$z_1^{iv}$$
, для которого можем записать:

$$z_1^N = y_1^N + x_1^N + \mathcal{N}(0, \sigma^2)$$

Задача состоит в декодировании информационных векторов (u_1^N,v_1^N) по полученному вектору $z_1^N.$



Рис. 4: Схема неортогонального множественного доступа для двух пользователей

Обозначим входные алфавиты двух пользователей как $\mathcal{X} \in \{0,1\}$ и $\mathcal{Y} \in \{0,1\}$. Тогда канал неортогонального множественного доступа для двух пользователей имеет вид:

$$W: \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$$

Обозначим за W(z|x,y) переходную вероятность канала. Процесс поляризации канала для двух пользователей разбивается на две процедуры.

1. Объединение каналов. Следуя аналогичной логике, что и в случае одного пользователя получим новый канал $W_N : \mathcal{X}^N \times \mathcal{Y}^N \to \mathcal{Z}^N$ с вероятностями передачи:

$$W_N(z_1^N | u_1^N, v_1^N) = \prod_{i=1}^{i=N} W(z_i | (uG_N)_i, (vG_N)_i) = W^N(z_1^N | u_1^N G_N, v_1^N G_N)$$

2. **Разделение канала.** Разобьем получившийся канал W_N на N виртуальных каналов. Для полученных каналов вероятности передачи:

$$W_N^{(i)}(z_1^N, u_1^{i-1}, v_1^{i-1} | u_i, v_i) = \sum_{u_{i+1}^N, v_{i+1}^N \in \mathcal{X}^{N-i}} \left(\frac{1}{2^N - 1}\right)^2 W_N(z_1^N | u_1^N, v_1^N)$$
(12)

3.2 Декодирование методом последовательного исключения

Идея декодирования методом последовательного исключения была предложена в [5], которая заключается в принятии решения декодером по парам (u_i, v_i) , исходя из значения для данного *i* вектора z_i . Переобозначим вектор z_1^N как y_1^N . Задача декодера вычислить векторы решений \hat{u}_1^N и \hat{v}_1^N . Формулы переходных вероятностей для случая двух пользователей примут вид:

$$W_{2N}^{(2i-1)}(y_1^{2N}, \hat{u}_1^{2i-2}, \hat{v}_1^{2i-2} | X, Y) =$$

$$\sum_{\hat{u}_{2i},\hat{v}_{2i}} \frac{1}{4} W_N^{(i)}(y_1^N, \hat{u}_{1,e}^{2i-2} \oplus \hat{u}_{1,o}^{2i-2}, \hat{v}_{1,e}^{2i-2} \oplus \hat{v}_{1,o}^{2i-2} | X \oplus \hat{u}_{2i}, Y \oplus \hat{v}_{2i}) W_N^{(i)}(y_{N+1}^{2N}, \hat{u}_{1,e}^{2i-2}, \hat{v}_{1,e}^{2i-2} | \hat{u}_{2i}, \hat{v}_{2i})$$

$$(13)$$

$$W_{2N}^{(2i)}(y_1^{2N}, \hat{u}_1^{2i-1}, \hat{v}_1^{2i-1} | X, Y) = \frac{1}{4} W_N^{(i)}(y_1^N, \hat{u}_{1,e}^{2i-2} \oplus \hat{u}_{1,o}^{2i-2}, \hat{v}_{1,e}^{2i-2} \oplus \hat{v}_{1,o}^{2i-2} | \hat{u}_{2i-1} \oplus X, \hat{v}_{2i-1} \oplus Y) W_N^{(i)}(y_{N+1}^{2N}, \hat{u}_{1,e}^{2i-2}, \hat{v}_{1,e}^{2i-2} | X, Y)$$

$$\tag{14}$$

Тогда, записав вероятности передачи, мы можем ввести понятие отношения правдоподобия как:

$$L_N^{(i)}(y_1^N, \hat{u}_1^{i-1}, \hat{v}_1^{i-1}, X, Y) = \frac{W(X, \ Y)}{W(0, \ 0)}$$

где

$$W(X, Y) = \begin{cases} W_N^{(2i-1)}(y_1^N, \hat{u}_1^{2i-2}, \hat{v}_1^{2i-2} | \hat{u}_{2i-1} = X, \ \hat{v}_{2i-1} = Y) &, \text{ для нечетных i} \\ W_N^{(2i)}(y_1^N, \hat{u}_1^{2i-1}, \hat{v}_1^{2i-1} | \hat{u}_{2i} = X, \ \hat{v}_{2i} = Y) &, \text{ для четных i} \end{cases}$$
(15)

Тогда, используя формулы 15, перепишем отношения правдоподобия в виде 7, 8:

- для четных индексов:

$$L_N^e(X, Y) = L_N^{(2i)}(X, Y) = \frac{W_N^{(2i)}(y_1^N, \hat{u}_1^{2i-1}, \hat{v}_1^{2i-1} | \hat{u}_{2i} = X, \ \hat{v}_{2i} = Y)}{W_N^{(2i)}(y_1^N, \hat{u}_1^{2i-1}, \hat{v}_1^{2i-1} | \hat{u}_{2i} = 0, \ \hat{v}_{2i} = 0)}$$
(16)

Для удобства записи, введем обозначение:

$$W_{N_1}^e(X, Y) = W_{N/2}^{(i)}(y_1^{N/2}, \hat{u}_{1,e}^{2i-2} \oplus \hat{u}_{1,o}^{2i-2}, \hat{v}_{1,e}^{2i-2} \oplus \hat{v}_{1,o}^{2i-2}|X, Y)$$
(17)

$$W_{N_2}^e(X, Y) = W_{N/2}^{(i)}(y_{N/2+1}^N, \hat{u}_{1,e}^{2i-2}, \hat{v}_{1,e}^{2i-2} | X, Y)$$
(18)

Из данных обозначений вытекают следующих два обозначения:

$$L_{N_1}^e(X, Y) = \frac{W_{N_1}^e(X, Y)}{W_{N_1}^e(0, 0)}$$
(19)

$$L_{N_2}^e(X, Y) = \frac{W_{N_1}^e(X, Y)}{W_{N_1}^e(0, 0)}$$
(20)

Тогда подставляя 19, 20 в 16:

$$L_{N}^{e}(X, Y) = L_{N_{2}}^{e}(X, Y) \frac{L_{N_{1}}^{e}(\hat{u}_{2i-1} \oplus X, \hat{v}_{2i-1} \oplus Y)}{L_{N_{1}}^{e}(\hat{u}_{2i-1}, \hat{v}_{2i-1})}$$
(21)

– для нечетных индексов

Проделаем аналогичные выкладки для случая нечетных индексов. Для нечетных индексов отношение правдоподобия может быть записано в следующем виде:

$$L_N^o(X, Y) = L_N^{(2i-1)}(X, Y) = \frac{W_N^{(2i-1)}(y_1^N, \hat{u}_1^{2i-2}, \hat{v}_1^{2i-2} | \hat{u}_{2i-1} = X, \ \hat{v}_{2i-1} = Y)}{W_N^{(2i-1)}(y_1^N, \hat{u}_1^{2i-2}, \hat{v}_1^{2i-2} | \hat{u}_{2i-1} = 0, \ \hat{v}_{2i-1} = 0)}$$
(22)

Введем обозначения, аналогичные 17, 18,19, 20, для случая нечетных индексов, тогда выражение для отношения правдоподобия в случае нечетных индексов имеет вид:

$$L_{N}^{o}(X, Y) = \frac{L_{N_{1}}(0 \oplus X, 0 \oplus Y) + L_{N_{1}}(0 \oplus X, 1 \oplus Y)L_{N_{2}}(0, 1) + L_{N_{1}}(1 \oplus X, 0 \oplus Y)L_{N_{2}}(1, 0)}{1 + L_{N_{1}}(0, 1)L_{N_{2}}(0, 1) + L_{N_{1}}(1, 0)L_{N_{2}}(1, 0) + L_{N_{1}}(1, 1)L_{N_{2}}(1, 1)} + \frac{L_{N_{1}}(1 \oplus X, 1 \oplus Y)L_{N_{2}}(1, 0) + L_{N_{1}}(1, 1)L_{N_{2}}(1, 1)}{1 + L_{N_{1}}(0, 1)L_{N_{2}}(0, 1) + L_{N_{1}}(1, 0)L_{N_{2}}(1, 0) + L_{N_{1}}(1, 1)L_{N_{2}}(1, 1)}$$
(23)

Алгоритм декодирования основанный на поиске большего значения $L_N^{(i)}(X,Y)$ неэффективен ввиду перебора. Можно заметить, что пара отношений правдоподобия $L_N^{2i-1}(y_1^N, u^{2i-2}, v^{2i-2}|u_{2i-1}, v_{2i-1})$ и $L_N^{2i}(y_1^N, u_1^{2i-1}, v_1^{2i-1}|u_{2i}, v_{2i})$ вычисляется из пары вероятностей: $W_{N/2}^{(i)}(y_{N/2+1}^N, \hat{u}_{1,e}^{2i-2}, \hat{v}_{1,e}^{2i-2}|X,Y), W_{N/2}^{(i)}(y_1^{N/2}, \hat{u}_{1,e}^{2i-2} \oplus \hat{u}_{1,o}^{2i-2}|X,Y). Аналогично [1], для вычисления одного <math display="inline">L_N^{(i)}(X,Y)$ требуется 3N (в отличие от N вычислений для одного пользователя) вычислений $L_{N/2}^{(i)}(X,Y)$. Дальнейшие вычисления по рекурсивной схеме сводятся к поэтапному подсчету отношений для длины $N \to \frac{N}{2} \to \frac{N}{4} \to \ldots \to 1$, где для единичной длины мы вычисляем:

$$L_1^{(1)}(y_i) = \frac{W(y_i|u_i, v_i)}{W(y_i|u_i = 0, v_i = 0)}$$

которое для канала с гауссовским шумом, вычисляется напрямую. Таким образом, мы добиваемся взаимосвязи при пересчете отношения правдоподобия, что уменьшает сложность подсчета набора $L_N^{(i)}(X,Y)$ до $\mathcal{O}((2^m - 1)NlogN)$, где m - количество пользователей.

Определение Логарифмический отношением правдоподобия для случая двух пользователей называется отношение:

$$LLR_N^{(i)}(y_1^N, \hat{u}_1^{i-1}, \hat{v}_1^{i-1}, X, Y) = \ln \frac{W(X, Y)}{W(0, 0)}$$

где W(X, Y) определено в 15. В отличие от случая для одного пользователя, формулы для логарифмического отношения правдоподобия не упрощаются. Переход к логарифмическому отношению правдоподобия имеет большое значение в практической реализации декодера. Рассмотрим на примере 23. Тогда:

$$LLR_N^o(X, Y) =$$

$$\ln \left(\frac{e^{LLR_{N_1}(0\oplus X, \ 0\oplus Y)} + e^{LLR_{N_1}(0\oplus X, \ 1\oplus Y)} e^{LLR_{N_2}(0, \ 1)}}{1 + e^{LLR_{N_1}(0, \ 1)} e^{LLR_{N_2}(0, \ 1)} + e^{LLR_{N_1}(1, \ 0)} e^{LLR_{N_2}(1, \ 0)} + e^{LLR_{N_1}(1, \ 1)} e^{LLR_{N_2}(1, \ 1)}} + \frac{e^{LLR_{N_1}(1\oplus X, \ 0\oplus Y)} e^{LLR_{N_2}(1, \ 0)} + e^{LLR_{N_1}(1\oplus X, \ 1\oplus Y)} e^{LLR_{N_2}(1, \ 1)}}{1 + e^{LLR_{N_1}(0, \ 1)} e^{LLR_{N_2}(0, \ 1)} + e^{LLR_{N_1}(1, \ 0)} e^{LLR_{N_2}(1, \ 0)} + e^{LLR_{N_1}(1, \ 1\oplus Y)} e^{LLR_{N_2}(1, \ 1)}}} \right)$$

где использованные обозначения логарифмических отношений правдоподобия соответствуют определенным выше отношениям правдоподобия 19, 20. LLR более устойчиво к переполнениям, однако в примере с 23 приведение выражений под знаком логарифма может привести к переполнению, что решается выносом высшей степени экспоненты.

3.3 Итеративное декодирование

Аналогично 3.2 выходом канала является вектор z_1^N - поэлементная сумма кодовых слов u_1^N и v_1^N пользователей X и Y соответственно с наложением белого аддитивного гауссовского шума. Принятый z_1^N отправляется на списочный декодер длины l пользователя X, где происходит декодирование вектора z_1^N . В качестве метрики используется логарифмическое отношение правдоподобия, которое для пользователя X на нулевом шаге рекурсии будет иметь вид:

$$LLR_1^{(1)}(z_i) = \ln \frac{W(z_i|x_i=0)}{W(z_i|x_i=1)} = \ln \frac{\sum_{y_i=\{0,1\}} P(y_i)P(z_i|x_i=0,y_i)}{\sum_{y=\{0,1\}} P(y_i)P(z^i|x_i=1,y_i)}$$

где вероятности $P(y_i)$ считаем равными 1/2. На выходе списочного декодера пользователя X имеем l наиболее вероятных слов $(\hat{u}_1^N(1), \hat{u}_1^N(2), ..., \hat{u}_1^N(l-1), \hat{u}_1^N(l))$ с вероятностями $(P_1, P_2, ..., P_{l-1}, P_l)$. Далее, на пользователя X происходит происходит происходит кодирование полученного этапом ранее списка длины l наиболее вероятных слов $(\hat{u}_1^N(1), \hat{u}_1^N(2), ..., \hat{u}_1^N(l-1), \hat{u}_1^N(l)) \to (\hat{x}_1^N(1), \hat{x}_1^N(2), ..., \hat{x}_1^N(l-1), \hat{x}_1^N(l))$, которые имеют вероятности $(P_1, P_2, ..., P_{l-1}, P_l)$. Введем обозначение $\hat{x}_1^N(1) \equiv \hat{x}(1)$, тогда *i*тый элемент вектора $\hat{x}_1^N(1)$ выражается как $\hat{x}(1)_i$. Для априорных вероятностей выражения имеют вид:

$$P(x_i = 1) = \frac{\sum_{n=1}^{l} \hat{x}(n)_i P_i}{\sum_{n=1}^{l} P_l}$$
$$P(x_i = 0) = 1 - P(x_i = 1)$$

Тогда вектор z_1^N , с учетом знания априорных вероятностей $P(x_i)$, полученных ранее, декодируется пользователем Y. На нулевом шаге рекурсии логарифмическое отношение правдоподобия будет иметь вид:

$$LLR_{1}^{(1)}(z_{i}) = \ln \frac{W(z_{i}|y_{i}=0)}{W(z_{i}|y_{i}=1)} = \ln \frac{\sum_{x_{i}=\{0,1\}} P(x_{i})P(z_{i}|y_{i}=0,x_{i})}{\sum_{x_{i}=\{0,1\}} P(x_{i})P(z_{1}^{N}|y_{i}=1,x_{i})}$$

Дальнейшее декодирование в точности повторяет данные два шага, пересчитывая вероятности $P(x_i)$ и $P(y_i)$. Сложность декодера $\mathcal{O}(lkN \log N)$, где k - количество итераций.

4 Численные результаты

Сравним две представленных в работе реализации декодера для полярных кодов в случае канала двухпользовательского неортогонального доступа с гауссовским шумом. Для итеративного декодера параметры зададим следующие: k = 10 - количество итераций, l = 8 - длина списка для однопользовательских списочных декодеров. Выбраны следующие параметры кодов: длины $N_1 = 64$ и $N_2 = 128$, скорости $R_1 = 0.2$, $R_2 = 0.25$, $R_3 = 0.375$. Построим графики зависимости ошибки на блок (ОНБ) от отношения сигнал-шум (ОСШ) в логарифмическом масштабе.



Рис. 5: Сравнение корректирующей способности:
а) на $N=64,\ R=0.2$ б) на $N=64,\ R=0.25$ в) на
 $N=64,\ R=0.375$ г) на $N=128,\ R=0.2$ д) на $N=128,\ R=0.25$ е) на
 $N=128,\ R=0.375$

На малых скоростях (R = 0.2) итеративный декодер обладает большей корректирующей способностью при одинаковых SNR. Однако, с увеличением скорости (R = 0.25, R = 0.375) можно заметить, что отличия в корректирующей способности становится меньше.

5 Заключение

В данной работе были исследованы методы декодирования схем неортогонального доступа на основе полярных кодов для случая двух пользователей. Был обобщен метод последовательного исключения для декодирования на случай двух пользователей и реализован декодер с общей сложностью $\mathcal{O}((2^m - 1)N \log N)$. Кроме того, была представлена реализация итеративного декодера, и проведено сравнение двух полученных методов путем сравнения их корректирующей способности на некоторых диапазонах конфигураций. Результаты, взятые за основу данной работы, были представлены на ИТИС 2017 [10], ESIT 2018 [11]. В дальнейших исследованиях планируется использовать списочный декодера последовательного исключения на случай *m*-пользовательского канала множественного доступа с последующей реализацией и оптимизации конструкции.

Благодарности

Исследование выполнено в ИППИ РАН при финансовой поддержке РФФИ в рамках научных проектов № 18-07-01409 А, 18-37-00322.

Список литературы

- E. Arıkan.: Channel Polarization: A Method for Constructing Capacity-Achieving Codes for Symmetric Binary-Input Memoryless Channels. IEEE Trans. Inf. Theory, vol. 55, no. 7, pp. 3051–3073, Jul. 2009.
- P. Trifonov.: Efficient Design and Decoding of Polar Codes. IEEE Trans. Inf. Theory, vol.60, no. 11, pp.3221-3227, Nov. 2012.
- 3. H. D. Pfister.: A Brief Introduction to Polar Codes. Supplemental Material for Advanced Channel Coding
- I.Tal, A. Vardy.: List Decoding of Polar Codes. IEEE Trans. Inf. Theory, vol.1, no. 5, pp. 2213-2226, May. 2015.
- 5. Saygun Önay Successive Cancellation Decoding of Polar Codes for the Two-User Binary-Input MAC. IEEE International Symposium on Information Theory, 2013.
- Petar Popovski, Kasper F. Trillingsgaard, Osvaldo Simeone, Giuseppe Durisi: 5G Wireless Network Slicing for eMBB, URLLC, and mMTC: A Communication-Theoretic View,2018.
- 7. Erdal Arikan Polar Coding for the Slepian-Wolf Problem Based on Monotone Chain Rules. IEEE International Symposium on Information Theory Proceedings, 2012.
- 8. Huazi Zhang, Rong Li, Jian Wang, Shengchen Dai, Gongzheng Zhang, Ying Chen, Hejia Luo, Jun Wang Parity-Check Polar Coding for 5G and Beyond, 2018.
- 9. Abdelaziz Amraoui, Sanket Dusad, Ruediger Urbanke Iterative Coding for Discrete-Time Multiple Access Channel, 2002.
- Г.Балицкий, А.Дзись, А.Фролов Сравнение методов построения полярных кодов для каналов с аддитивным белым гауссовским шумом. Труды 41-ой школы-конференции, стр.1-8, 2017.
- A.Dzis, G.Balitskiy, A.Frolov On Methods for Decoding Non-orthogonal Multiple Access Schemes Based on Polar Codes. IEEE European School of Information Theory, 2018.

Алгоритм планирования радиоресурсов в мультисервисных беспроводных сетях пятого поколения*

А.Н. Красилов, М.В. Любогощев, Е.М. Хоров

ИППИ РАН E-mails: {krasilov, liubogoshchev, khorov}@iitp.ru

Аннотация Одной из новых концепций, предложенных для построения сетей связи пятого поколения (5G), является концепция виртуализации сетевых ресурсов, позволяющая обеспечить совместное обслуживание потоков данных с различными требованиями к качеству обслуживания. Базовым алгоритмом разделения ресурсов беспроводного канала между виртуальными подсетями является фиксированное разделение частотно-временных ресурсов. Однако данный алгоритм неэффективно использует ресурсы сети. Например, если одна из виртуальных подсетей временно не будет иметь данных на передачу, то выделенные для ее обслуживания ресурсы не могут быть использованы для обслуживания других подсетей. В данной работе предложен новый алгоритм планирования радиоресурсов в мультисервисных сетях и показано, что использование данного алгоритма позволяет повысить качество обслуживания пользователей по сравнению с существующими алгоритмами, представленными в литературе.

1 Введение

В сентябре 2015 года сектор радиосвязи Международного союза электросвязи (ITU-R) опубликовал требования к сотовым сетям пятого поколения (5G) [1]. Согласно данным требованиям, сети 5G должны обеспечивать обслуживание различных типов трафика, обладающих различными и часто несовместимыми требованиями к качеству обслуживания. В частности, даже существующие типы генерируемого на мобильных устройствах трафика накладывают различные требования на качество обслуживания, не говоря уже о новых типах трафика, рассматриваемых в контексте сетей 5G. Например, при передаче голосовых потоков (Voice over IP) не требуется обеспечивать высокую пропускную способность сети, но при этом требуется выполнять строгие ограничения на задержку при передаче пакетов. С другой стороны, обслуживание веб-трафика не предполагает ограничений на время доставки отдельных пакетов, но вместо этого требует минимизации

^{*} Исследование выполнено в ИППИ РАН за счет гранта Российского научного фонда (проект № 16-19-10687)

времени загрузки всей веб-страницы, т.е. максимизации средней скорости передачи данных. При обслуживании видеопотоков, в свою очередь, требуется высокая и постоянная во времени скорость передачи данных для загрузки видеоизображения высокого качества и обеспечения непрерывного воспроизведения видеоизображения.

Для выполнения гетерогенных требований к качеству обслуживания различных типов трафика была предложена концепция виртуализации сетевых ресурсов, предполагающая возможность создания нескольких независимых виртуальных подсетей, использующих ресурсы общей физической сети. Данная концепция привлекла повышенное внимание исследователей. В частности, был опубликован ряд статей, описывающих архитектуру решений, позволяющих создавать виртуальные подсети и управлять их работой [2–4]. Однако в данных работах не предлагается конкретных алгоритмов для разделения ресурсов между созданными виртуальными подсетями, а от выбора данного алгоритма напрямую зависят как возможность выполнения требований к качеству обслуживания различных виртуальных подсетей, так и общая производительность сети (например, ее пропускная способность). Так как в беспроводных сетях узким местом при передаче данных чаще всего является беспроводной канал связи между базовой станцией и клиентским устройством, то значительное влияние на производительность сети передачи данных в целом оказывает эффективность алгоритма разделения радиоресурсов между виртуальными подсетями.

Базовым подходом к разделению радиоресурсов между виртуальными подсетями является фиксированное разделение: каждой из виртуальных подсетей выделяется фиксированный блок частотно-временных ресурсов, в котором осуществляется передача данных исключительно для данной подсети. Основным недостатком данного метода является его низкая спектральная эффективность: из-за частотно-селективных замираний сигнала в определенные моменты времени может быть не эффективно использовать для обслуживания выбранной подсети заранее выделенный ей частотно-временной блок. Кроме того, данный подход не учитывает того факта, что одна из подсетей может временно не иметь данных на передачу. Однако с точки зрения эффективности использования ресурсов беспроводного канала, было бы выгодно отдать оставшиеся ресурсы другой подсети.

Последняя проблема частично решается алгоритмом NVS (Network Virtualization Substrate), описанным в статье [5]. Данный алгоритм предполагает, что подсети i в долгосрочной перспективе выделяется определенная доля канальных ресурсов s_i^{res} . В момент времени t обслуживается подсеть $\hat{i} = \arg \max \frac{s_i^{res}}{s_i(t)}$, где $s_i(t)$ — это доля ресурсов, выделенная подсети i в некотором временном окне. Например, $s_i(t)$ может вычисляться, как экспоненциально взвешенное скользящее среднее доли ресурсов, выделявшейся подсети:

$$s_i(t) = (1 - \frac{\Delta t}{T_{win}})s_i(t - \Delta t) + \frac{\Delta t}{T_{win}}\frac{A(i)}{\sum_i s_i^{res}},$$
где T_{win} определяет характерный размер окна усреднения, A(i) — доля канальных ресурсов, выделенных подсети i в интервале $(t - \Delta t, t), \Delta t \leq T_{win}$.

В данной работе предложен новый алгоритм планирования радиоресурсов для совместного обслуживания двух типов трафика: веб-трафика и видеопотоков. Выбор данных типов трафика обусловлен тем, что, согласно исследованию компании Cisco Systems (см. [6]), данные типы трафика составляют порядка 90% от всего трафика, генерируемого мобильными устройствами, и их доля продолжает расти. При помощи имитационного моделирования в среде NS-3 показано, что по сравнению существующими алгоритмами разработанный алгоритм позволяет значительно повысить качество обслуживания пользователей в сетях с гетерогенным трафиком.

Дальнейшее изложение работы построено следующим образом. В разделе 2 описан предложенный алгоритм. Результаты сравнения предложенного алгоритма с существующими алгоритмами представлены в разделе 3. В разделе 4 резюмируются основные результаты работы.

2 Алгоритм планирования радиоресурсов

В данной работе рассматривается сеть с централизованным управлением, в которой передача данных осуществляется в слотах длительностью Δt (например, в сетях LTE $\Delta t = 1$ мс). В начале каждого слота на базовой станции запускается алгоритм планирования радиоресурсов, выбирающий поток данных для обслуживания в данном слоте. Задача алгоритма планирования радиоресурсов для совместного обслуживания нескольких типов трафика состоит из нескольких подзадач.

Во-первых, необходимо классифицировать потоки данных по типу трафика для выявления параметров потоков и их требований к качеству обслуживания. В данной работе считается, что классификация трафика осуществляется при помощи специального протокола уровня приложений (подобного протоколу SAND [7]), позволяющего организовать обмен служебной информацией о типе и требованиях трафика между клиентскими устройствами и различными сетевыми узлами (например, базовой станцией).

Во-вторых, необходимо назначить каждой из виртуальных подсетей долю канальных ресурсов s_i^{res} , которая будет выделена данной подсети в долгосрочной перспективе. В данной работе предполагается выделять каждой подсети долю ресурсов, пропорциональную среднему числу активных потоков в ней. Данное значение вычисляется, как экспоненциально взвешенное скользящее среднее доли активных потоков:

$$s_i^{res}(t) = (1 - \frac{\Delta t}{T_{win}})s_i^{res}(t - \Delta t) + \frac{\Delta t}{T_{win}}\frac{N_i}{\sum_j N_j},$$

где N_i — число активных потоков в подсети *i* в интервале $(t - \Delta t, t)$.

В-третьих, необходимо осуществлять планирование радиоресурсов таким образом, чтобы выполнить заданные долгосрочные ограничения на доли ресурсов, выделенные каждой подсети и при этом обеспечить выполнение требований к качеству обслуживания в каждой из подсетей.



Рис. 1: Разделение ресурсов между подсетями

2.1 Алгоритм разделения радиоресурсов между виртуальными подсетями

В данной работе рассматривается задача планирования радиоресурсов для обслуживания двух типов трафика: веб-трафика и видеопотоков, каждый из которых обслуживается внутри соответствующей подсети. Веб-страницы представляют собой файлы. Начальный размер файла для потока i равен $S_i^{(0)}$. Видеопотоки представляют собой последовательность видеосегментов определенной длительности воспроизведения.

Задача планирования формулируется, как следующая оптимизационная задача. Пусть заданы доли канальных ресурсов, выделенные подсети j в долгосрочной перспективе (s_j^{res}) , количество байт, оставшееся для загрузки соответствующих веб-страниц или видеосегментов S_i , канальные скорости каждого из клиентов r_i и объемы буфера видеоклиентов Q_i . Тогда задача планировщика заключается в минимизации среднего времени загрузки веб-страниц при условии загрузки видеосегментов клиентам до наступления дедлайнов, которые определяются, как некоторая функция объема буфера $D_i = D(Q_i)$ (в простейшем случае $D_i(Q_i) = Q_i$).

Утверждение. Пусть канальные скорости клиентов не меняются во времени, а также в сети не появляются новые потоки. Тогда планирование передач видеосегментов максимально близко к их дедлайнам позволяет минимизировать среднее время загрузки веб-страниц.

Доказательство. При условии, что новые потоки в сети не появляются, для минимизации среднего времени загрузки веб-страниц необходимо планировать обслуживание веб-подсети в приоритете над видеоподсетью до тех пор, пока остается возможным выполнить дедлайны загрузки видеосегментов. Пусть передачи видеосегментов спланированы так, чтобы они не пересекались между собой, и при этом все их дедлайны выполнены. Для простоты изложения рассмотрим случай одного видеосегмента и одной веб-страницы (см. рис. 1). Однако описанное далее утверждение легко распространяется на случай большего числа потоков. Пусть загрузка видеосегмента имеет длительность t_v и спланирована так, чтобы она закончилась за время τ до дедлайна D_v . При этом, чтобы дедлайн видеосегмента выполнялся,

 $0 \leq \tau \leq D_v - t_v$. Пусть время необходимое для загрузки веб-страницы $t_w = t_1 + t_2$. Итоговое время обслуживания веб-страницы зависит от момента загрузки видеосегмента, т.е.

$$T_w(t_w, \tau) = \begin{cases} t_w, t_w \le D_v - t_v - \tau, \\ t_w + t_v, t_w > D_v - t_v - \tau. \end{cases}$$

Пусть размер страницы t_w — время, необходимое для ее загрузки, представляет собой некоторую случайную величину, имеющую плотность распределения $\rho(t_w)$ ($\int_0^{\infty} \rho(t_w) dt_w = 1$). Тогда среднее время загрузки веб-страницы можно оценить следующим образом:

$$\bar{T}_{w}(\tau) = \int_{0}^{D_{v} - t_{v} - \tau} t_{w} \rho(t_{w}) dt_{w} + \int_{D_{v} - t_{v} - \tau}^{\infty} (t_{w} + t_{v}) \rho(t_{w}) dt_{w} =$$
(1)

$$= \int_0^\infty t_w \rho(t_w) dt_w + t_v \int_{D_v - t_v - \tau}^\infty \rho(t_w) dt_w.$$
⁽²⁾

Первое слагаемое в выражении 2 не зависит от τ . Второе слагаемое увеличивается с ростом τ . Следовательно $\bar{T}_w(\tau) \longrightarrow \min \min \tau \tau \longrightarrow 0 \blacksquare$.

При доказательстве приведенного выше утверждения предполагалось, что передачи видеосегментов спланированы так, что все их дедлайны выполнены. Данное условие зависит от величины и относительного расположения дедлайнов и определяется следующим образом. Пусть все видеосегменты пронумерованы в порядке невозрастания их дедлайнов, т.е. задано множество видеосегментов $\mathcal{D} = \{(D_i, T_i) : D_i \geq D_{i+1}, i \in 1..N\}$, где $T_i = \frac{S_i}{r_i}$ — время, требуемое для загрузки сегмента *i*, тогда дедлайны могут быть выполнены, если

$$\begin{cases}
D_1 - \sum_{i \ge 1} T_i > 0, \\
D_2 - \sum_{i \ge 2} T_i > 0, \\
\dots \\
D_N - T_N > 0
\end{cases}$$
(3)

Таким образом, используя доказанное выше утверждение, получаем следующий алгоритм разделения ресурсов между виртуальными подсетями. Каждый слот, если система неравенств 3 выполняется, то обслуживается веб-подсеть, а если не выполняется — видеоподсеть.

2.2 Алгоритмы планирования радиоресурсов внутри виртуальных подсетей

Одним из преимуществ концепции виртуализации сетевых ресурсов является возможность независимого планирования ресурсов внутри подсетей. Таким образом, становится возможным обслуживать потоки разного типа, оптимизируя соответствующие метрики. В данной работе для подсети, обслуживающей веб-трафик, предлагается использовать классический алгоритм SRPT (англ. Shortest Remaining Processing Time) [8], который минимизирует среднее время загрузки вебстраниц. В момент времени t данный алгоритм выбирает для обслуживания клиента

 $\hat{i} = \arg\min_{i} T_i(t).$

В случае видеотрафика предлагается использовать подход, предложенный в работе [9]. Обслуживание видеотрафика состоит из двух стадий. На стадии долгосрочного планирования происходит централизованный выбор битрейтов, загружаемых клиентами видеопотоков. Под битрейтом видеопотока в данной модели понимается отношение размера видеосегмента к его длительности в секундах. Для выбора битрейтов клиентам решается задача оптимизации, учитывающая доступную долю канальных ресурсов и функцию полезности, зависящую от назначенного клиенту битрейта. В данной работе предлагается использовать следующую функцию полезности:

$$U_i(b_i) = \begin{cases} 0, b_i < b_i^{(0)} \\ 1, b_i \ge b_i^{(0)}, \end{cases}$$

где b_i — битрейт видеоизображения, назначенный клиенту i, а $b_i^{(0)}$ — битрейт видеоизображения, требуемого клиентом i качества (в данной работе рассматривается битрейт, который соответствует разрешению 360р). Тогда оптимизационная задача централизованного выбора битрейтов видеоизображения для клиентов имеет следующий вид:

$$\begin{cases} \sum_{i} U_i(b_i) \longrightarrow_{\{b_i\}} \max\\ \sum_{i} s_i(b_i) \le s_{video}, \end{cases}$$

где $s_i(b_i) = \frac{b_i}{\bar{r}_i(t)}$ — доля ресурсов, выделенная клиенту *i* для загрузки видеосегмента с битрейтом *i*, $\bar{r}_i(t)$ — усредненная в некотором окне канальная скорость клиента *i*.

Вторая стадия — краткосрочное планирование радиоресурсов, т.е. выбор клиента для обслуживания в данном слоте. Задача алгоритма планирования на данной стадии — избежать прерываний воспроизведения видеоклипов из-за опустошения буфера на клиенте. Для этого предлагается использовать планировщик ресурсов M-LWDF (англ. Maximum-Largest Weighted Delay First) [10], который выбирает для обслуживания клиента

$$\hat{i} = \arg\max_{i} \frac{r_i(t)}{\bar{r}_i(t)} \frac{1}{D_i(t)},$$

где $r_i(t)$ — канальная скорость клиента *i* в момент времени *t*, $D_i(t)$ — время, оставшееся до наступления дедлайна передачи видеосегмента клиенту *i*. Выбор данного алгоритма обусловлен тем, что, во-первых, данный алгоритм нацелен на выполнение заданных ограничений на время загрузки сегментов, а, во-вторых, позволяет повысить эффективность использования радиоресурсов за счет учета текущей канальной скорости клиентов.

Параметр	Значение
Модель распространения сигнала	Okumura-Hata [11]
Канал	20 МГц @ 2 ГГц
Монон, замираний сигнала	Релеевское многолучевое замира-
модель замирании сигнала	ние (Extended Pedestrian A [12])
Мощность передатчика	30 дБм
Тип антенны базовой станции	Всенаправленная
Высота базовой станции	30 м
Высота клиентских станций	1 м
TCP congestion control	NewReno
TCP Maximum Segment Size (MSS)	1460 байт
TCP initial congestion window	10 MSS
Продолжительность эксперимента	1000c
Число прогонов эксперимента	10

Таблица 1: Параметры эксперимента

3 Численные результаты

3.1 Постановка экспериментов

Для оценки производительности разработанного алгоритма используется среда имитационного моделирования NS-3 [13]. Рассматривается сеть LTE, в которой N станций случайным образом размещены в круге радиусом R = 500 в зоне покрытия базовой станции. Базовая станция соединена с видео- и веб-серверами с помощью проводного соединения с пропускной способностью 10 Гбит/с и круговой задержкой передачи пакета 10 мс. Другие параметры экспериментов приведены в таблице 1.

Клиенты используют протокол MPEG-DASH [14] для просмотра видеоклипов. Поведение видеоклиентов моделируется следующим образом. Клиент выбирает видеоклип и начинает загружать его с сервера. Видеоклипы имеют случайною длительность T_v , равномерно распределенную на интервале (90, 110)с. Спустя время T_p после окончания просмотра видеоклипа клиент начинает загрузку следующего видеоклипа. T_p — случайная величина, имеющая экспоненциальное распределение со средним 30с, минимальным и максимальным значениями 10с и 60с соответственно. Продолжительность одного видеосегмента составляет 5120мс. Загружаемые клиентами видеоклипы выбираются случайным образом из базы данных, которая содержит 60 видеоклипов, собранных с сервиса YouTube и закодированных в следующих разрешениях: {144p,...,360p,...,1080p}.

Поведение веб-клиентов моделируется следующим образом. Клиент загружает веб-страницы с интервалом T_w . T_w — случайная величина, имеющая экспоненциальное распределение со средним 45с, минимальным и максимальным значениями 15с и 90с соответственно. Размер веб-страницы S_w — случайная величина, имеющая логнормальное распределение со следующими параметрами: $\mu = 14.46$, $\sigma = 0.327$, минимальное и максимальное значения — 100кБ и 5МБ соответственно. В экспериментах рассматривается сценарий, в котором половина клиентов в сети загружает веб-страницы, а половина — видеоклипы. Сравниваются 3 алгоритма управления радиоресурсами:

- Планировщик PF (Proportional Fair [15]) на базовой станции, обслуживающий всех клиентов без разделения на виртуальные подсети, и видеоклиенты, автономно выбирающие битрейт видеоклипов при помощи алгоритма, описанного в работе [16].
- Алгоритм NVS, выполняющий разделение канальных ресурсов между виртуальными подсетями, и планировщики, описанные в разделе 2.2, используемые для обслуживания потоков внутри подсетей.
- 3. Алгоритм ВАА (англ. Buffer-Aware Algorithm), описанный в разделе 2.1, и планировщики, описанные в разделе 2.2, используемые для обслуживания потоков внутри подсетей.

Для оценки производительности исследуемых алгоритмов используются следующие метрики. Для оценки качества обслуживания видеоклиентов определяется число удовлетворенных видеопотоков, где видеопоток считается удовлетворенным, если клиент смотрел более 95% видеоклипа в разрешении не меньше 360р, задержка начала воспроизведения не превышала 2 с, а доля прерываний воспроизведения составляла менее 0,5%. Для оценки качества обслуживания веб-клиентов используется средняя скорость загрузки веб-страницы, вычисляемая как отношение размера веб-страницы ко времени ее загрузки. Кроме того, оценивается коэффициент использования канальных ресурсов сети.

3.2 Анализ численных результатов

Результаты имитационного моделирования приведены на рис. 2. Из приведенных графиков видно, что, за счет использования концепции виртуализации сетевых ресурсов и планировщиков, специально разработанных для обслуживания определенного типа трафика, возможно повысить качество обслуживания обоих типов трафика. Сравнивая алгоритмы NVS и BAA, можно видеть, что учет информации об ограничениях на время загрузки видеосегментов, позволяет значительно повысить качество обслуживания веб-клиентов, не ухудшая при этом качество обслуживания видеоклиентов. В целом, по сравнению с базовым решением, использующим планировщик PF для всех типов трафика, предложенное в данной работе решение позволяет на 50% увеличить число удовлетворенных видеопотоков и в 2.5 раза увеличить среднюю скорость загрузки веб-страниц.

4 Заключение

Виртуализация сетевых ресурсов является одной из ключевых технологий, позволяющих обеспечить выполнение разнородных требований к качеству



Рис. 2: Результаты экспериментов

обслуживания, в сетях пятого поколения. Одной из важных задач, стоящих перед разработчиками данной технологии, является задача оптимального разделения ресурсов беспроводного канала между виртуальными подсетями.

В данной работе был предложен новый алгоритм разделения ресурсов между подсетями, обслуживающими видео- и веб-клиентов. С помощью

имитационного моделирования показано, что разработанный алгоритм позволяет значительно повысить качество обслуживание данных типов трафика по сравнению с существующими решениями, представленными в литературе.

Список литературы

- 1. "Framework and overall objectives of the future development of IMT for 2020 and beyond," ITU-R, Recommendation M.2083, September 2015.
- X. Zhou, R. Li, T. Chen, and H. Zhang, "Network slicing as a service: enabling enterprises' own software-defined cellular networks," *IEEE Communications Magazine*, vol. 54, no. 7, pp. 146–153, 2016.
- H. Zhang, N. Liu, X. Chu, K. Long, A.-H. Aghvami, and V. C. Leung, "Network slicing based 5g and future mobile networks: mobility, resource management, and challenges," *IEEE Communications Magazine*, vol. 55, no. 8, pp. 138–145, 2017.
- X. Li, R. Casellas, G. Landi, A. de la Oliva, X. Costa-Perez, A. Garcia-Saavedra, T. Deiss, L. Cominardi, and R. Vilalta, "5g-crosshaul network slicing: Enabling multi-tenancy in mobile transport networks," *IEEE Communications Magazine*, vol. 55, no. 8, pp. 128–137, 2017.
- R. Kokku, R. Mahindra, H. Zhang, and S. Rangarajan, "NVS: a virtualization substrate for WiMAX networks," in *Proceedings of the sixteenth annual international* conference on Mobile computing and networking. ACM, 2010, pp. 233–244.
- Cisco, "Cisco visual networking index: Global Mobile Data Traffic Forecast Update, 2016–2021 White Paper," 2016.
- ISO, ISO/IEC 23009-5:2017 Information technology Dynamic adaptive streaming over HTTP(DASH) – Part 5: Server and network assisted DASH (SAND). ISO.
- L. Kleinrock, Queueing Systems, Vol. II: Computer Applications. John Wiley & Sons, 1976.
- E. Khorov, A. Krasilov, M. Liubogoshchev, and S. Tang, "Sebra: SAND-enabled bitrate and resource allocation algorithm for network-assisted video streaming," in 2017 IEEE 13th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob), Oct 2017, pp. 1–8.
- M. Andrews, K. Kumaran, K. Ramanan, A. Stolyar, P. Whiting, and R. Vijayakumar, "Providing quality of service over a shared wireless link," *IEEE Communications magazine*, vol. 39, no. 2, pp. 150–154, 2001.
- M. Hata, "Empirical formula for propagation loss in land mobile radio services," IEEE transactions on Vehicular Technology, vol. 29, no. 3, pp. 317–325, 1980.
- 3GPP, 3GPP TS 36.104:Evolved Universal Terrestrial Radio Access (E-UTRA); Base Station (BS) radio transmission and reception (Release 8). 3GPP, 2008.
- 13. "Network simulator 3 (ns-3)." [Online]. Available: https://www.nsnam.org/
- ISO, ISO/IEC 23009-1:2014 Information technology Dynamic adaptive streaming over HTTP (DASH) – Part 1: Media presentation description and segment formats. ISO.
- A. Stolyar, "On the asymptotic optinlalily of the gradient scheduling algorithm for multi-user throughput allocation," *Operations Research*, vol. 53, no. 1, pp. 12–25, 2005.
- R. Luciano, "A Dynamic Adaptive http Streaming Video Service for Google Android," Ph.D. dissertation, Thesis Royal Institute of Technology Stockholm, 2011.

Моделирование передачи трафика межмашинного взаимодействия в сетях Wi-Fi HaLow с использованием механизма окна ограниченного доступа в режиме с пересечением границ *

Р.Р. Юсупов

ИППИ РАН yusupov@iitp.ru

Аннотация В новом стандарте Wi-Fi IEEE 802.11ah введён механизм доступа к каналу, называемый окном ограниченного доступа (англ. Restricted Access Window, RAW). Этот механизм позволяет уменьшать конкуренцию за доступ к каналу посредством распределения попыток передач станций по интервалу времени. Механизм RAW может работать в двух режимах: режиме без пересечения границ и режиме с пересечением границ. В статье рассмотрено применение механизма RAW в режиме с пересечением границ для передачи трафика межмашинного взаимодействия, предложена математическая модель и проведена её валидация.

Keywords: Wi-Fi HaLow, окно ограниченного доступа, cross slot boundary, межмашинное взаимодействие

1 Введение

В настоящее время набирает популярность концепция Интернета вещей (англ. Internet of Things, IoT). Данная концепция подразумевает большое количество устройств, которые обмениваются данными друг с другом и с внешними сервисами. Примерами таких устройств являются автомобили, бытовая техника, датчики, управляющие механизмы. Наиболее простым и экономичным способом обеспечения этих устройств связью и доступом в Интернет являются беспроводные сети, к примеру, сети Wi-Fi. Для Интернета вещей типичен сценарий, в котором большое число датчиков генерирует разреженный трафик межмашинного взаимодействия.

Обеспечить выполнение требований Интернета вещей к сетям Wi-Fi может новый стандарт IEEE (англ. Institute of Electrical and Electronics Engineers) 802.11ah [1], принятый в 2017 году. Сети этого стандарта поддерживают до

^{*} Исследование выполнено в ИППИ РАН при финансовой поддержке РНФ в рамках научного проекта № 14-50-00150.

8191 одновременно подключённых устройств в радиусе 1 километра. Больпое количество одновременно подключённых к точке доступа устройств приводит к большой конкуренции за доступ к каналу. Из-за этого увеличивается энергопотребление устройств, уменьшается время жизни устройств с автономным питанием, уменьшается пропускная способность сети. В стандарте IEEE 802.11ah предложен механизм доступа к каналу, позволяющий решить проблему высокой конкуренции за доступ к каналу – механизм окна ограниченного доступа (англ. Restricted Access Window, RAW).

В основе механизма RAW лежит идея деления всех подключённых к точке доступа станций на группы. Каждой группе выделяется зарезервированный интервал времени – RAW-слот. Внутри зарезервированного интервала времени станции используют базовый для Wi-Fi механизм доступа к каналу EDCA (англ. Enhanced Distributed Channel Access). EDCA основан на методе CSMA/CA (англ. Carrier Sense Multiple Access with Collision Avoidance, множественный доступ с прослушиванием несущей и избеганием коллизий), использующем механизм случайной экспоненциальной отсрочки (англ. binary exponential backoff). Используя механизм случайной отсрочки, перед каждой передачей кадра данных станция генерирует случайное целое число, равномерно распределённое на отрезке, называемом конкурентным окном (англ. Contention Window), от которого зависит, когда будет совершена попытка передачи. Если передача кадра неудачна, станция увеличивает конкурентное окно вдвое. Если передача кадра успешна, то конкурентное окно сбрасывается до начального значения. Конкурентное окно также сбрасывается в начале зарезервированного для станции RAW-слота. Более подробное описание механизмов RAW и EDCA может быть найдено в [2].

По-умолчанию механизм RAW работает в режиме без пересечения границ, при использовании которого станции передают до тех пор, пока их попытки передач завершаются внутри зарезервированного интервала времени. В режиме с пересечением границ (англ. Cross-Slot Boundary) станции могут завершить попытку передачи вне зарезервированного интервала времени. Так как внутри зарезервированного интервала времени доступ к каналу пытается получить меньшее число станций, конкуренция уменьшается. Станции реже попадают в коллизии, уменьшается энергопотребление станций, увеличивается время жизни станций с автономным питанием. Если часть станций не имеет кадров данных для передачи, то при делении станций на большое число групп могут возникать группы станций, не осуществляющих передачу данных. В таком случае часть канального времени не используется, и пропускная способность сети падает. Если станциям разрешено завершать попытку передачи вне зарезервированного интервала времени, то с одной стороны в сети увеличивается конкуренция за доступ к каналу, что приводит к увеличению энергопотребления станций, но с другой стороны увеличивается доля использования канального времени, что приводит к увеличению пропускной способности сети. Возникает вопрос выбора режима работы механизма RAW и параметров механизма RAW, таких как число групп, длительность зарезервированного интервала времени. Необходима математическая модель механизма RAW.

Дальнейшее изложение построено следующим образом. В разделе 2 выполнен краткий обзор предыдущих работ. В разделе 3 представлена постановка задачи. В разделе 4 кратко описана аналитическая модель. В разделе 5 приводятся численные результаты.

2 Предыдущие работы

Ещё до выхода окончательной версии стандарта IEEE 802.11ah [1] предпринимались попытки построения математической модели механизма ограниченного доступа. Приведём далее краткий обзор статей с математическими моделями, которые могут быть использованы для анализа механизма RAW.

В статье [3] описывается математическая модель базового механизма доступа DCF (англ. Distributed Coordination Function), которая легко модифицируется для описания механизма EDCA с одной категорией доступа. Это одна из первых работ о механизме доступа DCF, важность которой невозможно переоценить. Авторы статьи [3] рассматривают стационарный режим, предполагая длительную работу сети в стационарных условиях. Однако, согласно стандарту [1] в начале RAW-слота каждая станция, которой разрешено передавать в данном RAW-слоте, замораживает свой счётчик отсрочки и инициализирует новый. При выходе из RAW-слота станция возвращается к использованию старого счётчика отсрочки. Рассчитать пропускную способность при использовании механизма RAW с помощью модели из статьи [3] нельзя, так как эта модель не учитывает ни конечность RAW-слотов, ни реинициализацию счётчика отсрочки в начале RAW-слота.

Авторы статей [4,5] используют модификацию подхода из статьи [3] для анализа описанного ими метода доступа Group Synchronized DCF (GS-DCF), основанного на механизме окна ограниченного доступа. В статье [5] рассмотрен в том числе и режим с пересечением границ. Модель из статье [4,5] некорректно использовать для анализа механизма RAW, так как при построении модели GS-DCF авторы использовали модель из статьи [3] для подсчёта вероятностей передачи и коллизии в виртуальном слоте. Это оправдано только если по окончании каждого зарезервированного интервала времени станции запоминают значение счётчика отсрочки и используют его в начале следующего зарезервированного интервала времени. Такое поведение станций противоречит стандарту IEEE 802.11ah.

Реинициализация счётчика отсрочки и конечная длительность RAWслота учтены в [6], где строится математическая модель передачи данных межмашинного взаимодействия в RAW-слоте. В [6] рассматривается режим без пересечения границ. В данной статье рассматривается адаптация модели из [6] для анализа режима с пересечением границ. При адаптации модели используется предложенный в [2] подход, предполагающий анализ окна ограниченного доступа посредством анализа двух последовательных RAWслотов.

3 Постановка задачи

Рассматривается следующий сценарий: группа из M станций передаёт восходящий поток данных межмашинного взаимодействия с использованием окна ограниченного доступа длительностью T_{RAW} . В начале окна ограниченного доступа каждая станция имеет один кадр данных в очереди, новые кадры внутри окна ограниченного доступа не появляются. Предполагается что внешние помехи в радиоканале отсутствуют, т.е. передача кадра успешна если нет интерференции с другими станциями. Также предполагается что скрытые станции отсутствуют.

Точка доступа делит M станций на K групп, выделяя каждой группе станций RAW-слот длительностью $T_{slot} = \frac{T_{RAW}}{K}$. Станции передают неограниченно в течение выделенного им RAW-слота, а также могут совершить одну дополнительную попытку передачи, которая закончится вне выделенного станциям RAW-слота.

4 Модель

Для анализа процесса передачи данных в окне ограниченного доступа необходимо проанализировать передачу данных в каждом RAW-слоте. Для анализа процесса передачи данных в первом RAW-слоте может быть использована модель из статьи [6]. Для анализа процесса передачи данных в остальных RAW-слотах предлагается использовать модель для двух последовательных RAW-слотов, следуя описанному в [2] подходу. В рамках этого подхода предполагается, что пересекающие границы RAW-слотов попытки передач станций влияют только на станции, передающие в следующем RAW-слоте, а влиянием на остальные RAW-слоты можно пренебречь.

Модель передачи данных в двух последовательных RAW-слотах основана на адаптированной модели из [6]. Используя описанную в [6] модель, можно найти условную вероятность передачи в первом RAW-слоте со счётчиком числа попыток передач r по прошествии t виртуальных слотов $\mathbb{P}(TX|r,t)$ а также переходные вероятности цепи Маркова $(e, s, c, r, n)_t$, где e, s, c - число прошедших пустых, успешных и коллизионных виртуальных слотов, r счётчик числа попыток передач произвольным образом выбранной станции, n - число активных станций. Поглощающие состояния данной цепи Маркова следующие:

$$\mathfrak{A} = \{(e, s, c, r, n) : eT_e + sT_s + cT_c > T_{slot}\} \cup \{(e, s, c, r, n) : n = 0\} \cup \cup \{(e, s, c, r, n) : r = RL\},\$$

где T_e , T_s , T_c – длительности пустого, успешного и коллизионного виртуального слотов соответственно, RL – максимальное число попыток передач одного кадра данных. Можно оценить условную вероятность передачи на границе между первым и вторым RAW-слотами как:

$$\tau = \sum_{(e,s,c,r,n,t)\in\mathfrak{S}} \mathbb{P}(TX|r,t)\mathbb{P}(e,s,c,r,n)_{t}$$

где:

$$\mathfrak{S} = \{(e, s, c, r, n, t): eT_e + sT_s + cT_c > T_{slot}, \ e + s + c = t\}$$

Вероятность пересечения границы RAW-слота хотя бы одной станцией может быть вычислена следующим образом:

$$\mathbb{P}(exit) = \sum_{(e,s,c,r,n,t) \in \mathfrak{S}} \mathbb{P}(e,s,c,r,n)_t,$$

а распределение числа станций, активных на границе RAW-слота, может быть найдено как:

$$\mathbb{P}(n) = \sum_{(e,s,c,r,x,t)\in\mathfrak{S}(n)} \mathbb{P}(e,s,c,r,x)_t,$$
(1)

где:

$$\mathfrak{S}(n)=\{(e,s,c,r,x,t):eT_e+sT_s+cT_c>T_{slot},\,x=n,\,e+s+c=t\}.$$

Оценить вероятность $\hat{\mathbb{P}}(TX|t)$ передачи активных станций из первого RAW-слота во втором RAW-слоте можно как:

$$\hat{\mathbb{P}}(TX|t) = \frac{\tau}{\mathbb{P}(exit)} \left(1 - \frac{\tau}{\mathbb{P}(exit)}\right)^t.$$
(2)

Тогда для анализа процессов передачи данных во втором RAW-слоте можно использовать цепь Маркова $(e, s, c, n_1, n_2)_t$, где n_1 – число активных станций из первого RAW-слота, n_2 – число активных станций из второго RAW-слота. Начальное распределение этой цепи Маркова задаётся формулой (1), а переходные вероятности могут быть найдены с использованием формулы (2) и формул из статьи [6].

Получить среднее число успешно переданных кадров данных в течение второго RAW-слота можно как:

$$S(N_1, N_2, T_{slot}) = \sum_{(e, s, c, n_1, n_2) \in \hat{\mathfrak{S}}} s\mathbb{P}(e, s, c, n_1, n_2)_t,$$

где N_1 и N_2 – число активных станций, которым был выделен, соответственно, первый и второй RAW-слот, а

$$\hat{\mathfrak{S}} = \{ (e, s, c, n_1, n_2) : eT_e + sT_s + cT_c > T_{slot}, \ e + s + c = t \}.$$

Используя вышеописанный подход, можно найти число успешно доставленных кадров данных в любом RAW-слоте окна ограниченного доступа. Адаптируя описанный в [2] подход, найдём долю успешно доставленных в RAW кадров данных.

$$\begin{split} S_1 &= S\left(0, \lfloor M/K \rfloor, \frac{T_{RAW}}{K}\right), \\ S_2 &= S\left(0, \lfloor M/K \rfloor + 1, \frac{T_{RAW}}{K}\right), \\ S_{11} &= S\left(\lfloor M/K \rfloor, \lfloor M/K \rfloor, \frac{T_{RAW}}{K}\right), \\ S_{12} &= S\left(\lfloor M/K \rfloor, \lfloor M/K \rfloor + 1, \frac{T_{RAW}}{K}\right), \\ S_{21} &= S\left(\lfloor M/K \rfloor + 1, \lfloor M/K \rfloor, \frac{T_{RAW}}{K}\right), \\ S_{22} &= S\left(\lfloor M/K \rfloor + 1, \lfloor M/K \rfloor + 1, \frac{T_{RAW}}{K}\right). \end{split}$$

$$PDR(M, K, T_{RAW}) = \frac{1}{KM} \Big(S_1 \cdot (K - M \mod K) + S_2 \cdot M \mod K + (K - 1) \cdot (S_{12} + S_{21} + S_{11} \cdot (K - M \mod K - 1) + S_{22} \cdot (M \mod K - 1)) \Big).$$

5 Численные результаты

5.1 Параметры модели

В таблице 5.1 приведены параметры модели, соответствующие стандарту IEEE 802.11ah [1]. Параметры модели взяты из [2]. При расчёте параметров модели предполагалось, что станции не используют механизм RTS/CTS, т.е. $T_s = T_c$. Станции передают кадры данных длиной 100 байт с использованием сигнально-кодовой конструкции MCS8, самой скоростной сигнально-кодовой конструкции, доступной для канала шириной 2 МГц.

5.2 Численные результаты

На рис. 1 изображены результаты проверки аналитической модели с использованием методов имитационного моделирования. Рассматривалось окно RAW, состоящее из двух RAW-слотов с одинаковым числом станций. Как видно из графиков, при увеличении числа станций уменьшается доля успешно доставленных кадров данных. Это связано с тем, что увеличение

Параметр	Значение	Параметр	Значение	
T_e	$52\mu s$	T_s	$1064 \mu s$	
length	100	T_c	$1064 \mu s$	
T_{slot}^{\max}	$246140 \mu s$	RL	7	
CW_0	16	CWmax	1024	

Таблица 1. Параметры модели

числа активных станций увеличивает конкуренцию за доступ к каналу и нагрузку на сеть передачи данных. При фиксированном числе станций и увеличении длительности RAW-слота доля успешно доставленных кадров данных увеличивается. Это связано с тем, что при увеличении длительности окна ограниченного доступа увеличивается число попыток передач, которые могут совершить станции.

На рис. 2 изображены результаты проверки аналитической модели для окна RAW с пятью RAW-слотами. Увеличение числа RAW-слотов приводит к уменьшению точности построенной аналитической модели. Можно сделать вывод о том, что предположение, лежащее в основе двухслотовой модели, приводит к большой погрешности при анализе сценария передачи трафика межмашинного взаимодействия. Погрешность обусловлена тем, что попытки передач станций, пересекающих границы RAW-слотов, влияют на станции, передающие как в следующем RAW-слоте, так и в последующих RAW-слотах. Для анализа сценария передачи трафика межмашинного взаимодействия необходимо строить более точные и сложные математические модели, к примеру, последовательно рассматривать все RAW-слоты.

6 Заключение

Получены следующие результаты:

- Разработана модель передачи трафика межмашинного взаимодействия с использованием механизма окна ограниченного доступа в режиме с пересечением границ
- Проведена валидация разработанной модели с использованием методов имитационного моделирования
- Показано, что предложенный в [2] подход к анализу режима с пересечением границ не применим в рассмотренном сценарии передачи трафика межмашинного взаимодействия

Список литературы

1. IEEE P802.11ah/Standard for Information technology – Telecommunications and information exchange between systems - Local and metropolitan area networks – Specific requirements - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications – Amendment 6: Sub 1 GHz License Exempt Operation.

- Khorov Evgeny, Lyakhov Andrey, Yusupov Ruslan. Two-Slot Based Model of the IEEE 802.11ah Restricted Access Window with Enabled Transmissions Crossing Slot Boundaries // 2018 IEEE 19th International Symposium on "A World of Wireless, Mobile and Multimedia Networks" (WoWMoM) / IEEE. — 2018.
- 3. Bianchi Giuseppe. Performance analysis of the IEEE 802.11 distributed coordination function // IEEE Journal on selected areas in communications. 2000. Vol. 18, no. 3. P. 535–547.
- Performance analysis of grouping strategy for dense IEEE 802.11 networks / Lei Zheng, Lin Cai, Jianping Pan, Minming Ni // Global Communications Conference (GLOBECOM), 2013 IEEE / IEEE. — 2013. — P. 219–224.
- Performance analysis of group-synchronized DCF for dense IEEE 802.11 networks / Lei Zheng, Minming Ni, Lin Cai et al. // IEEE Transactions on Wireless Communications. — 2014. — Vol. 13, no. 11. — P. 6180–6192.
- Khorov Evgeny, Krotov Alexander, Lyakhov Andrey. Modelling machine type communication in IEEE 802.11 ah networks // Communication Workshop (ICCW), 2015 IEEE International Conference on / IEEE. — 2015. — P. 1149–1154.



Рис. 1. Валидация для окна ограниченного доступа с двумя RAW-слотами



Рис. 2. Валидация для окна ограниченного доступа с пятью RAW-слотами

Алгоритм управления радиоресурсами для плотных сетей Wi-Fi*

А.Г. Кирьянов, А.В. Кротов, Е.М. Хоров {kiryanov, krotov, khorov}@iitp.ru

ИППИ РАН

Аннотация В настоящее время постоянно возрастает количество беспроводных устройств. Для обслуживания всех устройств и обеспечения высокой пропускной способности сети требуется увеличивать число беспроводных точек доступа. Однако при повышении плотности точек доступа в сетях Wi-Fi возникает проблема несправедливого разделения канальных ресурсов между различными беспроводными соединениями. Для решения этой проблемы предлагается централизованный алгоритм управления радиоресурсами, обеспечивающий справедливое распределение канальных ресурсов, используемых для передачи данных различным устройствам.

1 Введение

В инфраструктурных сетях Wi-Fi, получивших наибольшее распространение в наши дни, передача данных осуществляется через точку доступа, которая одновременно выполняет роль шлюза в Интернет. Чтобы начать передачу данных, станция сети Wi-Fi должна получить доступ к каналу. В подавляющем большинстве случаев используется регламентированный в стандарте IEEE 802.11 [1] метод случайного доступа, согласно которому станции одной сети с равной вероятностью получают доступ к каналу. Однако при передаче пакетов данных одинакового размера различные станции могут занять канал на разное время в зависимости от выбранной сигнально-кодовой конструкции (СКК). Станции, находящиеся далеко от точки доступа, будут использовать медленные СКК и таким образом потреблять намного больше канальных ресурсов для передачи того же объема данных чем станции, находящиеся в зоне уверенного приема сигнала. Такое разделение оказывается несправедливым с точки зрения равенства получаемого различными станциями сети частотно-временного ресурса. Кроме этого, если несколько сетей Wi-Fi работают в одной области пространства и используют один и тот же частотный канал, может возникать сильно неравномерное распределение канального времени между различными станциями, если число клиентов разных точек доступа существенно отличается. Если в первом случае для

^{*} Исследование выполнено в ИППИ РАН за счет гранта Российского научного фонда (проект №16-19-10687)

равномерного распределения ресурса между станциями достаточно использовать некоторый алгоритм, работающий на одной точке доступа Wi-Fi, то во втором случае необходимо также некоторым образом осуществлять распределение канального ресурса еще и между различными точками доступа, что существенно усложняет поставленную задачу. В следующем разделе приводится краткий обзор решений, нацеленных на равномерное распределение канального ресурса в сети Wi-Fi.

2 Обзор литературы

Одним из первых решений, предложенных для решения обозначенной проблемы, был адаптивный выбор максимально допустимого размера пакета на сетевом уровне (англ.: Maximum Transmission Unit, MTU) для различных соединений [2,3]. Однако такое решение требует дополнительного взаимодействия между канальным уровнем и сетевым уровнем и ненадежно из-за того, что размер пакета может быть дополнительно ограничен любым оборудованием на пути следования пакета.

Стандарт IEEE 802.11е [4] предоставил возможность точкам доступа Wi-Fi настраивать параметры метода случайного доступа, такие как конкурентное окно (англ.: Congestion Window, CW) и ограничение на длительность передачи (TXOP Limit) [5]. Однако ограничение TXOP Limit, гарантирующее одинаковое время использования среды после получения доступа, как правило, используется только для передачи данных реального времени, таких как аудио- и видеопотоки. Для обслуживания остального трафика, такого как трафик веб-приложений, настройка ограничения TXOP Limit не предусмотрена.

Авторы [6] предложили и реализовали решение, в основе которого лежит планирование ресурсов на точке доступа Wi-Fi. Точка доступа разделяет трафик, предназначенный различным станциям, на отдельные очереди и каждый раз выбирает для передачи пакет согласно разработанной дисциплине обслуживания очередей, основная идея которой состоит в следующем. В каждый момент времени для обслуживания выбирается пакет, адресованный той станции, которая в последнее время использовала наименьшее количество канальных ресурсов. Данное решение подходит для сетей, в которых множество станций обслуживается одной точкой доступа, однако не гарантирует справедливого распределения ресурса между станциями различных точек доступа.

В [7] был предложен алгоритм одновременного управления мощностью передачи и планирования радиоресурсов, однако из-за высокой сложности алгоритма он оказался неприменим для большого числа соединений. Также был рассмотрен случай, когда используется только централизованное планирование ресурсов без управления мощностью, и показано, что в рассмотренных сценариях такое решение позволяет достичь результатов, сравнимых с результатами предложенного алгоритма. В [8] был предложен алгоритм одновременного управления мощностью и планирования ресурсов для сетей Wi-Fi, обладающий меньшей сложностью чем предложенный в [7] и учитывающий налагаемые стандартом ограничения: точка доступа не может вести передачу, если принимаемая ей мощность сигнала и шума превышает порог определения занятости среды (англ. Carrier Sense Threshold, CST). Рассмотрен сценарий, в котором к каждой точке доступа подключена одна станция, и показано, что алгоритм позволяет добиться справедливого распределения ресурсов между различными точками доступа.

В данной работе предлагается решение, позволяющее координировать передачу данных с использованием множества точек доступа и подключенных к ним станций таким образом, чтобы обеспечить равномерное разделение канальных ресурсов между различными станциями сети. В следующем разделе приводится формальная постановка решаемой задачи.

3 Постановка задачи

Имеется N станций, каждая из которых подключена к одной из M точек доступа. Задача управления мощностью передачи и планирования радиоресурсов состоит в том, чтобы в каждый момент времени выбрать вектор мощностей передачи p, каждая координата p_i которого соответствует соединению между станцией i и соответствующей точкой доступа, таким образом, чтобы максимизировать функцию полезности

$$U(\boldsymbol{r}) = \sum_{i}^{N} \log(r_i), \qquad (1)$$

где r_i — пропускная способность соединения со станцией i.

Как показано в [8], для этого нужно периодически выбирать мощности передачи таким образом, чтобы максимизировать

$$\sum_{i}^{N} \frac{r_i}{R_i},\tag{2}$$

где R_i — средняя пропускная способность, полученная станцией ранее. Связь между векторами мощностей p и векторами отношений сигнал-шум γ может быть описана с помощью системы линейных уравнений, а пропускные способности соединений монотонно зависят от соотношений сигнал-шум на приемниках. При этом должны выполняться ограничения на максимальную мощность передачи, а для точек доступа, которые ведут передачу, принимаемая мощность не должна превышать порог CST.

Заметим, что в случае одной точки доступа алгоритм из [6] решает эту задачу.

4 Алгоритм

В [8] рассмотрен случай, когда к каждой точке доступа подключена только одна станция. Для обобщения алгоритма на случай, когда к одной точке доступа может быть подключено несколько устройств, исключим из рассмотрения точки доступа, к которым не подключена ни одна станция, а каждую из оставшихся точек доступа разделим на несколько виртуальных точек доступа, так что одна виртуальная точка доступа будет обслуживать одну станцию. Чтобы при работе алгоритма исключить возможность одновременной передачи данных двумя виртуальными точками доступа, соответствующими одной реальной, положим коэффициенты передачи между данными виртуальными точками доступа равными $+\infty$. В принятых обозначениях коэффициент передачи показывает во сколько раз мощность, получаемая приемником, превышает мощность, излучаемую передатчиком (обычно коэффициент передачи менее 1). Так как точка доступа не может вести передачу, если принимаемая ей мощность превышает порог CST, то таким образом исключаются физически недостижимые в рассматриваемой архитектуре сети решения, в которых реальная точка доступа передает данные нескольким получателям одновременно.

Для решения проблемы несправедливого распределения канальных ресурсов в сетях с множеством точек доступа будем использовать решение, основанное на описанном в [8] алгоритме одновременного управления мощностью передачи и планирования ресурсов.

Предлагаемый алгоритм работает следующим образом. На первом шаге находится предварительное решение. Для этого перебираются все возможные подмножества виртуальных точек доступа. Для каждого подмножества составляется вектор мощностей передачи p, где $p_i = 0$, если точка доступа i не входит в рассматриваемое множество, и $p_i = \hat{p}_i$ иначе, где \hat{p}_i — максимальная мощность передачи. Затем мощности пропорционально уменьшаются таким образом, чтобы для всех точек доступа из выбранного подмножества выполнялось ограничение на CST. После этого полученное решение улучшается с помощью алгоритма, основанного на методе ветвей и границ, описанного в [8]. Такой подход позволяет существенно ускорить нахождение решения в сравнении в [8].

5 Численные результаты

Рассмотрим квадрат со стороной 50 м, в котором расположены M точек доступа Wi-Fi. Внутри рассматриваемого квадрата случайным образом разбросаны N = 10 станций, каждая из которых присоединена к ближайшей точке доступа. Точки доступа передают насыщенные потоки данных всем присоединенным к ним станциям.

Будем варьировать число точек доступа от 1 до 30. Чтобы обеспечить наилучшее покрытие данной области пространства, расположение точек доступа выбирается в соответствии с решением задачи об упаковке кругов в квадрат, для которой при $M \leq 30$ известны оптимальные решения [9]. Основные параметры сценария приведены в таблице 1.



Рис. 1. Зависимость среднего геометрического и среднего арифметического пропускной способности от числа точек доступа

Таблица	1.	Па	рамет	ры	сцена	рия
---------	----	----	-------	----	-------	-----

Параметр	Значение
Высота размещения станций	1 м
Высота размещения точек доступа	3 м
Максимальная мощность передачи	1000 мВт
Мощность шума	-174 дБм/Гц
Ширина канала	80 МГц
Коэффициент шума усилителя	7 дБ
Ограничение на CST	-82 дБм
Стандарт Wi-Fi	IEEE 802.11ac
Алгоритм выбора СКК	Minstrel HT [10]
Размер пакета данных	1500 байт
Агрегирование данных	A-MPDU

Сравним предложенное решение, т.е. использование централизованного управления мощностью передачи и планирования ресурсов, с использованием независимо работающих планировщиков на каждой отдельной точке доступа, а также с использованием только централизованного планировщика.

В случае использования независимо работающих планировщиков каждая точка доступа выбирает для передачи ту станцию, на обслуживание

которой к данному моменту времени она затратила наименьшее количество канальных ресурсов, как предложено в [6]. Заметим, что в случае M > 1это решение не гарантирует, что для всех станций будет использовано одинаковое количество канальных ресурсов, так как различные точки доступа могут занимать различную долю канального времени. При использовании централизованного планирования планировщик периодически выбирает подмножество станций, которым нужно вести передачу на максимальной мощности таким образом, чтобы максимизировать (2).

На рис. 1 показана зависимость среднего арифметического и среднего геометрического пропускной способности соединения сети в зависимости от числа точек доступа. Предложенный алгоритм представляет собой оптимальное решение с точки зрения функции полезности (1) и позволяет более чем в 1,5 раза повысить среднюю пропускную способность при увеличении числа точек доступа по сравнению с распределенным планированием ресурсов. Распределенное планирование является оптимальным только для случая одной точки доступа, а при увеличении числа точек доступа показывает худшие результаты, так как растут накладные расходы обусловленные использованием метода случайного доступа к среде множеством точек доступа. Отметим, что при дальнейшем увеличении числа точек доступа средняя пропускная способность становится постоянной, и в предельном случае каждая точка доступа обслуживает только одну станцию, которая находится в той же точке пространства. Централизованное планирование ресурсов позволяет избежать уменьшения пропускной способности при увеличении плотности точек доступа, но не позволяет улучшить пропускную способность в плотных сетях, так как в каждый момент времени в рассматриваемом сценарии передача ведется только одной из точек доступа.

6 Заключение

В работе предложен алгоритм централизованного распределения канальных ресурсов в сетях Wi-Fi, обеспечивающий справедливое распределение канальных ресурсов между станциями. Проведено сравнение предложенного алгоритма с существующими решениями, управляющими передачей данных независимо на различных точках доступа. Показано, что в плотных сетях Wi-Fi использование централизованных алгоритмов планирования радиоресурсов и управления мощностью передачи позволяет значительно повысить пропускную способность и обеспечить справедливое распределение канальных ресурсов.

Список литературы

 802.11-2016 - IEEE Standard for Information technology–Telecommunications and information exchange between systems Local and metropolitan area networks– Specific requirements - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications, ANSI/IEEE Std. 802.11 (2016) 1 – 3534doi:10.1109/IEEESTD.2016.7786995.

- S.-h. Yoo, J.-H. Choi, J.-H. Hwang, C. Yoo, Eliminating the performance anomaly of 802.11 b, in: International Conference on Networking, Springer, 2005, pp. 1055– 1062.
- J. Dunn, M. Neufeld, A. Sheth, D. Grunwald, J. Bennett, A practical cross-layer mechanism for fairness in 802.11 networks, Mobile Networks and Applications 11 (1) (2006) 37–45.
- 4. IEEE Standard for Information Technology–Local and metropolitan area networks–Specific requirements–Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications - Amendment 8: Medium Access Control (MAC) Quality of Service Enhancements (2005). doi:10.1109/IEEESTD.2005.97890.
- A. Banchs, P. Serrano, H. Oliver, Proportional fair throughput allocation in multirate ieee 802.11 e wireless lans, Wireless Networks 13 (5) (2007) 649–662.
- T. Høiland-Jørgensen, M. Kazior, D. Täht, P. Hurtig, A. Brunstrom, Ending the anomaly: Achieving low latency and airtime fairness in wifi, in: 2017 USENIX Annual Technical Conference (USENIX ATC 17). USENIX Association, Santa Clara, CA, 2017, pp. 139–151.
- L. P. Qian, Y. J. Zhang, S-MAPEL: Monotonic optimization for non-convex joint power control and scheduling problems, IEEE Transactions on Wireless Communications 9 (5) (2010) 1708–1719.
- E. Khorov, A. Kiryanov, A. Krotov, Joint power control and time division to improve spectral efficiency in dense Wi-Fi networks, in: Black Sea Conference on Communications and Networking (BlackSeaCom), 2018 IEEE International, IEEE, 2018, pp. 199–203.
- M. C. Markót, T. Csendes, A new verified optimization technique for the "packing circles in a unit square"problems, SIAM Journal on Optimization 16 (1) (2005) 193–219.
- D. Xia, J. Hart, Q. Fu, On the performance of rate control algorithm Minstrel, in: 2012 IEEE 23rd International Symposium on Personal, Indoor and Mobile Radio Communications - (PIMRC), 2012, pp. 406–412. doi:10.1109/PIMRC.2012.6362819.

Имитационное моделирование метода последовательных обновлений при рассылке служебной информации в децентрализованных беспроводных сетях*

Белогаев А.А.¹, Красилов А.Н.¹

ИППИ РАН, Москва, Россия, http://iitp.ru {belogaev, krasilov}@iitp.ru

Аннотация В децентрализованных беспроводных сетях для работы различных протоколов необходимо обеспечить своевременный обмен служебной информацией между станциями сети. Так, например, для корректной работы протоколов маршрутизации станции должны рассылать актуальную информацию об установленных соединениях и их качестве. Наличие неактуальной информации может привести к появлению циклов и потере пакетов. Для поддержания актуальности информации станции вынуждены рассылать большие объемы служебных данных, что в свою очередь снижает объем канальных ресурсов, доступных для передачи пользовательских данных. В данной работе проводится исследование метода последовательных обновлений, который широко применяется для рассылки служебной информации в существующих протоколах маршрутизации. Он заключается в том, что станции отправляют в служебных сообщениях не всю информацию, а только обновления, произошедшие с момента отправки предыдущего сообщения. В работе [1] построена аналитическая модель, позволяющая оценить объем рассылаемой служебной информации и степень ее актуальности. На основе построенной аналитической модели разработан алгоритм, позволяющий адаптивно выбирать параметры метода последовательных обновлений так, чтобы минимизировать объем рассылаемой служебной информации и в то же время выполнить заданное ограничение на степень актуальности. В данной работе проводится исследование этого алгоритма в среде имитационного моделирования NS-3. В отличие от анализа на основе аналитической модели, проведенного в [1], при имитационном моделировании учитываются особенности работы реальных алгоритмов доступа к каналу, а также используются более детальные модели канала и мобильности станций. Кроме того, разработанный алгоритм оценивает параметры сети на основе собранной статистики, тогда как в работе [1] они предполагались известными. Численные результаты показывают, что использование разработанного алгоритма выбора

^{*} Исследование выполнено в ИППИ РАН при финансовой поддержке РНФ в рамках научного проекта № 14-50-00150.

параметров позволяет существенно снизить объем рассылаемой служебной информации по сравнению со случаем выбора фиксированных значений параметров и при этом обеспечить заданную степень актуальности информации. В частности, показано, что при низкой мобильности станций и малой плотности сети можно на порядок снизить объем рассылки.

Список литературы

 Belogaev, A.A., Khorov, E.M., Krasilov, A.N., Lyakhov, A.I.: Analytical Study of Incremental Approach for Information Dissemination in Wireless Networks. In proc. of 10th Wireless Days Conference (WD 2018), 3-5 April 2018, Dubai, UAE.

Сигнально-кодовая конструкция для векторного дизъюнктивного канала, основанная на перемеженных кодах Рида-Соломона*

Иванов Ф. И.^{1,2}, Рыбин П. С.^{1,2}

¹ Институт проблем передачи информации им. А.А. Харкевича Российской академии наук, Москва, Россия

 $^2\,$ Национальный исследовательский университет «Высшая школа экономики», Москва, Россия

Аннотация В работе предложена сигнально-кодовая конструкция для специального класса векторного дизъюнктивного канала, в котором пользователи передают пакеты бит длины L. Данная конструкция основана на перемеженных кодах Рида-Соломона. Рассмотренные в статье методы кодирования и декодирования для данной конструкции с одной стороны имеют низкую вычислительную сложность, с другой стороны алгоритм совместного декодирования перемеженных кодов Рида-Соломона позволяет исправлять существенно большее число коллизий, чем классический декодер кодов Рида-Соломона. Также в работе представлены численные результаты для относительной суммарной скорости передачи и вероятности ошибки на блок при различном числе пользователей и различных параметрах канала и кодовых конструкций. Полученные результаты позволяют сделать вывод о том, что предложенная конструкция обладает хорошей корректирующей способностью даже в случае использования достаточно коротких кодов.

1 Введение

В настоящее время одним из ключевых особенностей беспроводных сетей является очень быстрое увеличение числа активных пользователей в системе, что вызвано активным развитием концепции интернета вещей и концепции умного дома. Ввиду этого взаимное влияние пользователей друг на друга становится все более ощутимым. В то же время требования, которые предъявляются к скорости передачи данных и вероятности успешной передачи, становятся все более жесткими.

В связи с вышесказанным, требуется, чтобы очень большое число пользователей могли вести некоординированную передачу в беспроводных сетях.

^{*} Работа выполнена при поддержке грантов Р
ФФИ 18-37-00319 мол_а и 18-07-01409 А

Среди общего числа пользователей следует особо выделить подмножество из U пользователей, которые активны в любой момент времени, при этом каждый из них ведет передачу коротких пакетов информации на базовую станцию. При этом задача базовой станции заключается только в восстановлении списка сообщений, которые были на нее переданы, при этом не происходит идентификация пользователя, который передал конкретное сообщение. Некоординированный способ передачи, а также короткая длина передаваемых сообщений существенно влияют на методы построения сигнальнокодовых конструкций, которые были бы практически применимыми в данных условиях.

Практически все широко известные сигнально-кодовые конструкции, которые используются в классических многопользовательских каналах, например системы на основе кодового разделения (CDMA), разделения скоростей [1], чередования [2], предполагают координирование передачи между пользователями. Кроме того, параметры таких схем, такие как разделяющая последовательность, кодовые скорости, графы Таннера кодов и т. д., зависят от числа пользователей в системе.

В данной работе рассматривается конструкция кодов для многопользовательского векторного дизъюнктивного канала, который был предложен Cohen, Heller и Viterbi в [3]. Данный канал является обобщением A-канала [4]– [7]. Анализируемая нами конструкция основана на перемеженных кодах Рида-Соломона (ПРС) с совместным декодированием, предложенным в [8]– [10]. Представлены численные результаты для относительной суммарной скорости передачи и вероятности ошибки на блок при различном числе пользователей и различных параметрах канала и конструкций кодов. Полученные результаты позволяют сделать вывод о том, что предложенная конструкция обладает хорошей корректирующей способностью даже в случае использования достаточно коротких кодов.

2 Модель канала

Вначале дадим описание канала, посредством которого осуществляется передача информации.

2.1 Описание канала

Рассмотрим векторный дизъюнктивный канал, который состоит из Q > 1 независимых подканалов (частот), которые мы будем называть элементарными подканалами. Предположим, что имеется U > 0 пользователей, которые используют канал для передачи. При классическом подходе произвольный *i*-ый пользователь ($1 \le i \le U$) в произвольный момент времени выбирает произольный *j*-ый подканал ($1 \le j \le Q$) и использует его для передачи.

Рассмотрим обобщение представленной выше модели. Предположим, что Q = LS, т. е. многопользовательский канал разделен на S независимых

каналов, каждый из которых состоит L = ml элементарных подканалов. Мы будем называть такие "обобщенные"каналы *L*-каналами или слотами. В произвольный момент времени τ каждый *i*-ый пользователь выбирает произволный *j*-ый *L*-канал ($1 \le j \le \frac{Q}{L}$) и передает в нем блок из *L* бит. Очевидно, что максимальное число пользователей, которые могут передавать ортогонально в данных условиях, равно *S*.

Если пользователю требуется передать K = Lk = mlk информационных бит, то выбирается k случайных чисел $i_1, i_2, ..., i_k, 1 \le i_t \le S$ и затем в каждый t-ый момент времени осуществляется передача L = ml бит в i_t -ом слоте

На рис. 1 представлена схема передачи.



Рис. 1. Модель передачи информации

2.2 Модель коллизии

В предыдущих рассуждениях мы не учли факт, что при описанном способе передачи могут возникать коллизии. Коллизии могут иметь место в том случае, когда два или более из U пользователей выбирают для передачи один и тот же L-канал в один момент времени. Очевидно, что вероятность P_s такого события может быть вычислена как:

$$P_s = 1 - \left(1 - \frac{1}{S}\right)^{U-1}.$$

Если обозначить через вероятность коллизии через P_{col} , то легко заметить, что $P_{col} \leq P_s$. Данное неравенство означает, что в некоторых случаях два и более пользователей могут передавать информацию в одном и том же слоте без коллизии для некоторого выбранного пользователя. Для того, чтобы посчитать вероятность коллизии более точно, рассмотрим случай

передачи двух пользователей *i* и *k*, которые выбрали общий *L*-канал для передачи в *j* момент времени.

Предположим, что пользователи *i* и *k* передают двоичные векторы длины *L*: $\mathbf{X}_{ij} = (x_{ij}^{(1)}, x_{ij}^{(2)}, \dots, x_{ij}^{(L)})$ и $\mathbf{X}_{kj} = (x_{kj}^{(1)}, x_{kj}^{(2)}, \dots, x_{kj}^{(L)})$ соответственно в одном и том же слоте. Ввиду нашего предположения о том, что канал является дизъюнктивным, то в общем слоте на приемник придет: $\mathbf{X}_{ij} \vee \mathbf{X}_{kj}$. Теперь рассмотрим передачу *i*-ого пользователя. Ввиду того, что канал бесшумный, то мы можем считать, что передача *k*-ого пользователя является шумом для *i*-ого пользователя. Учитывая это, можно сделать вывод о том, что для *i*-ого пользователя на приемнике будет наблюдаться коллизия в том и только том случае, когда вектор \mathbf{X}_{ij} не покрывает вектор \mathbf{X}_{kj} .

Рассмотрим данное событие более детально. Обозначим через $supp(\mathbf{X})$ носитель вектора \mathbf{X} , т. е.

$$supp(\mathbf{X}) = \{i : x_i \neq 0\}.$$

Таким образом, если *i*-ый и *k*-ый пользователи выбрали один *L*-канал для передачи, то коллизия для *i*-ого пользователя не будет иметь места, когда

$$supp(\mathbf{X}_{kj}) \subset supp(\mathbf{X}_{ij}).$$

Теперь рассчитаем вероятность P_{\subset} этого события. Обозначим через $[L] = \{1, 2, ..., L\}$. Таким образом, $supp(\mathbf{X}_{kj}) \subset [L]$ и $supp(\mathbf{X}_{ij}) \subset [L]$. Предположим, что элементы векторов \mathbf{X}_{kj} и \mathbf{X}_{ij} выбираются независимо и равновероятно: $p(0) = p(1) = \frac{1}{2}$. Таким образом, требуется рассчитать вероятность того, что случайное подмножество множества [L] является подмножеством другого случайно заданного подмножества [L]. Легко видеть, что данная вероятность P_{\subset} равна

$$P_{\subset} = \frac{1}{2^{2L}} \sum_{m=0}^{L} {\binom{L}{m}} 2^m = {\left(\frac{3}{4}\right)}^L.$$

Таким образом, можно заключить, что

$$P_{col} \approx P_s (1 - P_{\subset}).$$

2.3 Модель передачи

Ввиду того, что предложенная нами модель передачи допускает наличие коллизий, то для достоверной передачи информации требуется, чтобы пользователи вводили некоторую избыточность в передаваемые сообщения. Более детально: если информационный вектор *i*-ого пользователя имеет вид $\mathbf{X}_i = (\mathbf{X}_{i1}, \mathbf{X}_{i2}, \ldots, \mathbf{X}_{ik})$, где $\mathbf{X}_{ij} = (x_{ij}^{(1)}, x_{ij}^{(2)}, \ldots, x_{ij}^{(L)})$, $x_{ij}^{(\tau)} \in \{0, 1\}$, то для разрешения коллизий пользователь вводит проверочные векторы (векторы из проверочных символов) $\mathbf{P}_i = (\mathbf{P}_{i1}, \mathbf{P}_{i2}, \ldots, \mathbf{P}_{ih})$, где $\mathbf{P}_{ij} = (p_{ij}^{(1)}, p_{ij}^{(2)}, \ldots, p_{ij}^{(L)})$, $p_{ij}^{(\tau)} \in \{0, 1\}$ и затем отправляет в канал слово $\mathbf{C} = (\mathbf{X} \mathbf{P})$. Если n = k + h,

то для того, чтобы реализовать описанную выше передачу, пользователи должны использовать либо (nL, kL) двоичные коды, либо (n, k) коды над полем $GF(2^L)$, либо некоторые каскадные конструкции кодов.

Так как мы предполагаем, что рассматриваемый нами канал является бесшумным, то коллизии могут происходить только в том случае, когда несколько пользователей выбирают один и тот же слот для передачи в некоторый момент времени. Таким образом мы имеем дело с группированными по L символов коллизиями, причем в каждой такой группе количество коллизий в среднем L/2. Это наблюдение позволяет сделать вывод о том, что двоичные (nL, kL) не оптимальны в данной ситуации поскольку они предназначены для исправления независимых, а не группированных опибок.

На рис. 2 изображена модель передачи, представленная в данном разделе. Эллипсами выделены коллизии, произошедшие при передаче.



Рис. 2. Передача через канал - случай двух пользователей

3 Конструкция кодов

Как было отмечено ранее, q-ичные коды, где $q = 2^L$ являются более предпочтительным классом кодов для борьбы с коллизиями по сравнению с двоичными кодами. Данный факт позволяет применять широкий класс q-ичных кодов, например (n,k) кодов Рида-Соломона с минимальным расстоянием d = n - k + 1. Если использовать коды Рида-Соломона (PC) совместно с алгоритмом декодирования Берлекампа-Месси, то можно исправлять вплоть до $\lfloor \frac{n-k}{2} \rfloor$ независимых коллизий [11]. Если же требуется исправлять большее число коллизий, то вместо декодера Берлекампа-Месси следует использовать алгоритм Гурусвами-Судана, который исправляет вплоть до $1 - \sqrt{\frac{k}{n}}$ коллизий, имея при этом сложность $O(n^2m^4)$, где n – длина кода, а m – параметр, называемый "кратностью который имеет тот же порядок, что и n [12]. Ввиду высокой сложности алгоритм Гурусвами-Судана не нашел широкого практического применения.

В данной работе предлагается использовать другой подход к построению кодов: вместо применения обычного (n,k) кода Рида-Соломона над полем $GF(2^L)$ мы рассмотрим $l = \frac{L}{m}$ -кратное перемежение (n,k) кодов Рида-Соломона над $GF(2^m)$ (ПРС). Таким образом, наша схема имеет следующие параметры: $n = 2^m - 1$, $d = 2^m - k$. Такой подход позволяет использовать совместное декодирование l перемеженных кодов Рида-Соломона, которое позволяет исправлять вплоть до $t_{max} = \frac{l}{l+1}(d-1)$ коллизий [8]. Сложность такого алгоритма декодирования складывается из сложности l декодирований по алгоритму Берлекампа-Месси.

3.1 Перемеженные РС коды

Вначале дадим описание схемы кодирования. Для этого прежде всего требуется дать определение классических кодов Рида-Соломона. Существует два их альтернативных определения: на основе преобразования Фурье или на основе порождающего многочлена. Хотя первый метод значительно проще и естественнее, в результате его применения получается несистематический код. В данной работе для простоты рассматриваются только систематические коды. Именно поэтому мы дадим более "классическое"определение кодов Рида-Соломона – с точки зрения порождающего многочлена.

Определение 1 Пусть $GF^*(q) = \{\alpha_1, \alpha_2, ..., \alpha_{q-1}\}$ мультипликативная группа поля GF(q) с примитивным элементом α_1 , поэтому $\alpha_i = \alpha_1^i$. Рассморим информационный многочлен степени не выше чем k - 1: $u(x) = \sum_{i=0}^{k-1} u_i x^i$, $u_i \in GF(q)$. Тогда кодовое слово c(x) (n,k) кода Рида-Соломона с

минимальным расстоянием d = n - k + 1 имеет вид:

c(x) = u(x)g(x),

где $g(x) = \prod_{i=0}^{d-1} (x - \alpha_i)$ порождающий многочлен кода.

Хотя данное представление кода Рида-Соломона также задает несистематический код, достаточно легко преобразовать его к систематическому виду:

- 1. Умножим u(x) на x^{n-k} : $T(x) = u(x)x^{n-k}$
- 2. Найдем такой R(x)что: T(x) = Q(x)g(x) + R(x)
- 3. Получим слово $c_{sys}(x)$ систематического кода: $c_{sys}(x) = T(x) + R(x)$.

Так как каждый вектор $\mathbf{u} = (u_0, u_1, \ldots, u_{k-1}), u_i \in GF(q)$ может быть представлен в виде многочлена $u(x) = \sum_{i=0}^{k-1} u_i x^i$, то обозначим процедуру систематического кодирования вектора \mathbf{u} как $e(\mathbf{u})$.

Перемеженные коды Рида-Соломона могут быть получены из обычных PC кодов, если взять l слов классических PC кодов и представить их как строки $l \times n$ матрицы. Дадим более строгое определение таких кодов.

Определение 2 Рассмотрим последовательность $C^{(1)}, C^{(2)}, \ldots, C^{(l)}(n, k_i)$ РС кодов над полем GF(q). *l*-перемеженный РС код есть прямая сумма *l* РС кодов, т. е. если **C** – кодовое слово *l*-перемеженного РС кода, то

$$\mathbf{C} = \left\{ \begin{bmatrix} \mathbf{c}_1 \\ \mathbf{c}_2 \\ \dots \\ \mathbf{c}_l \end{bmatrix}, \mathbf{c}_i \in \mathcal{C}^{(i)} \right\} \in GF^l(q).$$

Если все l PC кодов являются эквивалентными, т.е. $k_1 = k_2 = \ldots = k_l = k$, то соответствующий ПРС код называется гомогенным. Иначе ПРС код называется гетерогенным. В данной работе рассматриваются только гомогенные коды.

Кодирование перемеженных кодов представляет из себя последовательное кодирование l информационных векторов обычным PC кодом. Предположим, что информационный вектор имеет вид $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{lk}), \mathbf{x}_i \in GF^m(2)$. Ввиду того, что мы можем рассматривать произвольный вектор $\mathbf{x}_i \in GF^m(2)$ как элемент α_i поля $GF(2^m)$, то информационный вектор может быть записан в виде $l \times k$ матрицы над полем $GF(2^m)$:

$$\mathbf{I} = \begin{pmatrix} \alpha_1 & \alpha_2 & \dots & \alpha_k \\ \alpha_{k+1} & \alpha_{k+2} & \dots & \alpha_{2k} \\ \dots & \dots & \dots & \dots \\ \alpha_{kl-k+1} & \alpha_{kl-k+2} & \dots & \alpha_{kl} \end{pmatrix} = \begin{bmatrix} \mathbf{I}_1 \\ \mathbf{I}_2 \\ \dots \\ \mathbf{I}_l \end{bmatrix}$$

Для того, чтобы получить слово С ПРС кода, необходимо закодировать каждый q-ичный вектор $\mathbf{I}_1, \mathbf{I}_2, \ldots, \mathbf{I}_l$ обычным (n, k) РС кодом:

$$\mathbf{C} = \begin{bmatrix} e(\mathbf{I}_1) \\ e(\mathbf{I}_2) \\ \dots \\ e(\mathbf{I}_l) \end{bmatrix} = \begin{bmatrix} \mathbf{c}_1 \\ \mathbf{c}_2 \\ \dots \\ \mathbf{c}_l \end{bmatrix}.$$

С этой точки зрения l-перемеженный PC код можно рассматривать как кодпроизведение со строчным (n, k) PC кодов и столбцовым (l, l) кодом скорости 1.

3.2 Алгоритм декодирования и граница на вероятность неправильного декодирования

Рассмотрим многопользовательский канал, состоящий из Q = LS, L = ml независимых подканалов. Предположим, что в канале ведут передачу U активных пользователей. Рассмотрим произвольного *i*-ого пользователя, считая передачу всех остальных пользователей шумом. Кроме того, будем считать, что пользователи используют *l*-перемеженный код PC для кодирования информации. Тогда вместо того, чтобы посылать в канал вектор $\mathbf{X} = (x_1, x_2, ... x_{mlk}), x_i \in GF(2)$, пользователи отправляют слова *l*-ПРС кода, основанного на (n, k) PC коде:

$$\mathbf{C} = \begin{bmatrix} \mathbf{c}_1 \\ \mathbf{c}_2 \\ \dots \\ \mathbf{c}_l \end{bmatrix}$$

На приемнике после канала наблюдаем:

$$\mathbf{Y} = \mathbf{C} + \mathbf{E} = egin{bmatrix} \mathbf{c}_1 + \mathbf{e}_1 \ \mathbf{c}_2 + \mathbf{e}_2 \ \dots \ \mathbf{c}_l + \mathbf{e}_l \end{bmatrix},$$

где **E** это $l \times n$ матрица коллизий над полем $GF(2^m)$. Алгоритм декодирования ПРС кодов (известный как алгоритм совместного декодирования) детально описан в работе [8]. Этот алгоритм работает только с символами поля $GF(2^m)$, не используя операций в расширенном поле $GF(2^{ml})$.

Кроме того, в работе [8] получена граница на вероятность P_w ошибочного совместного декодирования ПРС кодов. Дадим ее краткое описание:

$$P_{w} \leq \sum_{t=t_{g}+1}^{n} \binom{n}{t} \bar{P}_{w}(t) P_{col}^{t} (1 - P_{col})^{n-t},$$

где

$$\begin{split} t_g = \lfloor \frac{n-k}{2} \rfloor, \\ \bar{P}_w(t) = \begin{cases} 0 & t \leq t_g \\ min\{\bar{P}_e(t) + \bar{P}_f(t), \ 1\} & t > t_g \end{cases}, \end{split}$$

$$\bar{P}_f(t) \le \left(\frac{q^l - \frac{1}{q}}{q^l - 1}\right)^t \frac{q^{-(l+1)(t_{max} - t)}}{q - 1},$$

$$\begin{split} \bar{P}_e(t) &\leq \frac{\sum\limits_{w=d}^{t+t_{max}} A_w \sum\limits_{p=0}^{\min\{t,t_{max}\}} U(q,t,w,p)}{\binom{n}{t} (q-1)^t}, \\ U(q,t,w,p) &= \sum\limits_{i=\lceil \frac{t+w-p}{2}\rceil}^{t+w-p} \binom{t}{i} \binom{p-(t+w)+2i}{i} \cdot \\ &\cdot \binom{n-t}{w-i} (q-2)^{p-(t+w)+2i} (q-1)^{w-i}, \end{split}$$

и A_w количество слов веса w ПРС кода (спектр):

$$A_w = \binom{n}{w} (q^l - 1) \sum_{i=0}^{w-d} (-1)^i \binom{w-1}{i} q^{l(w-d-i)}.$$

Для примера рассмотрим 8-перемеженный PC код, построенный на базе (31,11) PC кода с минимальным расстоянием d = 21. Верхняя граница на вероятность опшбочного декодирования в зависимости от вероятности опшбки в q-ом симметричном канале представлена на рис. 3:



Рис. 3. Верхняя граница на *P*_w для 8-перемеженного (31,11,21) РС кода

4 Численные результаты

Рассмотрим результаты моделирования ПРС кодов в многопользовательском канале, представленном в данной работе. Для примера возьмем век-
торный дизъюнктивный канал, состоящий из Q = 5120 независимых подканалов. Ввиду того, что такие параметры как n, k, l и U могут быть практически произвольными для заданного Q, в зависимости от их выбора будут получаться различные сценарии передачи. Поэтому, для того, чтобы оценить эффективность ПРС кодов, необходимо рассмотреть величину, которая бы показывала эффективность использования канала для различного числа пользователей. Рассмотрим следующее выражение:

$$R_{\sigma} = \frac{ULR}{Q}(1 - P_w) = \frac{UR}{S}(1 - P_w),$$

где U – число активных пользователей, l – порядок перемежения, $R = \frac{k}{n}$ – скорость кода, $\lfloor S \rfloor = \lfloor \frac{Q}{L} \rfloor$ – максимальное число пользователей, способных вести ортогональную передачу, и P_w – вероятность ошибочного декодирования. Величина R_σ показывает суммарную относительную скорость передачи в векторном дизьюнктивном канале. Будем для кратности называть ее общей скоростью (англ. sum - rate).

Рассмотрим следующие параметры:

Таблица	1.	Параметры	модели	передачи
---------	----	-----------	--------	----------

n	k	l	$\lfloor S \rfloor$	d
31	6	8	128	26
31	8	8	128	24
31	11	8	128	21

Прежде чем построить кривые для общей скорости передачи для параметров, представленных в данной таблице, рассмотрим зависимость между числом пользователей и вероятностью ошибки на блок для каждого пользователя. Результаты приведены на рис. 4.

Результы, представленные на рис. 4 показывают ожидаемую зависимость: с увеличением числа активных пользователей вероятность ошибки (неисправленной коллизии) для каждого пользователя увеличивается.

Теперь рассмотрим численные результаты для общей скорости передачи ПРС кодов с параметрами, взятыми из таблицы 1 (см. рис. 5).

На рис. 5 видно, что максимальные величины относительной суммарной скорости передачи практически одинаковы, хотя и незначительно убывают с уменьшением кодовой скорости R. Наибольшее из всех значений R_{σ} примерно равно 0.195. Кроме того, можно заметить, что чем выше скорость передачи, тем при меньшем числе пользователей величина R_{σ} достигает своего максимума. Таким образом, зная среднее число пользователей в системе, можно выбрать оптимальную конструкцию кодов так, чтобы величина R_{σ} достигала своего максимума.



Рис. 4. Зависимость между числом пользователей *U* и вероятностью ошибки на блок для каждого пользователя для параметров из таблицы 1.



Рис. 5. Зависимость между общей относительной суммарной скоростью передачи R_{σ} и относительным числом активных пользователей U/S для кодов с параметрами из таблицы 1.

5 Заключение

В работе представлена кодовая конструкция для многопользовательского векторного дизъюнктивного канала. Данная конструкция основана на перемеженных кодах Рида-Соломона. Для предложенной конструкции представлены численные результаты для относительной суммарной скорости передачи и вероятности ошибки (неисправленной коллизии) на блок при различном числе активных пользователей. Полученные результаты позволяют сделать вывод о том, что предложенная конструкция обладает хорошей корректирующей способностью даже в случае использования достаточно коротких кодов, что позволяет говорить о практической применимости данной системы.

Список литературы

- B. Rimoldi and R. Urbanke, "A rate-splitting approach to the Gaussian multipleaccess channel," IEEE Transactions on Information Theory, vol. 42, no. 2, pp. 364– 375, 1996.
- L. Ping, L. Liu, K. Wu, and W. K. Leung, "Interleave division multiple-access," IEEE Transactions on Wireless Communications, vol. 5, no. 4, pp. 938–947, 2006.
- A. R. Cohen, J. A. Heller and A. J. Viterbi, "A new coding technique for asynchronous multiple access communication," IEEE Transactions on Communication Technology, vol. 19, pp. 849–855, October 1971.
- S. C. Chang and J. Wolf, "On the t-User m-Frequency Noiseless Multiple-Access Channel with and without Intensity Information," IEEE Transactions on Information Theory, vol. 27, no. 1, pp. 41–48, Jan. 1981.
- L. Wilhelmsson and K. S. Zigangirov, "On the Asymptotic Capacity of a Multiple-Access Channel," Problems of Information Transmission, vol. 33, no. 1, pp. 9–16, 1997.
- L. A. Bassalygo and M. S. Pinsker, "Evaluation of the Asymptotics of the Summarized Capacity of an m-Frequency t-User Noiseless Multiple-Access Channel," Problems of Information Transmission, vol. 36, no. 2, pp. 91–97, 2000.
- A. Han Vinck and K. Keuning, "On the Capacity of the Asynchronous t-User m-Frequency Noiseless Multiple-Access Channel without Intensity Information," IEEE Transactions on Information Theory, vol. 42, no. 6, pp. 2235–2238, Nov. 1996.
- G. Schmidt, V. R. Sidorenko and M. Bossert, "Collaborative decoding of interleaved Reed–Solomon codes and concatenated code designs," IEEE Transactions on Information Theory, vol. 55, no. 7, pp. 2991–3012, 2009.
- S. Puchinger and J. R. Nielsen, "Decoding of interleaved Reed-Solomon codes using improved power decoding," Information Theory (ISIT), Proceedings of IEEE International Symposium on Information Theory (ISIT), pp. 356–360, 2017.
- A. Wachter-Zeh, A. Zeh and M. Bossert, "Decoding interleaved Reed–Solomon codes beyond their joint error-correcting capability," Designs, Codes and Cryptography, vol. 7, no. 2, pp. 261–281, 2014.
- J. L. Massey, "Shift-register synthesis and BCH decoding," IEEE Transactions on Information Theory, vol. 15, pp. 122–127, 1969.
- R. J. McEliece, "The guruswami-sudan decoding algorithm for Reed-Solomon codes," IPN progress report, pp. 42–153, 2003.

Об оценке пропускной способности специального класса векторных дизъюнктивных каналов*

Иванов Ф. И.^{1,2}, Рыбин П. С.^{1,2}, Афанасьев В. Б.¹

¹ Национальный исследовательский университет «Высшая школа экономики», Москва, Россия

² Институт проблем передачи информации им. А.А. Харкевича Российской академии наук, Москва, Россия

Аннотация В работе рассматривается векторный дизъюнктивный канал, в котором пользователи передают пакеты, состоящие из L бит. Рассмотрена пропускная способность данного канала и получена нижняя граница данной величины. Для данной оценки были получены численные результаты при различном числе активных пользователей, ведущих передачу, а также при различных параметрах канала.

1 Введение

В настоящее время одним из ключевых особенностей беспроводных сетей является очень быстрое увеличение числа активных пользователей в системе, что вызвано активным развитием концепции интернета вещей и концепции умного дома. Ввиду этого взаимное влияние пользователей друг на друга становится все более ощутимым. В то же время требования, которые предъявляются к скорости передачи данных и вероятности успешной передачи, становятся все более жесткими.

В связи с вышесказанным, требуется, чтобы очень большое число пользователей могли вести некоординированную передачу в беспроводных сетях. Среди общего числа пользователей следует особо выделить подмножество из U пользователей, которые активны в любой момент времени, при этом каждый из них ведет передачу коротких пакетов информации на базовую станцию. При этом задача базовой станции заключается только в восстановлении списка сообщений, которые были на нее переданы, при этом не происходит идентификация пользователя, который передал конкретное сообщение. Некоординированный способ передачи, а также короткая длина передаваемых сообщений существенно влияют на методы построения сигнальнокодовых конструкций, которые были бы практически применимыми в данных условиях.

^{*} Работа выполнена при поддержке грантов Р
ФФИ 18-37-00319 мол_а и 18-07-01409 А

Практически все широко известные сигнально-кодовые конструкции, которые используются в классических многопользовательских каналах, например системы на основе кодового разделения (CDMA), разделения скоростей [1], чередования [2], предполагают координирование передачи между пользователями. Кроме того, параметры таких схем, такие как разделяющая последовательность, кодовые скорости, графы Таннера кодов и т. д., зависят от числа пользователей в системе.

В данной работе мы рассматриваем специальный класс некоординированного многопользовательского канала – векторного дизъюнктивного канала, известного также, как Z-канал. Данный канал был предложен Cohen, Heller и Viterbi в работе [3]. Работы [4], [5] посвящены исследованию свойств такого канала. Кроме того, в работах [6], [7], [8] [9] исследуется пропускная способность такого канала при некоторых дополнительных допущениях.

Мы рассматриваем специальный класс векторного-дизъюнктивного канала, где каждый пользователь передает некоторый двоичный вектор длины $L \ge 1$. Мы оцениваем пропускную способность данного канала в естественном предположении о том, что векторы длины L являются равновероятными. Основываясь на данном предположении, мы получили нижнюю границу на пропускную способность канала. Полученная величина является функцией как от числа активных пользователей U, так и от других параметров канала.

2 Модель канала

2.1 Описание канала

Рассмотрим векторный дизъюнктивный канал, который состоит из Q>1 независимых подканалов (частот), которые мы будем называть элементарными подканалами. Предположим, что имеется U>0 пользователей, которые используют канал для передачи. При классическом подходе произвольный *i*-ый пользователь ($1 \le i \le U$) в произвольный момент времени выбирает произольный *j*-ый подканал ($1 \le j \le Q$) и использует его для передачи.

Рассмотрим обобщение представленной выше модели. Предположим, что Q = LS, т. е. многопользовательский канал разделен на S независимых каналов, каждый из которых состоит L элементарных подканалов. Мы будем называть такие "обобщенные"каналы L-каналами или слотами. В произвольный момент времени τ каждый *i*-ый пользователь выбирает произволный *j*-ый L-канал ($1 \le j \le \frac{Q}{L}$) и передает в нем блок из L бит. Очевидно, что максимальное число пользователей, которые могут передавать ортогонально в данных условиях, равно S.

2.2 Модель коллизии

В предыдущих рассуждениях мы не учли факт, что при описанном способе передачи могут возникать коллизии. Коллизии могут иметь место в том

случае, когда два или более пользователей выбирают для передачи один и тот же L-канал в один момент времени. Очевидно, что вероятность P_s такого события может быть вычислена как:

$$P_s = 1 - \left(1 - \frac{1}{S}\right)^{U-1}.$$

В данной работе мы рассматриваем бесшумный канал и некоординированную передачу. Поэтому для заданного пользователя передача остальных *U* – 1 пользователей будет рассматриваться как шум (источник коллизий).

Обозначим через κ случайную величину – кратность коллизии. Данная величина может быть рассмотрена как количество активных пользователей, которые выбрали заданный *j*-ый *L*-канал, в котором ведет передачу пользователь. Очевидно, что $0 \le \kappa \le U - 1$. Кроме того, легко посчитать вероятность события $\kappa = t$ (построить распределение случайной величины κ):

$$P(\kappa = t) = {\binom{U-1}{t}} \left(\frac{1}{S}\right)^t \left(1 - \frac{1}{S}\right)^{U-t-1}.$$

В самом деле, ввиду того, что каждый из U-1 пользователей независимо и равновероятно (с вероятностью $\frac{1}{S}$) выбирает фиксированный *L*-канал для передачи, то *t* пользователей выберут один и тот же канал с вероятностью $\left(\frac{1}{S}\right)^t \binom{U-1}{t}$. Остальные U-1-t пользователей могут выбрать любой из оставнихся S-1 *L*-каналов. Вероятность такого события равна $\left(1-\frac{1}{S}\right)^{U-t-1}$. Легко заметить, что

$$P_s = P(\kappa \ge 1) = \sum_{t=1}^{U-1} P(\kappa = t).$$

В некоторых случаях передача других пользователей в том же слоте, где ведет передачу заданный пользователь, не приводит к коллизии. Например, если пользователь передает (1, 1, ..., 1), то такая информация будет принята корректно вне зависимости от того, что передают другие пользователи в этом же слоте.

Рассмотрим случай передачи двух пользователей i и k, которые выбрали общий L-канал для передачи в j момент времени.

Предположим, что пользователи *i* и *k* передают двоичные векторы длины *L*: $\mathbf{X}_{ij} = (x_{ij}^{(1)}, x_{ij}^{(2)}, \dots, x_{ij}^{(L)})$ и $\mathbf{X}_{kj} = (x_{kj}^{(1)}, x_{kj}^{(2)}, \dots, x_{kj}^{(L)})$ соответственно в одном и том же слоте. Ввиду нашего предположения о том,что канал является дизъюнктивным, то в общем слоте на приемник придет: $\mathbf{X}_{ij} \vee \mathbf{X}_{kj}$. Теперь рассмотрим передачу *i*-ого пользователя. Ввиду того, что канал бесшумный, то мы можем считать, что передача *k*-ого пользователя является шумом для *i*-ого пользователя. Учитывая это, можно сделать вывод о том, что для *i*-ого пользователя на приемнике будет наблюдаться коллизия в том и только том случае, когда вектор \mathbf{X}_{ij} не покрывает вектор \mathbf{X}_{kj} . Рассмотрим данное событие более детально. Обозначим через $supp(\mathbf{X})$ носитель вектора \mathbf{X} , т. е.

$$supp(\mathbf{X}) = \{i : x_i \neq 0\}.$$

Таким образом, если *i*-ый и *k*-ый пользователи выбрали один *L*-канал для передачи, то коллизия для *i*-ого пользователя не будет иметь места, когда

$$supp(\mathbf{X}_{kj}) \subset supp(\mathbf{X}_{ij}).$$

Теперь рассчитаем вероятность P_{\subset} этого события. Обозначим через $[L] = \{1, 2, ..., L\}$. Таким образом, $supp(\mathbf{X}_{kj}) \subset [L]$ и $supp(\mathbf{X}_{ij}) \subset [L]$. Предположим, что элементы векторов \mathbf{X}_{kj} и \mathbf{X}_{ij} выбираются независимо и равновероятно: $p(0) = p(1) = \frac{1}{2}$. Таким образом, требуется рассчитать вероятность того, что случайное подмножество множества [L] является подмножеством другого случайно заданного подмножества [L]. Легко видеть, что данная вероятность P_{\subset} равна

$$P_{\subset} = \frac{1}{2^{2L}} \sum_{m=0}^{L} {\binom{L}{m}} 2^m = {\binom{3}{4}}^L.$$

3 Анализ свойств канала

Предыдущие рассуждения могут быть обобщены для произвольной кратности коллизии. Для лучшего понимания предложенного нами метода анализа канала, рассмотрим пример канала, представленный на рис. 1



Рис. 1. Возможные переходы для L = 3

На рис. 1 представлены все возможные переходы между входами и выходами канала. Как можно заметить, для входного вектора длины L и веса w существует 2^{L-w} возможных переходов в выходные векторы. Таким образом, количество переходов и вероятность перехода из входа в выход зависит только от веса входного вектора и не зависит от его конкретного представления (носителя). Более того, можно заметить, что вес выходного вектора увеличится по сравнению с весом входного вектора тогда только тогда, когда при передаче произошла коллизия. Таким образом, мы можем представить наш исходный канал с 2^L равновероятными входами и 2^L выходами, как канал с L + 1 входом и L + 1 выходом, которые соответствуют весам входных и выходных векторов (см. рис. 2).



Рис. 2. Возможные переходы для L = 3

Следует обратить внимание на то, что в данном случае входы канала перестали быть равновероятными. В частности, распределение на входе канала имеет вид:

$$P_{in}(w) = \binom{L}{w} 2^{-L}.$$

Очевидно, что входное распределение не зависит от кратности коллизии κ . В то же время, выходное распределение зависит от этой величины. Предположим, что $\kappa = t$. Если входной вектор **u** имеет вес w < L, то выходной вектор **v** имеет вес w' > w если хотя бы один из t пользователей (которые используют тот же слот для передачи, что и выбранный пользователь) передает 1 на любой из позиций из множества $[L] \setminus supp(\mathbf{u})$. Вероятность p(t) того, что хотя бы один из t пользователей позиции $k \in [L] \setminus supp(\mathbf{u})$, равна:

$$p(t) = 1 - (1 - p)^t,$$

где p(1) = p – вероятность передачи 1. Ввиду нашего предположения о том, что $p(1) = p(0) = \frac{1}{2}$, имеем:

$$p(t) = 1 - 2^{-t}.$$

Учитывая данный факт, легко посчитать условные вероятности P(w'|w,t) получить выходной вектор веса w' для заданного вектора веса w, если произошла коллизия порядка t:

$$P(w'|w,t) = \begin{cases} 0, & w' < w \\ \binom{L-w}{w'-w} p(t)^{w'-w} (1-p(t))^{L-w'}, & w' \ge w \end{cases}$$

Наконец, распределение выходов канала следующее:

$$P_{out}(w'|t) = \sum_{i=0}^{w'} P(w'|i,t) P_{in}(i)$$

4 Оценка пропускной способности

Оценим пропускную способность канала, описанного выше. Вначале мы получим оценку на пропускную способность для заданной кратности коллизии $\kappa = t$ а затем рассмотрим математическое ожидание данной величины при всех возможных кратностях коллизии, получив таким образом нижнюю границу на пропускную способность векторного дизъюнктивного канала, представленного в данной работе.

Вначале напомним определение пропускной способности канала. Пусть \mathcal{X} и \mathcal{Y} – случайные величины, описывающие вход и выход канал соответсвенно. Пропускная способность \mathcal{C} канала определяется следующим образом:

$$\mathcal{C} = \sup_{p_{in}(\mathcal{X})} I(\mathcal{X}; \mathcal{Y}),$$

где $I(\mathcal{X}; \mathcal{Y})$ — взаимная информация между \mathcal{X} и \mathcal{Y} . Супремум берется по всем возможным распределениям $p_{in}(\mathcal{X})$ случайной величины \mathcal{X} .

Если кратность коллизии фиксирована $\kappa = t$, то пропускная способность C(t) канала для коллизии кратности t оценивается как:

$$C(t) \ge \sum_{w=0}^{L} P_{in}(w) \sum_{w'=w}^{L} P(w'|w,t) \log P(w'|w,t) - \sum_{w''=0}^{L} P_{out}(w''|t) \log P_{out}(w''|t).$$

Таким образом, пропускная способность канала \mathcal{C} может быть оценена снизу математическим ожидаением величины $\mathcal{C}(t)$ по всем возможным кратностям коллизии t:

$$\mathcal{C} \ge C^{\star} = \sum_{t=0}^{U-1} P(\kappa = t)\mathcal{C}(t).$$

5 Численные результаты

В данном разделе мы анализируем численные результаты для полученной нижней границы на пропускную способность. Для начала фиксируем Q = 1024 и рассмотрим зависимость между C^* и числом активных пользователей U для различного числа ортогональных слотов S = Q/L: $S \in \{16, 32, 64, 128, 256, 512, 1024\}$. Данные зависимости представлены на рис. 3:



Рис. 3. Нижняя граница на пропускную способность при Q=1024 и различных S=Q/L

Как можно заметить, величина C^* для заданного числа пользователей растет вместе с S. Данное поведение наблюдается ввиду того, что вероятность коллизии уменьшается с ростом S, так как чем больше S, тем больше пользователей могут передавать информацию ортогонально.

Так как C^* можно рассматривать как долю информации, которая может быть передана корректно, то имеет смысл рассматривать "скорость передачи данных"в виде $T = C^*L$. Данная величина определяет долю корректно переданных бит из общего числа L бит, которые пользователь отправляет на приемник через канал.

На рис. 4 представлена зависимость между нижней границей на "скорость передачи данных"T и числом активных пользователей U для Q = 1024и различными значениями параметра S = Q/L. Можно заметить, что для некоторых диапазонов $[0; U_{th}]$ числа пользователей при меньших значениях S достигается большие значения T. Таким образом, для того, чтобы увеличить скорость передачи при относительно небольшом числе пользователей в системе, имеет смысл увеличивать число L (число переданных пользователем бит за одно использование канала), тем самым уменьшая S. При уве-



Рис. 4. Нижняя граница на "скорость передачи данных " $T=C^{\star}L$ для Q=1024и различных значений S=Q/L

личении же числа активных пользователей имеет смысл увеличивать число ортогональных слотов S.

Для того, чтобы оценить эффективность предложенной схемы передачи, рассмотрим относительную суммарную скорость передачи в векторном дизъюнктивном канале:

$$R_{\sigma} = \frac{ULC^{\star}}{Q} = \frac{UT}{Q}.$$

Данная величина является соотношением между суммарной скоростью передачи всех пользователей (UT) и максимальной достижимой скоростью передачи (Q) системы.

На рис. 5 представлена зависимость между R_{σ} и числом пользователей U при фиксированном Q = 1024 и значениях $S \in \{16, 32, 64, 128, 256, 512, 1024\}$. Очевидно, что чем меньше S, тем при меньшем числе пользователей величина R_{σ} достигает своего максимума. Более того, максимальные значения R_{σ} также убывают вместе с уменьшением S. В то же время существуют об-



Рис. 5. Нижняя граница на относительную суммарную скорость передачи $R_{\sigma} = UC^{\star}L/Q$ при Q = 1024 и различных значениях S = Q/L

ласти (когда число пользователей относительно невелико), где для меньших значений S достигаются большие значения R_{σ} .

Теперь рассмотрим "оптимальное" число пользователей, то есть такое число пользователей U_{opt} , что $U_{opt} = \underset{U}{\operatorname{argmax}} R_{\sigma}(U,Q,S)$. Очевидно, что U_{opt} является функцией от Q и S.

На рис. 6 представлена зависимость между U_{opt} и различными значениями Q и S. Можно заметить, что для фиксированного числа ортогональных слотов S, оптимальное количество пользователей системы практически не зависит от Q. Но еще более интересным фактом является то, что число U_{opt} всегда превосходит S – максимальное число пользователей, которые могут работать ортогонально в системе. Более того, чем больше S, тем больше относительная разница между U_{opt} и S, например если S = 128, то среднее значение U_{opt} равно 200, что примерно в 1,56 раз больше S.

Теперь рассмотрим максимальные значения R_{σ} , т. е. $R_{\sigma}(U_{opt})$ для $Q \in \{256, 512, 1024, 2048, 4096\}$ и $S \in \{16, 32, 64, 128\}.$



Рис. 6. Число пользователей, для которого достигается максимум R_{σ} для различных Q и S = Q/L.

На рис. 7 представлена зависимость между $R_{\sigma}(U_{opt})$ и различными значениями Q и S. Можно сделать вывод, что максимальное значение R_{σ} всегда убывает с ростом Q вне зависимости от значения S. Однако, в то же время легко видеть, что чем больше величина Q, тем меньше $R_{\sigma}(U_{opt})$ зависит как от Q, так и от S. Данное наблюдение, полученное численными методами, позволяет предположить о существовании теоретического асимптотического значения суммарной относительной скорости передачи для данной модели канала.

6 Заключение

В работе рассмотрен векторный дизъюнктивный канал, где пользователи передают двоичные векторы длины L. Была получена оценка на пропускную способность данного канала. Численные результаты для скорости передачи и суммарной относительной скорости передачи позволяют сделать заключение о том, что предложенная схема передачи может быть использована в том случае, когда число активных пользователей значительно превосходит число слотов, при котором возможна ортогональная передача данных.

Более того, представленная схема передачи обладает таким достоинствами как: простой метод приема информации от пользователей, некоорди-



Рис. 7. Максимальное значение R_{σ} для различныхQ и S = Q/L.

нированная передача, возможность динамически изменять число активных пользователей. Кроме того, относительная суммарная скорость передачи в системе остается ненулевой даже в том случае, когда число активных пользователей существенно превосходит то количество, которое допускает ортогональную передачу данных.

Список литературы

- B. Rimoldi and R. Urbanke, "A rate-splitting approach to the Gaussian multipleaccess channel," IEEE Transactions on Information Theory, vol. 42, no. 2, pp. 364– 375, 1996.
- L. Ping, L. Liu, K. Wu, and W. K. Leung, "Interleave division multiple-access," IEEE Transactions on Wireless Communications, vol. 5, no. 4, pp. 938–947, 2006.
- A. R. Cohen, J. A. Heller and A. J. Viterbi, "A new coding technique for asynchronous multiple access communication," IEEE Transactions on Communication Technology, vol. 19, pp. 849–855, October 1971.
- D. S. Osipov, A. A. Frolov and V. V. Zyablov, "Multiple access system for a vector disjunctive channel," Problems of information Transmission, vol. 48, no. 3, pp. 243– 249, 2012.

- A. A. Frolov, V. V. Zyablov, V. R. Sidorenko and R. Fischer, "On a multiple-access in a vector disjunctive channel," Information Theory (ISIT), Proceedings of IEEE International Symposium on Information Theory (ISIT), pp. 211–215, 2013.
- L. Wilhelmsson and K. S. Zigangirov, "On the Asymptotic Capacity of a Multiple-Access Channel," Problems of Information Transmission, vol. 33, no. 1, pp. 9–16, 1997.
- S. C. Chang and J. Wolf, "On the t-User m-Frequency Noiseless Multiple-Access Channel with and without Intensity Information," IEEE Transactions on Information Theory, vol. 27, no. 1, pp. 41–48, Jan. 1981.
- L. A. Bassalygo and M. S. Pinsker, "Evaluation of the Asymptotics of the Summarized Capacity of an m-Frequency t-User Noiseless Multiple-Access Channel," Problems of Information Transmission, vol. 36, no. 2, pp. 91–97, 2000.
- A. Han Vinck and K. Keuning, "On the Capacity of the Asynchronous t-User m-Frequency Noiseless Multiple-Access Channel without Intensity Information," IEEE Transactions on Information Theory, vol. 42, no. 6, pp. 2235–2238, Nov. 1996.

Etalog – a natural-looking knowledge representation formalism^{*,**}

Ivan Rygaev¹

¹ Laboratory of Computational Linguistics, A. A. Kharkevich Institute for Information Transmission Problems, Russian Academy of Sciences, Moscow, Russia irygaev@gmail.com

Abstract. In this paper we present Etalog – a formal knowledge representation language developed in the Laboratory of Computational Linguistics IITP RAS in scope of the semantic text analyzer SemETAP. Borrowing ideas from existing formalisms we created a formal language which is a) more expressive than Description Logic, b) more human-readable than Datalog and Turtle, c) allows capturing not only truth-conditional but also (a part of) communicative structure of the sentence.

Keywords: Semantics, Natural language understanding, Knowledge representation, Inference.

1 Introduction

The semantic text analyzer SemETAP, under development in the Laboratory of Computational Linguistics IITP RAS, is aiming at performing deep semantic analysis of natural language texts. Based on various linguistic and extra-linguistic resources (a combinatorial dictionary, an ontology, a fact base and an inference engine), SemETAP is able to build a semantic structure of the text and extend it with implicit knowledge using inference rules (Boguslavsky et al 2015, 2018).

Semantic structure is based on the ontology and is built in two steps. First a syntactic (dependency) tree of the sentence is produced, and then (roughly) words are replaced with semantic concepts from the ontology and syntactic relations – with semantic ones, which results in a set of binary predicates (triples) of the form relation (node1, node2) or a semantic graph.

Below is an example of a sentence, its syntactic tree, semantic graph and the corresponding set of triples:

⁶ This paper presents the results of a joint effort of the team of the Laboratory of Computational Linguistics IITP RAS. The team includes I. Boguslavsky, L. Iomdin, A. Lazursky, S. Timoshenko, T. Frolova, V. Dikonov, E. Inshakova, V. Sizov and others. I would like to thank my colleagues for their wonderful collaboration.

^{**} This work was supported by the RSF grant 16-18-10422, which is gratefully acknowledged.

(1) John sold an umbrella to Peter



hasAgent2 (Selling_1, Human_2)
hasObject (Selling_1, Umbrella_1)
hasGivenName (Human_2, "Peter")
before (Selling_1, SpeechTimeInterval)

In the semantic structure, each node represents an individual of a certain class (semantic concept). There are implicit triples that assign a class to each individual, such as isA(Human_1, Human) or Human(Human_1), but since the class name is encoded in the individual name these triples are omitted for the sake of simplicity.

A sale event could be represented as an n-place predicate: Selling (Human_1, Umbrella_1, Human_2), but its reduction to binary predicates as suggested by neo-Davidsonian semantics (Parsons 1990, Higginbotham 2000), has a number of advantages: a) the event individual (Selling_1) allows attaching time, location and other adjuncts to the event; b) unexpressed arguments (such as selling price) can be safely ignored; c) the semantic structure resembles the syntactic one in a more straightforward way; d) it makes the semantic structure compatible with Semantic Web standards – RDF and OWL.

On the other hand a big list of unsorted triples is hard to read or type manually (when writing inference rules). Etalog was created to solve this issue. While maintaining the underlying structure as a set of triples it allows presenting it in such a way that simplifies reading and typing. Etalog is used for two purposes:

- 1. to write inference rules;
- 2. to represent the semantic structure of the sentence.

2 Language Features

Etalog emerged as a language for inference rules, most of which in our system are concept decomposition rules. Concept decompositions are similar to word definitions in the explanatory dictionary, but are written in formal language. They are created manually by linguists.

Originally both the premises and the conclusion of the rules were written as a list of triples. In our rules one variable usually participates in many relations, so the variable name had to be typed multiple times. A small typo in the variable name could lead to hard-to-identify logical errors. In the example below variables ?selling and ?buying are repeated 7 times each, other variables – 3 times each.

```
(3) Rule Selling: // Sale
```

```
Selling (?selling) ->
hasAgent(?selling,?seller), Agent(?seller), //seller
hasAgent2(?selling,?buyer), Agent(?buyer), //buyer
hasObject(?selling,?thing), Thing(?thing), //product
hasPrice(?selling,?money), Money(?money), //money
// Sale is defined through Purchase
hasSyncEvent(?selling,?buying), Buying(?buying),
hasAgent(?buying,?buyer),
hasAgent2(?buying,?seller),
hasPrice(?buying,?thing),
hasPrice(?buying,?money),
hasSyncEvent(?buying,?selling).
```

To reduce the variable usage the following changes were made to the language:

- 1. Instead of functional PSO notation, triples are now written in SPO notation without parentheses: ?selling hasAgent ?seller. This is similar to RDF/Turtle (Beckett et al 2014) and offers a closer resemblance to natural language. Unary predicates (class assignments) are also written without parentheses but they precede the subject: Agent ?seller.
- 2. Two or more consecutive triples with the same subject can be joined together without repeating the subject: ?selling hasAgent ?seller hasAgent2 ?buyer is equivalent to ?selling hasAgent ?seller, ?selling hasAgent2 ?buyer. Turtle also supports this feature but requires triples to be separated by a semicolon. In Etalog no special

separator is needed, which resembles the behavior of prepositions in natural language, where we can attach multiple prepositions to the same verb: 'go from Moscow to Kazan'. This feature allows us to write a complex proposition with a single usage of the subject variable.

- 3. If the subject variable is used only once in the expression, it can be omitted altogether when preceded by a class name. Buying hasAgent ?buyer is equivalent to Buying ?x hasAgent ?buyer. The variable name is generated automatically for internal purposes and is not shown to the user.
- 4. Nested expressions are allowed. In the place of a variable it is possible to use a (bracketed) complex proposition in which this variable is a subject. ?selling hasAgent (Agent ?seller) is equivalent to ?selling hasAgent ?seller, Agent ?seller. Nested expressions also allow the variable name to be omitted, if it is not used anywhere else: ?selling hasAgent (Agent).
- 5. Native support for inverse relations introduced. Relations of the form hasX (such as hasAgent, hasObject, etc.) can be used in the inverse form of isXOf (isAgentOf, isObjectOf, etc.) even if the inverse relation is not defined in the ontology. ?seller isAgentOf ?selling is equivalent to ?selling hasAgent ?seller but has ?seller as a subject. So it can be combined with other triples where ?seller is a subject into a complex proposition: Agent ?seller isAgentOf (Selling hasObject (Umbrella)). This resembles the relative clause construction in natural languages: 'An agent who sold the umbrella'.

All these features greatly reduce the usage of variable names in Etalog expressions and make the Etalog text look more natural. Here is how the above rule will look like after all the features are applied:

```
(4) Rule Selling: // Sale
    Selling ?selling ->
    ?selling
      hasAgent (Agent ?seller) // seller
      hasAgent2 (Agent ?buyer) // buyer
      hasObject (Thing ?thing) // product
      hasPrice (Money ?money) // money
      hasSyncEvent
       ( // Sale is defined through Purchase
         Buying
           hasAgent ?buyer
           hasAgent2 ?seller
           hasObject ?thing
           hasPrice ?money
           hasSyncEvent ?selling
      ).
```

A semantic graph having a tree-like structure can be written in Etalog without mentioning any variable names whatsoever. This is similar to Description Logic (Baader et al 2003) and OWL Abstract Syntax (Patel-Schneider et al 2003). But unlike DL and OWL, which allow only tree-like expressions and do not make use of variables, in Etalog you can link different branches of the tree by introducing explicit variables and thus make a loop.

Loops are very useful in concept decomposition rules, where different parts of the decomposition are usually interconnected. For example, see below a semantic graph of rule (4). DL and OWL are not expressive enough to capture such links.



On the logical level Etalog is compatible with Datalog^{+/-} (Calì et al 2011) and supports the same set of logical constructions, namely:

- 1. Conjunction
- 2. Implication
- 3. Implicit universal quantification in the antecedent
- 4. Implicit existential quantification in the consequent

Negation, disjunction, explicit quantification and modality are not supported at the level of language. There is some partial support for negation and modality in SemETAP semantic analyzer, but they are modelled as classes in the ontology, not as logical constructions of the language. Introducing full-fledged support for these features is constrained by RDF model (which has native support only for conjunction) and by the reasoner capabilities (adding these features can greatly increase the reasoning complexity).

3 Inference Rules Application Process

Another advantage of Etalog is that it hides technical details of the rule application from the linguist who creates the rule.

As we can see from the example above, concept decomposition almost always requires new variables to be present in the consequent which are not there in the antecedent. These variables represent various parts of the concept decomposition, in particular the arguments of an event.

Every situation of selling involves a seller, a buyer, a product and money. If anything is missing from the situation it cannot be called 'selling'. Nevertheless not all of these four arguments are necessarily instantiated in a sentence. In (1) there is no mentioning of money. In '*The car is sold*' three of the four arguments are missing. They may be missing in a sentence but they are still there in the situation the sentence describes. So, on the semantic level we have to restore all the missing arguments.

Rules are applied in the forward chaining manner known as *chase* (Benedikt et al 2017), i. e. they extend the existing semantic structure by adding new propositions and individuals to it. But we cannot just create new individuals for all new variables in the consequent. First we need to check if any of them are already there in the original semantic structure.

The logic of these checks can be complicated (see Rygaev 2017 for details) but the bottom-line is that an Etalog rule is internally split into a number of smaller implications, which are applied independently. Each such implication checks the existence of one or more individuals and creates them only if they are missing. For example, there will be a separate implication for each functional relation mentioned in the rule text.

So, for an Etalog rule it is possible that some individuals will be accommodated from the existing data while others will be added. This happens invisibly for the linguists who create rules in Etalog, so they can concentrate on the concept decomposition and ignore technical aspects of the rule application.

As we mentioned functional relations we have to note that Etalog rule compilation and application processes import certain data from an OWL ontology. This includes information about:

- 1. Class hierarchy
- 2. Inverse relations
- 3. Functional and inverse functional relations

Although all this information can be written in Etalog directly, it is much easier to include it in the ontology using OWL editors such as Protégé (Gennari 2003).

4 Semantic Structure Representation

Since the semantic structure of the sentence is also a set or triples, it is more than natural to represent it as an Etalog expression as well for better readability. But unlike inference rules, which are created manually, this task requires automatic generation of the Etalog expression from the set of triples.

Given that Etalog has substantial expressive power, we see that there are now many different ways to represent the same set of triples. So the question arises: how to choose the most appropriate expression? How to group nodes into complex propositions? Where to use nested expressions and inverse relations?

We came up with the following solution:

1. Do not invert any relations unless they are explicitly inverted in the original set of triples.

- 2. Identify the closest head for each individual in the structure. Starting from a node in the semantic graph and following the reverse direction of arrows identify the closest node which does not have any incoming arrows.
- 3. In case of loops when we return to the same node without finding a head sort all individuals in the loop alphabetically and declare the first one to be the head.
- 4. Group all nodes by the closest head.
- 5. Express each group as a single complex proposition starting from the head (using the head as its subject).

Here is the Etalog expression for sentence (1)

```
(5) Selling #1
    hasObject (Umbrella #1)
    hasAgent (Human #1 hasGivenName "John")
    hasAgent2 (Human #2 hasGivenName "Peter")
    before SpeechTimeInterval
```

In this example we see another notation convention: a hash sign is used to shorten a variable name if it starts with the name of the class mentioned immediately before the variable. Selling #1 is equivalent to Selling ?selling_1. Since all the individuals in the semantic structure obtain such naming, the new notation was introduced in order to exclude names duplication for better readability but still keep the class assignment explicit.

5 Communicative Dependencies

The expressiveness of the language allows capturing not only truth-conditional meaning of the sentence but some part of the communicative (information) structure as well. Although introducing such categories as theme/rheme or given/new will require additional means of expressing, Etalog is already capable of capturing communicative (referential) dependencies between nodes.

Igor Mel'čuk (2001) defines the communicative dependency in the following way: "In a semantic configuration (σ_1 - σ_2), the semantic node (σ_2) is said to depend communicatively on the semantic node (σ_1) in a direct way, if this configuration can be reduced to (σ_1) (rather than to (σ_2)) such that the meaning conveyed is simply reduced but not distorted, the referent of (σ_1) remaining the same as that of the whole configuration (σ_1 - σ_2)".



For example, the referent of the semantic configuration in Fig. 4 can be either 'crying' (*the baby cries, crying of the baby*) or 'baby' (*the crying baby, the baby who is crying*) whereas the semantic dependency does not change ('baby' being the argument

of 'crying'). The referent is determined by the communicative head of the configuration, not the semantic head. On the other hand the communicative head always corresponds to the syntactic head of a phrase when this configuration is expressed linguistically. "Comm-dependency is, so to speak, a way of 'foreseeing,' on the semantic level, the future syntactic dependencies" (Mel'čuk, ibid).

Using inverse relations we can express this configuration in Etalog either as (Cry hasAgent (Baby)) or as (Baby isAgentOf (Cry)). So we can use Etalog syntax to represent the communicative dependencies between nodes while the underlying semantic structure (set of triples) remains the same.

```
(6) The writer burned the novel that he had written
    (Burning #1
      hasAgent (Writer #1)
      hasObject
       (Novel #1
           isObjectOf
            (Writing #1
               hasAgent ?writer 1
            )
       )
   )
(7) The novel was burned by the writer who had written it
    (Burning #1
      hasObject (Novel #1)
      hasAgent
       (Writer #1
           isAgentOf
            (Writing #1
               hasObject ?novel 1
            )
       )
   )
```

Sentences (6) and (7) describe the same situation but constitute different messages for the hearer. They answer different questions ('What did the writer do?' and 'Who burned the novel?' respectively) and thus are communicatively different. In accordance to that, both Etalog expressions represent the same set of triples (the same truth-conditional meaning), but the syntax of the expressions mirrors the syntax of the sentences and thus reflects (some) communicative differences between them. They are organized in tree-like structures which resemble the syntactic trees of the original sentences.

On the one hand such semantic representation is language-independent (it contains no language-specific features of the original sentence). On the other hand it contains enough information to restore the syntax of the original sentence (to some extent). So it is a good candidate (at least better than just a set of triples) to serve as a languageindependent interlingua to translate from one language to another.

6 Referring Expressions for Answers

Another application of Etalog is to present answers to questions. SemETAP is able to answer questions to a text. A natural language question is converted into a SPARQL query and is run against the semantic structure of the text. The query returns an individual which corresponds to a wh-word in the question.

Then we have to decide how to present it to the user. Ideally a natural language expression should be generated which allows the user to identify the object in the text. While we are not there yet, as an intermediate solution we present the answer expression in Etalog (Rygaev 2018). Here is an example:

(8) Text: After a pass by Kerzhakov into the penalty area Arshavin with a brilliant shot in the fall hammers the ball into the net.

Question: Which team does Arshavin play for?

Answer: (FootballTeam isObjectOf (PlaysFor hasAgent

(Human hasName "Kerzhakov")))

Potential translation: The team which Kerzhakov plays for (The same team with Kerzhakov)

These Etalog expressions also capture not only truth-conditional meaning, but also the communicative (referential) structure of the answer. The expression is tree-like, with the found individual being the head of the phrase, and can serve as a template for the syntactic tree of the surface realization of the answer in natural language. Roughly what needs to be done is to replace ontological concepts with corresponding words and semantic relations with syntactic ones. This process is the opposite of the process of semantic analysis, which SemETAP is already capable to perform.

7 Conclusions

Etalog is a formal knowledge representation language, which is created to simplify reading and typing by linguists who work with it. It is based on the existing formalisms (Datalog^{+/-}, Description Logic, RDF/Turtle) and borrows a couple of ideas from natural language.

Etalog can express both *fact* knowledge (ABox in Description Logic terminology) as part of the semantic structure of the text and *rule* knowledge (TBox respectively) as a set of inference rules.

In addition, Etalog is able to capture some elements of communicative (referential) structure of the sentence, which makes it suitable to use as a language-independent interlingua for translation from one natural language to another.

8 References

- Baader F., Calvanese D., McGuinness D.L., Nardi D., Patel-Schneider P.F. (2003). The Description Logic Handbook: Theory, Implementation, Applications. Cambridge University Press, Cambridge, UK, 2003.
- Beckett D., Berners-Lee T., Prud'hommeaux E., Carothers G. (2014). RDF 1.1 Turtle. Terse RDF Triple Language. W3C Recommendation 25 February 2014. https://www.w3.org/TR/turtle/
- Benedikt, M., Konstantinidis, G., Mecca, G., Motik, B., Papotti, P., Santoro, D., & Tsamoura, E. (2017). Benchmarking the chase. In Proceedings of the 36th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems (pp. 37-52). ACM.
- Boguslavsky I.M., Dikonov V.G., Iomdin L.L., Lazursky A.V., Sizov V.G., Timoshenko S.P. (2015). Semantic Analysis and Question Answering: a System Under Development. Computational Linguistics and Intellectual Technologies. Papers from the Annual International Conference "Dialogue" (2015), p.62-79.
- Boguslavsky I.M., Frolova T.I., Iomdin L.L., Lazursky A.V., Rygaev I.P., Timoshenko S.P. (2018). Semantic Analysis with Inference: High Spots of the Football Match. Computational Linguistics and Intellectual Technologies. Papers from the Annual International Conference "Dialogue" (2018), p.124-142.
- Cali, A., Gottlob, G., Lukasiewicz, T., & Pieris, A. (2011). Datalog+/-: A family of languages for ontology querying. In Datalog Reloaded (pp. 351-368). Springer, Berlin, Heidelberg.
- Gennari, J.H., Musen, M.A., Fergerson, R.W., Grosso, W.E., Crubézy, M., Eriksson, H., Noy, N.F. and Tu, S.W. (2003). The evolution of Protégé: an environment for knowledge-based systems development. International Journal of Human-computer studies, 58(1), pp.89-123.
- Higginbotham J. (2000). On events in linguistic semantics. In: J. Higginbotham, F. Pianesi & A. Varzi (eds.). Speaking of Events. New York, Oxford: Oxford University Press, 49–79.
- 9. Mel'čuk, I. A. (2001). Communicative organization in natural language. John Benjamins Publishing Company.
- Patel-Schneider P.F., Hayes P., Horrocks I. (2003). OWL Web Ontology Language. Semantics and Abstract Syntax. W3C Recommendation 10 February 2004. https://www.w3.org/TR/owl-semantics/
- 11. Parsons T. (1990). Events in the Semantics of English. A Study in Subatomic Semantics. Cambridge, MA: MIT Press.
- Rygaev I. (2017), Rule-based Reasoning in Semantic Text Analysis. Proceedings of the Doctoral Consortium, Challenge, Industry Track, Tutorials and Posters @RuleML+RR 2017 hosted by International Joint Conference on Rules and Reasoning 2017 (RuleML+RR 2017).
- Rygaev I.P. (2018). Referring Expression Generation for Question Answering and Graph Visualization. Computational Linguistics and Intellectual Technologies. Papers from the Annual International Conference "Dialogue" (2018), p.619-636.

Действие отбора на рождение и смерть преждевременных старт-кодонов в филогенетическом дереве приматов

Сегодин В.¹ и Денисов С.^{2,3,1}

¹ Национальный исследовательский университет "Высшая школа экономики", Москва, 101000, Россия ² Сколковский институт науки и технологий, Москва, 121205, Россия ³ ИППИ РАН, Москва, 127051, Россия

stepan@bioinf.fbb.msu.ru

Аннотация.

Данная работа посвящена действию отбора на короткие рамки считывания (sORF), находящиеся в 5'лидерной области (5'UTR) перед основной рамкой трансляции (CDS) белок-кодирующих генов. Анализ проводился на филогенетическом дереве приматов среди отобранных генов для трех типов коротких рамок: uORF (upstream open reading frame), iORF (in-frame open reading frame) и oORF (overlapping reading frame). В результате проделанной работы выяснилось, что появление преждевременных старт-кодонов в 5'UTR области ограничено из-за воздействия отрицательного отбора, причём конфигурация iORF является наиболее вредной. В то же время зафиксировавшиеся в популяции uORF и oORF поддерживаются стабилизирующим отбором, тогда как iORF вымываются из популяции положительным отбором. Также было проанализировано влияние таких факторов, как длина рамки, близость к основной рамке считывания (для uORF), длина N-концевого удлинения (для iORF) и сила контекста старт-кодона (для всех трех типов коротких рамок) на силу отбора.

Keywords: 5'UTR, sORF, uORF, iORF, oORF, oтбор, консервативность, мутация.

1 Вступление

В последнее время большое количество научных работ было посвящено апстримным открытым рамкам считывания (upstream open reading frame или uORF). Они располагаются в 5' нетранслируемой области мPHK (5'UTR) перед основной кодирующей областью (coding sequence или CDS). С помощью рибосомного профайлинга было обнаружено, что порядка 50% белоккодирующих транскриптов человека имеют признаки трансляции в 5'UTR [1,2]. Более того, uORF способны регулировать трансляцию CDS. Эти факты опровергают ранее господствовавшую идею о том, что в клетках эукариот в одной мPHK имеется единственная последовательность, которая подвергается трансляции [3].

Мы решили выяснить, как отбор действует на короткие рамки (sORF), расположенные в 5'UTR. Для дальнейшего анализа мы разделяем все множество sORF на три возможные конфигурации: uORF (upstream open reading frame – открытая рамка считывания с преждевременным старт-кодоном, имеющая стоп-кодон в 5'UTR), iORF (in-frame open reading frame – открытая рамка считывания, находящаяся в одной рамке с CDS) и оORF (overlapping open reading frame – открытая рамка, старт-кодон которой находится не в одной рамке с CDS). Мы вычислили скорости рождения и смерти каждой конфигурации sORF на каждой ветке филогенетического дерева из 12 приматов.

Далее, скорости рождения и смерти были подсчитаны на разбиениях sORF на группы: по длине sORF, расстоянию от стоп-кодона до CDS – для uORF конфигурации; по расстоянию от старт-кодона sORF до CDS (длина удлиненного N-конца) – для iORF конфигурации. Кроме того, вышеупомянутые статистики были подсчитаны на разбиении sORF каждой конфигурации по степени соответствия окружения преждевременного старт-кодона контексту Козак [10].

2 Механизмы трансляции коротких рамок

2.1 Типы коротких рамок

Преждевременные старт-кодоны (uAUG) могут быть представлены в трех конфигурациях (Puc. 1):



Рис. 1. Типы коротких рамок.

- 1. oORF. В данном случае, CDS не находится в одной рамке с uAUG, при этом oORF и CDS пересекаются по последовательности. Обычно стопкодон oORF'а располагается внутри CDS, однако возможно его расположение и за пределами CDS в 3' направлении. На рис. 1 изображены две пересекающиеся открытые рамки.
- 2. iORF. В отличии от oORF, uAUG в данной конфигурации находится в одной рамке с CDS.
- 3. uORF. В этой конфигурации uAUG и стоп-кодоны, располагаются в 5'UTR. В результате трансляции такой рамки образуется пептид.

2.2 Механизмы трансляции: реинициация и "протечка сканирования".

Сканирующая модель является стандартной моделью трансляции в эукариотах. Началом трансляции служит сборка преинициаторного комплекса (43 preinitiation complex 43 PIC) на 5'-конце мРНК, который осуществляет сканирование мРНК в направлении от 5' к 3' концу. Когда 43 PIC достигает старт-кодона, происходит присоединение 60S рибосомального комплекса, что приводит к началу синтеза цепочки аминокислот.



Рис. 2. Схематическое представление возможных механизмов трансляции (взято из [4]).

Однако, помимо вышеупомянутого механзима, существуют еще несколько его видов: "протечка сканирования" (leaky scanning) и реинициация (reinitiation). В случае сканирования с протечкой, малая рибосомальная субъединица может пропустить старт-кодон и начать трансляцию со следующего AUG. В случае реинициации, рибосомальный комплекс, отработав на одной из рамок, может возобновить трансляцию на следующем AUG. На протяжении всего процесса трансляции рибосомальный комплекс остается связанным с мРНК. Возникновение реинициации или протечки сканирования зависит от следующих факторов: наличия трансляционных факторов, вторичных и третичных структур на мРНК и других

цис-регуляторных элементов, расстояния между короткими рамками, расстояния от кэпа (CAP) до uAUG, близости расположения к CDS, контекста старт и стоп кодонов, и т.д. Вышеперечисленные факторы представлены на следующем рисунке:



Рис. 3. мРНК-специфичные регуляторные факторы (взято из [3]).

3 Данные

Для сравнительного анализа и вычисления скоростей рождения и смерти было использовано выравнивание 100 позвоночных, загруженное из UCSC браузера. Дерево с восстановленными предковыми геномами представлено на Рис. 4.

Координаты старта трансляции и транскрипции были взяты из аннотации Refseq. Только белок-кодирующие гены были отобраны для анализа. Кроме того, еще одним требованием была длина 5'UTR в пределе от 15 до 500 нуклеатидов, а также его полное нахождение в первом экзоне. В результате для анализа было отобрано 7382 гена.

4 Методы

Для реконструкции последовательностей во внутренних узлах филогенетического дерева (предковых последовательностей) мы использовали метод максимального правдоподобия реализованный в программном пакете PAML (Phylogenetic Analysis by Maximum Likelihood) [5]. В качестве модели нуклеотидных замен была использована HKY85 модель [6]. Главное преимущество метода максимального правдоподобия (по сравнению с методом наибольшей экономии) является возможность учитывать множественные замены в одном сайте, что является важным фактором при рассмотрении видов, расположенных на больших филогенетических расстояниях.

2.2.1. Рождение и смерть коротких рамок

Мы вычислили скорость рождения и смерти uAUG кодонов в различных конфигурациях для установления степени действия стабилизирующего и движущего отбора на филогенетическом дереве приматов. При подсчете этих характеристик статистики суммировались по всем веткам дерева. Восстановленные последовательности изображены на рис. 4.



Рис. 4. Филогенетическое дерево с восстановленными предковыми последовательностями (пронумерованы).

Скорость рождения считалась следующим образом:

$$V_{uAUG \rightarrow birth} = \frac{preAUG \rightarrow uAUG}{preAUG},$$

где preAUG – триплет, отличающийся от AUG на один нуклеотид. Эта статистика, как и скорость смерти (см. ниже) была вычислена для трех конфигураций sORF в 5'UTR: uORF, oORF и iORF. В случае uORF конфигурации, числитель представляет количество предковых рамок с preAUG, находящимися в одной рамке со стоп-кодоном (uSTOP), которые перешли в потомке в настоящую рамку с AUG кодоном. Знаменатель отображает общее количество

[preAUG, uSTOP] рамок в предках. Здесь мы требовали консервативность стоп-кодонов.

Замена триплета preAUG на AUG и отсутствие стоп-кодонов в рамке – два основных критерия, которые использовались при рассмотрении случаев рождения рамок в конфигурациях iORF и oORF.

Для удобства, назовем sORF с preAUG кодоном (вместо AUG) следующим образом: pre-uORF, pre-iORF и pre-oORF.

Скорость смерти была подсчитана следующим образом:

$$V_{uAUG \rightarrow death} = \frac{AUG \rightarrow notAUG}{AUG},$$

где notAUG — кодон, отличный от AUG. Числитель представляет количество предковых [uAUG, uSTOP] рамок, перешедших в [notAUG, uSTOP] рамку в потомке, в то время как знаменатель представляет общее количество [uAUG, uSTOP] рамок (т.е. uORF) в предках (для uORF конфигурации). Аналогичным образом требовалась консервативность uSTOP кодона.

Подсчет данной характеристики для iORF и oORF конфигурации осуществлялся аналогичным образом. Стоит отметить, что мы исключили случаи смерти старт-кодона за счет вставок или делеций.

Вышеперечисленные статистики также были вычислены на нейтральной области длиной в 500 нуклеотидов перед стартом транскрипции. Расстояние между нейтральной областью и стартом транскрипции составляло 2000 нуклеотидов.

5 Результаты и обсуждение

5.1 Рождение uAUG



Birth rates of uAUG



Мы наблюдаем заниженные скорости рождения в 5'UTR (gene) по сравнению с нейтральной областью (neutral) для всех конфигураций коротких рамок. Это означает, что в среднем рождение uAUG является вредной мутацией, появлению которой препятствует стабилизирующий отбор. Наибольшая относительная величина скорости рождения наблюдается для конфигурации uORF (отношение gene/neutral), что свидетельствует о том, что в среднем появление AUG кодона в упомянутой конфигурации носит менее вредный характер, нежели для iORF или oORF.

5.2 Смерть uAUG

Death rates of uAUG



Рис. 6. Скорости смерти uAUG. Числа над столбцами представляют относительные величины: отношение скорости смерти на 5'UTR (Gene) и нейтральной областях (Neutral).

Наибольшая относительная скорость смерти наблюдается в конфигурации iORF – 1.2495. Это означает, что уже просуществовавшие какое-то время iORF умирают в среднем быстрее, чем рамки в других конфигурациях. Высокая скорость смерти по сравнению с ожидаемой позволяет нам сделать вывод о вредности данной конфигурации. В то время как iORF находятся под воздействием движущего отбора, стремящегося их уничтожить, рамки в конфигурациях uORF и оORF защищены стабилизирующим отбором (скорости смерти sORF в 5'UTR меньше ожидаемых). Данные факты свидетельствуют о том, что среди uORF и оORF присутствуют функциональные рамки, которые сохраняются отрицательным отбором.

5.3 Скорости рождения в различных функциональных группах

Таблица 1. Относительная скорость рождения для uORF, iORF и oORF конфигураций, разбитых на группы по значениям признаков. Pre-ORF'ы поделены на группы по терцилям.

uORF					
	1 группа	2 группа	3 группа		
Длина sORF	[12;33]	(33;75]	(75;495]		
Относительная ско- рость рождения	0.6901	0.7466	0.5690		
uSTOP-CDS расстоя- ние	[0;54]	(54;140]	(140;499]		
Отноительная ско- рость рождения	0.5761	0.7608	0.7235		
uAUG контекст	[- 5.682;1.144]	(1.144;2.854]	(2.854;8.324]		
Относительная ско- рость рождения	0.7704	0.6520	0.5153		
	iOl	RF			
	1 группа	2 группа	3 группа		
uAUG-CDS расстоя- ние	[6;36]	(36;87]	(87;495]		
Относительная ско- рость рождения	0.7203	0.5417	0.5213		
uAUG контекст	[- 5.321;1.505]	(1.505;3.153]	(3.153;7.981]		
Относительная ско- рость рождения	0.6540	0.4704	0.4564		
	<i>o0</i>]	RF			
	1 группа	2 группа	3 группа		
uAUG контекст	[- 5.323;1.450]	(1.450;3.149]	(3.149;8.262]		
Относительная ско- рость рождения	0.5930	0.4945	0.3294		

При сравнении относительных скоростей рождения для uORF конфигурации, можно сделать вывод о том, что рождение sORF, близко расположенных к CDS, с большой длиной рамки или сильным контекстом uAUG более подвержены стабилизирующему отбору, что говорит о вредоносном характере таких рамок. Эти факты коррелируют с данными из литературы:

- чем ближе uORF к CDS, тем меньше вероятность реиницации со следующего старт-кодона [7]
- большая длина uORF может являться причиной потери важных реинициаторных факторов [8]
- сильный контекст преждевременного старт-кодона отрицательно коррелирует с уровнем трансляции основной рамки считывания (CDS) [9]

Появление подобных рамок может привести к нарушению сложившихся механизмов трансляции CDS.

Также мы утверждаем, что iORF с большой длиной N-удлиненного конца появляются в потомках с меньшей вероятностью. Мы предполагаем, что трансляция таких iORF может привести к возникновению белков с непредсказуемой конформацией.

Аналогичные результаты по зависимости отбора от контекста старт-кодона были получены для конфигураций iORF и oORF.

5.4 Скорости смерти в различных функциональных группах

Таблица 2. Относительная скорость смерти для uORF, iORF и oORF конфигураций, разбитых на группы по значениям признаков. sORF поделены на

uORF					
	1 группа	2 группа	3 группа		
Длина sORF	[12;27]	(27;63]	(63;492]		
Относительная скорость смерти	0.6449	0.7062	0.7552		
uSTOP-CDS расстояние	[6;42]	(42;93]	(93;456]		
Отноительная скорость смерти	1.4871	0.9602	0.8242		
uAUG контекст	[- 3.924;2.368]	(2.368;4.212]	(4.212;9.694]		
Относительная скорость смерти	0.8598	0.7279	0.5510		

группы по терцилям.
iORF			
	1 группа	2 группа	3 группа
uAUG-CDS расстояние	[6;36]	(36;87]	(87;495]
Относительная скорость смерти	0.7203	0.5417	0.5213
uAUG контекст	[- 2.372;2.492]	(2.492;4.135]	(4.135;8.342]
Относительная скорость смерти	0.9975	1.3850	1.0720
oORF			
	1 группа	2 группа	3 группа
uAUG контекст	[- 3.432;2.303]	(2.303;4.192]	(4.192;9.046]
Относительная скорость смерти	0.7642	0.7872	0.6401

Мы не нашли никаких свидетельств зависимости отбора на uORF от расстояния до CDS. Вполне возможно, что рассмотренные в анализе uORF, просуществовав определенное время, адаптировались к работе на разном расстоянии от CDS. Примечательно, что в исследовании влияния sORF на экспрессию CDS [2], не было найдено разницы между uORF расположенными на разных расстояниях от CDS. Данный результат может выглядеть противоречивым в сравнении с результатом, полученным при вычислении скорости рождения. Мы можем предположить, что выборка генов, использованных в [2], могла быть обогащена давно существующими, "устоявшимися" генами. Также мы выяснили, что uORF меньшей длины и с более сильным контекстом преждевременного старт-кодона в большей степени защищены стабилизирующим отбором, что позволяет нам сделать вывод о существовании большой доли функциональны рамок с вышеупомянутыми характеристиками. Для конфигурации оORF был получен тот же самый результат по контекстам uAUG.

Менее однозначные результаты были получены для iORF: относительно большое количество iORF с коротким N-концом устраняется движущим отбором. Мы предполагаем, что возможность смерти таких рамок за счет вставки или делеции мала, поэтому большая часть таких рамок может исчезать за счет замены в AUG кодоне. Скорости смерти iORF в 5'UTR практически равны ожидаемым, за ис-

ключением второй группы, где sORF со средним контекстом наиболее подвержены действию отбора. Однако, мы считаем, что необходимо более подробное изучение в случае iORF конфигурации.

6 Выводы

- 1. Рождение всех типов коротких рамок ограничено стабилизирующим отбором
 - Степень вредности sORF: uORF<oORF<iORF
- Уже существовавшие iORF находятся под воздействием движущего отбора, элиминирующего данный тип рамок. В то же время uORF и oORF находятся под защитой стабилизирующего отбора.
- 3. Протечка сканирования (Leaky scanning) и реинициация (reinitiation) влияют на силу отбора отбора:
 - Близость uORF к CDS и их длина влияет на скорость рождения, в то время как просущестовавшие sORF на разных расстояниях до CDS не показывают разницы в скоростях смерти
 - uORF и oORF с сильным контекстом имеют меньше шансов на появление в потомке, в то время как уже просущестовавшие sORF с сильным контекстом имеют наименьшую относительную скорость смерти
 - iORF с коротким N-концом рождаются чаще, чем с большим, однако умирают с наибольшей вероятностью

References

- Lee S, Liu B, Huang SX, Shen B, Qian SB: Global mapping of translation initiation sites in mammalian cells at single-nucleotide resolution. Proc Natl Acad Sci USA 109: E2424– E2432 (2012).
- Calvo SE, Pagliarini DJ, Mootha VK: Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. Proc Natl Acad Sci USA 106:7507–7512 (2009).
- 3. Wethmar K: The regulatory potential of upstream open reading frames in eukaryotic gene expression. Wiley Interdisciplinary Reviews: RNA 5 (6): 765–768 (2014).
- 4. Andrews, Shea J, Rothnagel, Joseph A: Emerging evidence for functional peptides encoded by short open reading frames. Nature Reviews Genetics 15 (3):193–204 (2014).
- 5. Yang, Z: PAML 4: a program package for phylogenetic analysis by maximum likelihood. Molecular Biology and Evolution 24: 1586-1591 (2007).
- Hasegawa M, Iida Y, Yano T, Takaiwa F, Iwabuchi M: Phylogenetic relationships among eukaryotic kingdoms inferred from ribosomal RNA sequences. Journal of Molecular Evolution. 22 (1): 32-8 (1985).
- Kozak M: Constraints on reinitiation of translation in mammals. Nucleic Acids Res 29: 5226–5232 (2001).
- 8. Occhi G, Regazzo D, Trivellin G, et al: A novel mutation in the upstream open reading frame of the CDKN1B Gene Causes a MEN4 Phenotype. PLOS Genetics 9(3): e1003350 (2013).

- 9. Johnstone TG, Bazzini AA & Giraldez AJ: Upstream ORFs are prevalent translational repressors in vertebrates. EMBO Journal 35:706–723 (2016). 10. Kozak M.: Point mutations define a sequence flanking the AUG initiator codon that modu-
- lates translation by eukaryotic ribosomes. Cell 44(2): 283-292 (1986).

14

Генетический анализ ассортативного скрещивания в человеческих популяциях

Кузнецов Иван А.* ^{1,2} Кондрашов Алексей С. ^{1,2,3}

¹ Факультет биоинженерии и биоинформатики, Московский государствекнный университет им. М.В. Ломоносова, Ленинские Горы 1-73, 119234 Москва, Россия.

² Лаборатория эволюционной геномики, НИИ физико-химической биологии им. А.Н. Белозерского Московского государствекнного университета им. М.В. Ломоносова, 119992, Москва, Россия.

³Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI, 48109, USA.

* ikuznetsov@fbb.msu.ru

Abstract. Ассортативное скрещивание по сложным признакам широко распространено в человеческих популяциях. Как правило, корреляция между партнёрами положительна.

К настоящему времени получены противоречивые результаты, касающиеся существования избирательности по различным показателям здоровья при выборе партнёра. Наследуемая составляющая общего здоровья в значительной степени соответствует приспособленности генотипа. Приспособленность генотипа связана, в том числе, с мутационным грузом.

В данном исследовании мы предлагаем рассчитывать мутационный груз генотипа как число вредных мутаций, которое он в себе содержит. Многие вредные аллели в популяции находятся на очень низких частотах и не имеют качественной аннотации в специальных базах данных. Поэтому перед нами стояла необходимость выбрать способ, которым можно оценить степень вредности каждого генетического варианта. Другой проблемой является структурированность популяций, что может служить причиной наличия ложноположительного сигнала.

Нами протестировано несколько возможных способов расчёта числа вредных мутаций в генотипе. Найдены доводы в пользу существования ассортативного скрещивания по этому признаку в некоторых популяциях по данным из проекта «1000 Genomes». Также найдены закономерности, связанные со сходством геномов супругов при их сравнении с геномом шимпанзе.

Keywords: Ассортативное скрещивание, мутационный груз, человеческие популяции.

Ассортативное скрещивание по сложным признакам широко распространено в человеческих популяциях. Как правило, корреляция между партнёрами положительна. Сходство наблюдается по многим социокультурным признакам, по этническому происхождению, личностным характеристикам, количественным фенотипическим признакам и показателям здоровья. Действительные причины и генетические последствия ассортативного скрещивания остаются не изученными в должной мере. Трудности заключаются в том, что существует множество механизмов, по которым может возникать сходство между партнёрами. Оно может быть следствием избирательности по фенотипу, социокультурным признакам. Также сходство может возникать и увеличиваться из-за конвергенции признаков партнёров с течением времени.

Следствием ассортативного скрещивания по наследуемым признакам является возникновение и поддержание отклонения от равновесия Харди-Вайнберга в сторону увеличения доли гомозигот в популяции. При этом может усиливаться генетическая стратификация популяции. Напротив, генетические последствия отсутствуют, если сходство партнеров возникает из-за близких условий окружающей среды при социальной гомогамии или из-за длительного взаимодействия между партнёрами.

К настоящему времени получены противоречивые результаты, касающиеся существования избирательности по различным показателям здоровья при выборе партнёра. Наследуемая составляющая общего здоровья в значительной степени соответствует приспособленности генотипа. Приспособленность генотипа связана, в том числе, с мутационным грузом, который он в себе содержит. Знание о том, существует ли ассортативное скрещивание в человеческих популяциях по генетической составляющей общего здоровья важно как для фундаментального понимания пологого поведения человека, так и для медицинских приложений.

Нами было обнаружено, что полногеномные данные из проекта «1000 Genomes» не могут быть использованы для анализа распределения числа редких аллелей по генотипам в отдельных популяциях, так как между числом редких аллелей в геноме и качеством его секвенирования существует монотонная зависимость. При этом для экзомных данных значимой корреляции между качеством секвенирования и числом редких аллелей в большинстве популяций не обнаружено. Выбор аллелей по частоте приводит к высоким значениям корреляции между супругами в некоторых популяциях, что может быть связано с их внутренней структурой.

Более тонкий анализ, включающий в себя предсказывание эффекта, который однонуклеотидная замена может оказывать на свойства кодируемого геном белка, показал наличие ассортативного скрещивания по числу предположительно вредных аллелей в экзоме, при учёте возможной структуры популяции.

При сравнении экзомов человека с экзомом шимпанзе оказалось, что количество аллелей, отличных от вариантов шимпанзе, в различных надпопуляциях отличается слабо. При этом мы наблюдаем корреляцию по данному признаку между супругами внутри европейских и американских популяций.

Квантово-химические расчеты процесса переноса H⁺ между H₄SiO₄ и πN-имидазольного кольца гистидина с использованием теории функционала плотности

Изотова Е.Д.^{1[0000-0002-4449-8715]} и Багаева Т.В.¹

¹ Казанский (Приволжский) федеральный университет, 420008, Казань, Россия Izotova.e.d@gmail.com

Абстракт. Методом теории функционала плотности рассчитана поверхность потенциальной энергии реакции переноса H^+c H_4SiO_4 на πN -имидазольного кольца His. Расчеты проводились с помощью пакета GAMESS-US в базисе 6-31G(d)//6-311++G(2d,p) в газовой фазе и с использованием модели поляризованного континуума с включением в систему трех явных молекул воды. Энергетический барьер прямой реакции составил 25 ккал/моль, в то время как барьер обратной реакции – 18 ккал/моль.

Ключевые слова: Поверхность потенциальной энергии, Ортокремниевая кислота, Гистидин, GAMESS-US.

1 Введение

Одним из ключевых ферментов, принимающим участие в образовании структурированных фрагментов - спикул у стеклянных губок является катепсино - подобный белок — силикатеин — альфа. Активный центр силикатеина представлен тремя аминокислотными остатками - His 163, Ser 25, Gln 19. По одной из гипотез катализ силикатеина- альфа протекает по дву- шаговому механизму с образованием ковалентного интермедиата [1], подобно механизмам протеолитических ферментов. Fairched с соавторами на основе данных ренгеноструктурного анализа выдвинули другой механизм, в котором ключевая роль активации молекулы ортокремниевой кислоты (H_4SiO_4) принадлежит аминокислотному остатку His 163, которая принимает на себя H^+ с H_4SiO_4 [2], в результате отрицательно заряженный кислород совершает нуклеофильную атаку на кремний второй молекулы H_4SiO_4 , что приводит к конденсации двух молекул, по щелочному типу катализа и образованию димера — пирокремниевой кислоты ($H_6Si_2O_7$).

Первой реакцией активации H_4SiO_4 является перенос H^+ с кислорода молекулы H4SiO4 на πN -имидазольного кольца His. Перенос H^+ в среднем протекает в десять раз быстрее, чем такие же реакции с участием атома водорода [3]. В данной работе рассчитывается реакция переноса H^+ с кислорода H4SiO4 на πN -имидазольного кольца His.

2 Материалы и методы

Для изучения процесса переноса H⁺ с кислорода H₄SiO₄ на азот имидазольного кольца His была подготовлена и оптимизирована структура His — H₄SiO₄ и построена поверхность потенциальной энергии (ППЭ) переноса H⁺ вдоль вектора π N-O, рисунок 1. Все расчеты произведены с помощью пакета GAMESS-US [4]. Оптимизация геометрии молекул проводилась методом DFT с функционалом X3LYP, в валетно-расщепленном базисе 6-31G(d), уточнение энергий и расчет энергий в точках осуществлялся в базисе 6-311++G(2d,p). Атомы азота и кислорода были зафиксированы по трем координатам, на удалении 2.96 Å. Расчеты проводились в газовой фазе и с использованием модели поляризованного континуума (PCM) с включением в систему трех явных молекул воды.



Рис. 1. Схема вектора переноса H^+ с H_4SiO_4 на πN -имидазольного кольца His.

3 Результаты и обсуждение

Энергетический профиль реакции переноса H⁺ представлена на рисунке 2. Профиль изменения энергии выведен на ноль по минимальному значению системы, полученной в результате пошагового переноса H⁺. Выделяют два стационарных состояния: первое характерно для связи O-H молекулы H₄SiO₄, второе вблизи атома π N - имидазольного кольца His. Расчеты с использованием модели PCM приводит к сохранению общего профиля ППЭ, однако наблюдается уменьшение энергетического барьера прямой реакции на 8.6 ккал/моль и составляет 25 ккал/моль. Так же смещается положение второго минимума, и расстояние N-H уменьшается с 1.13 Å до 1.0 Å и снижается энергия второго стационарного состояния с 14.45 ккал/моль до 6.95. В то время энергия и геометрия первого стационарного следионарного положения остаются неизменны при расчете с использованием модели PCM.



Рис. 2. Поверхность потенциальной энергии пепеноса H^+ с H_4SiO_4 на πN -имидазольного кольца His.

Список литературы

.

- Schröder, Heinz C.: Hierarchical architecture of sponge spicules: biocatalytic and structure-directing activity of silicatein proteins as model for bioinspired applications. Bioinspiration & Biomimetics 11, 041002 (2016).
- 2. Fairhead, M.: Crystal structure and silica condensing activities of silicatein a–cathepsin L chimeras. Chemical Communications, 1765–1767 (2008).
- 3. Гаммет, Л.: Основы физической органической химии. «Мир», Москва (1972).
- 4. Gordon, M.S.: Advances in electronic structure theory: GAMESS a decade later. Theory and Applications of Computational Chemistry: the first forty years, 1167-1189 (2005).

Реконструкция межхромосомных транслокаций в геномах рода Vibrio

Кристина Перевощикова¹, Елена В. Мороз², Ольга О. Бочкарева^{2,3}

¹ ФББ МГУ имени М.В.Ломоносова, 119234, Москва, Россия,

² Институт проблем передачи информации им. А.А. Харкевича, 127051, Москва, Россия,

³ Сколковский институт науки и технологий, 121205, Москва, Россия.

kristina.ferryman@yandex.ru

Вибрионы – род грамотрицательных бактерий, относящийся к классу Gammaproteobacteria. Характерной особенностью представителей этого рода является наличие двух хромосом. Первая хромосома несет большую часть генов, необходимых для жизнедеятельности бактерии, а большая часть генов второй хромосомы является нишеспецифичной и необходима для адаптации к конкретным условиям среды.

В данной работе были реконструированы и описаны транслокации генов между хромосомами, произошедшие в ходе эволюции вибрионов. На основе полученных данных проверена гипотеза, что при межхромосомных траслокациях гены чаще сохраняют свое положение относительно лидирующей/запаздывающей цепи, что может объясняться отбором. Также показано, что для блоков из нескольких генов отбор на сохранение положения на цепи выражен сильнее.

Keywords: вибрионы, геномные перестройки, отбор

Введение

Vibrio – род грамотрицательных бактерий, представители которого имеют форму изогнутых палочек и обитают преимущественно в морях [1]. Большинство видов имеет две хромосомы [2]. Некоторые виды вибрионов способны вызывать серьезные заболевания человека, сопровождающиеся гастроэнтеритом, диареей, рвотой, мышечными коликами, сепсисом. Заражение чаще всего происходит при употреблении моллюсков или при контакте с морской водой. Основными патогенами человека являются виды Vibrio cholerae, Vibrio vulnificus, Vibrio parahaemolyticus [1].

Анализ доступных последовательностей полных геномов бактерий показал, что около 10% из них принадлежат видам и штаммам, обладающим несколькими крупными и потенциально значимыми репликонами [3]. Одно из возможных преимуществ генома, состоящего из нескольких хромосом – это способность к быстрой репликации за счет наличия нескольких ориджинов. Так *Vibrio spp* имеют одну из наибольших скоростей клеточного деления [2]. Обычно помимо размера и количества генов хромосомы отличаются GC составом, составом синонимичных кодонов, распределением генов по лидирующей и отстающей цепям, функциональным составом генов, уровнем транскрипции. Такая организация генома не является случайной и сформирована давлением отбора [3].

Инверсии, проходящие через ориджин, перепредставлены по сравнению с инверсиями, лежащими внутри реплихоры [4]. При этом, инверсии, лежащие внутри реплихоры, значимо короче инверсий, проходящих через точку начала репликации [5]. Действие отрицательного отбора на длинные внутриреплихорные инверсии может 5062 ясняться тем, что такие перестройки меняют участки лидирующей и запаздывающей цепи и тем самым нарушают G\C и A\T сдвиг, а также меняют положение генов относительно ориджина, что может влиять на уровень экспрессии генов [5].

В этой работе для 114 штаммов рода Vibrio была выполнена реконструкция межхромосомных транслокаций и проверена гипотеза о действии отбора, сохраняющего положение транслоцированного участка относительно лидирующей/запаздывающей цепи.

Методы

В работе были использованы последовательности полных геномов вибрионов, доступные в базе данных genbank [6]. Геном однохромосомного штамма Vibrio был исключен из анализа. Были исправлены ошибки в идентификации хромосом.

Ортологические ряды были построены с помощью программы ProteinOrtho [7]. Затем для генов каждого ортологического ряда при помощи BioPython были получены характеристики, такие как: аминокислотная последовательность, координаты в геноме от точки начала репликации, положение на цепи, положение на хромосоме.

Для аннотации ортологических рядов была использована программа InterProScan v5 [8]. Выравнивания ортологических рядов были выполнены с помощью программы Mafft v7.273 [9], затем полученные выравнивания были собраны в один файл для построения филогении.

Филогенетическое дерево были построены с помощью PhyML v 3.0 [10] с 100 повторениями для вычисления бутстреп-поддержки узлов.

Межхромосомные транслокации по заданной топологии филогенетического дерева были реконструированы на основе следующей модели. Для каждого ортологического ряда был составлен вектор положения на хромосоме входящих в ряд генов (1 – первая, 2 – вторая) и события перехода гена с одной хромосомы на другую реконструированы с помощью алгоритма PAML 4.6 [11]. Далее гены, транслокация которых произошла на одной ветке филогенетического дерева, были объединены в синтенный блок на основе данных об их расположении в геномах.

Разметка участков хромосом, составляющих лидирующую и запаздывающую цепь, была сделана по данным о положении точек начала и конца репликации, которые в свою очередь были получены из анализа графиков GC-skew с помощью программ на BioPython.

Результаты

Первые хромосомы разных штаммов рода Vibrio имеют длину около 3*106 пар нуклеотидов, в то время как длины вторых хромосом существенно различаются у штаммов разных видов (Рис. 1а). Доля генов, лежащих на лидирующей цепи, для первых хромосом всех видов вибрионов лежит в узком диапазоне 58-62%, в то время как для вторых хромосом перепредставленность генов на лидирующей цепи менее выражена. (Рис. 16) Так как положение гена на лидирующей цепи связано с высоким уровнем транскрипции такого гена, сильная перепредставленность генов на лидирующей цепи для первых хромосом может говорить о большей значимости данных генов для бактерии и большем давлении отбора на гены, локализованные там, относительно генов, лежащих на вторых хромосомах.

2



Рис. 1. (а) Распределение длин 1 и 2 хромосом (б) Доля генов, лежащих на лидирующей цепи 1 и 2 хромосом

Было выявлено 196 генов, в истории которых происходили транслокации между хромосомами (Рис.2). 29 из них транслоцировались несколько раз. Среди параллельных транслокаций не выявлено блоков, содержащих несколько генов. Границы самых крупных транслокаций - 16 генов у *Vibrio alginolyticus* и 79 генов у *Vibrio parahaemolyticus*, сформированы генами 16S rRNA и 23S rRNA.



Рис. 2. Транслокации, реконструированные в истории рода Vibrio

Мы не выявили перепредставленности случаев сохранения цепи среди всех реконструированных транслокаций, однако транслоцированные участки, содержащие несколько генов, чаще сохраняют цепь, чем в случае транслокации одного гена (Рис. 3). Не было выявлено перепредставленности каких-либо GO категорий для генов, в истории которых Фроисходили транслокации.



Рис. 3. Реконструированные транслокации синтенных блоков. Фиолетовым цветом показаны гены, которые не сохранили положение на цепи, зеленым — гены, сохранившие свое положение на запаздывающей цепи, красным – гены, сохранившие свое положение на лидирующей цепи.

Благодарности

Эта работа начата на Школе молекулярной и теоретической биологии 2018 при поддержке Zimin Foundation, далее выполнялась в УНЦ Биоинформатика под руководством М.С. Гельфанда (поддержана грантом РНФ 18-14-00358).

Литература

1. Nicholas A. Daniels, MD, MPH, Alireza Shafaie, MD.: A Review of Pathogenic Vibrio Infections for Clinicians. Infect Med 17(10):665-685 (2000).

2. Cooper VS, Vohr SH, Wrocklage SC, Hatcher PJ.: Why Genes Evolve Faster on Secondary Chromosomes in Bacteria. PLoS Comput Biol 6(4):e1000732 (2010).

3. diCenzo GC, Finan TM.: The divided bacterial genome: structure, function, and evolution. Microbiol Mol Biol Rev 81:e00019-17 (2017).

4. Bochkareva OO, Dranenko NO, Ocheredko ES, Kanevsky GM, Lozinsky YN, Khalaycheva VA, Artamonova II, Gelfand MS.: Genome rearrangements and phylogeny reconstruction in Yersinia pestis. PeerJ 6:e4545 (2018)

5. Darling AE, Miklo's I, Ragan MA.: Dynamics of Genome Rearrangement in Bacterial Populations. PLoS Genet 4(7): e1000128. (2008)

6. GenBank Homepage, https://www.ncbi.nlm.nih.gov/genbank/

7. Lechner et al.: Proteinortho: Detection of (Co-) orthologs in large-scale analysis. BMC Bioinformatics 12:124 (2011).

8. Jones P, Binns D, Chang H-Y, et al.: InterProScan 5: genome-scale protein function classification. Bioinformatics.;30(9):1236-1240 (2014)

9. Kazutaka Katoh, John Rozewicki and Kazunori D. Yamada.: MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Briefings in Bioinformatics, 1–7 (2017).

10. Guindon S., Dufayard J.F., Lefort V., Anisimova M., Hordijk W., Gascuel O.: New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. Systematic Biology, 59(3):307-21, (2010).

11. Yang, Z.: PAML 4: a program package for phylogenetic analysis by maximum likelihood. Molecular Biology and Evolution 24: 1586-1591 (2007)

Эволюция операторных участков в лямбдоидных фагах

Быкова Дарья¹, Суворова Инна²

¹ Московский государственный университет им. М. В. Ломоносова, Москва, Россия

² Институт проблем передачи информации им. А.А. Харкевича Российской академии наук, Москва, Россия

db.bykova@gmail.com

Резюме. В работе рассматриваются операторные участки бактериофага лямбда, участвующие в переключении между литическим и лизогенным циклами его развития. Работа направлена на поиск аналогичных генетических переключателей в других фагах и восстановление возможного пути эволюции данной системы.

Ключевые слова: бактериофаг лямбда, генетический переключатель

1 Введение

Бактериофаг лямбда способен инициировать два цикла развития: литический и лизогенный [1]. Литический цикл предполагает репликацию фаговой хромосомы после попадания в клетку хозяина, сборку множества новых фаговых частиц и затем лизис хозяйской клетки. При лизогении фаг встраивается в геном хозяина и реплицируется вместе с хозяйской хромосомой. В таком пассивном состоянии фаг может существовать на протяжении множества поколений, все его гены, кроме одного, в это время находятся в выключенном состоянии. При наступлении определённых условий гены фага вновь активируются, и он переходит к литическому циклу развития.

Интерес представляет устройство генетического переключателя, при помощи которого происходит переход от одного цикла развития к другому. У бактериофага лямбда он устроен следующим образом [1,2]: имеется два оператора - правый (O_R) и левый (O_L). Каждый из них состоит из трёх частей, представляющих собой палиндромные последовательности длиной 17 п.н., с которыми могут связываться регуляторные белки сI и Cro.

Правый оператор расположен между генами сI и Cro, и все его участки перекрываются с промоторами этих генов P_{RM} и P_R соответственно. Оба белка являются активаторами собственной транскрипции и репрессорами друг для друга, при этом они имеют немного разное сродство ко всем трём операторным участкам. С левым оператором связывается только сI, он является репрессором не только для Cro, но и для всех остальных белков фага. За счёт различных комбинаций связывания сI и Cro с операторными участками и осуществляется переключение между литическим циклом и лизогенным.

Целью данной работы был поиск аналогичных генетических переключателей в других фагах, а также восстановление возможного пути эволюции данной системы.

2 Промежуточные результаты

В ходе работы было рассмотрено 11 межгенных участков, гомологичных правому оператору фага лямбда (последовательности взяты из геномов 31 фага, удалены одинаковые). Среди них нашлось 5, близкородственных исходному. По ним была построена позиционная весовая матрица, и была сделана попытка при помощи неё найти регуляторные палиндромные последовательности внутри оставшихся шести участков. Для этой же цели были построены выравнивания. Согласно полученным данным, 4 межгенных участка сохранили общую структуру правого оператора, хотя и с изменениями, в двух других участках отчётливо выражена только одна палиндромная последовательность (предположительно гомологична $O_R 2$ в одном случае и $O_R 1$ в другом).

В дальнейшем предстоит проделать сходные действия с левым оператором и оценить биологический смысл полученных данных.

Благодарности

Работа выполнена совместно с М.С. Гельфандом. Работа поддержана грантом РНФ 18-14-00358.

Список использованной литературы

- 1. Ptashne M., A genetic switch: gene control and phage λ . Blackwell Scientific Publications and Cell Press (1986).
- Court DL, Oppenheim AB, Adhya SL. A new look at bacteriophage lambda genetic networks. Journal of Bacteriology; 189:298-304 (2007).

Технология идентификации редких аллелей, определяющих предрасположенность к раку почек

Мухамадеева Р.А.¹, Серебрийский И.Г.^{1,2}, Arora S.²

¹Казанский (Приволжский) федеральный университет, Казань, Россия ²Fox Chase Cancer Center, Филадельфия, США

r.mukhamadeeva@yandex.ru

Abstract. Определение аллелей, определяющих предрасположенность к спорадическому раку почек, является важной задачей. Причиной образование рака в подобных случаях может быть сочетание факторов риска с неизвестными нам на данный момент генетическими изменениями. Для определения функциональной важности подобных генетических изменений уже разработан обширный программный инструментарий. В связи с этим целью работы являлась выработка четкого технологического процесса идентификации редких вредоносных аллелей, предположительно определяющих предрасположенность к раку почек, и их описания с помощью биоинформатических инструментов.

В ходе работы мы составили предварительный список геновкандидатов, функционально связанных с раком почек, на основе интеграции данных литературного поиска и поиска по базам данных, доступным онлайн (Ingenuity, HGMD Professional, cBioPortal, driverDB, COSMIC, ICGC, DisGeNet, OMIM, GeneCards),. После дальнейшего отбора и приоритизации с помощью построения сетей взаимодействия (программное обеспечение Cytoscape) общий список генов-кандидатов составил 479 генов. На основе анализа последовательности экзома 6 пациентов с помощью биоинформатических инструментов предсказания функциональной значимости мутаций *in silico* (PolyPhen-2, SIFT, PROVEAN, MutationAssessor, SIFT-Indel, PROVEAN, MutationTaster) выявлены 18 редких вредоносных аллелей, предположительно определяющих предрасположенность к раку почек. В их число входит 10 миссенс-мутаций, 3 мутации сдвига рамки и 5 делеций.

В ходе работы выработан четкий технологический процесс первичного анализа и идентификации редких вредоносных аллелей, включающий в себя определение списка генов-кандидатов, экзомное секвенирование и анализ с помощью биоинформатических инструментов предсказания функциональной значимости мутаций *in silico*. Эта технология первичного анализа будет использоваться в дальнейшем исследовании рака почек.

Keywords: transcriptomics, kidney cancer, single nucleotide polymorphism (SNP), транскриптомика, рак почек, однонуклеотидные полиморфизмы

1 ВВЕДЕНИЕ

Рак почек считается сравнительно редким онкологическим заболеванием. Но хотя рак почек встречается всего лишь в 3% случаев из всех раковых заболеваний, он является шестым по смертности, а количество больных раком почек за последние несколько десятилетий увеличилось на 43%.

Существует несколько факторов риска заболевания рака почек, такие как: курение, ожирение, гипертония. Но также большое влияние на образование данного вида рака оказывает наследственность. Существует несколько наследственных синдромов, которые ведут к раку, но подобные синдромы не всегда диагностируются у больных раком почек. Случаи, которые при которых не найдены мутации, связанные с наследственными синдромами, обычно называют спорадическими. Причиной образование рака в подобных случаях может быть сочетание факторов риска с неизвестными нам на данный момент генетическими изменениями. Определение подобных редких аллелей, определяющих предрасположенность к раку почек, является важной задачей.

Для решения данной проблемы могут быть использованы современные технологии геномного секвенирования, в развитии которых в последние годы сделаны огромные шаги. Это позволяет нам сравнительно быстро и недорого определить все вариации генома индивидуума. Для определения функциональной важности подобных генетических изменений уже разработан обширный программный инструментарий. В относительно скором времени выявление и предсказание функциональной значимости аллели in silico будет рутинной задачей персонифицированной медицины.

Целью настоящей работы являлась выработка технологического процесса идентификации редких вредоносных аллелей, возможно определяющих предрасположенность к раку почек, и их описания с помощью биоинформатических инструментов.

2 МАТЕРИАЛЫ И МЕТОДЫ

2.1 Определение генов-кандидатов

Для составления списка генов-кандидатов, которые определяют предрасположенность к раку почек, был проведен всесторонний поиск по литературным данным и базам данных, доступным онлайн (see Fig. 1).

Литературный поиск был осуществлен в статьях, описывающих наследственные синдромы, имеющие отношение к раку почек, а также полиморфизмы, которые анализировались в исследованиях рака почек.

Поиск по базам данных был проделан как в коммерческих базах данных, так и в бесплатных. Коммерческие базы данных: Ingenuity [12], HGMD Professional [19]. Бесплатные базы данных: cBioPortal [10], driverDB [3], COSMIC [2], ICGC [11], DisGeNet [14], OMIM [8], GeneCards [16].

2

После того, как был составлен список генов, полученный из списков литературного поиска и поиска по базам данных, была проведена приоритизация генов. Мы разделили все гены из списка на две группы. Первая группа – «Основной приоритет» (see Fig. 1). Эта группа была составлена из генов, взятых из:

базы данных Ingenuity (только гены с более, чем двумя найденными публикациями);

базы данных DisGeNet (только гены, имеющие оценку более или равную 0.3) базы данных HGMD Professional;

базы данных ОМІМ;

базы данных GeneCards.

Все остальные гены было временно помещены в группу «Низкий приоритет». Данная группа составила 2468 генов.



Fig. 1. Схема исследования. А – Определение предварительного списка геновкандидатов, Б – Приоритизация и окончательный отбор генов, В – Секвенирование экзомов и получение списка аллелей, Г – Идентификация редких вредоносных аллелей. Звездочкой отмечен этап работы, проделанной не нами.

2.2 Построение сетей взаимодействия

Для дальнейшей сортировки была использовано программное обеспечение Суtoscape [18], позволяющее строить сети взаимодействия. Мы использовали плагин MiMi [5], задав группу «Основной приоритет» как запрос. После получения сети мы отфильтровали группу «Низкий приоритет». В этой группе мы оставили только гены, которые взаимодействуют как минимум с двумя генами из запроса. Все остальные гены были удалены из списка генов-кандидатов (see Fig. 1).

Полная схема отбора генов-кандидатов представлена на рисунке 2 (see Fig. 2).



Fig. 2. Схема отбора генов-кандидатов

2.3 Отбор пациентов

Были выбраны 6 пациентов, больные раком почек, которые проходили лечение в Fox Chase Cancer Center (Филадельфия, США). Отбор пациентов проходил по наличию в семейной родословной множественных случаев раковых заболеваний, как рака почек, так и других типов раков (see Fig. 3). ДНК периферической крови пациентов была послана на анализ в Ambry Genetics, Inc для анализа экзомов.



Fig. 3. Пример родословной пациента (отобранный пациент отмечен красным) с указанием возраста на момент смерти или возраста на момент взятие тканей пациента на анализ. Также указаны случаи раковых заболеваний с указанием возраста на время диагноза.

2.4 Секвенирование экзомов

Секвенирование экзомов было осуществлено компанией Ambry Genetics, Inc (Алисо Виехо, США) при 30X среднем покрытии используя VCRome kit для подготовки библиотеки, индексирования и обработки с помощью платформы Hiseq (Illumina, Хэйворд, США). Выравнивание и определение вариантов было выполнено с помощью программного обеспечения Illumina CASAVA. Процесс дальнейшей сортировки вариантов был выполнен для того, чтобы отсеять варианты, попадающие в некодирующие участки, синонимичные варианты или варианты, популяционные частоты которых были больше 1% в базах данных dbSNP, 1000 genomes и ESP (see Fig. 1).

2.5 Анализ экзомов

Из полученного списка вариантов были отобраны те, которые относились к генам из списка кандидатов. Далее варианты были проанализированы с помощью программ предсказания функциональной значимости мутаций in silico (see Fig. 1).

Для анализа списка однонуклеотидных полиморфизмов (SNP) были использованы следующие программы PolyPhen-2 [1], SIFT [13], PROVEAN [4], MutationAssessor [17]. Были выбраны следующие критерии отбора: все варианты должны были иметь оценку PolyPhen-2 более 0.95 и должны были иметь не нейтральную оценку как минимум в двух из остальных трех программ (для SIFT – "Damaging", для PROVEAN – "Deleterious", для MutationAssessor – "high" или "medium").

Список инсерций и делеций (Indel) был отобран только для вариантов, кратным трем (3N indels). Для анализа этого списка были использованы программы SIFT-Indel [9] (оценка – "Damaging"), PROVEAN (оценка – "Deleterious") и MutationTaster [7] (оценка – "Disease-causing"). Были выбраны варианты, имеющие как минимум одну не нейтральную оценку.

Мутации некратные трем мы автоматически занесли в список вредоносных.

Для составления запросов и фильтрации полученных данных было написано 2 скрипта. Первый скрипт на основе списков, полученных после секвенирования, составляет файлы запроса для ввода в программы предсказания функциональной значимости мутаций in silico. Второй скрипт был написан для фильтрации полученных из программ данных. Для написания скриптов был использован язык программирования R [6] в программной оболочке RStudio [15].

Для определения популяционных частот вариантов была использована база данных ExAC – Exome Aggregation Consortium.

3 РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

3.1 Предварительный список генов-кандидатов

Для составления списка генов-кандидатов, которые определяют предрасположенность к раку почек, был проведен всесторонний поиск по литературным данным и базам данных, доступным онлайн (Ingenuity, HGMD Professional, cBioPortal, driverDB, COSMIC, ICGC, DisGeNet, OMIM, GeneCards).

Количество найденных генов в каждой баз данных: Ingenuity – 961, HGMD Professional – 25, cBioPortal – 169, driverDB – 72, COSMIC – 20, ICGC – 40, Dis-GeNet – 960, OMIM – 10, GeneCards – 38.

Общее количество генов, найденных в базах данных и литературе – 2697.

После процедуры поиска мы получили предварительный список геновкандидатов, которые согласно литературным данным или базам данных относятся к раку почек.

3.2 Расстановка приоритетов в списке генов-кандидатов и последующий отбор

Так как список генов, полученный после поиска был сравнительно большим, было необходимо произвести дальнейший отбор.

Первичный отбор был проделан по базам данных. Мы выделили группу «Основной приоритет», в которой присутствовали гены из 5 базах данных. Мы выбрали две самые крупные базы данных – Ingenuity и DisGeNet – и, отфильтровав наименее надежные из них (мы выбрали гены с более, чем 2 публикациями из Ingenuity, и гены с оценкой более или равную 0.3 из DisGeNet), получили два списка генов. Также мы взяли гены из баз данных OMIM, HGMD Professional и GeneCards. Данную группу составили 202 гена. Все остальные гены были временно помещены в группу с меньшим приоритетом.

Дальнейший отбор в группе с меньшим приоритетом был осуществлен по поиску белок-белковых взаимодействий. Для этого было использовано программное обеспечение Cytoscape, позволяющее строить сети взаимодействия. К этой программе существует плагин MiMi, ищущий все взаимодействия белков, которые кодируются генами из запроса.

Как запрос для плагина MiMi мы использовали гены из группы «Основной приоритет». Критерием отбора генов для сети было наличие белковых взаимодействий с как минимум двумя генами из запроса. Таким образом мы выделили еще одну группу генов, так называемую «Низкий приоритет». В эту группу вошли только те гены, которые были найдены в сети взаимодействия. Эта группа состоит из 277 генов. Все остальные гены были отсеяны из списка геновкандидатов.

Общий список генов-кандидатов составил 479 генов (see Fig. 4).



Fig. 4. Сеть взаимодействия генов-кандидатов. Красные круги - гены из группы «Основной приоритет», синие треугольники - гены из группы «Низкий приоритет». Размер красных маркеров пропорционален количеству баз данных из нашего списка, в которых найден этот ген.

3.3 Анализ экзомов и отбор вариантов

Анализ экзомов был проведен на ДНК периферической крови шести пациентов, больных раком почек.

После секвенирования экзомов мы получили список вариантов, отвечающих следующим критериям: все варианты находились в экзомах (кодирующей части), все варианты были несинонимичные, и популяционные частоты всех вариантов были меньше 1% в базах данных dbSNP, 1000 genomes и ESP.

Далее мы отсортировали список вариантов относительно списка геновкандидатов. Мы получили список в среднем из 35 вариантов на каждого пациента.

Затем варианты были проанализированы с помощью программ предсказания функциональной значимости мутаций in silico.

Список, использованных программ:

для анализа однонуклеотидных полиморфизмов (SNP) – PolyPhen-2, SIFT, PROVEAN, MutationAssessor;

для анализа инсерций и делеций (indel) – SIFT-Indel, PROVEAN, MutationTaster.

Используя написанные скрипты, были получены списки однонуклеотидных полиморфизмов для каждого пациента, отвечающие критериям: оценка PolyPhen-2 более 0.95 и не нейтральная оценка как минимум в двух из остальных трех программ (для SIFT – "Damaging", для PROVEAN – "Deleterious", для MutationAssessor – "high" или "medium"). Для некоторых вариантов была определена популяционная частота европеоидной популяции с помощью базы данных ExAC – Exome Aggregation Consortium. Все однонуклеотидные полиморфизмы имели в среднем 50% покрытия (отношения количества чтений с данным аллелем к общему количеству чтений данной последовательности).

Также, используя скрипты, были получены списки инсерций и делеций, кратных трем, для каждого пациента, отвечающие следующим критериям: не нейтральные оценки, полученной из одной из трех программ: SIFT-Indel – "Damaging", PROVEAN – "Deleterious" и MutationTaster – "Disease-causing". Также для некоторых вариантов была определена популяционная частота европеоидной популяции с помощью базы данных ExAC – Exome Aggregation Consortium.

Помимо этого, в список были включены инсерции и делеции, которые не кратны трем, а значит приводят к сдвигу рамки считывания. Данные полиморфизмы мы автоматически приписывали к вредоносным.

Инсерции, делеции и мутации сдвига рамки считывания имели в среднем 25% покрытия. Также нами не было найдено инсерции или делеции, которая имела бы больше одной не нейтральной оценки. Мы считаем, что причиной

малого покрытия и неоднозначного результата, полученного из программ предсказания, может быть тот факт, что при экзомном секвенировании происходит большое количество ошибок при анализе мутаций большого размера.

В общей сложности нами было найдено 18 различных полиморфизмов в 15 генах. В их число входит 10 миссенс-мутаций, 3 мутации сдвига рамки и 5 делеций.

Данные гены были найдены как и в группе «Основной приоритет», так и в группе «Низкий приоритет». Какой-либо статистически значимой разницы между двумя группами нами найдено не было, в первую очередь из-за малень-кой выборки генов.

В результате данной работы нами был выработан четкий технологический процесс первичного анализа данных экзомного секвенирования и идентификации редких вредоносных аллелей, также с помощью биоинформатических инструментов были описаны редкие мутации, возможно определяющие предрасположенность к раку почек. Технологический процесс включает в себя:

Составление списка генов-кандидатов с помощью литературного поиска, поиска по доступным базам данных и приоритизации генов путем построения сетей взаимодействия.

Выявление вредоносных мутаций с помощью биоинформатических инструментов предсказания функциональной значимости мутаций in silico на основе экзомного анализа пациентов.

Описание найденных мутаций.

References

- 1. Adzhubei, I. a *et al.* A method and server for predicting damaging missense mutations. *Nat. Methods* **7**, 248–249 (2010).
- Bamford, S. *et al.* The COSMIC (Catalogue of Somatic Mutations in Cancer) database and website. *Br. J. Cancer* **91**, 355–8 (2004).
- 3. Cheng, W. C. *et al.* DriverDB: An exome sequencing database for cancer driver gene identification. *Nucleic Acids Res.* **42**, (2014).
- 4. Choi, Y., Sims, G. E., Murphy, S., Miller, J. R. & Chan, A. P. Predicting the Functional Effect of Amino Acid Substitutions and Indels. *PLoS One* **7**, (2012).
- Gao, J. *et al.* Integrating and annotating the interactome using the MiMI plugin for cytoscape. *Bioinformatics* 25, 137–138 (2009).
- Gentleman, R. & Ihaka, R. R: A language for data analysis and graphics. J. Comput. Graph. Stat. 5, 299–314 (1996).
- Grimm, D. G., Azencott, C., Aicheler, F., Gieraths, U. & Daniel, G. The evaluation of tools used to predict the impact of missense variants is hindered by two types of circularity. 1–37 (2014). doi:10.1002/humu.22768.This
- Hamosh, A., Scott, A. F., Amberger, J. S., Bocchini, C. A. & McKusick, V. A. Online Mendelian Inheritance in Man (OMIM), a knowledgebase ofhumangenes and genetic disorders. *Nucleic Acids Res.* 33, D514-517 (2005).

- 9. Hu, J. & Ng, P. C. SIFT Indel: Predictions for the Functional Effects of Amino Acid Insertions/Deletions in Proteins. *PLoS One* **8**, (2013).
- Jianjiong Gao, Bülent Arman Aksoy, Ugur Dogrusoz, Gideon Dresdner, Benjamin Gross, S. Onur Sumer, Yichao Sun, Anders Jacobsen, Rileen Sinha, E. L. & Ethan Cerami, Chris Sander, and N. S. Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the cBioPortal. *Sci. Signal.* 6, 1–34 (2013).
- Joly, Y., Dove, E. S., Knoppers, B. M., Bobrow, M. & Chalmers, D. Data sharing in the post-genomic world: The experience of the international cancer genome consortium (ICGC) data access compliance office (DACO). *PLoS Comput. Biol.* 8, (2012).
- 12. Krämer, A., Green, J., Pollard, J., Tugendreich, S. & Tugendreich, S. Causal analysis approaches in Ingenuity Pathway Analysis. *Bioinformatics* **30**, 523–30 (2014).
- Kumar, P., Henikoff, S. & Ng, P. C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat. Protoc.* 4, 1073–1081 (2009).
- 14. Piñero, J. *et al.* DisGeNET: A discovery platform for the dynamical exploration of human diseases and their genes. *Database* **2015**, (2015).
- 15. Racine, J. S. RSTUDIO: A platform-independent IDE for R and sweave. *Journal of Applied Econometrics* 27, 167–172 (2012).
- Rebhan, M., Chalifa-Caspi, V., Prilusky, J. & Lancet, D. GeneCards: A novel functional genomics compendium with automated data mining and query reformulation support. *Bioinformatics* 14, 656–664 (1998).
- Reva, B., Antipin, Y. & Sander, C. Predicting the functional impact of protein mutations: Application to cancer genomics. *Nucleic Acids Res.* 39, 1–14 (2011).
- Shannon, P. *et al.* Cytoscape : A Software Environment for Integrated Models of Biomolecular Interaction Networks Cytoscape : A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* 2498–2504 (2003). doi:10.1101/gr.1239303
- Stenson, P. D. *et al.* Human Gene Mutation Database (HGMD). *Hum. Mutat.* 21, 577– 581 (2003).

10

Кластеризация хроматина по данным о модификации гистонов

Маргасюк С. Д.¹; Миронов А. А.²

¹ Факультет биоинженерии и биоинформатики МГУ имени М. В. Ломоносова ² Факультет биоинженерии и биоинформатики МГУ имени М. В. Ломоносова; Институт проблем передачи информации РАН; Сколковский университет науки и технологий smargasyuk@gmail.com

Посттрансляционная модификация гистонов – один из способов регуляции структуры хроматина. Функции этих модификаций можно упрощенно разделить на две крупные категории: поддержание глобальной СТРУКТУРЫ хроматина (доменов) и управление связанными с ДНК биологическими механизмами [2]. Изучение модификаций основано, в первую очередь, на анализе данных экспериментов ChIP-Seq. При том, что показана связь отдельных модификаций с участками генов (например, TSS) и биологически значимыми параметрами (например, уровнем экспрессии по данным mRNA-Seq) [4] с помощью анализа корреляции, для описания функционального эффекта совокупности меток (то есть разделения хроматина функционально значимые классы) используются более на сложные статистические подходы, такие как иерархическая кластеризация [3] и скрытая марковская модель [1]. Цель данной работы – улучшить существующие методы кластеризации хроматина, основанные на скрытой марковской модели, и создать программные инструменты, реализующие следующие процедуры.

Во-первых, широко используемые программы (например, ChromHMM, примененный для аннотации в проекте ENCODE) кластеризуют хроматин на основании бинаризованных данных (считается, что модификация присутствует в позиции, если fold-change эксперимента ChIP-Seq превышает заданное значение). Такая подготовка данных приводит к потере информации о силе сигнала; в нашей работе бинаризация не проводится, и экспериментальные после квантильной трансформации считаются выборкой данные многомерного нормального распределения. (такой подход уже применен, например, в программе MRHMMs [5]). В этом случае необходимо каким-то образом определить эмиссионную вероятность при непрерывном распределении, и в существующих инструментах она считается равной значению функции плотности в соответствующей точке (в частности, может быть больше 1). Таким образом, нужно разработать альтернативные способы эмиссионных вероятностей и использовать их для залания оценки правдоподобия существующих моделей и, возможно, обучения других. Выбор различных методов оценки может привести, в частности, к различным результатам при выборе оптимального количества состояний (выбор в зависимости от правдоподобия модели со штрафом за количество параметров: критерии BIC и AIC): выбор числа состояний по мере правдоподобия, описанной выше, не дает результатов (минимум критерия не достигнут при числе состояний менее 30).

Во-вторых, вычисление корреляции между данными ChIP-Seq для различных модификаций показывает положительную корреляцию между любыми двумя из них, что противоречит предположению об их различной функции. По нашей гипотезе, этот эффект вызван зависимостью экспериментально определяемого уровня всех модификаций от плотности нуклеосом, различной в разных геномных позициях. Для выделения плотности нуклеосом как отдельного показателя применен метод главных компонент: векторы данных проецируются на пространство, ортогональное первой главной компоненте.

Обучение скрытой марковской модели исходя из этих принципов приводит к набору состояний, передающих основные классы хроматина: в частности, существуют активные и репрессированные промоторные состояния, энхансерные области и несколько классов гетерохроматина. Показана корреляция между длиной проекции на первую главную компоненту для состояния и средним уровнем экспрессии по данным RNA-Seq для состояния. Не показана возможность выбора количества состояний исходя из кластеризации состояний модели с избыточным количеством состояний.

Эта работа поддержана грантом РНФ 14-50-00150.

References

- 1. Ernst, J., Kellis, M.: Chromatin-state discovery and genome annotation with ChromHMM. Nature Protocols volume 12, pp. 2478–2492 (2017)
- 2. Kouzarides, T.: Chromatin modifications and their function. Cell. 2007 Feb 23;128(4):693-705
- Noureen, N., Qadir, M. A. et al.: ChromClust: A semi-supervised chromatin clustering toolkit for mining histone modifications interplay. Genomics Volume 106, Issue 6, December 2015, pp. 355-359
- Stavrovskaya, E. D., Mironov, A.A. et al.: StereoGene: Rapid Estimation of Genomewide Correlation of Continuous or Interval Feature Data. *Bioinformatics*, Volume 33, Issue 20, 15 October 2017, pp. 3158–3165
- Lee, Y., Ghosh, D., Hardison, RC., Zhang Y.: MRHMMs: Multivariate Regression Hidden Markov Models and the variantS. *Bioinformatics*, Volume 30, Issue 12, 15 June 2014, pp. 1755–1756

Связь разнообразия местообитаний и размера периферии пангенома у бактерий

Николаева Дарья Дмитриевна Факультет биоинженерии и биоинформатики daranikolaeva@gmail.com

Пангеном - это совокупность белок-кодирующих генов, присутствующих в наборе геномов одного вида или рода бактерий [1]. Традиционно в структуре пангенома выделяют "универсальный геном" - гены, которые встречаются почти во всех рассматриваемых штаммах, и "периферию" - гены, встречающиеся у небольшого количества штаммов. Соотношение размера периферии к размеру универсального генома отличается у разных бактерий, встречаются как виды с обширной периферией, так и виды, у которых количество уникальных для штаммов генов крайне мало. Какие именно факторы определяют это соотношение, до сих пор остается непонятным. Мы предположили, что одним из определяющих факторов может являться разнообразие местообитаний, в которых встречается данный вид. С одной стороны, если вид бактерий живет в большом количестве разнообразных местообитаний (назовем такой вид генералистом), то разным подгруппам внутри этого вида требуются специфичные гены для приспособления к разным физико-химическим условиям в разных экологических нишах. А, с другой стороны, существуют виды-специалисты, которые привязаны к единственной экологической нише, хорошо приспособлены к ней и поэтому, вероятно, генетически более однородны. Напрямую гипотеза о взаимосвязи между структурой пангенома и разнообразием местообитаний вида бактерий, насколько нам известно, ни разу не проверялась. Получить информацию о структуре сообществ можно из данных метагеномики, когда совместно секвенируются все нуклеотидные последовательности, выделенные из данного местообитания. Такой подход позволяет определить как качественный, так и количественный состав организмов разных экосистем.

Таким образом, целью данной работы было установить, существует ли связь между соотношением элементов структуры пангенома (универсального генома и периферии) и

количеством местообитаний, в которых данный вид бактерий встречается. Для проверки этой гипотезы были использованы метагеномные данные из 61 различного местообитания, сгруппированных в 11 кластеров по сходству условий [2]. Виды бактерий, обитающие в пяти и более кластерах были названы генералистами, а виды, встречающиеся только в одном кластере, - специалистами. Из этих данных для построения пангеномов были отобраны все доступные виды, для которых отсеквенировано как минимум семь геномов. Сравнение структуры пангенома генералистов и специалистов показало, что более общирная периферия скорее характерна для видов-генералистов. Понимание того, какое влияние оказывает разнообразие местообитаний на структуру пангенома, позволит получать более информативные результаты при использовании этого инструмента в исследованиях бактериальной эволюции.

Слова благодарности

Автор выражает благодарность научным руководителям С. К. Гарушянц и М. С. Гельфанду. Данная работа поддержана грантом РНФ 18-14-00358.

Литература

Tettelin H. et al. Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: implications for the microbial "pan-genome" //Proceedings of the National Academy of Sciences of the United States of America. – 2005. – T. 102. – №. 39. – C. 13950-13955.
Sriswasdi S., Yang C., Iwasaki W. Generalist species drive microbial dispersion and evolution //Nature communications. – 2017. – T. 8. – No. 1. – C. 1162.

Репертуар белков, связанных с эпигенетикой

Ильницкий Иван¹, Анастасия Жарикова¹, Андрей Миронов¹

¹ МГУ им. М.В. Ломоносова, факультет биоинженерии и биоинформатики, Москва, Россия nfsus96@gmail.com

Abstract. Эпигенетические модификации (также называемые метками), такие как метилирование ДНК и модификации гистонов, меняют структуру хроматина и доступность ДНК, тем самым регулируя паттерны генной экспрессии. Для оставления каждой специфической метки используются разные факторы и белковые комплексы. Другой белок, у которого была найдена связь с ремоделлингом хроматина – это PIWI. Белки, находящиеся в семействе PIWI, участвуют в сайленсинге транспозонов на ранних стадиях эмбрионального развития. Для борьбы с ними они используютс специальные малые нкРНК – ріРНК. Этот способ похож на PHK-интерференцию. Перед тем, как ріРНК формирует комплекс с PIWI, они процессируются группой белков. Новые исследования указывают на связь системы ріРНК-РІWI с эпигенетикой. С целью изучить коэволюционные паттерны между белками системы ріРНК—РІWI и факторами эпигенетики нами была собрана коллекция белков, ассоциированных с этими группами.

Ключевые слова: эпигенетика, PIWI, транспозоны.

Эпигенетические процессы важны для функционирования живых систем и поддержания стабильности генома, влияют на экспрессию белков и РНК. Сегодня термин "эпигенетика" используется для обозначения наследуемых изменений генома, не связанных с преобразованием последовательности ДНК. Взаимодействие эпигенетических факторов и хроматина к настоящему времени достаточно широко изучено. Эпигенетические модификации (также называемые метками), такие как метилирование ДНК и модификации гистонов, меняют структуру хроматина и доступность ДНК, тем самым регулируя паттерны генной экспрессии. В 2006 году был открыт новый класс малых некодирующих РНК — РНК, взаимодействующие с белками класса PIWI (PIWI-interacting RNA, piPHK), отмеченных в процессах модификации хроматина. piPHK транскрибируются со специальных кластеров, после чего их созревание сопровождается взаимодействием с группой белков. Связываясь с белком PIWI, они способны лимитировать экспрессию ретротранспозонов на посттранскрипционном уровне с помощью эпигенетического сайленсинга (рекрутирование ДНК-метилтрансфераз) в развивающихся эмбриональных клетках [1]. Также исследования на дрозофилах показывают, что белок PIWI активирует фактор HP1, стимулирующий формирование гетерохроматина [2]. Наконец, был предложен механизм регуляции гена CREB2 у моллюска рода Aplysia при котором комплекс piPHK—PIWI связывается с формирующимся транскриптом гена. Комплекс рекрутирует неизвестный модификатор, метилирующий CpG нуклеотиды в промоторной области [3]. Таким образом, piPHK—PIWI система имеет много аспектов в эпигенетике, однако ее взаимодействие с основными факторами ремоделирования хроматина на сегодняшний день недостаточно исследованы. С целью изучить коэволюционные паттерны между белками системы piPHK—PIWI и факторами эпигенетики нами была собрана коллекция белков, представляющих каждую из данных групп, а также были написаны скрипты на языках программирования Python и SPARQL для автоматизации последующего анализа.

Эта работа поддержана грантом РНФ 14-50-00150.

Литература

- Aravin A.A., Sachidanandam R., Bourc'his D., Schaefer C., Pezic D., Toth K.F., Bestor T, Hannon GJ. A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. Mol Cell. 2008 Sep 26;31(6):785-99. doi: 10.1016/j.molcel.2008.09.003. PubMed PMID: 18922463; PubMed Central PMCID: PMC2730041.
- Brower-Toland B, Findley SD, Jiang L, Liu L, Yin H, Dus M, Zhou P, Elgin SC, Lin H. Drosophila PIWI associates with chromatin and interacts directly with HP1a. Genes Dev. 2007 Sep 15;21(18):2300-11. PubMed PMID: 17875665; PubMed Central PMCID: PMC1973144.
- Rajasethupathy, P., Antonov, I., Sheridan, R., Frey, S., Sander, C., Tuschl, T., & Kandel, E. R. (2012). A role for neuronal piRNAs in the epigenetic control of memory-related synaptic plasticity. Cell, 149(3), 693–707. doi: 10.1016/j.cell.2012.02.057

2

Консервативность неконсенсусных позиций в сайтах связывания факторов транскрипции.

Белоусова Е. А.¹

¹Факультет Биоинженерии и Биоинформатики, МГУ им. М. В. Ломоносова, Москва, Россия bilyius@mail.ru

Краткий обзор

Ранее [1] был замечен феномен консервативности неконсенсусных позиций в сайтах связывания факторов транскрипции: некоторые неконсенсусные основания сайта медленнее эволюционируют в конкретных группах бактерий. Это явление хорошо заметно только для глобальных регуляторов, для которых существуют большие регулоги – группы регулонов из близких видов.

Чтобы оценить, насколько значима консервативность неконсенсусных позиций, эти позиции можно сравнить с третьими позициями четырехвырожденных кодонов генов, транскрипция которых регулируется данным фактором, что и было сделано в 2005 году [1]. Оказалось, что изучаемые неконсенсусные основания, действительно, более консервативны, чем третьи позиции. Этому может существовать несколько объяснений. Вопервых, возможно перекрывание сайтов связывания регуляторов транскрипции, и тогда консервативные «неконсенсусы» относятся к еще неизвестным регуляторным элементам. Во-вторых, замена неконсенсусного основания на консенсусное меняет сродство регулятора к сайту, а значит, и уровень транскрипции гена. То есть, если неконсенсусное основание однажды начало обеспечивать нужный уровень регуляции, то оно уже не может «переместиться» на другую позицию.

Благодаря развитию полногеномных технологий, число известных сайтов регуляторов транскрипции в последние годы существенно возросло. В данной работе стоит задача найти хорошо выраженные консервативные неконсенсусы в ортологичных сайтах, сравнить их консервативность с нейтрально (или почти нейтрально) эволюционирующими элементами и объяснить это явление. Мы начали с регулятора СсрА. Этот фактор обеспечивает транскрипционный ответ на появление быстроусвояемых углеводородов в среде [2]. Логотип сайта связывания СсрА, полученный с помощью WebLogo [3], представлен на рисунке 1. В данном сайте были обнаружены позиции с консервативным неконсенсусным основанием в нескольких группах ортологичных сайтов. По предварительной грубой оценке они, действительно, более консервативны, чем третьи основания четырехвырожденных кодонов. Однако, это наблюдение требует дальнейшей проверки.

Данная работа производится совместно с М. С. Гельфандом. Исследование поддержано грантами РНФ 18-14-00358 и РФФИ 18-34-01006.



Рис. 1. Логотип сайта связывания СсрА, полученный с помощью WebLogo.

Ссылки

- 1. Kotelnikova, E., Makeev, V., Gelfand, M.: Evolution of transcription factor DNA binding sites. Gene 347(2):255-63 (2005)
- 2. RegPrecise, http://regprecise.lbl.gov/RegPrecise/index.jsp
- 3. WebLogo, http://weblogo.berkeley.edu/logo.cgi

Mutational patterns in *E.coli* evolution

Sofya K.Garushyants^{1,2} and Maria V. Selifanova³

¹ Kharkevich Institute for Infromation Transmission Problems, Moscow, Russia ² Skolkovo Institute of Science and Technology, Moscow, Russia ³ Lomonosov Moscow State University, Moscow, Russia

garushyants@iitp.ru

Abstract. The concept of mutational signatures is widely applied in eukaryotic studies, especially in cancer genomics. To date there is no work applying this concept to the bacteria. We use the data from evolutionary experiments in *Escherichia coli* (LTEE and others) to determine mutational signatures in bacteria. We were able to identify three signatures associated with MMR system (*mutH*, *mutL*, *mutS*, *uvrD*), *dnaQ* and *mutY* imperfections. We calculate the divergence profile for each *E.coli* phylogroup, and then estimate the contribution of the different mutational processes identified above to *E.coli* divergence. We found that imperfection of *mutY* is contribute the most to *E.coli* evolution.

Keywords: Escherichia, evolution, mutational.

1 Introduction

Bacteria are perfect model objects for the evolution studies because they can be easily cultivated, have short generation time and small genomes. The Long term evolutionary experiment is ran on *Escherichia coli* for more than 30 years, which is approx. 60000 generations. Once in a few hundred generations the bactria are sequenced, which allow to observe the accumulation of mutations over time. Most of these mutations are single nucleotide substitutions. During this experiment the few strains with mutation that affect DNA repair or removal of oxidized nucleotides were identified. These strains accumulated 100-fold more mutations than wild type lines and were called hypermutators.

Studies of mutations in bacteria normally concentrate only on the change in the particular nucleotide, while the context in which this change occured is not acknowledged. This approach allows to identify the changes, but does not allow to study more complex mutational mechanisms.

On the other hand, in cancer genomics the nucleotide context information is widely used. It was shown, that knockouts of different genes lead to the change in mutational spectra, and especially the changes in the context in which this mutation occurs. To describe this observation more accurately the concept of mutational signatures was developed. The profile of each signature is displayed using the six substitution subtypes: C>A, C>G, C>T, T>A, T>C, and T>G. Further, each of the substitutions is examined by incorporating information on the bases immediately 5' and 3' to each mutated base generating 96 possible mutation types. Different signatures characterize differnt mutational processes in the cell. To date 30 such signatures were identified from the cancer sequencing data.

To date mutational signatures were not identified in bacteria.

2 Results

We show here that the mutational signatures approach can be used to study bacterial evolution and identify here three mutational signatures from mutator or known knockout lines. We show that these signatures are similar between different experiments, and that is why not an experimental artifacts. We calculate the divergence profile for each *E.coli* phylogroup, and then estimate the contribution of the different mutational processes identified above to *E.coli* divergence. We found that imperfection of *mutY* is contribute the most to *E.coli* evolution.

3 Acknowledgements

This work was done under the supervision of Prof. Gelfand (IITP RAS). This work was supported by the Russian Science Foundation under grant 18-14-00358.

Accumulation of mutations in experimental evolution of basidiomycete fungus Schizophyllum commune

Aleksandra Bezmenova^{1,2}, Elena Zvyagina³, Anna Fedotova², Tatiana Neretina⁴,

Aleksey Penin^{5,2}, and Alexey Kondrashov⁶

¹ Skolkovo Institute of Science and Technology, Moscow, Russia
² Lomonosov Moscow State University, Moscow, Russia
³ Yugansky Nature Reserve, Russia
⁴ N. A. Pertsov White Sea Biological Station, Russia
⁵ Institute for Information Transmission Problems, Russia
⁵ University of Michigan, USA

aleksandra.bezmenova@skoltech.ru

Abstract. Basidiomycete fungus Schizophyllum commune is a unique organism with the highest genetic diversity among studied species being, which can reach 20%, and a high mutation rate of $2.0*10^{-8}$ substitutions per nucleotide per generation. We designed an experiment which allows us to study the process of somatic mutations accumulation during vegetative growth of haploid monospore cultures of S. commune in long tubes with fixed diameter of the section for long period of time. We have been cultivating 24 derived lines of four S. commune cultures in tubes of two different diameters. 12 lines were cultivated in very thin tubes with diameter being 0.8 - 0.9 mm. Such system is aimed minimize the effect of natural selection, so that the mutation rate can be measured. 12 lines were cultivated in medium tubes with diameter being 4 mm, where the population size of the mycelium is effectively large. Cultures have grown to the length of 75 - 100 cm in thin tubes and 2 - 2.5 m in medium tubes, and were sequenced after every 15-20 cm of growth. We detected a total of 296 de novo mutations, 234 of which have reached high (70% and more) frequency in the mycelium. We saw no evidence of natural selection, in particular neither prevalence nor deficiency of nonsynonymous substitutions among coding mutations in thin tubes. The estimation of the rate of mutations accumulation was 3.4*10⁻¹¹ substitutions per nucleotide per cell division in thin tubes and 1.4*10⁻¹¹ substitutions per nucleotide per cell division in medium tubes, which is consistent with per generation estimations. We also observed 10 events of parallel mutations and 7 events of clonal interference.

Keywords: Experimental evolution, Schizophyllum commune, Somatic mutations, Mutation rate, Natural selection

1 Introduction

Basidiomycete fungus *Schizophyllum commune* is a unique organism with the highest genetic diversity among studied species being, which can reach 20% [1], and a high mutation rate of $2.0*10^{-8}$ substitutions per nucleotide per generation [2]. The life cycle of *S. commune* includes a mononuclear haploid stage called monospore culture, which originates from a single spore and can be easily cultivated on solid medium. This makes *S. commune* a promising object to study the somatic mutation process during vegetative growth of the fungus.

We designed an experiment which allows us to cultivate monospore cultures of *S. commune* in long tubes with fixed diameter of the section for long period of time. By varying the diameter of the tube, we can regulate the size of population of growing hyphae, thus effecting the efficacy of natural selection acting in the mycelium.

We have been cultivating 24 derived lines of four *S. commune* cultures in tubes of two different diameters. 12 lines were cultivated in very thin tubes with diameter being 0.8 - 0.9 mm. Such system is aimed minimize the effect of natural selection, so that the mutation rate can be measured. 12 lines were cultivated in medium tubes with diameter being 4 mm, where the population size of the mycelium is effectively large. Cultures have grown to the length of 75 - 100 cm in thin tubes and 2 - 2.5 m in medium tubes and were sequenced after every 15-20 cm of growth.

2 Results

We detected a total of 296 de novo mutations, 234 of which have reached high (70% and more) frequency in the mycelia. 67 mutations have reached some significant frequency but then disappeared, including 7 events of clonal interference. We also observed 10 events of parallel mutations (Table 1).

To estimate the influence of natural selection on mutations accumulation, we performed permutations of coding substitutions keeping reference and derivative variants and found no evidence of excess or deficiency of nonsynonymous substitutions in thin tubes compared to the expected number.

The estimation of the rate of mutations accumulation was $3.4*10^{-11}$ substitutions per nucleotide per cell division in thin tubes and $1.4*10^{-11}$ substitutions per nucleotide per cell division in medium tubes, which is of the same order as that estimated from per generation mutation rate.

2
	Culture								
	Small tubes				Medium tubes				
	sh01	sh02	sh03	sh04	sh01	sh02	sh03	sh04	Total
Total High frequency	25	30	20	57	12	26	30	96	296
(>70%)	17	27	20	36	12	23	26	73	234
Low frequency	8	3	0	21	0	3	4	23	62
Dissapeared	4	4	0	29	0	3	5	22	67
Parallele Clonal Interfer-	0	1	0	8	0	0	1	0	10
ence	0	0	0	3	0	0	1	3	7

Table 1. Number of different types of *de novo* mutations and clonal interference events.

References

_

- Baranova M. et al., "Extraordinary Genetic Diversity in a Wood Decay Mushroom," Mol. Biol. Evol., vol. 32, no. 10, pp. 2775–2783 (2015).
- 2. Seplyarskiy V. et al., "Crossing-over in a hypervariable species preferentially occurs in regions of high local similarity," Mol. Biol. Evol., vol. 31, no. 11, pp. 3016–3025 (2014).

Funding

This work was supported by the Russian Science Foundation, grant No 16-14-10173.

Population genomic data reveal signatures of genetic exchange in the bdelloid rotifer *Adineta vaga*

Olga A. Vakhrusheva^{1,2}, Elena A. Mnatsakanova³, Yan Galimov⁴, Evgeny S. Gerasimov³, Tatiana V. Neretina⁵, Aleksey A. Penin⁶, Maria D. Logacheva^{1,6}, Georgii A. Bazykin^{1,2}, Alexey S. Kondrashov^{7,8}

¹Skolkovo Institute of Science and Technology, Moscow 143026, Russia ²Sector for Molecular Evolution, Institute for Information Transmission Problems of the RAS (Kharkevich Institute), Moscow 127051, Russia e-mail: vakh57@gmail.com

³Faculty of Biology, M.V. Lomonosov Moscow State University, Moscow 119991, Russia ⁴Koltsov Institute of Developmental Biology, Moscow 119334, Russia

⁵N.A.Pertsov White Sea Biological Station, Faculty of Biology,

M.V. Lomonosov Moscow State University, Moscow 119991, Russia ⁶Belozersky Institute of Physico-Chemical Biology,

M.V. Lomonosov Moscow State University, Moscow 119991, Russia ⁷Faculty of Bioengineering and Bioinformatics,

M. V. Lomonosov Moscow State University, Moscow 119991, Russia

⁸Life Sciences Institute and Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, USA

Abstract. Sexual reproduction is considered indispensable for the long-term survival of a species, as it enables efficient selection against deleterious mutations. Transitions to asexual reproduction happen frequently in eukaryotes, but usually lead to quick extinction of the involved species, attributed to rapid accrual of harmful mutations. Therefore, obligate asexuality is often regarded as an evolutionary dead end. However the mere existence of ancient asexual lineages challenges this point of view [1]. Bdelloid rotifers are a large group of putative ancient asexuals that presumably abandoned sexual reproduction tens of millions of years ago [2]. To address the possibility of recombination in the bdelloid rotifer *Adineta vaga* on a whole-genome scale, we sequenced genomes of 11 *A. vaga* clonal lineages. Using several types of analysis we show that the observed patterns of population structure in *A. vaga* are incompatible with obligate asexuality, thus providing compelling evidence of genetic exchange in this 'ancient asexual' species.

Keywords: Evolution of Asexuality, Bdelloid Rotifers, Recombination.

1 Introduction

The reasons for the persistence of sexual reproduction in eukaryotes have become a focal point for debate in evolutionary biology in the twentieth century. One of the most widely accepted theories explaining the prevalence of sexual reproduction states

that recombination is crucial for elimination of deleterious mutations. Transition to asexual reproduction usually results in a quick extinction of a lineage; young evolutionary ages of the majority of asexual lineages provide strong support to this theory. However the existence of ancient asexual lineages, with bdelloid rotifers being the most striking example of such a lineage, contradicts this concept [1].

Bdelloid rotifers are a large group of microscopic aquatic invertebrates. Based on the long-term observations and fossil records, bdelloid rotifers were presumed to have completely abandoned sexual reproduction tens of millions of years ago [2]. If this hypothesis held true, this would make bdelloid rotifers the largest ancient asexual eukaryotic taxon. It is not clear how they managed to survive in spite of the apparent absence of meiosis. On these grounds, bdelloid rotifers are often referred to as 'ancient asexual scandals'.

However, the possibility of some form of cryptic sexual reproduction occurring in bdelloid rotifers has not been refuted. To address the possibility of recombination in the bdelloid rotifer *Adineta vaga* on a whole-genome scale, we sequenced genomes of 11 *A. vaga* clonal lineages. We carry out population genomic analysis to look for signatures of genetic exchange in the nuclear genomes of *A. vaga*.

2 Results

Clonal cultures of *A. vaga* were started from single rotifers isolated from moss samples collected at different locations in Zvenigorod forest area. We have successfully established 11 clonal lineages started from 11 *A. vaga* individuals originating from different geographical locations. All samples were subjected to DNA extraction followed by construction of genomic libraries which were subsequently sequenced on the Illumina HiSeq platform. We additionally sequenced one of the lineages on the MiSeq platform, which allowed us to generate a high quality de novo genome assembly for this lineage. The obtained assembly had an N50 of 22 kilobases, with 19068 contigs longer than 500 bp covering 197 megabases of sequence. The resulting assembly was used as a reference for downstream analysis of the other sequenced samples. All samples were mapped against the reference assembly and single nucleotide polymorphisms were called with SAMtools mpileup.

Next, we proceeded to look for potential evidence of recombination in the population of *A. vaga*. One of the most characteristic signatures of recombination is a decay of linkage disequilibrium (LD) with increasing physical distance. However, if there is no recombination, the extent of LD is not expected to depend on the distance between genomic loci. We investigated the patterns of LD in the population of *A. vaga*. Contrary to what would be expected if *A. vaga* propagated strictly asexually, we found that linkage disequilibrium rapidly declines with increasing distance between polymorphic sites, indicative of genetic exchange in the population of this species.

Besides maintaining complete linkage between different genomic loci, obligate asexual reproduction should affect population frequencies of homozygous and heterozygous genotypes, leading to significant deviations from Hardy-Weinberg equilibrium. However, only 1.2% of the polymorphic sites were found to deviate from HardyWeinberg expectations in *A. vaga* population. The distribution of ratios of the observed number of homozygous genotypes to the expected one is concentrated around 1 (median ratio = 0.97), suggesting that the studied *A. vaga* population at large is at Hardy-Weinberg equilibrium, hardly consistent with obligate asexuality.

Finally, we show that different loci of the *A. vaga* genome have different evolutionary histories reflected in incongruence among phylogenetic trees built from different loci, again pointing to genetic exchange in *A. vaga*.

Overall, several lines of evidence suggest that the diversity in the population of the bdelloid rotifer *A. vaga* has been shaped by genetic exchange and recombination, inconsistent with obligate asexuality previously thought to be the only mode of reproduction in bdelloid rotifers.

References

- 1. Smith, J.M.: Evolution: contemplating life without sex. Nature 324, 300-301 (1986).
- Birky, C.W.: Positively negative evidence for asexuality. J.Hered. 101 Suppl 1, S42–45 (2010).

Funding

This work was supported by the Russian Foundation for Basic Research, research project No. 16-34-01303 мол_а,

by the Russian Science Foundation, grant No. 16-14-10173 and by a grant from the RAS program "Molecular and Cellular Biology".

Senescence and entrenchment in evolution of amino acid sites

A. V. Stolyarova¹, E. R. Nabieva², V. V. Ptushenko², G. A. Bazykin^{1,3}

¹ Skolkovo Institute of Science and Technology, Moscow, Russia

² Faculty of Bioengineering and Bioinformatics, Moscow State University, Russia

³ Kharkevitch Institute for Information Transmission Problems of the Russian Academy of Sciences, Moscow, Russia

anastasiia.stoliarova@skoltech.ru

Abstract. Amino acid propensities at an amino acid site change with time, leading to amino acid substitutions and to molecular evolution. However, the causes of these changes are unclear. They may arise due to changes elsewhere in the genome, in a process known as entrenchment, whereby the fitness of the currently present allele increases with time since its origin. Alternatively, they may arise due to environmental changes that are external to the genome. While entrenchment is a prevalent force shaping protein evolution, how much the fitness landscape at the genome scale is affected by the changes in the environment remains unknown. Here, using analytical modelling and simulations, we show that environmental fluctuations are expected to cause a pattern opposite to that caused by entrenchment: namely, the average fitness of the currently present allele decreases with time since its origin - a process we term senescence. To elucidate the relative roles of these two processes - entrenchment and senescence - in protein evolution, we design a computational method for analysis of the distribution of substitutions at a single amino acid site across the phylogeny. By applying it to substitutions at amino acid sites in nuclear genomes of vertebrates, we show that among the 78% of the most conservative sites nearly all are susceptible to strong entrenchment. By contrast, among the ~0.7% of sites that experience positive selection, >50% experience senescence. Alongside the increased substitution rate, senescence of the current allele is a facet of positive selection.

Keywords: molecular evolution, fitness landscape, positive selection

1. Introduction

Fitness landscape is the key concept of evolutionary biology. It is necessary to study the shape and the structure of fitness landscapes to understand the processes of adaptive evolution and speciation. Because of the large dimension of the landscapes it is impossible to measure the fitness of all possible genotypes experimentally. However, comparative genomics methods can be used to distinguish the global features of fitness landscapes. For each single amino acid position we can define a single-position fitness landscape (SPFL): a vector of fitness values of all available alleles in this position. SPFL can change with time, leading to amino acid substitutions and to molecular evolution. The changes may arise due to environmental fluctuations; alternatively, even in a constant environment, they may arise due to changes in interacting sites elsewhere in the genome under epistasis. (reversions, convergent evolution, correlated substitutions, bursts of substitutions). However, the contribution of each of these factors to the SPFL dynamics remains unclear.

The prevailing factor shaping the SPFL changes was shown to be epistasis. Under prevalent negative selection, the interaction of amino acid substitutions in different sites results in an increase of the fitness of the current allele - the process called entrenchment. The other force shaping the fitness landscapes is positive selection, which leads to higher rate of alleles' replacements. However, how positive selection translates into dynamics of the relative fitness of a single allele remains poorly studied.

2. Results

First, we ask how fluctuating selection affects the relative fitness of different alleles at a site. If the changes of the SPFL are random with respect to the identity of the allele currently residing at the site, we expect that they, on average, will reduce its relative fitness. Indeed, due to natural selection, the relative fitness conferred by the current variant is, on average, higher than that of a random variant at this site, so that a random change to the SPFL will, on average, reduce it. An episode of positive selection triggered by an SPFL change may then cause the spread of a novel variant which would confer high fitness till the next SPFL change. We demonstrate this effect using simulations on randomly changing fitness landscape.

Since environmental fluctuations and genomic changes affect the SPFL in the opposite ways, we can distinguish between them by looking on the direction of current allele fitness changes. The increase of the current allele's fitness with time is called entrenchment, and the opposite process we term senescence. Alleles preferences in a single genomic position specify the probability of substitutions in the course of evolution of some group of species, so we can use substitution patterns along the phylogeny to examine whether senescence or entrenchment takes place in some genomic dataset.

To estimate the rate of the current allele's fitness change and the fraction, we used ABC-SMC (Approximate Bayesian Computation Sequential Monte Carlo) approach, allowing to infer parameters without likelihood function based on simulations of evolution on dynamic landscapes iteratively. In these simulations the fitnesses of all possible alleles are drawn randomly from gamma distribution with rate = shape parameter alpha, and the fitness of the current allele changes linearly with coefficient k (k < 0)

correspond to senescence, and k > 0 correspond to entrenchment). We used two model functions for ABC-SMC. The first one represented mixture of sites with static SPFL (k = 0) and sites under senescence (k < 0) or entrenchment (k > 0). It took three parameters to estimate: *alpha* parameter of gamma fitness distribution, current allele change parameter k and fraction of sites with dynamic SPFL w. The second model function allows to mix sites with static SPFL with both sites under senescence, entrenchment. Therefore, is requires five parameters: the same *alpha* gamma distribution parameter, $k_senescence < 0$, $k_entrenchment > 0$ and the two fractions of corresponding sites w senescence and w entrenchment.

Parameters inference was performed separately for sites with different dN/dS estimations (dN/dS > 1 corresponds to positive selection acting in these sites). The posterior distributions for *k* and *w* parameters are shown in fig. 1. In both models nearly all the sites under negative selection (with dN/dS < 1) are shown to be under strong entrenchment, and >50% of the positively selected sites (with dN/dS > 1) are shown to demonstrate strong senescence effect. The first model confirms that nearly neutral sites show no effect at all. We also see that the fraction of sites under entrenchment or senescence is the main factor changing with dN/dS rather than the rate of the current allele's fitness change.

3. Acknowledges

This work was supported by the Russian Science Foundation, grant No 16-14-10173.



Fig. 1. Posterior distributions for parameters k (rate of current allele's fitness change) and w (fraction of sites with k) for the first model (two parameters shown, A) and the second model (four parameters shown, B). The error bars correspond to 95% posterior probability interval in the last iteration of ABC-SMC.

Предсказание приспособленности вируса гриппа H3N2

Сафина Ксения^{1,2}, Двуреченский Павел^{2,3}, Неверов Алексей⁴, Гасников Александр^{2,5}, Фаворов Александр⁶, Базыкин Георгий^{1,2}

¹ Сколковский институт наук и технологий, Сколково, Россия
² ИППИ РАН, Москва, Россия
³ Weierstrass Institute for Applied Analysis and Stochastics, Berlin, Germany
⁴ ФБУН ЦНИИЭ, Москва, Россия
⁵ МФТИ, Долгопрудный, Россия
⁶ Johns Hopkins University School of Medicine, Baltimore, USA

k.r.safina@gmail.com

Абстракт. Вирус гриппа H3N2 является основной причиной сезонного гриппа, ежегодно убивающего до полумиллиона человек. Вакцины против гриппа нуждаются в регулярном обновлении в связи быстрой эволюцией вирусных поверхностных антигенов (нейраминидазы и гемагглютинина), вызванной давлением коллективного иммунитета. Каждый год Всемирная Организация Здравоохранения рекомендует три штамма гриппа (по одному для каждого из подтипов A/H1N1, A/H3N2 и B/ гриппа) в качестве кандидатов на включение в состав вакцины. Выбранные штаммы должны быть антигенно близки к штаммам, которые, как ожидается, будут доминировать в предстоящем сезоне. Эффективность вакцин оказывается низкой, если ВОЗ выбирает неправильные штаммы. Процедура выбора штаммов для вакцины может быть улучшена путем моделирования эволюции гриппа. Мы разрабатываем новую модель, которая прогнозирует сезонную динамику популяций штаммов гриппа на филогенетическом дереве H3N2, учитывая информацию об эпистатических взаимодействиях между сайтами гемагглютинина.

Ключевые слова: предсказание эволюции, вирус гриппа, эпистаз

1 Введение

Вирус гриппа H3N2 является основной причиной сезонного гриппа. Вакцины против гриппа нуждаются в регулярном обновлении в связи быстрой эволюцией вирусных поверхностных антигенов (нейраминидазы и гемагглютинина), вызванной давлением коллективного иммунитета. Каждый год Всемирная Организация Здравоохранения рекомендует три штамма гриппа в качестве кандидатов на включение в состав вакцины. Выбранные штаммы должны быть антигенно близки к штаммам, которые, как ожидается, будут доминировать в предстоящем сезоне. Эффективность вакцин оказывается низкой, если ВОЗ выбирает неправильные штаммы. Процедура выбора штаммов для вакцины может быть улучшена путем моделирования эволюции гриппа.

Методы

2

2

3

Мы разрабатываем модель, которая предсказывает приспособленность штаммов вируса гриппа H3N2, используя информацию о мутациях, накопленных на филогенетическом дереве гемагглютинина. Приспособленность штамма *i* является линейной функцией от независимых вкладов отдельных мутаций на линии *i* (первая компонента суммы) и вкладов эпистатических взаимодействий между парами мутаций (вторая компонента суммы):

$$f_{i} = \sum_{\forall \alpha: 0 < \alpha \leq (i)} \left(\sum_{k=1}^{L} h_{k}(a_{k}, A_{k}) \delta_{k}(\alpha) + \sum_{\forall (j,k): j \neq k} J_{j,k} \Delta_{j,k}(\alpha) \delta_{k}(\alpha) \right)$$

Для получения матрицы эпистатических взаимодействий *J* были посчитаны эпистатические статистики, как описано в [1]. Эти статистики затем были переведены в матрицу ковариации, из которой были получены частичные корреляции.

Предсказанная частота штамма і в следующий сезон, t+1, имеет вид:

$$\tilde{x}_i(t_i+1) = Z(t_i) x_i e^{f_i}$$

где *Z* – нормирующий множитель.

Мы обучаем нашу модель на данных, доступных до сезона t+1, и делаем предсказания для сезонов t+2 и t+3. Для оценки качества предсказания мы используем клады — более крупные филогенетические единицы, поскольку отдельные штаммы сильно подвержены шуму.

Результаты

Наша модель способна корректно предсказывать рост низкочастотных клад (синие точки на Рис. 1):





Учет эпистатических взаимодействий улучшает качество предсказания (Рис. 2), если судить по суммам квадратов разностей между предсказанными и наблюдаемыми частотами:



Рис. 2. Суммы квадратов разностей для эпистатической, неэпистатической (эпистатическая матрица J=0) и наивной модели (x(t+i) = x(t)), посчитанные для клад после 2011 года.

4 Выводы

Учет эпистатических взаимодействий улучшает предсказательную силу модели. Наша модель может быть использована при выборе штаммов-кандидатов для ежесезонной вакцины.

Эта работа была поддержана грантом Президента РФ МК-1806.2017.9 и грантом Российского научного фонда 16-14-10173.

Список литературы

- 1. Sergey Kryazhimskiy, PLoS Genet 7(2):e1001301 (2011).
- 2. Marta Łuksza & Michael Lässig. Nature 507(7490):57-61 (2014)

Поиск низкочастотных вредных аллелей в популяции Drosophila melanogaster с помощью отбора на пониженную приспособленность.

Клинк Г.В., Неретина Т.В., Ступникова А., Кондрашов А.С.

Ключевые слова: экспериментальная эволюция, отбор, полиморфизм

Экспериментальная эволюция — мощный и широко применяемый метод изучения эволюционных преобразований, происходящих в популяции под действием естественного отбора. В частности, этот метод используют для поиска аллелей, частота которых в популяции растёт под действием положительного отбора при адаптации к новым условиям. Мы провели отбор на пониженную приспособленность в аутбредной популяции плодовой мушки *Drosophila melanogaster*. Отбор проводился по двум параметрам: увличение периода времени между ссаживанием пары и спариванием и понижение плодовитости. В результате действия такого противоестественного отбора на 16том поколении мухи перестали размножаться. Мы отсеквенировали и проанализировали геномы мух из предковой популяции и трёх параллельных экспериментальных линий, что позволило нам найти понижающие приспособленность аллели, которые присутствовали в предковой популяции на низкой частоте, но стали доминировать в экспериментальных линиях под действием противоестественного отбора.

Целью нашего эксперимента был поиск самых вредных аллелей, встречающихся в природных популяциях. Для этого мы провели отбор на снижение приспособленности в трёх лабораторных линиях мух *D.melanogaster*, полученных от популяции мух, пойманных в природе. Идея экспериментальной эволюции дрозофил с отбором на пониженную приспособленность принадлежит Л.З.Кайданову, но для своих экспериментов он использовал инбредные линии мух, и приспособленность понижалась за счёт de novo мутаций. Мы же взяли в качестве исходной популяции природную московскую популяцию *D.melanogaster* и старались, чтобы экспериментальные линии оставались панмиктичными, давая возможность присутствующим в предковой популяции вредным аллелям подниматься в частоте под действием отбора.

Экспериментальные популяции продержались в течение 16ти поколений. Сейчас отсеквенировано 42 генома из предковой популяции и по 14, 18 и 14 геномов из трех экспериментальных линий и ведётся их анализ.

Работа поддержана грантом РНФ №16-14-10173.

Algorithmic correction of solvation improves the quality of molecular simulations

A. S. Zlobin¹, A. O. Zalevsky^{1,2,3} and A. V. Golovin^{1,2,3,4}

 ¹ Faculty of Bioengineering and Bioinformatics, Lomonosov Moscow State University, Moscow, Russia
² Sechenov First Moscow State Medical University, Institute of Molecular Medicine, Moscow, Russia
³ Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, Moscow, Russia
⁴ LLC Apto-Pharm

alexander.zlobin@fbb.msu.ru

Abstract. Despite fruitful and widespread use of molecular simulations nowadays, certain aspects of protocols implemented in popular modelling software may lead to dramatic errors in model preparations. One of such weak points is solvation. Misplacement of water molecules may result in erroneous dynamic behaviour and lead to false conclusions. Here we present general approach to score and filter algorithmically added solvent water and show both its predictive power in recognizing structurally conserved molecules and its beneficial effect on the quality of molecular dynamics simulations. Presented approach, once implemented in a handy tool, may become part of canonic protocol to set up a system for any molecular dynamics calculation task.

Keywords: Molecular Dynamics, Conserved Water Molecules, Structural Bioinformatics.

1 Introduction

Molecular simulations are widely recognised as one of the major tools used in modern biology and chemistry. These methods, implemented in variety of highly effective and scalable software packages, produced highly influential results such as Nobel 2013 insight into biomolecules dynamics and function [1–3], mechanistic investigations [4–6], works on protein and nucleic acid systems computational design [7]. Since in real cells molecules are solvated, canonic modelling protocol implies simulation of dozens to hundreds thousand of water molecules alongside the system under consideration. Normally very few information about molecule solvation is available from the start, and is usually acquired with the initial model from PDB database. This issue is then resolved by algorithmic water addition into the system, which is known to be straightforward and

always requires manual validation afterwards. Such procedure for average-sized model can take hours and is prone to errors due to being highly attention-requiring. On the other hand, algorithmic water addition is known to produce highly incorrect solvation pattern, for example placing water molecule in the center of a deeply buried hydrophobic core (see Fig. 1) or disrupting secondary structure (see Fig. 2). In some cases this problem can be solved by removing all newly placed waters inside model while keeping ones taken from crystal structure - but this approach is only applicable when no changes to this structure were made. However, this approach is not applicable when changes, for example substitutions or alternative loop conformations were introduced into the initial structure or it was acquired from low-resolution model, as a result of homology modelling procedure of other design approaches. Moreover, discarding all internal water molecules can result in totally erroneous model due to common abundance of so-called "structural" waters which are kept in specific conserved positions inside protein of nucleic acid structure and are crucial for correct conformation and dynamical behaviour of the system [8–11]. To date there is a plethora of methods that allow for scoring of co-crystallized water or prediction of putative water-binding sites [12], but none of them deals with solvation issues introduced by MD software. Thus, an automatic method to score and filter solvent water is definitely required - both to alleviate the need for manual assessment, this allowing for more systems to be dealt with simultaneously, and to improve the quality of solvation of models for which no initial information about water placement is available.

Here we present general approach to such filtering and validate it under Gromacs solvation scheme both for its ability to correctly retain co-crystallized water molecules in PDB entries and improve the overall molecular simulation quality.



Fig. 1. Misplaced water molecule disrupts hydrophobic core.



Fig. 2. Misplaced water molecule disrupts secondary structure.

2 Materials and Methods

All calculations presented were executed with in-house scripts in Python 2.7 with the help of Jupyter notebook environment [13] and a bunch of side packages: numpy [14], ProDy [15], and pmx [16]. Water addition and molecular simulations were performed with Gromacs 2018.1. We used parm99sb-ildn force field with corrections [17] to describe simulated systems. TIP3P water model was used, and the total charge was neutralized with Na⁺ or Cl⁻ ions. Water and ions were equilibrated around the protein by carrying out a 100-ps NVT simulation with position restraints, 100-ps NPT simulation with position restraints and then 50-ns unrestrained simulation. We utilized Parrinello-Rahman barostat and temperature coupling with Velocity Rescale. RMSD graphs were built based on coordinates of all heavy atoms of the protein and only of catalytic triad (Ser¹⁹⁸, His⁴³⁸, Glu³²⁵) heavy atoms.

All images of molecular systems were prepared with the help of PyMol software. Graphs were built with the help of matplotlib python package [18].

3 Results and Discussion

3.1 Method description

To score the correctness of added water molecules placement we developed a new approach based on existing structural data. We took all PDB structures with X-ray resolution better then 1.5Å filtered at 95% sequence similarity level yielding overall training set of 4 352 entries. For each of 2 090 407 water molecule oxygen atoms from

this set we extracted coordination parameters – distance and angle – relative to each surrounding biomolecular polar atom type (see Table 1). This information was circularly expanded to account for unobserved behaviour of protons in X-ray structures and used to finally construct 3D grid of water molecule position possibilities relative to each polar atom type and its pivot neighbour in structural element.

Table 1. Types of atoms analyzed in current study.

Element	Donor types	Acceptor types
Ν	N, NE, NH1, NH2, NZ, ND1, ND2, NE1, NE2 (HIS), NE2 (GLN)	ND1, NE2 (HIS)
0	OG, OG1, OH	O, OD1, OD2, OE1, OE2, OG, OG1, OH
S	SG	SG

For each water molecule being scored a set of polar atoms not farther then 5Å from it is first selected, and then for each such atom a corresponding position probability matrix is obtained. All these grids are then summed up and final superposed probability profile is then adjusted to account for VdW radii of surrounding atoms (fig. 3A). Final water molecule score is defined as the maximum value of a grid sub-sphere with center at the oxygen atom position and radius of 1Å (fig. 3B).



Fig. 3. Example of positional probability profile for one water molecule. A. Initial matrix. B. Scoring sub-sphere.

Since such procedure is time-consuming several rounds of pre-filtering are designed as well to reduce the number of water molecules to score. First of all we exclude from consideration all water molecules farther than 3.5Å from any non-water atom. Then we retain only those water molecules that have no more than 1 neighboring water thus excluding external hydration sphere of the system. Then if such water does not have at least one non-water polar atom in 3.5Å proximity it is explicitly removed from system as presumably placed inside hydrophobic core. All remaining water molecules are then scored as described before and either removed or kept in the system based on their score and specified threshold.

3.2 Correct detection of structural water molecules

For method to be reliable we need it to properly mark water molecules that have been placed into positions that are occupied by water in known X-ray structures. For this purpose we constructed test set of water molecules placed with Gromacs default solvation scheme into preliminary "dried" PDB entries not used in parameter extraction. These added water molecules were filtered as described and then labeled as "structural" if in initial PDB structure there is a water molecule oxygen in 0.5A proximity to the position of oxygen atom of molecule being labeled. Test set constructed this way consisted of 1335 water molecules. We then used it to validate the applicability of our scoring approach in predicting the correct class of added water molecule. We constructed ROC-curve (see fig. 4) which has an area under curve of 0.78 thus showing adequate predictive power.



Fig. 4. ROC-curve for reported scoring method

We note that actual performance of such classifier is expected to be higher due to the fact that not all "structural" waters are successfully resolved in X-ray structures.

3.3 Impact on quality of molecular dynamics simulations

Deleterious effects of water misplacement are naturally expected to be manifested in dynamics. Simulated can in principle make system properly solvated but each badly misplaced water molecule naturally makes proper equilibration last longer to reach desired RMSD convergence and has a potential to even disrupt the system and misguide the whole computational experiment. To test our filtering approach we conducted molecular dynamics simulations of the same system with different solvation setups. We chose

well-studied enzyme Butyrylcholinesterase and designed 5 its variations concerning solvation (see Table 2).

#	Crystallographic water	Added water
1	All molecules kept	Manually picked molecules kept
2	All molecules kept	Best scored molecules kept (94 discarded)
3	All molecules kept	All molecules kept
4	All molecules discarded	Best scored molecules kept (69 discarded)
5	All molecules discarded	All molecules kept

Table 2. Solvation setups for Butyrylcholinesterase system.

We then analyzed the behaviour of each system in terms of overall equilibration (see Fig. 5) and retention of initial catalytic site geometry (Fig. 6) crucial for the enzyme's function.

Systems 3 and 5 that both have all placed waters retained showed distinctively worse system behaviour in the course of 50-ns equilibration. They demonstrate dramatic initial displacement with RMSD of up to 0.18 nm, which only continues to grow, while for three other systems with corrected water placement RMSD shows emerging trend to stabilization. This data shows that while overall enzyme architecture is rigid enough not to be disrupted by misplacement of several dozens of water molecules, such imperfections of the starting model may have significant impact on time needed for equilibration run to converge under desirable threshold.



Fig. 5. All heavy-atom RMSD. Data shown in thin lines was smoothed by averaging with 1-ns wide sliding window, in thick lines – with 10-ns one to show general trend. Systems' setup described in Table 2.



Fig. 6. Retention of initial catalytic site geometry. Data shown in thin lines was smoothed by averaging with 1-ns wide sliding window, in thick lines – with 10-ns one to show general trend. Systems' setup described in Table 2.

4 Conclusion

Presented study describes general approach to correct automatized water addition implemented in modern MD software. While demonstrating its ability to keep important co-crystallized water molecules intact, our method, by deleting potentially misplaced ones, also improved the quality of molecular dynamics simulations. Once implemented in a handy tool presented approach may become part of canonic protocol of system setup for any molecular dynamics calculations.

5 Acknowledgements

Development of grid techniques and software in this study was supported by Ministry of Science and Education (RFMEF157617X0095).

6 References

1. Karplus M. Development of Multiscale Models for Complex Chemical Systems: From H H2to Biomolecules (Nobel Lecture). Angew Chem Int Ed. 2014;53:9992–10005.

- Levitt M. Birth and future of multiscale modeling for macromolecular systems (Nobel Lecture). Angew Chem Int Ed Engl. 2014;53:10006–18.
- Warshel A. Multiscale modeling of biological functions: from enzymes to molecular machines (Nobel Lecture). Angew Chem Int Ed Engl. 2014;53:10020–31.
- Piana S, Lindorff-Larsen K, Shaw DE. Atomic-level description of ubiquitin folding. Proc Natl Acad Sci U S A. 2013;110:5915–20.
- 5. Yin Y, Yang L, Zheng G, Gu C, Yi C, He C, et al. Dynamics of spontaneous flipping of a mismatched base in DNA duplex. Proc Natl Acad Sci U S A. 2014;111:8043–8.
- Luk LYP, Javier Ruiz-Pernía J, Dawson WM, Roca M, Loveridge EJ, Glowacki DR, et al. Unraveling the role of protein dynamics in dihydrofolate reductase catalysis. Proc Natl Acad Sci U S A. 2013;110:16344–9.
- Smirnov IV, Golovin AV, Chatziefthimiou SD, Stepanova AV, Peng Y, Zolotareva OI, et al. Robotic QM/MM-driven maturation of antibody combining sites. Sci Adv. 2016;2:e1501695.
- 8. Levinson NM, Boxer SG. A conserved water-mediated hydrogen bond network defines bosutinib's kinase selectivity. Nat Chem Biol. 2013;10:127–32.
- Wei D, Parkinson GN, Reszka AP, Neidle S. Crystal structure of a c-kit promoter quadruplex reveals the structural role of metal ions and water molecules in maintaining loop conformation. Nucleic Acids Res. 2012;40:4691–700.
- Maciag JJ, Mackenzie SH, Tucker MB, Schipper JL, Swartz P, Clark AC. Tunable allosteric library of caspase-3 identifies coupling between conserved water molecules and conformational selection. Proc Natl Acad Sci U S A. 2016;113:E6080–8.
- 11. Wei D, Wilson WD, Neidle S. Small-molecule binding to the DNA minor groove is mediated by a conserved water cluster. J Am Chem Soc. 2013;135:1369–77.
- Nittinger E, Flachsenberg F, Bietz S, Lange G, Klein R, Rarey M. Placement of Water Molecules in Protein Structures: From Large-Scale Evaluations to Single-Case Examples. J Chem Inf Model. 2018. doi:10.1021/acs.jcim.8b00271.
- Perez F, Granger BE. IPython: A System for Interactive Scientific Computing. Comput Sci Eng. 2007;9:21–9.
- 14. Oliphant TE. Python for Scientific Computing. Comput Sci Eng. 2007;9:10-20.
- 15. Bakan A, Meireles LM, Bahar I. ProDy: protein dynamics inferred from theory and experiments. Bioinformatics. 2011;27:1575–7.
- 16. Gapsys V, Michielssens S, Seeliger D, de Groot BL. pmx: Automated protein structure and topology generation for alchemical perturbations. J Comput Chem. 2015;36:348–54.
- 17. Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, et al. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX. 2015;1-2:19–25.
- Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, Dror RO, et al. Improved side-chain torsion potentials for the Amber ff99SB protein force field. Proteins. 2010;78:1950–8.
- 19. Hunter JD. Matplotlib: A 2D Graphics Environment. Comput Sci Eng. 2007;9:90-5.

Архитектура промоторов планарии Dugesia japonica

Reza Bagherzadeh^{1,2,3,5}, Elena Minkina⁷, Ruslan Deviatiiarov⁷, Ali Sharifi-Zarchi^{3,4}, Oleg Gusev^{6,7}, Hossein Bahrvand^{3,5}, Kiyokazu Agata^{1,2}

¹ Department of Biophysics, Kyoto University, Kyoto, Japan ² Department of Life Science, Gakushuin University, Tokyo, Japan

³ Department of Stem Cells and Developmental Biology, Cell Science Research Center,

Royan Institute for Stem Cell Biology and Technology, ACECR, Tehran, Iran

⁴ Department of Computer Engineering, Sharif University of Technology, Tehran, Iran ⁵ Department of Developmental Biology, University of Science and Culture, Tehran,

Iran

⁶ Division of Genomic Technologies, RIKEN Center for Life Science Technologies (CLST), Yokohama, Japan

⁶ Extreme Biology Laboratory, Kazan Federal University, Russia * rezabagherzadeh.email@gmail.com

Аннотация. Взрослые стволовые клетки являются недифференцированными клетками и необходимы для регенерации тканей. Регуляция транскрипции и конформационные изменения хроматина являются фундаментальными механизмами, учавствующими в поддержании плюрипотентности и дифференцировки клеток. Однако для популярных животных моделей, используемых в исследованиях регенерации, полногеномная характеристика промоторов на данный момент отсутствует. Данная работа ставит своей целью определение промоторов у планарии *Dugesia japonica*, являющейся модельным организмом во многих исследованиях механизмов регенерации.

Ключевые слова: регенерация, стволовые клетки, *Dugesia japonica, CAGE*.

Транскрипция и изменение конформации хроматина являются фундаментальным механизмом, который поддерживает плюрипотентность, дифференцировку и установление специфичных для клеточного типа профилей экспрессии генов. Непаразитические плоские черви, такие как *Dugesia japonica*, обладают большим количеством взрослых стволовых клеток, необластов [1], что делает планарий популярной моделью для изучения механизмов регенерации. Подобные исследования зависят от точной аннотации сайтов начала транскрипции. Тем не менее, всесторонние полногеномные исследования регуляторных элементов генома в контексте регенеративных способностей планарий на данный момент отсутствуют. Целью работы была характеристика цис-регляторного ландшафта у Dugesia japonica, с использованием кэп-анализа экспрессии генов (CAGE). Данные CAGE были получены из PHK двух популяций клеток Dugesia japonica (пролиферативные клетки X1, постмитотические дифференцированные клетки Xis), выделенных с использоваанием метода сортировки клеток с активированной флуоресценцией (fluorescence-activated cell sorting, FACS) [2]. Очистка от pPHK и картирование ридов на геном D. japonica были проведены с помощью программ TagDust2 и bwa, сайты старта тэгов CAGE (CTSS) были определены с использованием скриптов RIKEN Promoter Pipline [3]. Консенсусные регионы инициации транскрипции определялись с помощью пакета R CAGEr. Поиск дифференциально экспрессируемых промоторов был выполнен с помощью пакета R edgeR. Поиск мотивов был выполнен с использованием MEME Suite tool.

Идентифицированные промотеры были использованы для создания первого в своем роде Атласа промотеров планарий. Анализ формы промотеров и профилей экспрессии выявил и некоторые особенности регуляторных участков и случаи альтернативного использования планарных промоторов. Сравнение активности промоторов в вышеупомянутых клеточных популяциях позволило идентифицировать группы дифференциально экспрессируемых генов и регуляторные мотивы, дающие представление ою особенностях регуляции и дифференцировки стволовых клеток. Атлас промоутеров *Dugesia japonica* представляет собой ценный ресурс для сравнения цис-регуляторных регионов у многоклеточных, а также источник фундаментальных данных для исследований транскрипции, генной регуляции и эпигенетики.

Список литературы

- Fraguas S., Umesono Y., Agata K., Cebrià F. Analyzing pERK Activation During Planarian Regeneration. In: Jimenez G. (eds) ERK Signaling. Methods in Molecular Biology, vol 1487. Humana Press, New York, NY (2017)
- Hayashi T., Agata K. A Subtractive FACS Method for Isolation of Planarian Stem Cells and Neural Cells. In: Rink J. (eds) Planarian Regeneration. Methods in Molecular Biology, vol 1774. Humana Press, New York, NY (2018)
- Forrest ARR, et al. A promoter-level mammalian expression atlas. Nature 507, 462–470 (2014). doi: 10.1038/nature13182.

Phenobiome approach enabling genomics-based prediction and comparative analysis of metabolic phenotypes in human gut microbiome

German A. Ashniev, Stanislav N. Iablokov, Aleksandr A. Arzamasov, Matvei S. Khoroshkin,

Pavel S. Novichkov, Andrei L. Osterman, Dmitry A. Rodionov

Institute for Information Transmission Problems, Russian Academy of Sciences, Moscow, Russia

Abstract

Human microbiome is a set of all symbiotic, commensal and parasitic microorganisms that inhabit the human body. The total number of bacterial cells in the human body can vary from 10 to 100 trillion pcs, and the majority of microbiota dwells in the gastrointestinal tract (GIT). The GIT microbiota is extraordinarily important for human health. One of the most important functions of gut microbiota organisms is metabolism of complex polymers, polysaccharides, and peptides, as well as synthesis of short-chain fatty acids, an important source of energy for colonocytes. The importance of GIT microbiota for physiology and medicine indicates a strong need for a detailed study of its representatives with the methods of modern biology. However, the presence of a large number of different bacteria living in the human GIT makes it almost impossible to experimentally study all members of the microbiota. At the same time, the existence of large amounts of genomic data allows using of computational methods of massive genomic analysis. Comparative genomic analysis provide advantage for bioinformatics reconstruction of metabolic pathways and regulons in bacterial genomes.

We have developed the Phenobiome approach enabling predictive phenotype profiling of human gut communities based on 16S rRNA sequencing data. It includes three essential components: (i) full-genome reconstruction of metabolism of GIT bacteria with complete genomes (dependence on vitamins, capability to digest sugars and polysaccharides); (ii) building reference database of genes, metabolic pathways and phenotypes of GIT microbiota; and (iii) development of bioinformatics algorithm for automatic prediction of cumulative phenotypic properties of human stool samples using either 16S rRNA or whole metagenome sequencing data. We performed genomic analysis and reconstruction of pathways for metabolism of carbohydrates, amino acids, vitamins and other biologically active compounds produced by human microbiota in ~2,300 bacterial genomes representing GIT microbiota. As result, we created a database of metabolic genes and pathway variants, as well as phenotypic rules and binary phenotypes for over 70 metabolic pathways.

We further developed the Phenobiome bioinformatics pipeline for functional analysis of human GIT microbiota samples using their taxonomic abundance profiles. The Phenobiome pipeline can also start from raw 16S rRNA sequencing results to calculate taxonomic abundance profiles for each GIT sample using the QIIME2 package and the RDP database of reference 16S sequences. The obtained taxonomic profiles are mapped to the analyzed reference genomes with predicted binary metabolic phenotypes at three taxonomic levels (species, genus, family). Using the binary phenotype matrix and the results of weighted mapping procedure, the Phenobiome profiler calculate a set of community phenotype indices (CPIs) for each sample. The obtained CPI values give probabilistic estimate of a fraction of cells in the community with specific metabolic phenotypes reflecting major nutrient requirements and metabolic capabilities of the analyzed microbiome samples.

The Phenobiome pipeline was applied for analysis and functional classification of a large set of stool samples from Human Microbiome Project (HMP) and American Gut Project (AGP). Application of the Phenobiome approach to large human gut microbiota datasets will enable correlation analysis of obtained phenotypic indices with available metadata (geography, diet, disease) and biochemical screening results in large populations. The results of this project will be useful for disease prevention via diet correction and the usage of vitamins and other dietary supplements.

Aknowledgments

This research was supported by the Russian Science Foundation (grant #14-14-00289).

Novel approach for structure-based prediction of zinc binding sites in biomolecules

M. F. Buyanova¹, A. O. Zalevsky^{1,2,3} and A. V. Golovin^{1,2,3}

 ¹ Faculty of Bioengineering and Bioinformatics, Lomonosov Moscow State University, Moscow, Russia
² Sechenov First Moscow State Medical University, Institute of Molecular Medicine, Moscow, Russia
³ Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, Moscow, Russia

emkeller@yandex.ru

Abstract. Among all the metalloproteins deposited in the Protein Data Bank [1] zinc-binding proteins are the most abundant. Zn^{2+} cations are crucial for protein function as they are involved in enzyme catalysis, help maintain the structure of a protein and can play a regulatory role. Therefore, an accurate prediction of zinc binding sites may be useful in context of functional inference. However, available analytical procedures for detecting the presence of zinc in biomolecules are greatly limited due to its electron configuration and isotope instability. That's why a huge variety of computational approaches have been proposed to address this task. But the majority of such prediction tools have only been described in papers and either are not longer maintained or not publicly available. Here, we present SiteFinder, a novel computational approach to predicting zinc coordination spheres based solely on a structure of a given protein. It combines basic graph theory and geometry to determine the possible coordinates of Zn^{2+} cations. The proposed tool is open-sourced and available in both CLI format and a web-server (https://vsb.fbb.msu.ru/finder/).

Keywords: Zinc-binding, Metal Binding Sites, Structural Bioinformatics

1 Introduction

Zinc is the most abundant metal coordinated by proteins and it has been shown to be involved in numerous biological processes. As a part of a protein, Zn^{2+} cations possess a multifunctional role. The two most important of its functions are (1) structural and (2) catalytic.

The first relates to the case when zinc is required for the correct folding of polypeptide chain. For example, the largest class of transcription factors contain special zinc finger domain whose stability is maintained only in the presence of Zn^{2+} ions. Furthermore, when located at the surface of a protein, zinc may influence the formation of quaternary

structures. This effect can be illustrated by zinc-induced aggregation of amyloid- β , which is considered to be the key mechanism underlying the development of Alzheimer's disease [2].

The catalytic function in turn is performed when enzyme molecules contain Zn^{2+} cations (usually in the active site region) which influence various reactions, both directly and indirectly. The most common known members of zinc-binding enzymes include alcohol dehydrogenase, alkaline phosphatase, organophosphate hydrolase and superoxide dismutase. The search for such proteins led to the discovery of zinc-binding proteins in all six IUBMB classes, which renders its crucial importance for catalysis.

The reason of such prevalence of Zn^{2+} in enzyme structures stems from its completely filled d-orbitals (electron configuration of $[Ar]3d^{10}$). Zinc cation exists in redox-inert state and thus does not take part in reduction–oxidation reactions. Instead, it is capable of accepting an electron pair. This property makes Zn^{2+} a perfect cofactor for catalytic processes which require redox-inert Lewis acid.

However, due to such electron configuration zinc cation is diamagnetic and thus invisible for detecting with EPR method. In addition, out of five known zinc isotopes only ⁶⁷Zn is considered stable. But NMR technique becomes practically infeasible because this isotope requires special experimental conditions which do not allow for analysis of the whole protein. Moreover, zinc complexes have no absorbance in ultraviolet, visible and microwave regions of spectrum [3].

All the described difficulties constitute obstacles for applying analytic methods to detect Zn^{2+} cations in protein structures. However, the need for zinc detection is still there because an accurate prediction can be useful not only to infer the function of the protein of interest, but also to design one with desired properties. This is the reason why a plethora of computational approaches have been recently developed to address this task.

Existing *in silico* zinc binding prediction methods generally fall into two categories: (1) based on amino acid sequence information and (2) three-dimensional structure of a protein. Most of them make use of machine learning techniques such as SVM (Support Vector Machine) [4, 5], Random Forest [6, 7, 8], clustering algorithms [6] and neural networks [9]. However, these sophisticated predictors are either not open-sourced or do not have currently working web-servers.

Here, we conduct coordination spheres analysis in available protein structures from PDB and propose a new tool capable of not only predicting coordinating residues, but also the possible coordinates of zinc cation in the given structure. Moreover, we make the realization of our method publicly available.

2 Materials and methods

Both the analysis and the SiteFinder predictor tool development was performed in Python 2.7.15 with the following packages:

- 1. ProDy for manipulating structural PDB data;
- 2. matplotlib for visualization purposes and data analysis;
- 3. NumPy and pandas for performing efficient operations on matrices and other high-level data structures;
- 4. SciPy for solving optimization tasks;
- 5. SymPy for dealing with geometric objects;
- 6. NetworkX for handling graph structures.

3 Results and discussion

Figure 1 presents an outline of the proposed SiteFinder method discussed in detail in subsequent sections.



Schematic pipeline of the present work

3.1 Zinc coordination spheres analysis

Protein Data Bank, as of July 20, 2018, was queried for protein structures known to contain zinc which also had an acceptable resolution of 2Å or less. At the time of accession there were 6398 matching entries.

The data was split into two parts: training and test sets of sizes 5120 and 1278 structures respectively. Only training test was used for estimating typical geometric properties of amino acid coordination environments. The possible range of coordinators was defined as a list of the following types of atoms: SG, ND1, NE2, OD1, OD2, OE1, OE2, OG, OG1, OH and backbone oxygen (O). An atom was considered to be coordinating if it was closer to the zinc ion than 3\AA . This distance threshold corresponds to the maximal first-shell coordination radius for Zn^{2+} cation.



4

example of zinc cation coordinated by backbone oxygens (PDB entry 3M2R)

The most common coordination number (CN) was found to be four, which is consistent with previous works. However, the cases with only three amino acid coordinators were also viewed as worthy of consideration due to their special role in enzymes. In such cases the fourth "missing" coordination site is likely to be occupied by oxygen from H2O which can be easily displaced by another ligand in the catalytic process [3].

We also analyzed the contributions of different coordinating amino acid ligands to the binding of Zn^{2+} . Cysteine (Cys, C), histidine (His, H), glutamate (Glu, E) and aspartate (Asp, D) have been found to dominate coordination spheres. This observation is also consistent with earlier surveys and, furthermore, some of the previously reported prediction algorithms are even strictly limited to this "CHED" set of possible coordinators [6]. Our approach, however, does not use this assumption. It is also worth noting that in fact every amino acid can potentially serve as a ligand due to the presence of oxygen atom in the backbone (O). This case of so called backbone coordination is relatively rare but nevertheless there are examples of such binding sites with only O atoms as coordinating (see Fig. 2).

The distributions of observed zinc-to-ligand distances were built for each atom type separately and then subjected to background noise removal procedure. This step involved detection of histogram peaks and plotting a Gaussian curve which would best fit an initial distribution in a specified area around the peak. Such technique allowed for excluding distance information from "noisy" atoms which happened to be located closer than 3Å to zinc just by accident and thus became misinterpreted as coordinating. The

Atom name	Zn-to-ligand distance (mean \pm 1 standard deviation), Å	Count
SG	2.33 ± 0.05	13567
NE2	2.07 ± 0.06	11020
ND1	2.08 ± 0.07	5064
OD2	2.03 ± 0.11	3565
OD1	2.05 ± 0.12	3083
OE2	2.04 ± 0.12	2579
OE1	2.06 ± 0.13	2501
OG	2.15 ± 0.16	75
OG1	2.1 ± 0.15	63
ОН	1.69 ± 0.72	46

amounts, mean and standard deviation parameters of the distributions for different amino acid atoms are summarized in Table 1.

Table 1. Observed zn-to-ligand distances.

Further analysis was applied to ligand pairs in known coordination spheres. Here, an atom from 3Å sphere around Zn²⁺ was considered to be coordinating if its actual zinc-to-ligand distance fell within three standard deviations from the mean according to the basic values estimated at the previous step. During this stage the minimal and the maximal limits of distance values within each type of coordinating atom pairs were obtained. Pairs with extremely low number of observations (less than 100) were not taken into account and their typical ligand-to-ligand range of distance was defined as "from 1.29Å to 5.24Å", which are the extreme values for all the observed ligand-to-ligand distances in general.

The obtained parameters later served as a basis for the proposed rule-based hand-crafted SiteFinder method described in details the next subsection.

3.2 **Prediction algorithm**

Given a structure of a protein of interest, the algorithm captures all the possible coordinating atoms which then are treated as vertices of an undirected graph. The edge is created if the corresponding pair of atoms meets the condition that the distance between them falls within the ideal range found during the analysis stage.

Then SiteFinder searches for all the maximal cliques in this graph. These sets of vertices represent the prototypes of zinc binding sites since all the vertices can serve as ligands and are all connected by edges, meaning that the distance constraints are met.

The filtration of cliques is then performed to leave only such sets which contain 3 or 4 vertices. If the size is 5 and larger, four vertices are chosen according to how commonly the corresponding atom types are observed in known zinc coordination spheres.

The next stage is trilateration, which is applied to every set of atoms that make up a clique. The algorithm tries to place Zn^{2+} cation in such a point of space where zinc-to-ligand distances within the sphere would be as close as possible to the ideal ones. This is accomplished through Levenberg-Marquardt algorithm for solving non-linear least squares problems. It gives not only the coordinates of a point to place zinc in, but also a vector of residuals, which euclidean norm is used as an initial estimate of prediction reliability. Results with values of this estimate more than 1.5 are considered unacceptable.

It's worth noting that least squares optimization provides only one point as the result. However, in cases with CN = 3 both the result and its reflection (in the plane determined by three ligand atoms) would have the same quality. This observation is used if there is not enough space around the original point to place zinc cation in.

At the final stage a special scoring technique is applied to rank all the predictions. Every atom in a coordination sphere is assigned a score equal to the value of the probability density function of a normal distribution at a given point (observed zn-to-ligand distance). Gaussian parameters are those calculated during the analysis of typical zinc-to-ligand distances. The distributions are all additionally scaled to make peak values get a score of 100. The score of a certain prediction is simply the mean value of all the scores of atoms in a predicted sphere. One can adjust score threshold parameter to get more or less reliable zinc predictions.

As the result of its work SiteFinder outputs (1) a PDB file with the structure of an input protein with predicted Zn^{2+} cations and (2) a log file with a list of binding sites and assigned scores.

3.3 Performance evaluation

The performance of the algorithm was evaluated on the structures from the test set. PDB files were prepared in the following manner. At first, zinc environments (i.e. every residue with a coordinating atom located closer than 3\AA to the Zn2+ cation) were extracted from the whole protein structure. Next, a special filter was applied to remove unpredictable zincs from the testing set. Unpredictability in this case is defined as the presence of less than 3 coordinating atoms of allowed types (see Subsection 3.1).

The performance metric used is simply the ratio of predictions which fall within 1.5Å distance from the original position of zinc. This value can also be referred to as accuracy score.

The prepared test set contained 2375 zinc ions to predict. Our algorithm achieved 83% accuracy on the test set with 1970 positions correctly identified.

Another widely known metric is AUC-ROC (area under receiver operating characteristic curve) which reflects the diagnostic ability of a predictor with varying thresholds. The reason why this approach was not applied originates from the nature of our task which can not be viewed as a typical binary classification. The problem lies in defining false positive predictions. Even if the original structure lacks zinc nearby the

predicted coordinates, there is still a chance it was simply not identified by X-ray. Moreover, due to the dynamic manner of binding process there can be multiple sites capable of binding zinc in theory, but in practice they may almost always be free from the cation as they are for some reason less preferable.

What is even more important in the context of evaluating performance, majority of previously reported approaches do not predict coordinates of Zn^{2+} cations and are limited to identifying just the residues involved in binding. The issue discussed above makes any comparison technically and ideologically incorrect.

4 Conclusion

The key results of the present work can be summarized as follows:

- 1. Currently available zinc metalloproteins present in the Protein Data Bank have been analyzed. Key properties of Zn2+ binding sites, such as coordination number and amino acid composition, turned out to be consistent with the previously reported ones.
- 2. A special pipeline has been developed to deal with distance distributions which has made it possible to obtain the main parameters (mean values and standard deviations). These values have been exploited for both coordinate prediction and scoring steps.
- 3. A novel structure-based approach SiteFinder has been proposed which addresses the task of predicting zinc binding sites. This method is a combination of basic graph theory concepts and geometric manipulations performed on the three-dimensional structures. SiteFinder outputs the initial protein structure with an addition of predicted zinc cations and assigns scores to each of them. The scoring technique makes use of values obtained at the analysis stage. These scores can render the level of confidence in a certain prediction.
- 4. The accuracy score of our algorithm reached 83% on the test set.
- 5. SiteFinder is a publicly available tool for researchers and it is implemented in both the command-line interface format and the on-line web-server (https://vsb.fbb.msu.ru/finder/).

References

- 1. RCSB PDB: Homepage, https://www.rcsb.org/, last accessed 2018/07/29.
- 2. Miller, Y., Ma, B., Nussinov, R.: Zinc ions promote Alzheimer A β aggregation via population shift of polymorphic states. Proc Natl Acad Sci U S A 107(21), 9490–9495 (2010).
- Mikko, L., Jarkko, V., Janne, J.: Zinc Coordination Spheres in Protein Structures. Inorganic Chemistry 52, 10983–10991 (2013).

- 4. Lin, H.: Prediction of the functional class of metal-binding proteins from sequence derived physicochemical properties by support vector machine approach. BMC Bioinformatics 7(Suppl. 5): S13 (2006).
- 5. Passerini, A., Andreini, C., Menchetti, S.: Predicting zinc binding at the proteome level. BMC bioinformatics 8(1), 39 (2007).
- Zhen, C., Wang, Y., Zhai, Y., Song, J., Zhang, Z.: ZincExplorer: an accurate hybrid method to improve the prediction of zinc-binding sites from protein sequences. Mol. BioSyst. 9, 2213-22 (2013).
- 7. Bordner, A.: Predicting small ligand binding sites in proteins using backbone structure. Bioinformatics 24, 2865–2871 (2008).
- Zheng, C., Wang, M., Takemoto, K., Akutsu, T., Zhang, Z., Jiangning, S.: An Integrative Computational Framework Based on a Two-Step Random Forest Algorithm Improves Prediction of Zinc-Binding Sites in Proteins. PLoS One 7(11), e49716 (2012).
- 9. Lin, H.: Protein metal binding residue prediction based on neural networks. Int. J. Neural Syst. 15, 71–84 (2005).

Предсказание субстратной специфичности протеаз методами молекулярного моделирования

Елизарова Евгения, Залевский Артур

Московский Государственный Университет им. М.В.Ломоносова zheneliz147@gmail.com

Аннотация. В данном исследовании, на примере сериновой протеазы тромбина человека и его взаимодействия с хромогенным субстратом S2238, мы предложили и протестировали вычислительный подход предсказания субстратной специфичности протеаз, основанный на методах молекулярного моделирования.

Ключевые слова: субстратная специфичность, тромбин, молекулярное моделирование

1 Введение

Протеазы представляют собой важный класс ферментов, участвующих в различных биохимических процессах. Данные о субстратной специфичности фермента играют ключевую роль в создании и разработке ингибиторов и различных препаратов, основанных на действии фермента. Более того, с развитием геномных методов получения различных последовательностей белков, было обнаружено огромное количество протеаз, особенно у растений, с неизвестной функцией. Однако определение субстратной специфичности экспериментальными методами является трудоемким и ресурсозатратным процессом, так как необходимо провести скрининг большого количества потенциальных пептидных субстратов. Таким образом, изучение субстратной специфичности различных классов ферментов становится актуальным и важным предметом для исследований. В рамках данной работы была поставлена следующая цель: разработать подход предсказания субстратной специфичности протеаз

2 Субстратная специфичность тромбина

Реакция протеолиза пептидного субстрата осуществляется посредством расщепления пептидной связи между остатками, расположенными в Р1 и Р1' позициях.[1] Для представителей семейства сериновых протеаз область кармана вокруг Р1 сайта состоит из полярных и частично гидрофобных аминокислотных остатков боковых цепей (Ser217, Thr184 и Tyr219). Это позволяет оптимально связывать не только заряженные, но и гидрофобные остатки в составе субстрата, такие как лейцин, тирозин. [2] Кроме того, глутамин и тирозин, также могут взаимодействовать с образованием водородных связей с полярными аминокислотными остатками на дне кармана с сохранением гидрофобных взаимодействий с остатками боковых цепей. Это обеспечивает оптимальное расположение субстрата, и тем самым способствует протеканию реакции расщепления.

Исследования показали, что в положении P1 пептидного субстрата наиболее вероятно располагается аргинин, который образует солевой мостик с Arg189, который в свою очередь расположен в глубине кармана специфичности. Гидрофобный карман S2, сформированный остатками 60й петли: Туг60А, Pro60B, Pro60C, Trp60D имеет сродство к небольшой гидрофобной боковой цепи остатка в позиции P2. В данном положении в 71% встречается пролин, остальные 29 % приходятся на алифатические аминокислоты, такие как: глицин, валин, изолейцин. Наличие алифатических остатков в P2 присутствуют в ряде природных субстратов тромбина, включая фибриноген A α и B β . Арил-связывающий сайт, образован остатками Leu99, Ile174 и Trp215, которые взаимодействуют с ароматическими и гидрофобными остатками в позиции P4, в то время, как остаток субстрата в позиции P3 обращен кнаружи от поверхности белка. В положение P1' с большой вероятностью встречаются такие аминокислотные остатки как: серин, треонин, глицин и аланин.[3]

Показано, что другие представители семейства сериновых протеаз имеют некоторые отличия в составе расщепляемого субстрата, в том числе в P1 позиции.

Ингибиторы тромбина представляют широкий интерес в связи с их применением в качестве антикоагулянтов (лекарственные средства, блокирующие ферментативные реакции, приводящие к образованию фибрина) в медицинской практике. Взаимодействия между пептидными лигандами, оказывающими ингибирующее действие также позволяет получать данные для изучения субстратной специфичности фермента. Одними из синтезированных ингибиторов в S1 сайте являются Trx-d-Cha-l-Aze-d-Tyr-l-homoArg (8-1), и его модифицированный вариант с гуанилиновым остатком на N конце.(8-5). Ингибитор 8-5 образует водородную связь между гидроксильной группой D-Туг и карбоксильным кислородом Asp 189. Из-за пространственной конформации лиганда, боковая цепь содержащая тирозин только частично проникает в карман,в отличие от положения аргинина, обычно находящегося в P1 положении субстрата. [4]

3 Результаты и обсуждение

3.1 Построение гомологичной модели

В связи, с возможным отсутствием нативных или релевантных к ним структур ферментов, появляется задача в первую очередь создать подобную модель. С помощью пакета Modeller, была построена модель гомологичного белка из последовательности бычьего тромбина PDB ID:1AVG (разрешение 2.6 Å) на основе структуры человеческого тромбина PDB ID:3DUX. При визуализации полученных результатов, были выявлены отличия в положении субстрата относительно каталитического центра между гомологичной моделью и исходной структурой человеческого тромбина, что было связано с отсутствием важного для расположения субстрата аминокислотным остатком Glu 192. Далее докинг проводился в структуру имеющий данный аминокслотный остаток.

3.2 Множественный докинг

Для получения репрезентативных данных был проведен независимый множественный докинг хромогенного субстрата в нативную и гомологичные модели. По результатам которого было выявлено, что первые 7 моделей из 20 отражают сходное расположение лиганда относительно активного центра фермента в нативной и гомологичных структурах.

3.3 Докинг трипептидов

Для определения аминокислотного состава расщепляемых субстратов были созданы 8000 уникальных трипептидов (по комбинаторным вычислениям), которые в дальнейшем использовались в качестве субстрата. По результатам их докинга в гомологичную структуру тромбина, был построен мотив, отражающий вероятность нахождения аминокислот в позициях пептидного субстрата. (Puc.1)



Рис.1 Сайт расщепления субстрата построенный по данным множественного докинга трипептидов.

Построенный паттерн выявил положение пролина и глицина в P2 позиции, аргинина и лизина в P1 сайте, а также глицина в P' позиции, что подтверждается ранее известными данными. Также было выявлено положение гистидина и тирозина в Р1 сайте, что также согласуется с ранее проведенными экспериментальными данными. Положение этих аминокислотных остатков в Р1 сайте имеет биологическую роль. Согласно литературным данным, гистидин также встречается в данном положении в структурах естественных субстратов, в том числе в структуре антигемофильного глобулина (фактора VIII). Данный фактор вырабатывается в печени, селезенке, клетках эндотелия, лейкоцитах, почках. Протеолиз, осуществляемый тромбином при взаимодействии аминокислот каталитического центра с Arg372 фактора VIII, необходим для активации кофактора. Однако у больных гемофилией показана мутация по данному положению. Вместо аргинина там располагается гистидин, цистеин или пролин. При этом показано, что замена аргинина на пролин или цистеин приводит к более тяжелой форме гемофилии, нежели гистидиновый вариант. При этом наблюдается образование тромбов, что свидетельствует о протекании реакции протеолиза, осуществляемого тромбином. Структура фактора VIII при замене аргинина на гистидин не изменяется. Для оценки изменения скорости расщепления был искусственно создан мутантный фактор VIII (R372H). Согласно литературным данным, при сравнении с контрольной группой (дикий тпи) скорость расщепления у R372H снизилась на 80%. Однако, протекание протеолиза свидетельствует о наличии взаимодействий между His372 фактора VIII и активным центром фермента.[5]

Помимо данных аминокислот обнаружено наличие триптофана в P1 положении субстрата, что можно считать результатом неточности пакета AutoDock Vina. В P1' позиции было обнаружено оптимальное расположение глицина, что полностью согласовывается с данными базы данных MEROPS. (Рис.2)



Рис.2 Сайт расщепления субстрата по данным базы MEROPS
4 Выводы

- Подобраны и оптимизированы параметры для докинга пептидных субстратов в структуру тромбина.
- Множественный докинг в гомологичную и нативную структуры выявил, что наиболее оптимально располагаются первые 7 из 20 моделей.
- В результате анализа положений хромогенного субстрата, полученных в результате независимого докинга в гомологичную и нативную структуры, было обнаружено значительное влияние GLU 192 на определение расположения субстрата, относительно каталитического центра. Это позволяет сделать вывод, о его роли в определении субстратной специфичности тромбина.
- Мотив, построенный по данным, полученным в результате докинга трипептидов, выявил положение пролина/глицина в P2, аргинина/лизина в P1 и глицина в P1' позициях, что согласуется с данными базы данных MEROPS. Наличие гистидина и тирозина в P1 положении также имеет биологическое значение. Появления триптофана в P1 позиции связано с особенностями молекулярного докинга.

Список литературы

- 1. Maike Gallwitz, Mattias Enoksson, Michael Thorpe, Lars Hellman "The Extended Cleavage Specificity of Human Thrombin .
- 2. Pratap Singh,1 * Aaron M. LeBeau,2 Hans Lilja,3 Samuel R. Denmeade,1,4 and John T. Isaacs1 "Molecular insights into substrate specificity of prostate specific antigen through structural modeling".
- Le Bonniec BF¹, Myles T, Johnson T, Knight CG, Tapparelli C, Stone SR. "Characterization of the P2' and P3' specificities of thrombin using fluorescence-quenched substrates and mapping of the subsites by mutagenesis."Biochemistry. 1996 Jun 4;35(22):7114-22
- 4. Rainer Friedrich a , Daniel Riester b , Peter Gottig a , Marcel Thurk c , Andreas Schwienhorst b , Wolfram Bode "Structure of a novel thrombin inhibitor with an uncharged D-amino acid as P1 residue" European Journal of Medicinal Chemistry 43 (2008)
- 5. Keiji Nogami, Qian Zhou, Hironao Wakabayashi, and Philip J. Fay "Thrombin-catalyzed activation of factor VIII with His substituted for Arg372 at the P₁ site" 2005 Feb 10

Transcriptional response and structural features of chicken embryo promoterome under cold torpor

Deviatiiarov Ruslan1 and Gusev Oleg2

¹ Kazan Federal University, Kazan, 420021, Russia
² RIKEN-KFU Translational Genomics Unit, RIKEN Innovation Center, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan devyatiyarovrm@stud.kpfu.ru

Abstract. Molecular mechanisms of induced hypometabolic state of the organism are still hidden from researchers in a broad range of species from very rare and exotic cases to unexpectedly general ones. Inducible cold torpor is an obligate adaptation of reptiles, which leave their eggs to fate and helps to survive decrease of environmental temperature. The adaptation was partially mixed with parental care in birds, and finally eliminated in mammals by evolution. Right in between of human and reptiles there is a convenient model of chicken embryo useful for studying developmental processes in amniotes because of its availability on all stages from egg lying to hatching. Therefore, in our study we combined advantages of the chicken model with a frontier transcriptomics approaches as cap-analysis of gene expression (CAGE) in a goal to identify key pathways and networks responsible for cold torpor adaptation in the amniote egg.

Keywords: Cold Torpor, Chicken, Promoterome.

1 Introduction

Hibernation as a kind of hypometabolism a feature of heterothermic animals with consistent body temperature when active and unstable at rest. There are plenty of examples, starting from winter time torpor of dormouse (*G. glis*), daily torpor of fruit bats (*M. minimus*), short night time torpor of some bird species. The bird's eggs in general have characteristics similar to poikilotherm animals, because its metabolic rate, heart beats and blood circulation ratios are not enough to keep temperature stable even if ambient temperature drops on 2° C, including late developmental stages. Therefore, environmental changes of temperature have negative effect on embryos, its body temperature, metabolic rate, blood pH etc [1]. The amniotic fluid supposed to participate as a buffer but does not affect temperature changes according to van 't Hoff equation, supposing presence of specific regulatory mechanisms. The torpid state of adult birds is followed by nervous system, while its final differentiation and activation occurs right before hatching, excluding a possibility of neural regulation in eggs. Regulatory network of this process still remains unclear, but seems to include an activity of particular genes, promoters and related transcription factors [2].

2 Materials and methods

Chicken embryo were incubated at 38.5° C and control samples (A, in two replicates) were collected at four developmental Hamburger & Hamilton stages HH2, 4, 16, and 28 meaning 6 and 18 hours, 2.5 and 6 days embryos. Experimental samples were collected after cooling to 15° C for 24 hours (B) and after 1 h recovery at 38.5° C (C). CAGE libraries were constructed according to recent protocol, by applying 5 µg of total RNA extracted with Trizol [3]. After sequencing libraries on Illumina HiSeq2500 SE 50 bp, reads were trimmed, cleaned from rRNA and adapters, mapped by bwa and hisat2 to the recent galGal5 genome assembly of chicken. Further signal clustering, counting and TPM normalization done with DPI and PromoterPipeline protocols of FANTOM5 project [4]. Additional TMM normalization step done in R (edgeR), followed by differential expression, GO enrichment analysis. Scanning for overrepresented transcription factors binding motifs (JASPAR2018 database) done in +/-250 bp regions around defined TSS (transcription start site) positions.

3 Results and discussion

In our study we were able to collect samples from various stages, including early embryo (HH2, 4, 16, and 28). Assuming all CAGE signal in current experiment we received 18304 predicted and active promoter regions, and 63% of which cover previously annotated promoters of entire chicken development [5]. Significant number of defined TSS (82.5%) was associated with annotated transcripts and related genes (n=8052). Using predicted promoter clusters, we found that cooling affects strong response on promoter level on all studied embryo stages. Number of differentially expressed (DE) promoters varied from 239 to 584 in down-regulated group and 143 to 341 within activated. Among DE candidates 144 showed consistent response on all stages and were associated to 112 genes. Top 5 enriched functional terms related to cooling-responding genes for each studied developmental stage presented in Table 1. In small group of common GO tags "mitochondrion" and "aryl hydrocarbon receptor complex" point the most conservative parts of cell constantly responding to cooling on all stages. Ahr participates in normal developmental processes, proliferation and differentiation of numerous cell types, but also known to participate in toxic response. DE promoter regions have relatively higher GC% in comparison to all defined promoters (p.adj<0.05), except day 2.5. From motif enrichment analysis we found 10 union TFs (ZNF740, SP2, SP1, KLF5, KLF16, EGR4, SP3, YY2, ELK1, YY1) and almost all of them are GC rich. Especially notable SP1, which is known to have important role in differentiation, response to DNA damage and oxidative stress. Recovery for 1 hour resulted in few DE promoters from 6 to 11, but enough to capture genes with reversible expression profile, like SLIT2, NNT, RRS1 in 6 hrs., HSPA8 in 18 hrs., LOC421740, HSD17B2, PCK1, SLIT2, ACTA2 in 2.5 days and FGA, CWC27, WSX1, SLIT2 in 6 days. Our results for the first time uncovered candidate genes and TFs participating in cold torpor adaptation in chicken embryo. From this point further studies became accessible, considering next steps of verification by genome editing tools.

2

Term	background/activated	P.value	Embryo age
midbody	13/3	0.000366	6 hrs.
erythrocyte development	3/2	0.000375	6 hrs.
mitochondrial large ribosomal subunit	4/2	0.000745	6 hrs.
organellar large ribosomal subunit	4/2	0.000745	6 hrs.
Ras protein signal transduction	22/3	0.001834	6 hrs.
mitochondrial large ribosomal subunit	4/2	0.000501347	18 hrs.
organellar large ribosomal subunit	4/2	0.000501347	18 hrs.
large ribosomal subunit	20/3	0.000775417	18 hrs.
mitochondrion	123/6	0.00092561	18 hrs.
mitochondrial ribosome	6/2	0.001238384	18 hrs.
ion binding	572/18	8.76E-05	2.5 d
generation of precursor metabolites and energy	29/4	0.000329896	2.5 d
cytoplasmic part	661/18	0.000529905	2.5 d
cytoplasm	965/23	0.000542044	2.5 d
mitochondrion	123/7	0.000554363	2.5 d
organic substance catabolic process	133/13	1.10E-05	6 d
catabolic process	158/14	1.56E-05	6 d
cytoplasmic part	661/33	1.63E-05	6 d
cytoplasm	965/41	6.24E-05	6 d
negative regulation of cellular protein catabolic process	5/3	0.000117	6 d

 Table 1. Top enriched functional terms in cold torpor response gene groups at four developmental stages of chicken embryo

References

- 1. Tazawa H.: Response of egg temperature, heart rate and blood pressure in the chick embryo to hypothermal stress. *J. Comp. Physiol. B* (1985).
- 2. Nichelmann M., Tzschentke, B.: Ontogeny of thermoregulation in precocial birds. *Comp. Biochem. Physiol. A Mol. Integr. Physiol.* (2002).
- 3. Deviatiiarov, R., Lizio, M., Gusev, O.: Application of a CAGE method to an avian development study. Methods in Molecular Biology 1650, (2017).
- 4. Forrest A.R.R., *et al.*: A promoter-level mammalian expression atlas. *Nature* 507, 462–470 (2014).
- 5. Lizio, M., *et al.*: Systematic analysis of transcription start sites in avian development. *PLoS Biol.* 15, (2017).

Кольцевые последовательности ДНК в геноме *Dicyema* sp.

Исаев С. В.^{1, 2}, Зверков О. А.², Михайлов К. В.^{2, 3}, Алёшин В. В.^{2, 3} и Панчин Ю. В.^{2, 3}

¹ Московский государственный университет им. М. В. Ломоносова, факультет биоинженерии и биоинформатики, Москва, Россия

² Институт проблем передачи информации им. А. А. Харкевича РАН, Москва, Россия

³ Московский государственный университет им. М. В. Ломоносова, НИИ физико-

химической биологии им. А. Н. Белозерского, Москва, Россия

ypanchin@yahoo.com

Аннотация. Дициемиды — клада беспозвоночных животных, являющихся паразитами головоногих моллюсков. Ранее эта группа животных рассматривалась как базальная для дерева Metazoa, однако позже было показано, что дициемиды — это вторично упрощённые представители Lophotrochozoa [1], которые в связи с паразитизмом утратили большинство типичных для билатерий систем: кровеносную, нервную, мышечную [2]. Ещё одна интересная особенность дициемид заключается в том, что в них были обнаружены экстрахромосомные кольцевые ДНК ядерного и митохондриального происхождения [3, 4]. В ходе работы впервые был произведён анализ генома Dicyema sp., который показал наличие кольцевых ДНК двух типов («длинных» и «коротких»). Характеристики обнаруженных кольцевых ДНК согласуются с приведёнными ранее данными электронной микроскопии и ПЦР. Показано, что митохондриальные гены, кодирующие белки и рРНК, содержатся в кольцевых ДНК по правилу «одна молекула — один ген». Нами не было обнаружено свидетельств того, что последовательности из кольцевых митохондриальных ДНК содержатся в больших линейных участках, что ставит под сомнение образование колец из каких-либо некольцевых предшественников. Таким образом, в данной работе было показано, что для дициемид характерно нетипичные для билатерий особенности устройства митохондриального генома.

Ключевые слова: Mesozoa, Dicyemida, диминуция хроматина, митохондриальная ДНК.

1 Кольцевые контиги

В качестве материала для анализа были использованы секвенированные и собранные геном и транскриптом *Dicyema* sp. Идентификация кольцевых ДНК основывается на поиске таких геномных контигов, у которых первые и последние 77 нуклеотидов совпадают — такой критерий является следствием параметров сборки генома (77 – длина использованного *k*-мера) [5]. Стоит отметить, что при помощи методов анализа последовательностей сложно отличить кольцевые молекулы ДНК от прямых повторов. Однако существует ряд исследований, подтверждающих наличие колец ДНК у дициемид [3, 4]. В результате поиска было обнаружено, что в сборке присутствует 24 065 контигов с 77нуклеотидными повторами по концам. Распределение их длин представлено на рис. 1.



Рис. 1. Гистограмма распределения длин контигов с 77-нуклеотидными повторами по концам. Взяты последовательности длиной более порога 77+77 нуклеотидов. На гистограмме представлены длины контигов за вычетом концевых повторов. Видны два максимума — в начале и в районе длины в 700 нуклеотидов.

На гистограмме с рис. 1 видно, что распределение получается бимодальным — чётко видны два кластера, различающиеся по длине. Экспериментальные данные, в которых при помощи электронной микроскопии и методов базовой молекулярной биологии было показано наличие фракции кольцевых ДНК длиной около 700 нуклеотидов (с учётом поправки на то, что молекулы кольцевые) [4], верифицируют полученные результаты. С учётом вышесказанного в ходе работы было решено сосредоточить внимание именно на фракции ранее описанных «тяжёлых» колец (длиной более 500 нуклеотидов) — их оказалось 3220. Распределение длин контигов «тяжёлых» колец представлено на рис. 2.

Для подтверждения того, что это именно кольцевые последовательности, а не прямые повторы, был произведён анализ парных чтений секвенирования, которые картируются на рассматриваемые кольца. Нами не было обнаружено таких случаев, когда только одно чтение из пары ложится на кольцо, а другое – на какой-то фланкирующий фрагмент ДНК.

2



Рис. 2. Гистограмма распределения длин кольцевых контигов «тяжёлой» фракции.

Сравнением частот вхождений коротких слов в кольцевые и некольцевые сборки контиги генома обнаружено, что последовательность 5'-АGGGAGCACCCT-3' (и комплементарная ей 5'-AGGGTGCTCCCT-3') встречается в «длинных» «кольцевых» контигах значительно чаще, чем в других участках генома. Так, в длинных контигах мы видим одно вхождение на 26905 нуклеотидов, по всему геному — одно вхождение на 8162571 нуклеотид. По критерию хи-квадрат уровень значимости отличия частоты встречаемости р < 0.001. Стоит отметить, что практически во всех кольцах сигнал встречается исключительно один раз (за исключением одного случая). Несмотря на обнаруженную статистически значимую закономерность, сигнал в кольцах встречается редко: на 3220 колец приходится 104 кольца с сигналом.

2 Митохондриальная ДНК

Ранее было показано, что у дициемид гены *cox1*, *cox2* и *cox3* находятся в индивидуальных кольцевых молекулах ДНК [4]. Авторами упоминаемой работы была предложена гипотеза «один ген — один репликон», согласно которой все гены митохондриального генома закодированы в одиночных молекулах ДНК.

С целью проверки этой гипотезы последовательности, соответствующие генам митохондриальной ДНК, были идентифицированы на кольцах «тяжёлой» фракции с использованием веб-сервиса MITOS [6], а также дополнительно верифицированы при помощи программ blastn и blastx [7].

В ходе анализа на кольцах были обнаружены последовательности белоккодирующих генов cox1, 2, 3, cob, nad1, 2, 3, 4, 5, а также atp6. Все они находятся на контигах с концевыми поавторами. На таких же контигах найдены гены рибосомных PHK (*rrnL* и *rrnS*), однако достоверность находок pPHK с помощью других подходов подтвердить сложно: выравнивание нуклеотидных последовательностей не подтверждает гомологии, в то время как в основе алгоритмов поиска MITOS лежит структурный подход. Стоит упомянуть, что распознать ген *rrnL* по последовательности трудно, например, в мтДНК *Intoshia linei*, имеющих мтДНК типичного для животных строения. Мы предполагаем, что трудности распознавания связаны с тем, что *Dicyema* sp., как и *I. linei*, является организмом с «высокой скоростью эволюции», что видно на филогенетическом дереве как длинные ветви [9]. В линейных контигах копии исследуемых генов обнаружены не были.

Все найденные митохондриальные гены присутствуют в транскриптоме. Интересными представляются случаи, когда длина обнаруженного транскрипта больше, чем длина кольца, с которого данная последовательность транскрибируется. Такое наблюдается, например, в случае с парой кольцо-транскрипт для гена *cox2*. Точечная матрица локальных выравниваний показала, что транскрипт является прямым повтором и копирует кольцо. Это же справедливо и для случаев с генами *nad1* и *rrnS*. Такие ситуации можно объяснить ошибками сборки; другое же объяснение заключается в том, что транскрипция может проходить «по второму кругу» кольца. Есть ли в этом какая-то биологическая значимость, неизвестно.

Таким образом, геномный анализ подтверждает ранее высказанную гипотезу о нетипичном устройстве митохондриального генома у дициемид.

Список литературы

- 1. Lu, T.-M., Kanda, M., Satoh, N. & Furuya, H. The phylogenetic position of dicyemid mesozoans offers insights into spiralian evolution. Zoological letters 3, 6 (2017).
- 2. Furuya, H. & Tsuneki, K. Biology of dicyemid mesozoans. Zool. Sci. 20, 519-532 (2003).
- Noto, T., Yazaki, K. & Endoh, H. Developmentally regulated extrachromosomal circular DNA formation in the mesozoan Dicyema japonicum. Chromosoma 111, 359–368 (2003).
- Watanabe, K. I., Bessho, Y., Kawasaki, M. & Hori, H. Mitochondrial genes are found on minicircle DNA molecules in the mesozoan animal Dicyema. Journal of molecular biology 286, 645–650 (1999).
- Compeau, P. E., Pevzner, P. A. & Tesler, G. How to apply de Bruijn graphs to genome assembly. Nature biotechnology 29, 987 (2011).
- Bernt, M. et al. MITOS: improved de novo metazoan mitochondrial genome annotation. Mol Phylogenet Evol 69(2), 313-9 (2017).
- Altschul, S. F. et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic acids research 25, 3389–3402 (1997).
- Altschul, S. F. et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic acids research 25, 3389–3402 (1997).
- 9. Robertson, H. et al. The mitochondrial genomes of the mesozoans Intoshia linei, Dicyema sp., and Dicyema japonicum. bioRxiv (2018).
- Schiffer, P. H. et al. Orthonectids Are Highly Degenerate Annelid Worms. Current Biology 28, 1970–1974.e3 (2018).

Неаннотированные сайты сплайсинга в геноме человека: эволюционные свидетельства функциональности

Степан Денисов^{1,2,3}, Алексей Миронов³, Дмитрий Первушин^{1,3,4}

¹Сколковский институт науки и технологий, ул. Нобеля, д. 3, Москва 121205, Россия
 ²ИППИ РАН, Б. Каретный пер., д. 19, стр. 1, 69121 Москва, Россия
 ³НИУ ВШЭ, ФКН, Кочновский проезд, д. 3, 125319 Москва, Россия
 ⁴МГУ им. М.В. Ломоносова, ФББ, Ленинские горы, д. 1, стр. 73, 119234 Москва, Россия

S.Denisov@skoltech.ru

Резюме. Мы определили неаннотированные сайты сплайсинга как сайты сплайсинга (СС), подтверждённые данными высокопроизводительного секвенирования РНК (RNAseq) при этом не представленные в базе данных GENCODE и находящиеся недалеко от известных аннотированных СС. Возникает вопрос: эти сайты представляют собой функциональные элементы или же они являются следствием случайных ошибок сплайсосомы (шум сплайсинга)? Мы попытались подойти к этому вопросу с эволюционной точки зрения: если это ошибки сплайсосомы, то мы не ожидаем, что они консервативны, напротив, если они функциональны, то они должны быть консервативны. Мы применили разработанный ранее тест на положительный и отрицательный отбор в СС и обнаружили, что на неаннотированные СС действует отбор схожий по силе с таковым, действующим на аннотированные сайты. Интересно, что на неаннотированные донорные СС действует положительный отбор, способствующий заменам неконсенсусных нуклеотидов на консенсусные, сила которого даже больше, чем в аннотированных СС. Также мы проверили, как такие характеристики, как частота использования неаннотированных СС по сравнению с аннотированными, сохранение рамки считывания влияют на силу отбора.

Ключевые слова: шум сплайсинга, сайты сплайсинга, естественный отбор, консервативность.

Введение и мотивировка

Согласно современным оценкам более 95% генов в геноме человека альтернативно сплайсируются [1]. Однако открытым остаётся вопрос какая доля из предсказанных изоформ мРНК действительно функциональна? Так Gonzàlez-Porta и соавт. показали, что большинстве белок-кодирующих генов человека доминирует одна изоформа: около 85% тотальной РНК приходится на мажорные изоформы [2]. С использованием данных белковой масс-спектрометрии в работе [3] утверждается, что только лишь малая часть пептидов подтверждают альтернативный сплайсинг, однако, вероятнее всего это связано с недостаточной чувствительностью метода.

Функциональность изоформы транскрипта белок-кодирующего гена может пониматься в разных смыслах. В широком смысле рассматриваемая изоформа должна существовать в клетке в достаточном количестве и с неё должен транслироваться белок. В ещё более узком понимании, данный белок должен выполнять определённую функцию в клетке. Однако в этой схеме существуют свои подводные камни. Например, известно, что изоформы мРНК с преждевременным стоп-кодоном подвергаются нонсенсенс-опосредованному распаду (nonsencemediated decay, NMD) [4]. Однако, это не значит, что изоформы, распадающиеся за счёт NMD не несут никакой функции. Показано, что за счёт переключения на такую изоформу может осуществляться тканеспецифическое подавление экспрессии гена [5]. Таким образом, функциональные исследования могут упустить из виду те аспекты функциональности изоформ или сайтов сплайсинга, на которые не рассчитан данный тип эксперимента. Альтернативой является эволюционный подход, где наличие положительного и отрицательного отбора свидетельствует о функциональности данной изоформы или СС.

В данной работе мы исследовали неаннотированные СС (т.е. подтверждённые данными высокопроизводительного секвенирования РНК (RNAseq) при этом не представленные в базе данных GENCODE и находящиеся на расстоянии < 30 нт от известных аннотированных) и сравнили действие отбора на них и на аннотированные СС, а также СС разных функциональных классов.

Данные и методы

В качестве исходных служили данные RNAseq проекта GTEx (8555 образцов представляющих 752 индивидуума и 53 ткани) [6]. Для определения координат аннотированных CC мы использовали аннотацию GENCODE version 19 [7]. Все координаты определялись по сборке генома GRCh37 (hg19).

Для картирования ридов RNAseq (из GTEx) на геном использовалась программа IPSA [8] с параметрами по умолчанию. СС с низкими показателями качества и неканоническими сайтами сплайсинга (GT/AG) были удалены из итоговой выборки.

Неаннотированные сайты сплайсинга технически определялись как CC, полученные из данных GTEx (с помощью IPSA), но не встречающиеся в аннотации GENCODE. При этом накладывалось дополнительное условие: неаннотированные CC должны находится на расстоянии менее 30 нт от аннотированных.

В таблице 1 представлено количество найденных неаннотированных СС.

Таблица 1. Неаннотированные СС*

,		1				
	CDS	Некод.	CDS/некод.	FS (CDS)	FP (CDS)	Всего
ДСС	4975	3484	7011	1526	3449	15470
ACC	11262	4911	17534	4433	6829	33707
Всего	16237	8395	24545	5959	10278	49177

*Сокращения: ДСС – донорные СС, АСС – акцепторные СС, CDS – кодирующая область гена (CoDing Sequence), FS – сдвигающие рамку считывания (Frame-Shifting), FP – сохраняющие рамку считывания.

Для оценки положительного и отрицательного отбора, действующего на различные классы СС использовался разработанный нами ранее метод [9]. Главной статистикой в этом методе является посчитанная для каждой позиции СС частота переходов между консенсусными (Cn) и неконсенсусными (Nc) нуклеотидами на линии человека (от ближайшего общего предка человека и макаки до человека). При этом таковая частота рассчитывается собственно в сайтах сплайсинга – это наблюдаемая частота переходов (*obs*) и в области гена, на которую почти не действует отбор (в участке интрона рядом с рассматриваемым CC) – это ожидаемая частота переходов (*exp*). На графиках далее представлено отношение этих величин *obs/exp*. Если *obs/exp* > 1, это свидетельствует о положительном отборе на данный тип переходов (например Cn→Nc), напротив *obs/exp* < 1 говорит об отрицательном отборе.

На всех рисунках представлены 95% доверительные интервалы.

Результаты и обсуждение

Мы сравнили действие отбора на разные классы СС. Мы рассматривали замены (переходы), которые случились на линии человека (после расхождения с макакой). При этом мы смотрели на замены двух классов: когда консенсусные нуклеотиды заменяются на неконенсусные (Cn→Nc) и наоборот (Nc→Cn).

В первую очередь нас интересовало отличается ли отбор между аннотированными и неаннотированными СС. Результаты сравнения представлены на рис. 1-4.

Как видно из рис. 1 и 2 на Cn нуклеотиды неаннотированных CC действует отрицательный отбор (obs/exp < 1). При этом сила отбора (значение obs/exp) приблизительно совпадает с таковой у аннотированных CC. Это говорит о том, что доля функциональных сайтов среди неаннотированных примерно такая же как и у аннотированных. Таким образом мы не видим свидетельств того, что выборка неаннотированных CC обогащена шумовыми событиями сплайсинга.

Более интересная картина наблюдается, если мы изучим отбор, действующий на Nc нуклеотиды (рис. 3 и 4). Рассмотрим донорные CC (рис. 4). Во многих позициях как неаннотированных, так и аннотированных CC мы наблюдаем



Рис. 1. Отбор, действующий на замены $Cn \rightarrow Nc$ в акцепторных CC. Сравнение аннотированных и неаннотированных CC. Ось X – позиция в CC, ось Y – obs/exp (см. данные и методы).



Рис. 2. Отбор, действующий на замены $Cn \rightarrow Nc$ в донорных CC. Сравнение аннотированных и неаннотированных CC. Ось X – позиция в CC, ось Y – obs/exp (см. данные и методы).



Рис. 3. Отбор, действующий на замены Nc \rightarrow Cn в акцепторных CC. Сравнение аннотированных и неаннотированных CC. Ось X – позиция в CC, ось Y – obs/exp (см. данные и методы).



Рис. 3. Отбор, действующий на замены Nc \rightarrow Cn в донорных CC. Сравнение аннотированных и неаннотированных CC. Ось X – позиция в CC, ось Y – obs/exp (см. данные и методы).

положительный отбор, способствующий заменам Nc \rightarrow Cn (obs/exp > 1). Причём в позициях -2, -1 и +5 отбор в неаннотированных CC существенно выше, чем в аннотированных. Этот неожиданный результат еще требует дополнительных проверок, но велика вероятность того, что выборка неаннотированных CC обогащена новыми CC, оптимизирующими соответствие последовательности CC консенсусу (т.е. силу CC).

В акцепторных CC характер отбора на Nc нуклеотиды зависит от позиции (рис. 3). Так позициях -7...-5 в неаннотированных CC (в отличие от аннотированных) мы видим отрицательны отбор против переходов Nc — Cn. Вероятно, это объясняется присутствием в выборке неаннотированных сайтов большего количества альтернатив (сдвигов) типа NAGNAG [10] (А и G технически являются неконсенсусными нуклеотидыми в этих позициях).

Мы также сравнили зависит ли отбор на неаннотированные CC от уровня использования неаннотированного по сравнению с ближайшим аннотированным. Для донорных сайтов удалось показать, что на Cn нуклеотиды CC с малым уровнем использования действует более слабый отбор. Именно такие сайты вероятно чаще появляются вследствие ошибок сплайсосомы. Однако для акцепторных CC результат оказался статистически не значимым (данные не показаны). Сравнение CC, которые сдвигают рамку считывания (относительно аннотированного CC) с теми, которые её сохраняют не дало статистически достоверных результатов. Вероятно расширение выборки и использование более далёких видов для сравнения позволит прояснить этот вопрос.

Таким образом, среди неаннотированных СС встречается сравнимое с аннотированными число функциональных сайтов. Сила отбора на неаннотированные сайты зависит от их уровня использования по сравнению с аннотированными. Повышенный уровень положительного отбора в неаннотированных донорных СС относительно аннотированных требует дальнейшего анализа.

Литература

- Tress M.L., Abasca F., Valencia A.: Alternative Splicing May Not Be the Key to Proteome Complexity. Trends in Biochemical Sciences 42(2), 98-110 (2017).
- Gonzàlez-Porta M. et al.: Transcriptome analysis of human tissues and cell lines reveals one dominant transcript per gene. Genome Biology 4(7), R70 (2013).
- Abascal F. et al.: Alternatively Spliced Homologous Exons Have Ancient Origins and Are Highly Expressed at the Protein Level. PLOS Computational Biology 1(6), e1004325 (2015)
- Isken O., Maquat L.: Quality control of eukaryotic mRNA: safeguarding cells from abnormal mRNA function. Genes & Development 21(15), 1833-1856 (2007).
- Mendell J. et al.: Nonsense surveillance regulates expression of diverse classes of mammalian transcripts and mutes genomic noise. Nature Genetics 36(10), 1073-1078 (2004).
- Melé M. et al.: The human transcriptome across tissues and individuals. Science 348(6235), 660–665 (2015).
- Harrow J. et al.: GENCODE: the reference human genome annotation for The ENCODE Project. Genome Research 22(9), 1760-1774 (2012).
- 8. https://github.com/pervouchine/ipsa, last accessed 2018/05/01.

- 9. Denisov S. et al.: Weak negative and positive selection and the drift load at splice sites. Genome Biology and Evolution 6(6), 1437-1447 (2014).
- Hiller M. et al.: Widespread occurrence of alternative splicing at NAGNAG acceptors contributes to proteome plasticity. Nature Genetics 36(12), 1255-1257 (2004).

Изучение молекулярной эволюции в игровой форме: настольная игра "Адаптация"

А. В. Безменова¹, К. Р. Сафина¹, В. О. Бурская¹, А. В. Столярова¹, Г. В. Клинк²

¹ Сколковский институт науки и технологий, Москва, Россия ² Институт проблем передачи информации РАН, Москва, Россия

aleksandra.bezmenova@skoltech.ru

Абстракт. Молекулярная эволюция - сложный для понимания процесс, происходящий под действием множества факторов. Цель данного проекта - разработать настольную игру, призванную помочь игрокам разобраться в основных принципах эволюции последовательностей белков, изучить основные механизмы и факторы, действующие в процессе эволюции. Мы планируем отладить игру во время практики студентов факультета Биоинженерии и Биоинформатики МГУ в августе 2018 года.

Игра начинается с того, что каждый игрок получает случайную исходную аминокислотную последовательность и карточку организма, за который он будет играть. Суть игрового процесса заключается в том, что игрок должен совершить адаптивную прогулку до пика приспособленности - то есть преобразовать свою исходную последовательность в некую заданную наперед наиболее приспособленную - свою для каждого игрока. Преобразовывать последовательность игрок может, используя карты, представляющие основные эволюционные механизмы случайный и направленный мутагенез (точечные замены, инсерции и делеции, инверсии, транслокации), рекомбинацию и др. Игра будет усложняться действием разных факторов, таких как различная скорость мутирования для разных организмов, поломки систем рекомбинации, репарации и др., внешние мутагены, эпистатические взаимодействия, смена адаптивного ландшафта и т. п.

Ключевые слова: Молекулярная эволюция, Настольная игра

Applying phylogenetic tree-based approach to genomewide association studies in Mycobacterium tuberculosis

Valentina Burskaia¹, Gennady Fedonin², Georgii Bazykin^{1,2}, Alexey Neverov³

¹ Skolkovo Institute of Science and Technology, Nobel street 3, Moscow 121205, Russia

² Institute for Information Transmission Problems, Bolshoy Karetny per. 19, build.1, Moscow 127051, Russia

³ Central Research Institute for Epidemiology, Novogireevskaya 3a, Moscow 111123, Russia

Phylogenetic approach could significantly improve Genome-Wide Association Studies (GWAS) in microbial organisms, especially in cases of recurrent antibiotic resistance evolution. Phylogenetic approach allows to describe simultaneous appearance and coexistence of phenotype and genotype traits on independent branches of the tree. We use TreeWAS software as well as our original algorithms to find dependencies between genotype and phenotype traits. 5000 full genomes of Mycobacterium tuberculosis are used; the alignment contains 200 000 SNPs; 13 types of drugs are analyzed. Mycobacterium tuberculosis is almost ideal model organism for our goal: as there is almost no recombination in it, independent appearance of similar mutations in different lineages is nearly impossible (And so we never deal with hemiplasious allels).

Defecation motor program in nematodes

Victor P. Kuznetsov^{1,2[0000-0001-7264-142X]}, Georgy A. Slivko-Koltchik^{1,2}, Dmitry A. Voronov¹ and Yuri V. Panchin¹

¹ Kharkevich Institute for Information Transmission Problems, Russian Academy of Sciences, Moscow 127994, Russian Federation

² Faculty of Bioengineering and Bioinformatics, Lomonosov Moscow State University, Mos-

cow 119991, Russian Federation

scikvp@gmail.com

Abstract. The nervous system controls most rhythmic behaviors, with a remarkable exception. The defecation motor program (DMP) in rhabditean nematodes is initiated by a pacemaker activity of intestinal cells without participation of nervous system, mostly driven by calcium concentration oscillations. In the nematode Caenorhabditis elegans DMP is a periodical 45-second process, which is controlled by the cell-autonomous oscillations that propagate between intestinal cells via gap junctions. In the related nematode Heterorhabditis megidis it was shown that DMP is accompanied by electrical membrane potential oscillations in the intestinal cells and that central pattern generator (CPG) is associated with an unusual all-or-none hyper-polarization "action potential". In C. elegans the molecular mechanism of DMP includes about 20 various proteins which are studied in details. In the midgut cells of a distant from Rhabditida nematode Enoplus brevis (Enoplida) we discovered membrane potential cycling which resemble oscillations in H. megidis. We also found that the gut cells in E. brevis are electrically connected through gap junctions. Electrophysiological and bioinformatics data analysis of this complex process in distant nematodes in comparison with H. megidis and C. elegans shows results that can lead to expanding the knowledge of the non-neuronal behavior control mechanisms. We suggest that distant nematode species possess the same non-nervous rhythmical DMP generation mechanism and that this mechanism was present in the nematode common ancestor.

Keywords: nematode defecation, gap junction, nematode evolution.

1 Introduction

Rhythmic activity plays an important role in a physiology of various species. The period of rhythmic oscillations in different systems can vary from seconds to years. Cellular and molecular mechanisms of circadian (about 24 hours) and fast (with period of seconds) rhythms are well studied, while less attention has been paid to ultradian rhythms with shorter periods (minutes to hours) [1]. CPGs are usually considered to be neural networks that generate periodic motor commands for rhythmic muscular contractions

in the absence of external rhythmic inputs [1,2]. A remarkable exception is the defecation motor program (DMP) in *Caenorhabditis elegans*.

Nematode studies, especially on the model species *C. elegans* have seeded numerous fundamental discoveries of general biological significance [3]. Studies of the model rhythmic systems are important as they allow to deeply understand the mechanisms of their functioning. While eating, *C. elegans* hermaphrodite defecates about once per minute. The genetics of this process have been studied for quite a long time [4-6]. It is known that the defecation motor program (DMP) is controlled by calcium concentration oscillations in the cytoplasm of intestinal cells [7,8]. The defecation process consists of three consecutive series of muscle contraction (aBoc) and expulsion step (Exp) [4,5]. The posterior body wall muscle contraction (pBoc), anterior body wall muscle contraction (pBoc) in the defecation cycle is regulated by pH changes in the pseudocoelomic space caused by Na⁺/H⁺ transporter activity in the intestinal cells [9,10].

Intracellular ion (H⁺ and Ca²⁺) concentration changes are involved in DMP operation and could influence on electrical properties such as voltage and conductance of the intestinal cell membrane. Heterorhabditis megidis that belongs to the same order Rhabditida defecation cycle is driven by a CPG associated with unusual all-or-none hyperpolarization "action potential" with a fixed duration of about one minute, period of about four minutes and amplitude of about 60 mV [11]. The unique "non-nervous" CPG governing rhythmic defecation was so far found only within the order Rhabditida (C. elegans and H. megidis) It is not clear if this mechanism is a specific feature of Rhabditids characterized by their rapid evolution [12], or also exists in other nematodes. Enoplida are probably the most evolutionary distant nematodes from Rhabditida. We studied the electrophysiological properties of the intestine cells in Enoplus brevis with a special focus on the membrane potential oscillations and electrical connections between the intestinal cells in combination with the analyses of several dozen genes important for DMP [6] from genomic sequences of many nematode species. This makes it possible to study elements of machinery in phylogenetically distant nematode E. brevis and consider its origin in nematodes.

2 Results and Discussion

Enoplus brevis Bastian, 1865 is a free-living littoral and sub-littoral marine nematode with the body length of an adult animal about 6-8 mm and the maximum diameter 0.13-0.17 mm (Fig. 1A). *E. brevis* were extracted from sandy littoral at the White Sea biological station of the Russian Academy of Sciences "Kartesh" in the summer of 2017 and kept in seawater at 4^oC. The experiments were carried out at 15-17^oC. Only adult males, females, or forth stage juveniles were taken for experiments.

E. brevis is a nematode distant from Rhabditida and its intestine consists of significantly larger number of cells. (6000 cells in *E. brevis*, 20 cells in *C. elegans*, and about 24-28 cells in *H. megidis*). As a result, the intestine cell size of *E. brevis* is quite small, that restricts the use of some electrophysiological techniques.

Quantitative calculations about cell-to-cell coupling needs accurate electrical coupling coefficient measurements with simultaneous impaling of each cell by two intracellular microelectrodes. However only one electrode per cell could be inserted which allowed us to obtain only qualitative data. Two cells lying relatively close to each other were simultaneously penetrated by microelectrodes. The first microelectrode was used to inject electrical current into the cell, and the second microelectrode was used to identify the presence of electrical coupling by corresponding membrane potential changes. (see Fig. 1). Recorded deviation of the membrane potential shift in a response to injected current in the neighbor cell confirms the presence of electrical coupling in *E. brevis*. In control experiments, where one of the electrodes was inserted into the intestinal cell and the second remained in the extracellular milieu, no electrical coupling was observed.



Fig. 1. The presence of electrical coupling between neighbor cells in *E. brevis*. An impulse of negative current of -20 nA is applied through I_2 (red). Deviation of membrane potentials in adjacent cell ΔV_1 is recorded by V_1 microelectrode (blue).

For *H. megidis*, the presence of an unusual all-or-none hyperpolarization "action potential" with reversed polarity was previously observed [11]. *E. brevis* has the similar regular membrane potential oscillations in the intestinal cells (see Fig. 2). The oscillation amplitude in *E. brevis* is lower and vary in a range of 15-25 mV, while the mean period is longer and was about 20 minutes. It is possible that the DMP frequency is adapted to the mechanical properties of the animal. Similar to *H. megidis* the potential in *E. brevis* starts at the low positive level and transits to a rapid all-or-none hyperpolarization "action potential", which later slowly tends back to previous values. As a result, midgut (endoderm) cells of *E. brevis* are able to generate stable oscillations of

the membrane potential without any other cell type involvement, including nervous cells. In both species, the oscillation cycle can be decomposed into a relatively long positive "plateau" and a shorter hyperpolarization "pit". In both species, the intestinal cells are electrically connected and could be synchronized through gap junctions. The cyclic membrane potential changes similarity in intestinal cells of *E. brevis* and *H. me-gidis* makes it plausible that the defecation motor program in both species is based on the same mechanism.



Fig. 2. Intracellular recordings of the spontaneous voltage oscillations in an isolated *E. brevis* gut preparations.

Although the endoderm is considered to be non-excitable tissue, the self-supporting or regenerative electrical wave in nematode gut is similar to the action potential in excitable cells. The described phenomena calls for a reevaluation of the definition of excitable cells.

Nematode	PBO-5	NLP-40	AEX-2	EXP-1
Caenorhabditis elegans	+	+	+	+
Heterorhabditis bacteriophora	+	+	+	+
Pristionchus pacificus	+	+	+	+
Ditylenchus destructor	+	+	+	+
Meloidogyne incognita	-	+	+	+
Bursaphelenchus xylophilius	+	+	+	+
Syphacia muris	+	+	+	+
Enterobius vermicularis	+	+	+	+
Acaris suum	-	+	+	+
Thelazia callipaeda	-	-	+	+
Brugia malayi	-	-	+	+
Dracunculus medinensis	-	-	+	+

Table 1. DMP proteins in various nematode species.

Gongylonema pulchrum	-	-	+	+
Enoplius brevis	+	+	+	+
Trichuris suis	-	-	+	+
Trichinella spiralis	-	-	+	+
Romanomermis culicivorax	-	-	-	-
Soboliphyme baturini	-	-	+	+

To check the possibility of the same machinery presence in other nematodes we compared the existence of DMP related proteins by bidirectional best hit (BBH) method. We considered four proteins involved in DMP to be a markers in our investigation that participate in signal transmission from intestine on muscles in 17 nematode species with full-sequenced genomes. These proteins are present only in Nematode phylum. Presence of these proteins in various nematode species is illustrated on Table 1. All these proteins are absent in *Romanomermis culicivorax*, the only considered species with reduced intestine. We suggest that distant nematode species possess the same nonnervous rhythmical DMP generation mechanism and that this mechanism was present in the nematode common ancestor.

References

- Arshavsky, Y. I., Deliagina, T. G., Orlovsky, G. N: Pattern generation. Curr. Opin. Neurobiol. 7, 781–789 (1996).
- Kuo, A. D.: The Relative Roles of Feedforward and Feedback in the Control of Rhythmic Movements. Motor Control 6, 129–145 (2002).
- 3. Wormbook Homepage, http://www.wormbook.org
- Thomas, J. H.: Genetic analysis of defecation in Caenorhabditis elegans. Genetics 124, 855– 872 (1990).
- 5. Liu, D. W. C., Thomas, H.: Regulation of a Periodic Motor Program in C . elegans z. J. Neurosci. 14, (1994).
- Branicky, R., Hekimi, S.: What keeps C. elegans regular: the genetics of defecation. Trends Genet. 22, 571–579 (2006).
- Teramoto, T., Iwasaki, K.: Intestinal calcium waves coordinate a behavioral motor program in C. elegans. Cell Calcium 40, 319–327 (2006).
- Espelt, M. V, Estevez, A. Y., Yin, X., Strange, K.: Oscillatory Ca²⁺ signaling in the isolated caenorhabditis elegans intestine : role of the inositol-1,4,5-trisphosphate receptor and phospholipases C β and γ. J. Gen. Physiol. 126, 379–392 (2005).
- 9. Pfeiffer, J., Johnson, D. & Nehrke, K.: Oscillatory transepithelial H(+) flux regulates a rhythmic behavior in C. elegans. Curr. Biol. 18, 297–302 (2008).
- Beg, A. A., Ernstrom, G. G., Nix, P., Davis, M. W., Jorgensen, E. M.:Protons act as a transmitter for muscle contraction in C. elegans. Cell 132, 149–160 (2008).
- Kuznetsov, V. P., Slivko-Koltchik, G. A., Voronov, D. A., Panchin, Y. V.: Electrophysiology of the rhythmic defecation program in nematode Heterorhabditis megidis. Sci. Rep. 1–9

6

(2017).

 Blaxter, M. L. et al. A molecular evolutionary framework for the phylum Nematoda. Nature 392, 71–75 (1998).

Algorithms for Clustering and the Comparison of Hierarchies within the Spatial Organization of Chromatin

Pushkareva O.,^{#1} Kurashenko A.,^{#2} Moskvina U.,^{#3} Gelfand M.,^{#4} Rubinov A.^{#4}

¹ NRU HSE, Moscow, Russia

² MSU, Moscow, Russia

³ Skolkovo Institute of Science and Technology, Skolkovo, Russia

⁴ IITP RAS, Moscow, Russia

#contributed equally

Abstract. The recent development of high-throughput 3C-based technologies made it possible to study the spatial organization of chromatin inside the nucle⁻ us, as well as reveal new biological structures -- Topologically Associating Domains (TADs). Each such domain is comprised of DNA regions actively contacting with each other and having fewer contacts outside the domain. TADs are conserved between cell lines and even between related species. While there exist numerous bioinformatic tools for the TADs identification, the choice of the method and its manually set parameters influences the number and characteristics of the identified domains. One more common disadvantage is that most methods may leave gaps between the TADs. This study aimed to develop a stable and universal tool which would overcome these limitations. We have developed and implemented an algorithm, which builds a hierarchical tree on given Hi-C contact matrices, and methods to calculate the differences between such trees.

Keywords: Chromatin, Hi-C, Topologically Associating Domains (TADs).

1. Introduction

A unique feature of the genome organization in eukaryotes is the packaging of the DNA into chromatin (from the Greek chroma, "color") and its location in the cell nucleus. The DNA of eukaryotes is divided between a set of chromosomes, which keep the long DNA strands in a very confined nuclear volume. Thus, the human genome has a size of about 3×109 base pairs, one human cell contains tentatively 2 meters of DNA, yet the size of a nucleus is only about 10 µm in diameter. [5] DNA of Drosophila melanogaster has a length of 5 centimeters and is located in a volume of only 5 micrometers. This knowledge together with the recent development of Chro-

mosome Conformation Capture techniques made it possible to study chromatin as a three-dimensional structure.

Topologically Associating Domains — TADs, — are hierarchically embedded domains of genome organization which represent tightly packed DNA regions with a high local density of contacts. Being conserved across cell lines and organisms, TADs may denote functional units of a genome and be responsible for the distribution of epigenetic marks. [1] In perspective, TADs can shed new light on the connection between the structural and functional organization of a genome. Thus, it can be applied while studying diseases — Lupianez et al. (2015) showed that genomic rearrangements involving some neighboring TADs caused limb malformations. [2] Various tools for the TADs prediction exist. However, all of them are defined by specific readjustable parameters, and therefore, show various results which are nearly impossible to compare. This leads to a problem of finding a unique algorithm that can provide the results that do not fluctuate depending on a threshold or a parameter. [3] We propose an algorithm that does not search for the domains but builds hierarchical trees of the chromatin interactions. This algorithm does not rely on any changeable parameters, which allows it to be a universal tool for building trees on the Hi-C data. Thus, the current research aims to develop a new algorithm (suggested by A.Rubinov in 2017) for building hierarchical trees and their comparison, and apply it on the different cell lines and conditions.

2. Materials and Methods

2.1. Materials

The algorithm was tested on the Hi-C data from several Drosophila melanogaster cell lines (S2, BG3, KC). Drosophila is a model organism, widely used in studying genetics, development and disease. Drosophila has four chromosomes: a pair of sex chromosomes (XX or XY), which are designated Chromosome 1, and three pairs of autosomes (2, 3, and 4). Chromosome 4 represents approximately only 2% (1.35Mb) of the genome, thus, we did not use it in our analysis.

The data was taken partially from the collaborators (S.V.Razin lab, Institute of Gene Biology RAS), partially from the open sources. The resolution for all data is 20 kb, there are 39 files with data in total. The data is represented in the hdf5 format. Prior to analysis all data were subjected to the iterative correction. Empty bins on the heatmap, appearing due to the inaccuracy of mapping were removed from the matrices.

2.2 Method

The algorithms of one-dimensional hierarchical clustering and comparison of hierarchies were implemented in Python programming language using both standard libraries (pandas, Seaborn, etc) and the libraries specific to the topic (lavaburst, h5py).

3. Results

The algorithm was suggested by A. Rubinov in 2017. The algorithm takes as an input a symmetric Hi-C matrix C (i, j) (size $n \times n$). Entries of the matrix correspond to the contact frequencies between loci i and j. The main diagonal of the matrix contains only zero values. Each bin of a matrix has a particular value assigned to it. According to the Hi-C matrix characteristics the bins with higher values are located close to the main diagonal (they correspond to higher contact frequencies), and vice versa - remote bins have smaller values. Since the matrix is symmetric, in the algorithm described below we are only working with the upper triangle values.

The first part of the Rubinov algorithm proposes the following: let's assign each bin from the first nonzero diagonal a specific direction. To do so, we need to calculate the sum of contact frequencies to the left (all values in the column j) and the right (all values in row i) of each bin. However, as the matrices for a Drosophila melanogaster chromosome in 20 kb resolution have a size of approximately 1000-1300 bins and most of the domains are usually located in proximity to the main diagonal, we need to reduce the noise from remote bins. To do so and, moreover, increase the weight for the cells which lie close to the zero diagonal, we raise the value of each bin to the power of 4.



Now we can get back to calculating the sum of contact frequencies and assigning each bin a direction. In a case when the sum of contacts on the left (column) is greater than the one on the right (row), we call this bin left-oriented, and vice versa.

When each bin has a direction, we can begin the clustering. The Rubinov algorithm is based on the idea that bins located inside a topological domain closer to its left border will always be right-oriented, as they have most of the contacts on the right. The same idea is true for the bins on the right border - they always "look" to the left, where the sum of contacts is the greatest. This leads us to the main theorem of the algorithm: if a cluster i_0 is right-oriented, and the cluster i_1 is left-oriented, then we can merge them and call this union a cluster. In the algorithms, each bin in the initial matrix is also a cluster.



Fig. 2. Second iteration of the Rubinov algorithm

Thus, after assigning each bin the orientation, the algorithm starts merging the first set of co-directional bins.

After the first iteration of the algorithm, we have a matrix which looks as shown on Fig. 1. Arrows stand for the directions of the bins (markup is not shown).

Now each new cluster gets a new orientation - the sum on the left and the right is recounted based on the sums of each bin in a cluster. Next iteration, based on the reassigned orientations, gives us a matrix that looks shown on the Fig. 2.

After each next step, the overall number of clusters is increasing. By the last iteration, the entire matrix becomes a hierarchical tree. We are going to show the trees computed on real data in the next section of the chapter.

The most important feature of our algorithm is that, unlike the others, it does not rely on any given parameters. Therefore, there is one and only one hierarchical tree for each input contact matrix.

Comparison of vectors



Fig. 3 Binary Tree

Binary tree shown at the Fig. 3 will be an example of finding a vector for a hierarchical tree build on a Hi-C matrix. With each letter representing a bin, we can find a distance between two neighboring bins. The distance between a&b (b&c, c&d, etc.) equals the number of branches to be passed to

get from bin "a" to the bin "b". Thus, for a&b it is 1, for b&c it is 3, for d&e it is 5 etc. The vector we get for the tree from Fig. 21 is the following:

 $a = \{2, 4, 2, 6, 2, 4, 2\}$

After obtaining a vector for each of the trees, we compare them in pairs using the following norms:

$$0 \le \frac{\sum_{i} |a_{i} - b_{i}|}{\sum_{i} (|a_{i}| + |b_{i}|)} \le 1 \text{ or } 0 \le \frac{\sum_{i} |a_{i} - b_{i}|^{2}}{\sum_{i} |a_{i}^{2} + b_{i}^{2}|} \le 1 \quad (1, 2)$$

where a value closer to 0 shows the trees similarity. Let the first norm be $\rho 1$, and the second norm be $\rho 2$.

To illustrate the work of our algorithm we run it on the real Hi-C maps from fruit fly *Drosophila melanogaster* genome, different cell lines. First, we show how our algorithm builds hierarchical trees on the Hi-C heat map slice for a BG3 cell line, chromosome 3R.



Fig. 4. Hierarchical Tree for X chromosome (slice), BG3 cell line

On the image 4 we can see the hierarchical tree. Bright red color stands for the high contacts enrichment. Numbers "1" and "-1" on the empty diagonal represent the

directions of each bin (i,j). The constructed clusters almost entirely correspond to the domains on the heatmap, hence the algorithm shows a rather high precision.

Let us now turn to the comparison of the hierarchical trees. As we only assessed the trees represented above visually, we need to have numerical characteristics to help us understand the ability of the algorithm to distinguish between the cells lines and to be reliable for analyzing biological data.

The Rubinov vectors comparison

39 files with Drosophila genome (including different cell lines, replicates, both successful and unsuccessful experiments, control groups) form 741 possible pairwise combinations. For each of these pairs we calculated a pair of vectors, representing the distance between each 2 neighboring bins. Calculating 741 pair of vectors and comparing them using ρ 1 and ρ 2 norms gave us a result shown on Fig. 5a, b.

For both ρ 1 and ρ 2 have multimodal distribution, which indicates we have several different groups.

As further analysis shown, the peaks of the values higher than 0.3 stand for the pairs which included unsuccessful experiments. This indicates that the algorithm does not just detect the differences between the cell lines and conditions, but is also precise enough to catch anomalies.

To confirm that we did not get random distributions, we made a control experiment. We shuffled the vectors in each of 741 pair, and calculated ρ 1 and ρ 2 norms again. The results are shown on fig. 5b, c. As was expected, random shuffle of the real data looks like Gaussian distribution. Real vectors are more similar to each other than random vectors.





Fig. 5 (a,b) Distribution of $\rho 1$ and $\rho 2$ on real data; (c, d) distribution of $\rho 1$ and $\rho 2$ on shuffled real data

We assumed that the peaks seen on Fig. 5a, b represent several groups - replicates (0.15 vicinity for 5a and 0.1 for 5b) and all the other experiment (0.2 and 0.15 for 5a and 5b), respectively).

To check the assumption, we took 11 pairs of replicates and 11 random pairs (non-replicates) from our data and built the distribution (Fig.6, Fig.7).



Fig-s. 6,7 Distribution of $\rho 1$ and $\rho 2$ values for replicates (green) and non-replucates (blue)

Indeed, due to their similarity, in both cases the replicate pairs showed values closer to zero, than the randomly taken pairs.

Neighbour Joining

In the further analysis, we run the obtained $\rho 1$ and $\rho 2$ distance matrices through the Neighbour Joining clustering algorithm.

Neighbour Joining clearly showed the outliers (Vas2 and S2DL-3 experiments) as well as united the replicates and clustered different experiments on the same cell line together (for ex. Bam, Vas). It also clustered some of the experiments and their controls (Lamin-1, Lamin-2,Lacz-1,Lacz-2) This gives us preliminary biological results of the method.



Fig. 8 NJ tree for chromosome X, build on p1 distance matrix

4. Discussion and Conclusions

The body of work presented in this thesis contributes towards the development of the novel tool for Hi-C data analysis. Specifically, the primary goal of the study was to devise an algorithm, which partitions chromosomes into hierarchical trees, and does not rely on any predetermined parameters. Then, the secondary task was to find methods for comparing the hierarchies.

The major result of the study is the implementation of the algorithm of onedimensional clustering, proposed by A. Rubinov. This algorithm, which does not contain any changing parameters, lets us build a hierarchical tree on the entire length of a chromosome. Trees, delivered by the algorithm, correspond to Hi-C heat maps with very high precision. The algorithm identifies both small and large domains across the chromosome. The major feature of the algorithm is that, unlike the currently existing tools, it does not rely on any predefined value.

As for the comparison of hierarchical trees, we compared the vectors derived by counting the number of tree branches between each two neighboring bins. The method showed the difference between the trees build for replicates and for nonreplicate pars. We also run neighbor joining algorithm which shows us that the algorithm is able to cluster the replicates and cell lines.

The study briefly covers how the algorithm works with the real data and does not deliver many significant biological results. Recommendations for further research would be to apply the algorithm for more biological data and scrutinize how hierarchical trees behave for different cell lines and conditions. Also, more benchmarking of the already existing tools and the Rubinov algorithm can be provided.

Summarizing, the study provides a strong and elegant tool for the analysis of Hi-C data. The algorithm does not contradict the biological intuition, moreover, confirms it. The advantages of the algorithm are its independence on any given parameters, robustness, and the ability to discover the domains on any scale and unite them into hierarchical trees.

References

- Weinreb, C., & Raphael, B. J. (2016). Identification of hierarchical chromatin domains. Bioinformatics, 32(11), 1601–1609. http://doi.org/10.1093/bioinformatics/btv485
- Lupiáñez, D. G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., ... Mundlos, S. (2015). Disruptions of Topological Chromatin Domains Cause Pathogenic Rewiring of Gene-Enhancer Interactions. Cell, 161(5), 1012–1025. http://doi.org/10.1016/j.cell. 2015.04.004
- Forcato, M., Nicoletti, C., Pal, K., Livi, C. M., Ferrari, F., & Bicciato, S. (2017). Comparison of computational methods for Hi-C data analysis. Nature Methods, 14(7), 679–685. http://doi.org/10.1038/nmeth.4325
- Belton, J.-M., McCord, R. P., Gibcus, J., Naumova, N., Zhan, Y., & Dekker, J. (2012). Hi-C: A comprehensive technique to capture the conformation of genomes. Methods (San Diego, Calif.), 58(3), 10.1016/j.ymeth.2012.05.001. http://doi.org/10.1016/j.ymeth. 2012.05.001
- Dekker, J., Rippe, K., Dekker, M., & Kleckner, N. (2002). Capturing Chromosome Conformation. Science, 295(5558), 1306–1311. https://doi.org/10.1126/science.1067799

Об эволюционном происхождении механизма контроля взаимно исключающих экзонов конкурирующими структурами РНК

Тимофей Иванов¹, Степан Денисов^{1,3,4}, и Дмитрий Первушин^{1,2,}

¹ Сколковский институт науки и технологий, улица Нобеля 3, Москва 121205

² Факультет биоинженерии и биоинформатики, Московский государственный университет 1-73, Москва 119899

³ Факультет компьютерных наук, Высшая школа экономики, Кочновский проезд, д. 3 Москва, 125319

⁴ИППИ РАН, Б. Каретный пер., д.19, стр.1, 69121 Москва, Россия

tim.ivanov.920gmail.com

Аннотация. Взаимно исключающие экзоны (ВИЭ) представляют из себя частный случай событий альтернативного сплайсинга, при котором в зрелую мРНК включается один и только один из группы следующих друг за другом в геноме экзонов. В ряде случаев выбор включаемого экзона обеспечивается конкурирующими вторичными структурами внутри премРНК, состоящими из докерного сайта, который расположен перед или после кластера ВИЭ, и группы селекторных сайтов, расположенных рядом с каждым из экзонов. Независимо от этого известно, что внутри кластера ВИЭ экзоны часто гомологичны друг другу, и поэтому представляется вероятным, что они произошли в результате тандемных дупликаций одного предкового экзона. В данной работе мы предлагаем эволюционный механизм происхождения докерных и селекторных сайтов в результате тандемных дупликаций, затрагивающих не только экзоны, но и части фланкирующих интронов. Мы рассматриваем этот механизм на примере эволюционной истории конкретных генов, таких как MRP1, DSCAM, и МНС и указываем вероятные события дупликации и делеции экзонов, ассоциированные с ВИЭ.

Ключевые слова: взаимно исключающие экзоны, альтернативный сплайсинг, структура РНК

Введение

Выбор одного экзона из кластера ВИЭ широко распространен среди генов эукариот [1]. Некоторые случаи такой регуляции были подробно описаны на

модельных организмах [2], однако новейшие данные RNA-seq экспериментов свидетельствуют о наличии такого типа сплайсинга у высших позвоночных, растений, нематод [4].

Последовательности экзонов, принадлежащих кластерам ВИЭ обычно обладают высокой схожестью, что может объясняться их происхождением в результате геномной дупликации [5]. Исследования таких генов животных как MRP показывают независимые случаи дупликаций экзонов в разных филумах, что свидетельствует о конвергентной эволюции такого типа сплайсинга. Однако существуют примеры негомологичных ВИЭ, что говорит о возможности происхождения такого типа альтернативного сплайсинга не только в результате геномной дупликации.

Были предложены различные молекулярные механизмы, объясняющие взаимно исключающий выбор экзона из кластера. Один из таких механизмов включает два специальных регуляторных элемента: докинг-сайт и селекторсайт. В одном из фланкирующих кластер ВИЭ интронов содержится консервативная последовательность (докинг сайт), комплиментарная группе селектор-сайтов, расположенных во внутренних интронах кластера ВИЭ. При складывании пре-мРНК во вторичную структуру, докинг-сайт слипается с одним из селектор-сайтов, выпетливая весь промежуточный участок для дальнейшего сплайсирования. Наличие больше чем одного селектор-сайта приводит к их конкуренции за докинг-сайт. Такие регуляторные элементы были обнаружены в генах DSCAM, 14-3-3*z*, Mhc, и MRP1 насекомых. [6]

Наличие таких конкурирующих структур в кластерах ВИЭ может быть объяснено тандемными дупликациями экзонов: если дупликация затронула не только экзон, но и участки окружающих его интронов, содержащих селекторсайты, то образуются две последовательности, комплиментарные докинг-сайту. Только одна из этих последовательностей при сплайсировании сможет слипнуться с докинг-сайтом, приводя к образованию конкурирующих структур.

Материалы и методы

Геном Drosophila melanogaster (сборка 2006 год) был загружен с Berkeley Drosophila Genome Project (BDGP, R5). Аннотация была взята с сайта Flybase. Геном человека (февраль 2009) был взят из Genome Reference Consortium. Аннотация v19 была загружена из GENCODE.

Консервативность участков была оценена с помощью трека phastConsElements, взятого из геномного браузера UCSC.

В качестве контроля брались две группы: интроны, близкие к исследуемым по длине (но лежащие в другом гене) и интроны, полученные из исходных с помощью динуклеотидного шаффлинга.

Для оценки схожести нуклеотидных последовательностей использовался модуль Biopython, функция локального выравнивани Смита-Ватермана pairwise2.align.localms с параметравми: match = 1, mismatch = 0.2, gap opening =

-1, gap extension = -0.5. Последовательности длинней 10000 нуклеотидов не рассматривались.

Из-за относительно удаленного друг от друга расположения докинг и селектор последовательностей, структуры, образованные спариванием этих сайтов, относят к структурам дальнего взаимодействия и поэтому они рассматриваются как интермолекулярные [4]. Для оценки потенциала образования таких структур использовались программы из ViennaRNA package [5]. Все используемые программы запускались с параметром -noLP.

Обсуждение

На примере человека и плодовой мухи, в работе показано, что интроны внутри кластеров, ассоциированных со взаимно исключающими событиями сплайсинга, обладают значимо большей степенью консервативности по сравнению с интронами из контрольных групп.

На примере дрозофилы было также показано, что степень схожести интронов и экзонов кластера ВИЭ коррелируют. (см. рис 1)



Рис 1: pairplot схожести интронов и экзонов

References

- 1. Jin, Y.; Dong, H.; Shi, Y.; Bian, L. Mutually exclusive alternative splicing of pre-mRNAs. Wiley Interdiscip Rev RNA 2018, 9, e1468
- Graveley, B.R. Mutually exclusive splicing of the insect Dscam pre-mRNA directed by competing intronic RNA secondary structures. Cell 2005, 123, 65–73
- Kondrashov, F.A.; Koonin, E.V. Origin of alternative splicing by tandem exon duplication. Hum. Mol. Genet. 2001, 10, 2661–2669
- 4. Pervouchine, D.D. Towards long-range RNA structure prediction. Genes 2018, 9(6), 302
- Yue, Y.; Hou, S.;Wang, X.; Zhan, L.; Cao, G.; Li, G.; Shi, Y.; Zhang, P.; Hong,W.; Lin, H.; Liu, B.; Shi, F.; Yang,Y.; Jin, Y. Role and convergent evolution of competing RNA secondary structures in mutually exclusive splicing. RNA Biol 2017, 14, 1399–1410
- Yang, Y.; Zhan, L.; Zhang, W.; Sun, F.; Wang, W.; Tian, N.; Bi, J.; Wang, H.; Shi, D.; Jiang, Y.; Zhang, Y.; Jin, Y. RNA secondary structure in mutually exclusive splicing. Nat. Struct. Mol. Biol. 2011, 18, 159–168.
Splicing aberrations in human tissue transcriptomes

Aleksei Mironov¹, Stepan Denisov^{1,2}, Dmitri Pervouchine^{1,2}

¹ National Research University Higher School of Economics, Moscow Russia 125319 ² Skolkovo Institute for Science and Technology, Moscow Russia 145023 magmir71@gmail.com

Abstract. Recent advances in genomic studies and high-throughput sequencing have uncovered massive and heterogeneous repertoire of RNAs produced by alternative splicing. Although many alternative splicing isoforms are annotated and present in public databases, a growing number of unannotated isoforms is reported in each study. While some of them can represent natural variation of splicing across tissues and individuals, i.e., be splicing noise, others might be associated with disease. Functional alternative splicing in protein-coding RNA sequences is expected to preserve reading frame and follow tissue-specific patterns. Non-functional alternative splicing products are expected to be eliminated through nonsense-mediated decay (NMD) response.

In this work we study splicing aberrations, defined as shifts of splice sites from one splice site to a nearby splice site by not more than 30 nucleotides. We analyze a large panel of human RNA-seq data from Genotype-Tissue Expression Project (GTEx) and find distinct patterns of splicing aberrations in proteincoding and in non-coding regions. As expected, shifts by a multiple of three nucleotides preserving the coding frame are more frequent in coding regions compared to non-coding regions. Contrary to expectations, there are fundamental differences in the propensity of reading-frame-preserving aberrations among donor and acceptor splice sites, as well as different preferences in the upstream and the downstream shifts. Additionally, we analyze tissue-specificity of splicing aberrations and explore data from knock-down of nonsense-mediated decay pathway factors to distinguish functional alternative splicing from splicing noise.

Keywords: splicing noise, splicing sites, tissue specificity, nonsense-mediated mRNA decay.

1 Introduction

Genetic information transforms into a function in the process called gene expression, one of the key steps of which is splicing.

Splicing can generate many RNA isoforms from the same gene - then it is called alternative splicing. A lot of AS isoforms are annotated and are present in public databases like GENCODE [1]. However, current studies report a growing number of unannotated isoforms mostly without confirmation from proteomic data [2]. This raises a fundamental question: which fraction of isoforms is functional, and which represents the splicing noise caused by the stochasticity of the spliceosomal machinery?

In this study we try to shed light on this issue with deep diving into large amounts of human RNA-seq data and specifically studying aberrant splice site shifts, which we define as small shifts from major optimal splice sites to aberrant suboptimal sites.

We expect that the translated protein-coding regions fundamentally differ from the non-coding regions by the presence of splice shifts which lead to meaningful changes in amino acid sequence or do not cause any changes at all. In other words, we expect that aberrant shifts representing function and not the noise, lead to insertion, deletion or substitution of several amino acids without disruption of the other parts of the protein sequence. We refer to this feature of aberrant shifts as reading-framepreservation. Such splice shifts are expected to participate in gene regulation and demonstrate tissue-specific patterns.

In contrast, aberrant splice shifts in non-coding regions are not expected to demonstrate preferences in the preservation or alteration of the reading frame and tissuespecificity and thus are expected to mostly represent splicing noise whose RNA products may be eliminated by nonsense-mediated mRNA decay response [3].

2 Methods

We distinguish donor and acceptor shifts and measure them with a positive or negative number of nucleotides depending on the location of an aberrant site upstream or downstream from the annotated site. Thus, shifts to within upstream region are negative and shifts to within downstream region are positive (see Fig 1).



Fig 1. Donor and acceptor shift definition

Donor and acceptor shifts have opposite effects on exon length: positive donor shift leads to the extension of the exon while positive acceptor shift leads to shortening and vice versa. To indicate how efficiently sequences with aberrant shifts are spliced into transcripts we use specific PSIs inclusion measure which is the ratio between reads supporting shifted aberrant splice site (denoted as S) and reads supporting major annotated site (denoted as A) (see Fig 2). Only uniquely aligned reads are used in the calculation.



Fig 2. Calculation of PSIs

Exact determination of the reading-frame-preservation or alteration is complex in general and requires de novo reconstruction of the transcripts. However, we can easily estimate the contribution of each splice site shift into the preservation or the alteration of the reading frame: we classify splice site shift as reading-frame-preserving if its nucleotide value is a multiple of 3 and as reading-frame-altering otherwise (see Fig 3).



Fig 3. Estimation of the reading frame preservation

In RNA-seq data we don't have in vivo information on which mRNA transcripts are translated into proteins and which are not. Instead, we use information on major annotated splice sites provided by GENCODE.

If the annotated splice site is a boundary of a CDS exon we classify it as CODING, otherwise NON-CODING. However, the same gene may have multiple transcript isoforms generated through alternative promoters and through alternative splicing.

This means that the same annotated position may be an exon-intron boundary of the CDS in some transcripts and may be referred to the untranslated or excised intronic regions in other transcripts. Shifts of such positions are represented by the UTR/CODING class. The full list of classes also includes lincRNAs and other gene types which also undergo splicing and can harbor splice shifts.

Tissue-specificity of an aberrant shift is assessed using the Cramer's V measure [4] and bootstrap estimation of its confidence interval.

3 Data

We used large sample of human RNA-seq data from GTEx project, comprising about 8500 samples, each sample representing one tissue in one individual [5]. This data was processed with IPSA software for de novo identification of all the splice sites.

Then we used GENCODE version 19 annotation of the human transcriptome [1] and found all distinct shifted aberrant splice sites in pairs with the nearest annotated sites. Each distinct aberrant splice site is uniquely determined by its position in the genome and by its type as donor or acceptor.

We used RNA-seq data from UPF1 knock-down experiment in cell lines HepG2 and K562 to identify non-functional aberrant shifts that yield premature stop codons.

4 Results

In total we found about 43 thousand aberrant sites, including 3 times as many acceptors as donors.

In general, no specific tissue harbors much more or much less aberrations than other tissues, except for the testis, which harbors at least 1.5 times more aberrant shifts than any other tissue (see Fig 4).

4



Fig 4. Representation of aberrant shifts in tissues

The distribution of shift values among all aberrant donor sites demonstrates a notable abundance of positive shifts by 4 nt within intron region which alter the reading frame. No prevalence of reading-frame-preservation in CODING context is observed (see Fig 5).



Fig 5. Relative frequencies of donor shift values

We observed an even more interesting pattern among acceptor shifts. Shifts by 3, 4 and 5 nt to within exon region are the most abundant in all the contexts. Again, we did not observe the prevalence of reading frame preservation in the CODING context in comparison with intermediate UTR/CODING and non-coding contexts (see Fig 6).



We analyzed the distribution of inclusion level among donors and acceptors in coding versus non-coding context. The inclusion of reading-frame preserving acceptor shifts by 3 nt within exon region is slightly higher than the inclusion of other positive acceptor shifts. Negative acceptor shifts are noisier, but again we observed a slight surplus in the median inclusion level for -3 and -6 nt which preserve the reading frame.

Generally, the signal in the non-coding context is much noisier than in the coding context (see Fig 7).



Fig 7. Inclusion level distribution

We also analyzed splicing aberrations on the level of splice junctions to investigate the presence of the compensatory shifts preserving the reading frame. For this purpose, we imposed strict filters on the signal quality, read coverage, inclusion level and representation in samples.

Most of the splice junction shifts have only one of the splice sites shifted. Compensatory effect of reading frame preservation is rare and is observed mostly in the form of intron sliding when both acceptor and donor shifts have the same absolute value and the same sign with the zero net length change. Intron sliding leads to a substitution of one or several amino acids in the protein sequence without changes in the length of the transcript (Fig 8).



Numbers in cells indicate the fraction (%) of all aberrant junctions which have donor shift and acceptor shift values determined by the intersection

Fig 8. Distribution of the junction shifts

We further investigated the association of the reading frame preservation or alteration with the coding and non-coding contexts. Indeed, the effect of reading frame preservation is significant in CODING context in contrast to other contexts but there is still a significant fraction of aberrations with which alter the reading frame (see Fig 9).



Preservation / Alteration class non-coding contexts.

Error bars show 99% confidence interval based on distribution

Fig 9. Reading frame preservation in coding and non-coding contexts

5 Conclusion

Thousands of aberrant splice shifts are found across healthy human transcriptomes but most of them probably represent the splicing noise.

Aberrant shifts preserving the reading frame are more frequent in protein-coding regions compared to non-coding regions. Most of them occur in acceptor sites and follow a well-known NAGNAG pattern. Splice shifts in non-translatable regions have no tendency to preserve or alter the reading frame. Donor shifts by 4 nt within intron region are surprisingly frequent but seem to undergo massive elimination through NMD mechanism, which is to analyze accurately with UPF1 knock-down experiment data.

Some reading-frame-altering splice shifts may participate in specific gene regulation. To shed light on this issue, detailed analysis of tissue-specificity may be helpful.

References

- 1. Harrow J. et al. GENCODE: the reference human genome annotation for The ENCODE Project //Genome research. - 2012. - T. 22. - №. 9. - C. 1760-1774.
- 2. Tress M. L., Abascal F., Valencia A. Alternative splicing may not be the key to proteome complexity //Trends in biochemical sciences. - 2017. - T. 42. - №. 2. - C. 98-110.
- 3. Isken O., Maquat L. E. Quality control of eukaryotic mRNA: safeguarding cells from abnormal mRNA function //Genes & development. - 2007. - T. 21. - №. 15. - C. 1833-3856.
- 4. Cramer H. Mathematical methods of statistics Princeton University Press USA. 1946.
- 5. Melé M. et al. The human transcriptome across tissues and individuals //Science. 2015. -T. 348. – №. 6235. – C. 660-665.

Эволюция вторичной структуры мРНК Escherichia coli

Е. Ходжаева¹ и З. Червонцева^{2,3}

 ¹Биологический факультет МГУ им. Ломоносова, Москва 119234, Ленинские горы, д. 1, стр. 12, Россия
² ИППИ РАН, Москва 127051, Большой Каретный переулок, д.19 стр. 1, Россия

³ Сколтех, Москва 12105, Нобеля, д. 3, Россия

z.chervontseva@skoltech.ru

Синтез белка является одним из самых энергозатратных процессов в клетке, и его регуляция осуществляется на нескольких уровнях. Один из механизмов регуляции связан со вторичной структурой мРНК. В этой работе проведен анализ закономерностей эволюции вторичной структуры мРНК на основе полногеномных данных о вторичной структуре мРНК и данных о мутациях в 350 штаммах *E. coli*.

Keywords: мРНК, регуляция трансляции, Escherichia coli.

1 Введение

Из сравнения данных DMS-Seq и рибосомного профайлинга следует, что более структурированные мРНК хуже транслируются. Частично этот эффект обусловлен тем, что рибосомы расплетают структуру в активно транслируемых областях. Однако и при инактивации рибосом *in vivo* закономерность сохраняется: те мРНК, которые в норме больше транслируются, сворачиваются в менее стабильные структуры. [1]

Некоторые участки мРНК имеют обособленные регуляторные структуры определенной архитектуры: рибопереключатели, термосенсоры, терминаторы и т.п. Участки, мРНК, кодирующие плотные белковые домены одного белка, могут быть отделены друг от друга участками повышенной структурированности, вероятно, тормозящими рибосому и способствующими правильному сворачиванию доменов. [2] Однако основная часть последовательностей мРНК не имеет ни консервативной структуры, ни определенного паттерна распределения структурированных участков. В этой работе мы рассматриваем общие тенденции в эволюции структуры мРНК, оценивая точечные эффекты наблюдаемых замен на общую стабильность молекулы.

2 Результаты

Анализ частот мутаций показал, что, в противоположность ожидаемому, спаренные нуклеотиды мутируют чаще неспаренных (см. Рис. 1А). Этот эффект особенно выражен для наиболее частых мутаций из С в Т и, в какой-то степени, для менее частых мутаций из С в А. Напротив, вторые по частоте мутации, из А в G, реже происходят в спаренных участках (см. Рис. 1Б). Остальные типы мутаций относительно редки и их связь со структурой менее выражена.



Рис. 1. Число мутировавших позиций мРНК каждого типа зависит от вероятности замененного нуклеотида быть спаренным (Б). Все мутации показаны на левой гистограмме (А).

Известно, что мутации чаще происходят в определенных контекстах. Мы обнаружили, что вероятность нуклеотида быть спаренным тоже зависит от контекста (см. Рис 2А). Симуляции по методу Монте-Карло показали, что связь мутаций со спаренностью нуклеотида может быть обусловлена его контекстом (см. Рис. 2Б).

2



Рис. 2. Вероятность нуклеотида быть спаренным зависит от его нуклеотидного контекста (**A**), и перемешивание мутаций с сохранением контекстов дает такую же зависимость числа мутаций от вероятности спаривания, как и на рис. 1 (**Б**).

Для того, чтобы отделить закономерности эволюции, связанные именно со вторичной структурой, будет проведен анализ коэволюции структур мРНК генов эквимолярных субъединиц белковых комплексов, закодированных в одном опероне. Ранее мы показали, что уровни структурированности таких генов скоррелированы, что предположительно означает тонкую настройку уровня трансляции на уровне вторичной структуры мРНК, и таким образом, гены субъединиц белковых комплексов представляются хорошей моделью для изучения коэволюции вторичной структуры.

Это совместная работа с М. С. Гельфандом. Работа поддержана грантом РНФ 18-14-00358.

3 Литература

- 1. Burkhardt DH, Rouskin S, Zhang Y, Li GW, Weissman JS, Gross CA. Operon mRNAs are organized into ORF-centric structures that predict translation efficiency. Elife. 2017;6
- Faure G, Ogurtsov AY, Shabalina SA, Koonin EV. Role of mRNA structure in the control of protein folding. Nucleic Acids Res. 2016;44(22):10898-10911.

Поиск событий пост-транскрипционной перестановки экзонов в транскриптоме человека

Ольга Васюткина¹ и Дмитрий Первушин^{1,2}

¹ Факультет биоинженерии и биоинформатики, МГУ им. М.В. Ломоносова, Ленинские горы 1-73, 119899 Москва, Россия ² Сколковский институт наук и технологий, Нобеля 3, 121205 Москва, Россия d.pervouchine@skoltech.ru

Аннотация. Помимо канонического сплайсинга, у эукариот иногда встречается обратный сплайсинг, при котором донорный сайт сплайсинга расположен ближе к 3'-концу молекулы РНК, чем акцепторный сайт сплайсинга, а также транс-сплайсинг, при котором донорный и акцепторный сайты исходно принадлежат различным молекулам премРНК. Неканонический сплайсинг приводит к появлению РНК, порядок экзонов у которых не совпадает с порядком следования экзонов в геноме.

В литературе описаны случаи как перестановок, так и повторов экзонов в РНК относительно генома. Самым распространенным классом являются кольцевые РНК, а транскрипты с нециклическими перестановками (пост-транскрипционные перестановки) и удвоениями экзонов встречаются намного реже. Целью данной работы является идентификация пост-транскрипционных перестановок в транскриптоме человека с использованием данных высокопроизводительного секвенирования. На данный момент был произведен поиск РНК с перестановками экзонов в данных РНК-секвенирования человека (RNA-Seq) ENCODE и ICGC.

Ключевые слова: Сплайсинг, Пост-Транскрипционные Перестановки Экзонов.

У эукариот встречаются мРНК, которые не могут получиться при помощи канонического сплайсинга. Например, бывает «обратный» сплайсинг (backsplicing), в котором участвуют акцепторный сайт сплайсинга на 5'-конце экзона и донорный сайт сплайсинга на 3'-конце экзона. Если акцепторный сайт сплайсинга расположен ближе к 5'-концу мРНК, чем донорный сайт, то в результате такого сплайсинга образуются кольцевые РНК [1].

В литературе описаны мРНК, в которых порядок экзонов отличается от порядка тех же экзонов в геноме. Встречаются и перестановки, и повторы экзонов [2–4]. В линейной мРНК перестановка экзонов может быть следствием сплайсинга между двумя молекулами мРНК (транс-сплайсинга). Однако можно предположить, что сплайсинг одной пре-мРНК тоже способен привести к линейной мРНК с переставленными экзонами.

Рассмотрим ген с четырьмя экзонами (рис. 1). Заметим, что каждый донорный сайт сплайсинга потенциально может вступить в реакцию с тремя акцепторными сайтами. При каноническом сплайсинге интроны последовательно удаляются с образованием линейной зрелой мРНК (рис. 1А). При взаимодействии U1 мяРНП на донорном сайте первого экзона с U2/U2AF мяРНП на акцепторном сайте последнего экзона образуется линейная мРНК с пропуском двух экзонов. Пропущенные экзоны далее могут вступить в реакции сплайсинга внутри лассо, что приведет к образованию одной или двух кольцевых РНК (рис. 1В). Или наоборот, вначале образуются кольцевые структуры из средних экзонов, а затем соединяются крайние. Возможно представить и случай, когда один из средних экзонов пропущен и образует кольцо (рис. 1С).

При сплайсинге по схеме, представленной на рисунке 1D, получается одна линейная молекула мРНК с перестановкой средних экзонов вида 1-3-2-4, которая может быть полиаденилирована. Это гипотетический пример того, что «обратный сплайсинг» не обязательно приводит к образованию кольцевой РНК, и что перестановка экзонов может быть результатом цис-сплайсинга. Поиску именно таких транскриптов с пост-транскрипционными перестановками экзонов (Post-Transcriptional Exon Shuffling, PTES) в данных РНК-секвенирования человека и посвящена данная работа.

На данный момент разработан пайплайн, который позволяет найти риды с перестановками фрагментов вида 1-3-2-4 в данных секвенирования РНК и отличить их от кольцевых РНК. Поиск проведен в данных РНК-секвенирования ENCODE и ICGC. Некоторые из найденных ридов картируются на различные экзоны одних и тех же генов, другие – на интроны, 5'- или 3'-некодирующие области. Не обнаружены риды, у которых происходит разрыв только по аннотированным сайтам сплайсинга. Большинство находок в результатах ENCODE обнаружено в полиА+ фракции РНК в ядре. Среди находок в данных ICGC большинство ридов найдено в образцах раковой ткани. Эта работа выполнена в коллаборации с М.С. Гельфандом. Работа поддержана грантом РНФ 14-50-00150.



Рисунок 1. Соединение различных пар «донор, U1 – акцептор, U2» сайтов сплайсинга (красные стрелки) приводит к разным продуктам: линейной РНК (**A**), пропуску двух экзонов и образованию одной или двух кольцевых РНК (**B**), пропуску одного экзона, который также способен образовать кольцо (**C**), линейной РНК с перестановкой экзонов

(**D**). Темно-зеленым показано событие «обратного сплайсинга». В случаях В и С пунктирные стрелки обозначают альтернативные варианты сплайсинга.

Список литературы

- L. Szabo, J. Salzman, Detecting circular RNAs: bioinformatic and experimental challenges, Nat. Rev. Genet., 17 (2016) 679–692.
- H.H. Al-Balool, D. Weber, Y. Liu, M. Wade, K. Guleria, P.L.P. Nam, J. Clayton, W. Rowe, J. Coxhead, J. Irving, D.J. Elliott, A.G. Hall, M. Santibanez-Koref, M.S. Jackson, Post-transcriptional exon shuffling events in humans can be evolutionarily conserved and abundant, Genome Res., 21 (2011) 1788–1799.
- R. Rigatti, J.-H. Jia, N.J. Samani, I.C. Eperon, Exon repetition: a major pathway for processing mRNA of some genes is allele-specific, Nucleic Acids Res., 32 (2004) 441– 446.
- C.-Y. Yu, H.-J. Liu, L.-Y. Hung, H.-C. Kuo, T.-J. Chuang, Is an observed non-co-linear RNA product spliced in trans, in cis or just in vitro?, Nucleic Acids Res., 42 (2014) 9410– 9423.

От трансмиссивных раков к новым видам

Панчин А.Ю.¹ Алешин В.В.² и Панчин Ю.В.^{1,2}

¹ Институт проблем передачи информации им. А.А. Харкевича РАН ² НИИ физико-химической биологии им. А.Н. Белозерского МГУ аlexpanchin@yahoo.com ypanchin@yahoo.com Aleshin@genebee.msu.su

Абстракт. Лицевая опухоль тасманийского дьявола и венерическая саркома собаки – наиболее известные примеры трансмиссивных раковых опухолей. Подобное превращение раковых клеток в независимых паразитов известно также у хомяков и четырех видов моллюсков. Мы предполагаем, что некоторые упрощенные родственники более сложных современных животных могли эволюционировать из клеток трансмиссивных опухолей. Эта гипотеза предсказывает массовые потери ключевых генов, связанных с апоптозом у видов, упростившихся через стадию трансмиссивных раковых клеток. Из четырех рассмотренных нами таксонов упрощенных животных (dicyemida, orthonectida, myxosporea и trichoplax) паразитические книдарии myxosporea оказались наиболее подходящими кандидатами на подобное гипотетическое происхождение. В частности у всех пяти изученных видов этой группы отсутствуют все гомологи белка р53, а также все белки, содержащие каспазные домены или домены Bcl2. По количеству утраченных универсальных доменов, ассоциированных с апоптозом или раком, *myxosporea* обходят все контрольные виды, включая три вида одноклеточных родственников животных. Для дальнейшей проверки гипотезы мы предлагаем сравнить транскриптомы myxosporea и транскриптомы различных типов клеток неупрощенных родственных книдарий.

Ключевые слова: рак, видообразование, эволюция.

1 Введение

Опухоли можно рассматривать как результат эволюции клеток внутри организма [1, 2]. Иногда клетки приспосабливаются и выходят за пределы отведенных им тканей в форме метастаз, а иногда и за пределы исходного организма – в виде трансмиссивных раковых опухолей. Наиболее известные примеры трансмиссивных раковых опухолей – независимо возникшая не менее двух раз лицевая опухоль тасманийского дьявола [3] и венерическая саркома собаки, которая, по-видимому, возникла 11000 лет назад [4]. Подобные трансмиссивные раковые клетки, ставшие самостоятельными паразитами, находили у хомяков [5] и четырех видов моллюсков, причем один из них (*Polititapes aureus*) заражается клетками, ставшими раковыми в другом виде моллюска [3, 6]. Все это указывает на распространенность трансмиссивных раковых опухолей в природе и на их способность выживать в течении достаточно длительных сроков. Скорее всего подобные процессы имели место и раньше, а значит можно предположить, что некоторые современные упрощенные виды животных являются потомками трансмиссивных раковых клеток.

Согласно этой гипотезе, у таксонов, прошедших через такую необычную стадию видообразования, мы бы ожидали увидеть массовые потери ключевых генов, связанных с апоптозом, исчезновение систем, функций и клеточных процессов, нарушающихся при канцерогенезе [7], в том числе и универсальных для всех животных.

С этой точки зрения мы рассмотрели четыре таксона, которые являются сильно упрощенными, согласно современным филогенетическим представлениям [8]: *dicyemida*, *orthonectida*, *myxosporea* и *trichoplax*. Первые три представляют собой паразитов.

2 Методы

Мы использовали программу HMMER (http://hmmer.org/) и проанализировали на наличие доменов PFAM (https://pfam.xfam.org/) восемь видов, являющихся кандидатами на происхождение из трансмиссивных раковых клеток. Среди них пять видов myxosporea (Thelohanellus kitauei, Kudoa iwatai, Myxobolus cerebralis, Sphaeromyxa zaharoni, Enteromyxum leei), a также Trichoplax adhaerens, Dicyema sp. и Intoshia linei. В качестве контроля мы взяли 29 видов животных: Hydra vulgaris, Nematostella vectensis, Acropora digitifera, Polypodium hydriforme, Echinococcus granulosus, Gyrodactylus salaries, Schistosoma mansoni, Adineta vaga, Drosophila melanogaster, Eurytemora affinis, Strigamia maritime, Ixodes scapularis, Peripatopsis capensis, Gordionus alpestris, Hypsibius dujardini, Ascaris suum, Toxocara canis, Caenorhabditis elegans, Romanomermis culicivorax, Priapulus caudatus, Saccoglossus kowalevskii, Strongylocentrotus purpuratus, Ciona Homo sapiens, Branchiostoma floridae, Mnemiopsis leidyi, intestinalis, Pleurobrachia bachei, Amphimedon queenslandica, Oscarella carmela, a также 3 вида одноклеточных Holozoa – Monosiga brevicollis, Capsaspora owczarzaki и Sphaeroforma arctic.

Мы провели два типа анализа. Во-первых, мы проверили наличие в каждом геноме доменов, ассоциированных с раком. Такие домены определились с использованием списка "census of human cancer genes" [9] и по ключевому слову "apoptosis" в PFAM. Всего в PFAM нашлось 409 доменов, которые подходят хотя бы по одному из двух критериев и при этом встречаются не менее, чем в 15 из 29 видов животных (то есть является универсальными). Во-вторых, мы посмотрели универсальные для животных домены, которые отсутствуют у видов, рассматриваемых как кандидатов на происхождение из трансмиссивных раковых клеток.

3 Результаты

Пять видов *Myxosporea* утратили наибольшее количество доменов PFAM, ассоциированных с раком и апоптозом у других видов животных (Рис. 1). Это сравнение включает родственную для *Myxosporea* паразитическую книдарию *Polypodium* и три вида одноклеточных *Holozoa*. В частности, в отличие от всех остальных рассмотренных организмов, у всех пяти видов Myxosporea отсутствуют домены Caspase и Bcl2, играющих ключевую роль в апоптозе [10]. Поскольку у рассмотренных представителей Dicyemida, Orthonectida и Trichoplax мы нашли гены, связанные с апоптозом, содержащие данные домены, мы заключаем, что по отношению к этим таксоном, наша гипотеза, скорее всего фальсифицирована. Поэтому далее мы будем подробно рассматривать только группу Myxosporea.



Рис. 1. Число утраченных PFAM доменов у разных групп животных и одноклеточных Holozoa. Паразитические виды отмечены латинской буквой Р (Parasite).

У всех пяти исследованных видов Муховрогеа нет доменов Peptidase_C14 (PF00656) Caspase, Apoptosis regulator proteins domain Bcl-2 family (PF00452) и Caspase recruitment CARD domain (PF00619), которые встречаются в белке Apoptotic protease activating factor – 1 (Apaf-1), универсального и ключевого игрока, связанного с апоптозом как у позвоночных, так и беспозвоночных. Также у них отсутствует домен Calpain family cysteine protease (PF00648), вовлеченный в клеточную смерть и некроз [11] и домен Death domain (PF00531), присутствующий во многих белках, которые взаимодействуют с каспазами и регулируют апоптоз и воспаление. Домен NF-kappaB- RHD_DNA_bind domain

(PF00554) не найден у Мухоѕрогеа, как и P53 DNA-binding domain (PF00870), входящий в состав белка p53 – известного супрессора опухолей у человека. Также можно отметить отсутствие домена TB2/DP1 (deleted in polyposis) (PF03134). Этот домен встречается в белке HVA22, отсутствие которого приводит к семейному аденоматозному полипозу, аутосомному доминантному наследственному онкологическому заболеванию у людей. На основании геномного анализа можно предположить, что у Мухоѕрогеа отсутствует апоптоз, хотя это предположение требует экспериментальной проверки.

Сам факт утраты большого количества генов, в том числе связанных с апоптозом, не обязательно является следствием видообразования через переходную форму в виде трансмиссивной раковой опухоли. Возможно, что это просто следствие упрощения Myxosporea в результате их адаптации к паразитическому образу жизни. Действительно, геномы этих организмов очень маленькие и содержат небольшое количество генов [12].

Для дальнейшей проверки нашей гипотезы можно попробовать привлечь данные сравнительной транскриптомики. Например, известно, что на транскрипционном уровне клетки HeLa по-прежнему больше всего похожи на клетки шейки матки, из которых они произошли [13], а транскриптом клеток лицевой опухоли тасманийского дьявола больше всего похож на транскриптом швановских клеток [14]. Поскольку у Муховрогеа имеются книды – стрекательные органеллы [15], более правдоподобным кажется их происхождение из тотипотентных мигрирующих стволовых клеток (i-cells) родственных книдарий или более специализированных предшественников книдобластов или самих книдобластов, вернувших себе способность к пролиферации [16]. Анализ транскриптомов отдельных клеток книдарий помог бы проверить нашу гипотезу.

Мухоѕрогеа принадлежат к группе Мухогоа, в которой присутствует второй под-класс Malacosporea, который морфологически устроен гораздо сложнее [17]. Например, Buddenbrockia из этого подкласса, имеют сложную червеобразную жизненную стадию [18] и по-видимому обладают миоцитами [18], а так же вероятно имеют стадию похожую на балстулу и процесс похожий на гаструляцию [19]. У Мухоѕрогеа никаких признаков бластулы, эмбрионального развития или гаструляции пока не обнаружено. Возможно, у представителей Malacosporea будут обнаружены недостающие у Мухоѕрогеа гены, связанные с апоптозом, что было бы аргументом в пользу гипотезы о происхождении последних из трансмиссивного рака.

Стоит отметить, что хотя примеров трансмиссивных раков у книдарий еще не обнаружено, рак у представителей этой группы все же встречается [20]. Ближайшим родственником к Мухогоа внутри книдарий является Polypodium [12], который как и Мухогоа является паразитом рыб, но воспроизводится с использованием сложных свободноживущих медузоидных форм. В отличии от Polypodium, Мухогрогеа могут передаваться, когда одна рыба съедает другую [21]. Это и могло бы быть механизмом передачи, который и позволил трансмиссивному раку распространиться. Известен случай, когда человек заболел раком, изначально возникщем в паразитической цестоде *Hymenolepis nana*, которой пациент был заражен [22], то есть сценарий, когда рак паразита стал раком хозяина теоретически возможен.

4 Выводы

Среди четырех рассмотренных таксонов упрощенных животных, Мухоsporea обладают наибольшим числом признаков, которые мы бы ожидали увидеть у группы, возникшей из трансмиссивного рака: отсутствие ключевых доменов, связанных с апоптозом, и признаков эмбрионального развития, а также высокая степень упрощения морфологии, нетипичная даже для паразитов. Дальнейшая проверка гипотезы возможна с помощью методов сравнительной транскриптомики и геномного анализа родственных видов из группы Malacosporea.

5 Благодарности

Авторы выражают благодарность Кириллу Михайлову за замечания и обсуждение статьи. Эта работа была поддержана грантом РФФИ 18-29-13014 мк.

Список литературы

- Podlaha O, Riester M, De S, Michor F: Evolution of the cancer genome. Trends Genet 2012, 28(4):155-163.
- 2. Yates LR, Campbell PJ: Evolution of the cancer genome. Nat Rev Genet 2012, 13(11):795-806.
- Metzger MJ, Villalba A, Carballal MJ, Iglesias D, Sherry J, Reinisch C, Muttray AF, Baldwin SA, Goff SP: Widespread transmission of independent cancer lineages within multiple bivalve species. Nature 2016, 534(7609):705-709.
- Strakova A, Murchison EP: The cancer which survived: insights from the genome of an 11000 year-old cancer. Curr Opin Genet Dev 2015, 30:49-55.
- Copper HL, Mackay CM, Banfield WG: Chromosome Studies of a Contagious Reticulum Cell Sarcoma of the Syrian Hamster. J Natl Cancer Inst 1964, 33:691-706.
- Metzger MJ, Reinisch C, Sherry J, Goff SP: Horizontal transmission of clonal cancer cells causes leukemia in soft-shell clams. Cell 2015, 161(2):255-263.
- Hanahan D, Weinberg RA: Hallmarks of cancer: the next generation. Cell 2011, 144(5):646-674.
- Aleshin VV, Petrov NB: [Molecular evidence of regression in evolution of metazoa]. Zh Obshch Biol 2002, 63(3):195-208.
- 9. Futreal PA, Coin L, Marshall M, Down T, Hubbard T, Wooster R, Rahman N, Stratton MR: A census of human cancer genes. Nat Rev Cancer 2004, 4(3):177-183.
- Kim R: Unknotting the roles of Bcl-2 and Bcl-xL in cell death. Biochem Biophys Res Commun 2005, 333(2):336-343.

- 11. Goll DE, Thompson VF, Li H, Wei W, Cong J: The calpain system. Physiol Rev 2003, 83(3):731-801.
- Chang ES, Neuhof M, Rubinstein ND, Diamant A, Philippe H, Huchon D, Cartwright P: Genomic insights into the evolutionary origin of Myxozoa within Cnidaria. Proc Natl Acad Sci U S A 2015, 112(48):14912-14917.
- Landry JJ, Pyl PT, Rausch T, Zichner T, Tekkedil MM, Stutz AM, Jauch A, Aiyar RS, Pau G, Delhomme N et al: The genomic and transcriptomic landscape of a HeLa cell line. G3 (Bethesda) 2013, 3(8):1213-1224.
- Murchison EP, Tovar C, Hsu A, Bender HS, Kheradpour P, Rebbeck CA, Obendorf D, Conlan C, Bahlo M, Blizzard CA et al: The Tasmanian devil transcriptome reveals Schwann cell origins of a clonally transmissible cancer. Science 2010, 327(5961):84-87.
- 15. Canning EU, Okamura B: Biodiversity and evolution of the Myxozoa. Adv Parasitol 2004, 56:43-131.
- Muller WA, Teo R, Frank U: Totipotent migratory stem cells in a hydroid. Dev Biol 2004, 275(1):215-224.
- 17. Okamura B, Gruhl A: Myxozoa + Polypodium: A Common Route to Endoparasitism. Trends Parasitol 2016, 32(4):268-271.
- Jimenez-Guri E, Philippe H, Okamura B, Holland PW: Buddenbrockia is a cnidarian worm. Science 2007, 317(5834):116-118.
- Canning EU, Curry A, Hill SL, Okamura B: Ultrastructure of Buddenbrockia allmani n. sp. (Myxozoa, Malacosporea), a parasite of Lophopus crystallinus (Bryozoa, Phylactolaemata). J Eukaryot Microbiol 2007, 54(3):247-262.
- 20. Robert J: Comparative study of tumorigenesis and tumor immunity in invertebrates and nonmammalian vertebrates. Dev Comp Immunol 2010, 34(9):915-925.
- 21. Yokoyama H, Grabner D, Shirakashi S: Transmission Biology of the Myxozoa. Health and Environment in Aquaculture 2012.
- 22. Muehlenbachs A, Bhatnagar J, Agudelo CA, Hidron A, Eberhard ML, Mathison BA, Frace MA, Ito A, Metcalfe MG, Rollin DC et al: Malignant Transformation of Hymenolepis nana in a Human Host. N Engl J Med 2015, 373(19):1845-1852.

6

Inference of epigenomic momentum in single cells

Alexandra Pogorelskaya¹, Ruslan Soldatov² and Andrey Mironov¹

¹ Moscow State University, Moscow, 119991, Russian Federation ² Harvard Medical School, Boston, MA 02446, USA

Single-cell experiments are now widely used to determine cellular heterogeneity in complex tissues. While cell location in the expression or epigenomic space provides static snapshot, methods to discover dynamics are missing. Cell movement across the manifold is of high interest in differentiation and any other dynamic context analysis.

All cellular processes are guided by gene expression patterns which in turn is strictly regulated by chromatin accessibility with following transcription intensity. Therefore, current cell status could be well characterized by chromatin accessibility landscape provided for instance by ATAC-seq (assay for transposase-accessible chromatin by sequencing) [1]. Open chromatin regions may refer to regulatory sites and could be divided into promoter (P) regions (typically no further than 500 base pairs from transcription start site) and distal regulatory elements (DRE). We used the balance between ATAC-seq intensities of these region types to get the insight into gene regulation trend.

Having the fact that gene transcription starts with large complex formation on distal regulatory elements [2], we hypothesized that during cell differentiation or another time-dependent process we would observe distal element activation or deactivation earlier than that for promoter. In the case of steady state process, when transcription intensity hasn't changed for sufficient time, previous DRE_{intensity} multiplication appears with corresponding P_{intensity} multiplication, thus exhibiting linear connection in log-log grid:

$$\Delta lnP_{intensity} = \alpha \cdot \Delta lnDRE_{intensity}$$
(1)

where α is unknown parameter that could be estimated from several data timepoints for the current gene.

If the gene is up-regulated in certain timepoint, we would observe enhanced signal on the DRE compared with the linear model (1) with residuals being the evaluation of subsequent promoter activation.

We first applied our model for bulk hemopoietic data [3]. Blood cells type from bone marrow or peripheral blood were isolated by fluorescence activated cell sorting, giving seven unique stem and progenitor and six differentiated cell types spanning the myeloid, erythroid, and lymphoid lineages. Each lineage from hematopoietic stem cells to differentiated ones consisted of five developmental stages. For all cell types predicted residuals correlated with ATAC-seq intensity fold-change for upcoming types and negatively correlated with that for ancestral ones.

Single cell ATAC-seq dataset for *Drosophila* embryo development provided chromatic accessibility data for 23 085 cells at three time points [4]. Our residuals recapitulated dynamics of cell trajectories. At the apparent root state, a small group of cells had negligible residuals, and were surrounded on each side by cells with those pointing outwards in opposite directions.

This work is supported by RSF grant 14-50-00150.

References:

1. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ: Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Methods. 2013 Dec;10(12):1213-8.

2. Chen J et al.: Single-molecule dynamics of enhanceosome assembly in embryonic stem cells. Cell. 2014 Mar 13;156(6):1274-1285.

3. Howard Y Chang et al.: Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. Nat Genet. 2016 Oct;48(10):1193-203.

4. Cusanovich DA et al.: The cis-regulatory dynamics of embryonic development at single-cell resolution. Nature. 2018 Mar 22;555(7697):538-542.

Why are selectively neutral loci attracting?

Nadezhda A. Potapova^{1,2}, Anastasia V. Stolyarova³, Shamil R. Sunyaev⁴ Alexey S. Kondrashov^{2,5}

¹ IITP RAS, Bolshoy Karetny per. 19, build. 1, Moscow, 127051, Russia

 2 Lomonosov Moscow State University, ul. Leninski
ye Gory 1, Moscow, 119991, Russia

³ Skolkovo Institute of Science and Technology, Moscow, 143026, Russia

 $^4\,$ Genetics Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115, USA

 $^5\,$ University of Michigan, 500 S State St, Ann Arbor, MI 48109, USA

email: nadezhdalpotapova@gmail.com

Abstract. A tendency of alleles located in the same genotype to stay together or avoid each other, called linkage disequilibrium, is widely known. This phenomenon can be defined in a variety of ways, but in any case it may seems obvious that alleles with no effect on fitness should be distributed independently of each other. We investigated directional linkage disequilibrium between two selectively neutral alleles, such as synonymous substitutions or intron polymorphisms. Surprisingly, we observed that in *Drosophila melanogaster* and *Schizophyllum commune* synonymous substitutions within the same gene, including even those that affect different exons, tend to occur together. This attraction has been observed only when polymorphisms were polarized according to their frequencies, but not to their ancestral vs. derived status. Computer models confirmed that random drift alone produces attraction between linked, rare alleles at selectively neutral loci.

Keywords: linkage disequilibrium; synonymous substitutions; intron polymorphisms; drift.

Сравнение алгоритмов поиска ТАДов в данных Ні-С

Дмитрий Е. Мыларщиков¹, Александра А. Галицына^{2,3,4}

Факультет Биоинженерии и Биоинформатики МГУ им. Ломоносова
² Сколковский Институт Наук и Технологий
³ Институт Биологии Гена РАН
⁴ Институт Проблем Передачи Информации им. Харкевича РАН

dmitrymyl@gmail.com

Аннотация. Топологически ассоциированные домены являются одной из важнейших особенностей хроматина многих организмов. Для их поиска по данным Hi-C существует множество алгоритмов и программ. Чаще всего у них имеется один или несколько управляющих параметров, однако не существует универсальной стратегии поиска оптимальных значений параметров. В данной работе мы предпринимаем попытку решения этой проблемы, а также производим масштабный анализ самых популярных программ поиска ТАДов. Наши выводы позволяют выбрать наилучший алгоритм для поиска ТАДов по биологическим и техническим критериям, а также по скорости работы.

Ключевые слова: Ні-С, ТАДы, хроматин.

ДНК ядра образует ДНК-белковый комплекс, хроматин, который имеет сложную пространственную организацию. Проводить полногеномный анализ структуры хроматина стало возможным с появлением метода Hi-C, высокопроизводительной фиксации конформации хромосом [1].

Так, одними из первых особенностей хроматина были обнаружены топологически ассоциированные домены (ТАДы) -- последовательные участки генома, локусы которых контактируют чаще между собой, чем с локусами извне [2]. Было разработано множество программ по обнаружению ТАДов в данных Hi-C, использующих разные подходы к сегментации карты контактов [3]. У некоторых алгоритмов есть численный управляющий параметр (один или несколько), который влияет на итоговую сегментацию.

Сравнение некоторых программ поиска ТАДов показало, что одни алгоритмы справляются с задачей лучше, чем другие (как по воспроизведению симулированных данных, так и по сходимости результатов между техническими репликами) [4, 5]. Однако в этих работах не производился подбор значений управляющих параметров в достаточно широком диапазоне значений, что влияло на эффективность результатов соответствующих программ.

В данной работе мы проводим масштабный анализ управляющих параметров программ поиска ТАДов и предлагаем набор утилит по их подбору исходя из нескольких стратегий: технических и биологических. В технические стратегии входят воспроизведение симулированных данных (максимизация TPR, минимизация FDR) и воспроизводимость сегментаций между техническими репликами для реальных данных (максимизация индекса Жаккара с учётом среднего размера ТАДов). В биологические стратегии входит обогащение границ ТАДов пиками СТСГ для данных Hi-C человека и резкие сдвиги обогащения гистоновых меток на границах ТАДов для Hi-C дрозофилы, что отражает биологические свойства ТАДов в указанных организмах [6, 7].

Мы провели сравнение следующих алгоритмов: Armatus в версии для CLI [8], набор алгоритмов в Python-модуле lavaburst (Armatus, Modularity, Corner, Variance) [9], Directionality Index [2] и Insulation Score в Python-модуле TADtool [10], программы HiCseg [11], MrTADFinder [12] и HiCExplorer [13]. Симулированные данные были взяты из [4]. Реальные карты Hi-C были получены на клеточных линиях человека A549, HEK293, HepG2, в каждой по 2 биологические и 2 технические реплики [14].

Мы показали, что алгоритмы Armatus и Modularity наилучшим образом воспроизводят симулированные данные и показывают наилучшую сходимость сегментаций между техническими репликами при правильном подборе управляющего параметра. При этом две разные имплементации Armatus показали частичное несоответствие между полученными сегментациями даже для лучшего значения управляющего параметра. Помимо этого, мы произвели оценку времени работы алгоритмов и сделали вывод, что два данных алгоритма работают быстрее Directionality Index, HiCseg, Corner и Variance. Быстрый метод Insulation Score тем не менее не дает точных результатов в нашем исследовании.

Мы произвели оценку устойчивости алгоритмов к шуму в симулированных данных. Хуже всего себя показали HiCseg, Insulation Score и Directionality Index, тогда как остальные алгоритмы продемонстрировали стабильно высокий уровень TPR и стабильно низкий уровень FDR.

Данная работа поддержана грантом РФФИ 17-00-00180, а также грантом "Systems biology Fellowship Program".

Ссылки

- Lieberman-Aiden, E., et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science, 326(5950), 289–93 (2009). https://doi.org/10.1126/science.1181369
- Dixon J. R., et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature, 485(7398), 376–380 (2012). https://doi.org/10.1038/nature11082
- Lajoie, B. R., et al. The Hitchhiker's guide to Hi-C analysis: Practical guidelines. Methods, 72, 65–75 (2015). https://doi.org/10.1016/J.YMETH.2014.10.031
- Forcato, M., et al. Comparison of computational methods for Hi-C data analysis. Nature Methods, 14(7), 679–685 (2017). https://doi.org/10.1038/nmeth.4325
- Dali, R. and Blanchette M. A critical assessment of topologically associating domain prediction tools. Nucleic Acids Research, 45(6), 2994–3005 (2017)
- Rao S. S. P., et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell, 159(7), 1665–1680 (2014)

- 7. Ulianov, S.V., et al. Active chromatin and transcription play a key role in chromosome partitioning into topologically associating domains. Genome research, 26(1), 70–84 (2015)
- 8. Filippova D., et. al. Identification of alternative topological domains in chromatin. Algorithms for Molecular Biology, 9:14 (2013)
- 9. Lavaburst, Python library for TAD calling https://github.com/nvictus/lavaburst/tree/master/lavaburst version from 07.02.2018
- Kruse, K., et al. TADtool: visual parameter identification for TAD-calling algorithms. Bioinformatics, 32(20), 3190–3192 (2016) https://doi.org/10.1093/bioinformatics/btw368
- Lévy-Leduc, C., et al. Two-dimensional segmentation for analyzing Hi-C data. Bioinformatics 30(17): i386-92 (2014) https://doi.org/10.1093/bioinformatics/btu443
- Yan, K., et al. MrTADFinder: A network modularity based approach to identify topologically associating domains in multiple resolutions. PLOS Computational Biology 13(7): e1005647 (2017) https://doi.org/10.1371/journal.pcbi.1005647
- Ramirez, F., et al. High-resolution TADs reveal DNA sequences underlying genome organization in flies. Nature Communications, 9(189) (2018) https://doi.org/10.1038/s41467-017-02525-w
- 14. Protocol of Hi-C data download from GEO and processing: https://github.com/agalitsyna/ChromatinCompartments version from 28.05.2018

Исследование структуры хроматина Dictyostelium discoideum

Цой О.В. $^{1,2},$ Галицына А.А. 1, Храмеева Е.Е $^{1,2},$ Ульянов С.В 3

¹Сколковский институт науки и технологии ² Институт проблем передачи информации им. А.А. Харкевича, РАН ³Институт биологии гена, РАН

Аннотация. Dictyostelium discoideum - клеточный слизевик является одним из важных модельных организмов. В ходе жизненного цикла он находится в одноклеточном и многоклеточном состоянии. В этой работе изучается структура хроматина этого организма в целом, а также ее изменение в ходе развития.

Ключевые слова: хроматин, Hi-C.

1 Введение

Dictyostelium discoideum может находится в двух формах - одноклеточной и многоклеточной. При неблагоприятных условия одиночные особи образуют сначала агрегаты, а после многоклеточные плодовые тела, в которых формируются споры для дальнейшего размножения распространения. В 2005 году была опубликована последовательность генома *D.discoideum*. Его геном обладает высокой концентрацией генов, а также очень низким GC-составом (22,43%) и большим содержанием повторов. Возникает вопрос, как организован хроматин *D. discoideum*, а также, как его структура может изменяться при переходе от одноклеточной формы к многоклеточной.

2 Материалы и методы

Из данных Hi-C экспериментов получены матрицы контактов. С помощью алгоритма Armatus предсказаны границы возможных топологически ассоциированных доменов (ТАДов). Петли предсказаны с помощью визуальной оценки, которая сравнивалась с автоматическими предсказаниями программами hiccups и dotfinder. Дальнейший анализ проводился с использованием python и его графических модулей. Данные по экспрессии генов получены с помощью RNA-Seq. Картирование ридов проводилось программой hisat-2 с параметрами по умолчанию, кроме максимальной длины интронов. Подсчет количества ридов проводился программами htseq и bedtools с предварительной фильтрацией на риды с уникальным картированием. Для поиска

дифференциально экспрессируемых генов использовался пакет DESeq2. Дальнейший анализ проводился также с помощью R, python и его графических модулей.

3 Результаты

Были получены карты контактов для разных стадий жизненного цикла *D.discoideum* - начиная с одноклеточного состояния (0 час), через переходные состояния (2 и 5 час), к многоклеточной форме (8 час). Визуальное изучение показало наличие структур, напоминающих так называемые топологически ассоциированные домены (ТАДы) и петли. Также получены данные по экспрессии генов для тех же стадий развития *D. discoideum*. Анализ дифференциальной экспрессии показал, что экспрессия генов, отвечающих за развитие, сигнальные каскады, адгезию и формирование плодовых тел, повышается на стадиях 5 и 8 часа.

Структуры, напоминающие ТАДы, занимают около 85-90% генома. Определены характеристики подобных структур (количество, средняя длина, распределение длин, изменчивость в ходе развития), а также связь с расположением и экспрессией генов.

Структуры, напоминающие петли, на карте контактов *D. discoideum* часто образуют регулярные скопления. Такие регулярные образования обнаружены впервые. К сожалению, существующие алгоритмы поиска петель (hiccups, dotfinder) в текущем состоянии не подходят для автоматического предсказания таких структур, поэтому необходима либо их оптимизация, либо создание новых инструментов. С помощью визуальной оценки определены границы петель и далее получены некоторые характеристики (количество, средний размер, консервативность, GC-состав). Сопоставление данных экспрессии генов и организации хроматина показало, что гены с высокой экспрессией чаще располагаются в петлях, по сравнению с их окружением.

Работа поддержана грантом РФФИ 17-00-00180.

2

Оптимальное количество технических реплик для оценки аллельного дисбаланса в эксперименте PHK-секвенирования

Ася Менделевич¹, Светлана Виноградова², Андрей Миронов^{1,3}, Александр Гимельбрант²

¹ Сколковский институт науки и технологий, Москва 143026, Россия ² Гарвардская Медицинская Школа, Бостон 02115, США

 3 Московский Государственный Университет, Москва 119991, Россия

a.mendelevich@skoltech.ru

Аннотация В этой работе мы исследовали уровень технического шума при определении аллельного дисбаланса в экспериментах РНКсеквенирования. Мы разработали метод построения эксперимента, позволяющий оценивать аллельный дисбаланс для данного уровня покрытия с заданной точностью. Метод основан на статистистическом анализе свойств большого количества реплик.

Ключевые слова: Аллельный дисбаланс, экспрессия генов, PHKсеквенирование.

Аллельный дисбаланс в экспрессии генов является распространённым и не вполне изученным феноменом [Gimelbrant et al.,Savova et al.]. Аллельным дисбалансом мы называем статистически значимое отклонение в экспрессии в сторону материнского или отцовского аллеля.

Существуют различные методы определения дисбаланса. Мы будем использовать данные PHK-секвенирования и оценивать аллельный дисбаланс на основе количества чтений, покрывающих однонуклеотидные полиморфизмы в гетерозиготном состоянии: нас интересует отношение количества чтений, покрывающих материнский аллель, к общему количеству чтений, покрывающих выделенный нуклеотид. Метод, используемый для оценки данного отношения при построении графиков ниже, основан на методе, описанном в [Castel et al.].

Несмотря на то, что РНК-секвенирование дает достаточно точные результаты, оно также подвержено различным шумам, которые можно разделить на три основные группы: биологические — представляющие собой естественную вариабельность экспрессии в клетках, экспериментальные появляющиеся на этапах подготовки образцов и секвенирования, и аналитические — происходящие при выравнивании и последующей обработке данных. Наша работа была направлена на уменьшение влияния экспериментального шума на результаты анализа. Одним из методов борьбы с экспериментальным шумом является создание так называемых технических реплик — нескольких библиотек из одного биологического образца. Технические реплики не уменьшают биологическую вариабельность, но позволяют получать более качественную оценку состояния РНК в клетках на момент приготовления образца.

Если рассматривать пары технических реплик, аллельный дисбаланс оказывается плохо скоррелированным (ниже приведён пример для двух технических реплик для F1 кросса двух имбредных линий мышей: 129S1 и CAST [Puc.1 A]). Однако, на текущий момент, большинство работ содержат в себе анализ секвенирования одной технической реплики на биологический образец. Например, в проекте GTEX был проведен масштабный аллельспецифический анализ без технических реплик [Battle et al.]. При этом даже при предположениях, что данные приходят с идеального распределения, и шум обеспечивается исключительно статистическими причинами, на одной технической реплике детектируется только сильно выраженный дисбаланс [Puc.1 C].



Рис.1 А. Корреляция аллельного дисбаланса по генам для двух технических реплик для F1 кросса мыши. В. Распределения корреляций между рандомными выборками 20% генов с объединением по всем техническим репликам, для всех возможных п-комбинаций из 6 технических реплик, с сохранением общей глубины покрытия. С. Чувствительность для комбинаций из разного числа технических реплик в 10% интервалах аллельного дисбаланса.

На технические характеристики результата секвенирования и последующего анализа влияет большое количество факторов, включая модель секвенатора, количество РНК в исследуемом образце и технологию приготовления библиотеки.

Очевидно, что увеличение количества реплик и средней глубины покрытия ведёт к улучшению точности результата [Puc.1 B], так как первое помогает получить достаточную выборку для оценок параметров распределения для каждого гена, а второе делает возможным делать статистически значимые выводы на большем числе генов.

В условиях ограниченных ресурсов, при построении эксперимента основными направлениями для оптимизации являются дизайн эксперимента, количество биологических и технических реплик, количество материала и глубина покрытия.

Отметим, что наш анализ предполагает некоторые допущения: мы неявно предполагаем поликлональную структуру ткани и приорное распределение Пуассона над позициями однонуклеатидных замен.

Мы остановились на нескольких потенциально полезных в построении эксперимента стратегиях использования оценки технического шума:

- Нахождение статистически значимо детектируемого "шага" в аллельном дисбалансе на имеющихся данных, при заданном уровне покрытия на таргетных генах и определённом уровне значимости. Оценка количества дополнительных реплик, необходимых для достижения желаемой точности на таргетных генах.
- Оценка технического шума в конкретном эксперименте путём экстраполяции результатов с множества технических реплик для малого количества образцов на другие биологические реплики.

Эта работа поддержана грантом РФФИ 17-00-00180.

Список литературы

- [Gimelbrant et al.] Gimelbrant, A., Hutchinson, J. N., Thompson, B. R., Chess, A.: Widespread monoallelic expression on human autosomes. Science (2007)
- [Castel et al.] Castel, S. E., Levy-Moonshine, A., Mohammadi, P., Banks, E., Lappalainen, T.: Tools and best practices for data processing in allelic expression analysis. Genome Biology (2015)
- [Battle et al.] Battle, A., Brown, C. D., Engelhardt, B. E., Montgomery, S. B. et al.: Genetic effects on gene expression across human tissues. Nature (2017)
- [Savova et al.] Savova, V., Vinogradova, S., Pruss, D., Gimelbrant, A. A., Weiss, L. A.: Risk alleles of genes with monoallelic expression are enriched in gain-of-function variants and depleted in loss-of-function variants for neurodevelopmental disorders. Molecular Psychiatry (2017)

Сравнительный анализ геномов бактерий рода Streptococcus

Павел Шелякин¹⁻²⁻³, Ольга Бочкарёва²⁻³ и Анна Карань⁴

¹ Институт общей генетики им. Н.И. Вавилова Российской академии наук, Москва, Россия

² Институт проблем передачи информаци им. А.А. Харкевича Российской академии наук, Москва, Россия

³ Сколковский институт науки и технологий, Москва, Россия

⁴ Факультет биоинженерии и биоинформатики, Московский Государственный Университет имени М.В.Ломоносова, Москва, Россия

f.serval@gmail.com

Аннотация. Секвенирование геномов многих штаммов одного вида бактерий показало, что даже у близких штаммов бактерий может сильно различаться генный состав, то есть в геноме одного штамма до 25-30% генов могут не иметь родственных генов в геноме другого штамма. На основании этого наблюдения для описания бактериальных видов (или более высоких таксономических единиц) была создана концепция пангенома, т. е. совокупности генов, которые могут наблюдаться у представителей данной таксономической единицы. В настоящей работе мы изучили структуру и состав пан-генома 3 видов бактерий из рода Streptococcus: S.pyogenes, S.pneumoniae и S.suis.

Ключевые слова: пан-геном, *Streptococcus*, горизонтальный перенос генов.

1. Введение

Секвенирование геномов многих штаммов одного вида показало, что геном одного штамма не может служить хорошим описанием для генома всего вида, так как разные штаммы могут отличаться друг от друга по генному составу на 20-35% [1]. Для описания видов была предложена концепция пан-геномов, то есть общей совокупности генов, которые могут быть у представителя данного вида [1,2,3]. Пан-геном состоит из нескольких фракций генов, включающих коровые гены, имеющиеся у всех представителей вида, вспомогательные гены, имеющиеся у доли представителей вида, и уникальные или штамм-специфичные гены. Пан-геном называют открытым, если при добавлении

нового штамма его размер продолжает расти, или закрытым, если после добавления некоторого количества штаммов его размер перестаёт меняться [1].

Фракции пан-генома могут отличаться не только размером, но также по функциональному составу [4]. Обычно коровые гены представляют собой гены "домашнего хозяйства", тогда как вспомогатедьные и уникальные гены, кодируют белки, способствующие приспособлению к конкретным экологическим нишам, например отвеающие за антибиотиковую устойчивость [5]. Поэтому, можно ожидать, что гены из разных фракций пан-генома эволюционируют в разных режимах, и будут характеризоваться разной подверженностью горизонтальному переносу, разной скоростью потери/ приобретения и разной силой действующиего на них отбора.

2. Методы

Для анализа из базы данных NCBI скачали полные геномы 25 штаммов Streptococcus suis, 50 штаммов Streptococcus pyogenes, и 28 штаммов Streptococcus pneumoniae. Все гены указаных бактерий были распределены по ортологическим рядам с помощью программы Proteinortho v5.13. Каждому ортологическому ряду была сопоставленна функция, сила действующего на него отбора, подверженность горизонтальным переносам, а также содержание генов вирулентности. Также была оценена сила отбора, действующего на межгенные участки генома.

3. Результаты

В итоге мы привели оценку размера пан-генома и показали, что он является открытым. Как и ожидалось разные его фракции оказались обогащены разными функциями и были подвержены отбору разной силы. Межгенные участки, расположенные около генов из разных фракций, также показали разную подверженность отбору.

Это совместная работа с Михаилом Гельфандом. Настоящая работа была поддержена грантом РНФ 18-14-00358.

Список литературы:

- Medini, D., Donati, C., Tettelin, H., Masignani, V., Rappuoli, R.: The microbial pangenome. Current opinion in genetics & amp; development 15(6), 589–594 (2005)
- Tettelin, H., Masignani, V., Cieslewicz, M.J., Donati, C., Medini, D., Ward, N.L., Angiuoli, S.V., Crabtree, J., Jones, A.L., Durkin, A.S., et al.: Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: implications for the microbial "pan-genome". Proceedings of the National Academy of Sciences of the United States of America 102(39), 13950–13955 (2005)

- 3. Hogg, J.S., Hu, F.Z., Janto, B., Boissy, R., Hayes, J., Keefe, R., Post, J.C., Ehrlich, G.D.: Characterization and modeling of the Haemophilus influenzae core and supragenomes based on the complete genomic sequences of rd and 12 clinical nontypeable strains. Genome biology 8(6), 103 (2007)
- Gordienko, E.N., Kazanov, M.D., Gelfand, M.S.: Evolution of pan-genomes of Escherichia coli, Shigella spp., and Salmonella enterica. Journal of bacteriology 195(12), 2786–2792 (2013)
- Muzzi, A., Masignani, V., Rappuoli, R.: The pan-genome: towards a knowledge-based discovery of novel targets for vaccines and antibacterials. Drug discovery today 12(11), 429–439 (2007)

Предковая реконструкция белка ОхуR

Наталия Драненко¹ и Ольга Бочкарёва^{2,3}

¹МФТИ, Долгопрудный 141700, Россия ² ИППИ РАН, Москва 143182, Россия ³ Сколковский институт науки и технологии, Инновационный центр Сколково, Москва 143026, Россия dranenko@phystech.edu

Аннотация. Предковая реконструкция является одним из методов биоинформатики. Получение предковой последовательности белка позволяет синтезировать этот белок, изучить его физико-химические свойства. Более того, на основе свойств полученного белка можно сделать вывод об условиях среды, для работы в которых был адаптирован этот белок. В данной работе была реконструирована предковая последовательность белка ОхуR, являющегося одним из белков ответа на окислительный стресс в бактериях. Предковая последовательность имеет длину в 343 аминокислоты, в полученном белке определяются два домена, ДНК-связывающий и субстрат-связывающий, что соответствует структуре белков-потомков.

Ключевые слова: ОхуR, предковая реконструкция

Введение

Растущее число белковых последовательностей, доступных в открытых базах данных вместе с развитием методов филогенетического анализа, позволяют восстанавливать предковые последовательности белков.

Окислительный стресс, вызванный реакционноспособными формами кислорода, является одним из неизбежных кризисных состояний лпя подавляющего большинства аэробных организмов, так как реакционносопособные кислородные являются метаболическими формы побочными продуктами клеточного дыхания. Большинство бактерий контролирует окислительный стресс с помощью белка ОхуR, чувствительного к H_2O_2 . Этот белок активирует экспрессию защитных генов [1]. В каждом геноме встречается только по одной копии гена, кодирующего этот белок, что позволяет провести предковую реконструкцию.

Получение предковой последовательности белка позволяет синтезировать этот белок, изучить его физико-химические свойства. Более того, на основе свойств полученного белка можно сделать вывод об условиях среды, для работы в которых был адаптирован этот белок [2].

Методы

В данной работе были использованы 653 последовательности белка OxyR из различных бактериальных таксонов, доступные в базе ортологов OMA (https://omabrowser.org).

<u>Фильтрация последовательностей</u> была произведена с помощью программы BLAST [2]. В анализ были включены последовательности, для которых e-value сходства составило не более 1*10⁻⁵¹.

Дополнительная проверка на наличие соответствующих <u>функциональных</u> <u>доменов</u> в каждой из отобранных последовательностей была произведена с помощью программы InterProScan [https://www.ebi.ac.uk/interpro/about.html].

<u>Оценка сходства</u> последовательностей и <u>разбиение их на кластеры</u> по степени сходства происходило с помощью программы CD-HIT [http://weizhonglilab.org/cd-hit/]. Было построено 2 разбиения, с порогом сходства в 90 и 80% при использовании версии алгоритма для аминокислот.

<u>Множественное выравнивание</u> последовательностей было выполнено с помощью MUSCLE [https://www.ebi.ac.uk/Tools/msa/muscle/].

Для построения <u>филогенетического дерева</u> использовалась программа PHYLIP и алгоритм Neighbor-Joining (метод присоединения соседей) [http://evolution.genetics.washington.edu/phylip/progs.algs.heur.html].

Для реконструкции предковой последовательности в узле дерева и уточнения филогении использовалась программа **MrBayes** [http://mrbayes.sourceforge.net]. При запуске для каждой группы последовательностей была указана одна фиксированная клада на дереве с максимальной поддержкой. Первоначальное число поколений при каждом запуске: 20000. Используемая матрица замен: BLOSUM. Используемая модель: invgamma. Критерий окончания анализа: ASDSF не превосходит 0.05. При превышении этого значения после 20000 поколений добавлялись дополнительные поколения. Производилось не меньше четырех запусков для контроля того, что каждый запуск приводит к восстановлению одной и той же предковой последовательности.

Результаты

Было отобрано 500 последовательностей, принадлежащих семейству ОхуR на основе функциональной аннотации и предсказания доменной структуры.

В силу большого расхождения последовательностей исследуемого белка и наличия в данных некоторого количества случаев неверной аннотации, множественное выравнивание всех последовательностей не является достаточно надежным для использования в предковой реконструкции. Поэтому было построено филогенетическое дерево по матрице парных расстояний, что позволило разбить имеющиеся последовательности на 7 групп, соответствующих крупным кладам этого дерева, и восстановить предка каждой из них и общего предка.
Было показано, что ДНК-связывающий домен является довольно консервативным, а отличие от субстрат-связывающего домена, который накопил большое число изменений (Рис. 1).

TayoGroupAncestor PseudoGroupAncestor CoryneGroupAncestor CoryneGroupAncestor CoryneGroupAncestor	ī				
common Ancestor 	ł	+	 H	H	15

Рис.1. Филогенетическое дерево (слева) и выравнивание (справа) полученных предковых последовательностей.

Следующим этапом работы будет анализ соответствия консервативных и быстро эволюционирующих участков элементам пространственной структуры белка.

Белок, имеющий предсказанную аминокислотную последовательность, будет синтезирован *in vitro* и охарактеризован по физико-химическим свойствам.

Благодарности

Эта работа была выполнена в УНЦ Биоинформатика под руководством М.С. Гельфанда и поддержана грантом РНФ 18-14-00358.

Список литературы

- Qing Wei Phu Nguyen Le Minh, Andreas Dötsch, Falk Hildebrand, Warunya Panmanee, Ameer Elfarash, Sebastian Schulz, Stéphane Plaisance, Daniel Charlier, Daniel Hassett, Susanne Häussler, and Pierre Cornelis. Global regulation of gene expression by OxyR in an important human opportunistic pathogen. Nucleic Acids Res. 2012 May; 40(10): 4320– 4333.
- Joy JB, Liang RH, McCloskey RM, Nguyen T, Poon AFY. Ancestral Reconstruction. *PLoS Computational Biology*. 2016;12(7):e1004763. doi:10.1371/journal.pcbi.1004763.

Раскраска хроматина с помощью стохастического автокодировщика

Кудрин Р.А., Миронов А.А.

Введение

ДНК в ядре эукариот упакована очень эффективно, о чём свидетельствует то, что в развёрнутом состоянии длина ДНК, например, генома человека составляет около 2-х метров [1], а средний диаметр ядра клеток млекопитающих - около 6 микрометров [2].

В компактизации ДНК участвуют многочисленные белки, связывающиеся с ДНК в ядре. Эти белки традиционно подразделяют на два широких класса: гистоны и негистоновые хромосомные белки. Комплекс белков обоих классов с ядерной ДНК клеток эукариот составляет хроматин. Хроматин — это очень динамичная структура, способная менять свою конфигурацию при получении множества различных сигналов.

Известно, что хроматин имеет два основных принципиально разных состояния. Первое - относительно релаксированное, «открытое». В таком хроматине расположено большинство транскрипционно активных генов, он называется эухроматином. Второе - довольно компактное, в каком находятся, например, центромеры и теломеры, в которых находятся транскрипционно неактивные гены. Это состояние называется гетерохроматином [3].

Помимо двух основных состояний, ученые предпринимают попытки выделить более частные состояния хроматина. В наиболее известной работе, была предложена скрытая Марковская модель для разметки хроматина в 9 человеческих клеточных линиях по бинаризованным (есть пик/нет пика) трекам гистоновых модификаций [4]. Авторы сфокусировались на модели с 15-ю состояниями, но построили также расширенную модель с 18-ю.

Результаты

В настоящей работе для поиска состояний ("раскраски") хроматина был использован относительно новое изобретение в области глубоких нейросетей - стохастический автокодировщик с дискретной скрытой переменной [5]. Это нейросетевой первый пример использования данной архитектуры исключительно для кластеризации. Автокодировщик обучался кодировать значения кратных изменений количества чтений по сравнению с контролем экспериментов ChIP-seq 6 гистоновых модификаций, которые по ИЗ спецификации проекта ENCODE должны быть во всех референсных эпигеномах, в дискретную скрытую переменную, а затем раскодировать обратно. Качество восстановление автокодировщика сравнимо с качеством линейного восстановления по 5 главным компонентам. Было проведено полученными ChromHMM, сравнение состояний с таковыми, ОНИ значительно отличаются. Полученные автокодировщиком состояния были проаннотированы по насыщенности гистоновыми модификациями, энхансерной активности STARR-seq, промотерной активности по RNA-seq, доступности хроматина по DNAse-seq, трекам генов, интронов и экзонов, а также А- и В-компартментам, полученных из экспериментов HiC.

Список Литературы

[1] Alberts et al, Molecular Biology of the Cell, 2002

[2] McGraw Hill Encyclopedia of Science and Technology. New York: McGraw Hill, 1997

[3] Encyclopedia of molecular cell biology and molecular medicine, 2nd ed.

Weinheim: Wiley-VCH Verlag, 2004.

[4] Ernst et al., Nature, 2014

[5] Jang et al, arXiv:1611.01144 [cs, stat], 2016

Исследование свойств РНК-ДНК контактов в хроматине

Анастасия А. Жарикова¹, Александра А. Галицына^{2,3,4}, Алексей А. Гаврилов⁴, Мария Д. Логачева¹, Сергей В. Разин^{4,5} и Андрей А. Миронов^{1,2}

¹ Факультет Биоинженерии и Биоинформатики МГУ им. Ломоносова
 ² Институт Проблем Передачи Информации им. Харкевича РАН
 ³ Сколковский Институт Наук и Технологий
 ⁴ Институт Биологии Гена РАН
 ⁵ Биологический Факультет МГУ им. Ломоносова

azharikova89@gmail.com

Аннотация. Данная работа посвящена масштабному анализу результатов экспериментов по определению РНК-ДНК взаимодействий в хроматине. Был выполнен анализ собственных экспериментов, а также экспериментов GRID-Seq на клетках человека. Дополнительно проведено сравнение с RNA-Seq. Мы выявили фракцию хроматин-ассоциированных РНК, провели анализ их свойств связывания с ДНК. Корректность протокола подтверждена для некоторых РНК с известными свойствами (XIST, MALAT1). Обнаружены ранее неаннотированные хроматинассоциированные РНК, отсутствующие в RNA-Seq. Мы произвели ассоциацию РНК-ДНК контактов с данными ДНК-ДНК взаимодействий хроматина, обнаружили предпочтение хроматина из областей вне топологически ассоциированных доменов (ТАДов) и из активного компартмента. Ключевые слова: РНК-ДНК взаимодействия, хроматин, некодирующие РНК, секвенирование.

Известно, что большая часть ДНК эукариот транскрибируется с образованием мРНК, рРНК, тРНК, а также разнообразных длинных и коротких некодирующих РНК. При этом РНК могут оставаться в ядре [1] и выполнять функции регуляции транскрипции, ремоделирования хроматина, а также поддерживать пространственную структуру [2].

Ранее был разработан арсенал экспериментальных методик, выявляющих взаимодействующие друг с другом молекулы РНК и локусы ДНК (RNA Antisense Purification [3], ChIRP-seq [4] и CHART [5]). Эти методы позволяют регистрировать единичные случаи взаимодействия или контакты нескольких заранее известных РНК. В последнее время появляются методы полногеномного анализа РНК-ДНК контактов, лишенные этого ограничения (MARGI [6], ChAR-Seq [7] и GRID-Seq [8], а также метод, предложенный нами ранее [9]).

В данной работе мы разработали протокол обработки полногеномных данных секвенирования РНК-ДНК контактов, включающий этапы: триммирования по качеству, удаления адаптеров, картирования, фильтрации по сайтам рестрикции

и аннотации данных. С его помощью мы провели анализ собственных данных для клеток K562 и фибробластов человека, а также данных GRID-Seq для клеток MDA-MB- 231 и MM.1S человека.

Мы обнаружили потенциальный вклад ДНК-ДНК контактов в результаты наших экспериментов, а также предлагаем способ удаления таких случаев (по критериям расстоянию до сайта рестрикции по РНК-части, близости позиции картирования РНК и ДНК частей).

Мы провели валидацию экспериментов и их обработки для РНК MALAT1 и XIST. Данные РНК-ДНК контактов ассоциированы с RNA-Seq.

Особое внимание при обработке данных уделено обнаружению фона, т.е. выявлению локусов ДНК, где регистрируются неспецифические РНК-ДНК контакты. Неспецифическими считались контакты мРНК с хромосомами, не несущими ген, кодирующий данную мРНК. Процедура нормировки на фон заключалась в делении значения сигнала на уровень фона для каждого контакта. Также для каждого гена был рассчитан хроматиновый потенциал – мера, с помощью которой можно идентифицировать такие РНК, которые имеют большое количество контактов с хроматином, но при этом обладают довольно низкой экспрессией.

Нами обнаружены ранее неаннотированные хроматин-ассоциированные РНК, отсутствующие в RNA-Seq. В работе описаны свойства этих РНК.

Мы сравнили данные РНК-ДНК контактов с данными хроматиновых ДНК-ДНК контактов клеток К562. Показано, что контакты ассоциированы с открытым и активным хроматином (компартментом активного хроматина A, областями между топологически ассоциированными доменами, или ТАДами).

В результате работы разработан конвейер обработки данных высокопроизводительного секвенирования РНК-ДНК контактов, предложены методы борьбы с экспериментальными погрешностями и шумом, продемонстрированы свойства хроматин-ассоциированных РНК.

Данная работа была поддержана грантом РФФИ 17-00-00180, грантом РНФ 18-14-00011, а также грантом "Systems biology Fellowship Program".

Ссылки

- Werner, S.M.: Nuclear Fractionation Reveals Thousands of Chromatin-Tethered Noncoding RNAs Adjacent to Active Genes. Cell Reports, 12, 1089–1098 (2015)
- Engreitz, J.M.: Long non-coding RNAs: spatial amplifiers that control nuclear structure and gene expression. Nature Reviews Molecular Cell Biology, 17, 756-770 (2016)
- Engreitz, J.M.: RNA-RNA Interactions Enable Specific Targeting of Noncoding RNAs to Nascent Pre-mRNAs and Chromatin Sites. Cell 159(1), 188-199 (2014).
- Chu, C.: Technologies to probe functions and mechanisms of long noncoding RNAs. Nature Structural & Molecular Biology 22, 29-35 (2015)
- 5. Simon, M.D.: The genomic binding sites of a noncoding RNA. PNAS 108(51), 20497-20502 (2011)
- Sridhar D.: Systematic Mapping of RNA-Chromatin Interactions In Vivo. Current Biology 27, 1-8 (2017)

- 7. Bell J.C.: Chromatin-associated RNA sequencing (ChAR-seq) maps genome-wide RNAto-DNA contacts. bioRxiv (2017)
- 8. Li, X.: GRID-seq reveals the global RNA-chromatin interactome. Nature Biotechnology, 35, 940–950 (2017) 9. Жарикова, А. А.: Взаимодействие ДНК и РНК в хроматине. Тезисы конференции
- ИТИС (2017)

Анализ данных РНК-ДНК взаимодействий

Рябых Г.К.¹, Жарикова А.А.¹ и Миронов А.А.^{1,2}

¹ Московский государственный университет имени М.В.Ломоносова, Факультет биоинженерии и биоинформатики, Москва, Россия ² Институт проблем передачи информации, Москва, Россия гуаbykhgrigory@gmail.com

Abstract.

Мы создали базу данных, которая предназначена для накопления данных РНК-ДНК контактов и их быстрого и удобного анализа. С ее помощью можно: считать сводную статистику по конкретным РНК или группам РНК (например, количество и локализация контактов по хромосомам, их распределение по цветам хроматина), сравнивать и объединять эксперименты ChIRP, CHART, RAP с MARGI и GRID-seq, возможен поиск корреляций с другими данными взаимодействий макромолекул: белок-ДНК (CHIP-seq) и белок-РНК (eCLIP).

Некодирующие РНК – это группа транскриптов, с которых не происходит трансляции белков. Они главным образом делятся на малые некодирующие РНК (<200 нуклеотидов в длину) и более длинные некодирующие РНК (>200 нуклеотидов в длину). В первую группу входят, например, микроРНК, малые интерферирующие РНК, малые ядерные РНК, пиРНК, транспортные РНК. А вторая группа включает рибосомальные РНК, антисенс транскрипты и длинные некодирующие РНК [1].

Известно, что некодирующие РНК регулируют транскрипцию, сплайсинг, трансляцию и эпигенетический ландшафт, а также изменяют сигнальные пути путем взаимодействия с отдельными белками или белковыми комплексами, липидами, ДНК и другими РНК [2]. Например, XIST отвечает за инактивацию одной из X хромосом в женских клетках млекопитающих [3], NEAT1 и MALAT1 играют важную роль в регуляции транскрипции генов. MALAT1 локализуется в ядерных спекулах (ядерные структуры, обогащенные серин- и аргинин-богатыми факторами сплайсинга) и предпочитает связываться с телами генов и TTS (transcriptional termination sites). В то время как NEAT1 привлекается в структуры, которые называются параспекулы (ядерные структуры, располагающиеся рядом с ядерными спекулами), и связывается как с TSS (transcriptional start sites), так и с TTS генов [4].

На сегодняшний день существует несколько методов для того, чтобы локализовать на хроматине одну определенную РНК: ChIRP [5], CHART [6], RAP [7]. Они все основаны на использовании комплементарных последовательностей, необходимых для захвата конкретной РНК, с последующим определением сайтов связывания с хроматином с помощью

глубокого секвенирования. Однако, используя эти методы, можно анализировать только одну известную РНК за один эксперимент, и, следовательно, они не позволяют полногеномно посмотреть на взаимодействия всех РНК с ДНК за один эксперимент.

В 2017 году появились первые работы, которые предлагают методы, позволяющие получить данные обо всех потенциальных РНК-ДНК контактах в клетке: MARGI [8] и GRID-seq [9].

Мы создали базу данных, которая предназначена для накопления данных РНК-ДНК контактов и их быстрого и удобного анализа. С ее помощью можно: считать сводную статистику по конкретным РНК или группам РНК (например, количество и локализация контактов по хромосомам, их распределение по цветам хроматина), сравнивать и объединять эксперименты ChIRP, CHART, RAP с MARGI и GRID-seq, возможен поиск корреляций с другими данными взаимодействий макромолекул: белок-ДНК (CHIP-seq [10]) и белок-РНК (eCLIP [11]).

Список литературы

- Wang H, Bei Y, Shi J, Xiao J, Kong X. Non-Coding RNAs in Cardiac regeneration. *Cell Physiol Biochem*. 2015;36(5):1679-1687. doi:10.1159/000430141
- 2. A. Donlic and A. E. Hargrove, "Targeting RNA in mammalian systems with small molecules," no. March, pp. 1–21, 2018.
- M. D. Simon, S. F. Pinter, R. Fang, K. Sarma, M. Rutenberg-Schoenberg, S. K. Bowman, B. A. Kesner, V. K. Maier, R. E. Kingston, and J. T. Lee, "High-resolution Xist binding maps reveal two-step spreading during X-chromosome inactivation," *Nature*, vol. 504, no. 7480, pp. 465–469, 2013.
- J. A. West, C. P. Davis, H. Sunwoo, M. D. Simon, R. I. Sadreyev, P. I. Wang, M. Y. Tolstorukov, and R. E. Kingston, "The Long Noncoding RNAs NEAT1 and MALAT1 Bind Active Chromatin Sites," *Mol. Cell*, vol. 55, no. 5, pp. 791–802, 2014.
- C. Chu, K. Qu, F. L. Zhong, S. E. Artandi, and H. Y. Chang, "Genomic Maps of Long Noncoding RNA Occupancy Reveal Principles of RNA-Chromatin Interactions," *Mol. Cell*, vol. 44, no. 4, pp. 667–678, 2011.
- M. D. Simon, C. I. Wang, P. V. Kharchenko, J. A. West, B. A. Chapman, A. A. Alekseyenko, M. L. Borowsky, M. I. Kuroda, and R. E. Kingston, "The genomic binding sites of a noncoding RNA," *Proc. Natl. Acad. Sci.*, vol. 108, no. 51, pp. 20497–20502, 2011.
- J. M. Engreitz, A. Pandya-Jones, P. McDonel, A. Shishkin, K. Sirokman, C. Surka, S. Kadri, J. Xing, A. Goren, E. S. Lander, K. Plath, and M. Guttman, "The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome," *Science (80-.).*, vol. 341, no. 6147, 2013.
- B. Sridhar, M. Rivas-Astroza, T. C. Nguyen, W. Chen, Z. Yan, X. Cao, L. Hebert, and S. Zhong, "Systematic Mapping of RNA-Chromatin Interactions In Vivo," *Curr. Biol.*, vol. 27, no. 4, pp. 602–609, 2017.
- X. Li, B. Zhou, L. Chen, L. T. Gou, H. Li, and X. D. Fu, "GRID-seq reveals the global RNA-chromatin interactome," *Nat. Biotechnol.*, vol. 35, no. 10, pp. 940–950, 2017.
- Park P.J. ChIP-seq: advantages and challenges of a maturing technology. Nat. Rev. Genet. 2009;10:669–680.

11. Van Nostrand EL, Pratt GA, Shishkin AA, Gelboin-Burkhart C, Fang MY, Sundararaman B, et al. Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced clip (eclip) Nat Methods. 2016;13(6):508–14. doi: 10.1038/nmeth.3810.

Влияние белка Tat вируса иммунодефицита человека на экспрессию генов в В-лимфобластах

Анна Валяева^{*1}, Анастасия Жарикова¹, Алексей Пенин^{2,3}, Анна Клепикова^{1,2,3}, Евгений Шеваль³, Андрей Миронов^{1,2} * valvaevaann@gmail.com

¹ Факультет Биоинженерии и Биоинформатики, Московский государственный университет имени М. В. Ломоносова, Москва, Россия

² Институт проблем передачи информации им. А.А. Харкевича, Москва, Россия ³ Научно-исследовательский институт физико-химической биологии имени

А.Н.Белозерского, МГУ имени М.В. Ломоносова, Москва, Россия

Аннотация ВИЧ-инфекция, приводящая к развитию СПИД, часто сопровождается развитием других ВИЧ-ассоциированных заболеваний, например, В-клеточных лимфом. Ключевую роль в данном процессе играет вирусный белок Tat, обладающий способностью проникать в неинфицированные вирусом клетки и участвующий в регуляции транскрипции вирусных генов и генов клетки. Данная работа направлена на поиск генов, которые находятся под регуляцией белка Tat вируса иммунодефицита человека в В-лимфобластах, и определение механизма этой регуляции. Используя данные PHKсеквенирования, было выявлено, что белок Tat оказывает влияние на профиль экспрессии клеточных генов в В-лимфоцитах, и был получен список генов, экспрессия которых изменяется под влиянием белка Tat.

Ключевые слова: вирус иммунодефицита человека, белок Tat, PHKсеквенирование, анализ дифференциальной экспрессии генов

Вирус иммунодефицита человека (ВИЧ) инфицирует клетки иммунной системы человека, что приводит к развитию синдрома приобретенного иммунного дефицита (СПИД). Применение высокоактивной антиретровирусной терапии (ВААРТ) позволяет предотвращать развитие основных симптомов СПИД. Однако у многих пациентов на фоне ВААРТ развиваются другие ВИЧ-ассоциированные заболевания, в том числе В-клеточные лимфомы: лимфома Ходжкина, лимфома Беркитта и диффузная В-крупноклеточная лимфома [1]. При этом, В-лимфоциты не инфицируются ВИЧ, то есть развитие этих лимфом не является следствием заражения клеток вирусом. В развитии этих лимфом ключевую роль играет вирусный белок Tat [2]. Белок Tat способен секретироваться из ВИЧ-инфицированных клеток, накапливаться во внеклеточном матриксе и проникать в незараженные клетки [3]. Находясь в клетке, Tat участвует в регуляции транскрипции генов ВИЧ, и модулирует экспрессию ряда клеточных генов [4]. Целью нашей работы было изучить влияние белка Tat на экспресию генов в В-лимфоцитах. Для этого мы использовали культивируемые лимфоциты (линия RPMI 8866), которые экспрессировали Tat белок, слитый с EGFP. Для контроля использовали исходную культуру и линию клеток, экспрессирующую EGFP. Анализ проводился при использовании данных PHK-секвенирования. В ходе работы была проведена оценка качества PHKридов программой FastQC, картирование ридов на геном человека с помощью HISAT2. Используя программу HTSeq-count, было подсчитано количество ридов, выровненных с каждым геном, и с помощью пакетов R DESeq2 проведен поиск дифференциально экспрессируемых генов.

В ходе работы было обнаружено, что экспрессия EGFP в В-лимфобластах оказывает влияние на профиль экспрессии генов: около 800 генов изменили свой уровень экспрессии более, чем в 2 раза. Данное наблюдение было учтено при подсчете генов, на экспрессию которых влияет наличие в клетке белка Tat: из всех дифференциально экспрессирующихся генов, полученных при сравнении В-лимфобластов, экспрессирующих Tat, слитый с EGFP, и исходной культуры В-лимфобластов, были вычтены те гены, которые находятся под влиянием экспрессии EGFP. В результате был получен список из 353 белок кодирующих генов, уровень экспрессии которых изменился изза наличия в клетке белка Tat. GO аннотация показала, что среди найденных генов присутствуют гены антивирусного ответа, сигнальных путей TNF и PI3K-Akt, и гены, участвующие в негативной регулляции пролиферации клеток.

Список литературы

- Gibson, T. M., Morton, L. M., Shiels, M. S., Clarke, C. A., and Engels, E. A., 2014. Risk of non-Hodgkin lymphoma subtypes in HIV-infected people during the HAART era: a population-based study. AIDS, 28(15).
- Kundu, R. K., Sangiorgi, F., Yu, L. Y., Pattengale, P. K., Hinton, D. R., Gill, P. S., and Maxson, R., 1999. Expression of the human immunodeficiency virus-Tat gene in lymphoid tissues of transgenic mice is associated with B-cell lymphoma. Blood, 94(1), pp. 275–282.
- 3. Schwarze, S. R., Hruska, K. A., and Dowdy, S. F., 2000. Protein transduction: unrestricted delivery into all cells? Trends in Cell Biology, 10(7), pp. 290–295.
- Clark, E., Nava, B., and Caputi, M., 2017. Tat is a multifunctional viral protein that modulates cellular gene expression and functions. Oncotarget, 8(16), pp. 27569–27581.

Antisense transcription in bacteria: a transcriptional noise, an instrument to fine-tune gene expression, or a source for exoRNAs?

Maria Tutukina^{1,2,3}, Nataliya Markelova^{1,4}, Tatiana Bessonova^{1,5}, Uliana Shvyreva¹, Mikhail Gelfand ^{2,5,6,7} and Olga Ozoline^{1,3}

 ¹ Institute of Cell Biophysics RAS, Institutskaya str. 3, 142290, Pushchino, Russia
 ² Institute for Information Transmission Problems RAS, Bolshoy Karetny per. 19, 127051, Moscow, Russia

 ³ Pushchino Scientific Centre RAS, Prospekt Nauki 3, 142290, Pushchino, Russia
 ⁴ Pushchino State Institute of Natural Sciences, Prospekt Nauki 3, 142290, Pushchino, Russia
 ⁵ M.V.Lomonosov Moscow State University, Vorobievy gory 1, 119991, Moscow, Russia
 ⁶ Skolkovo Institute of Science and Technology, Moscow, 143028, Russia
 ⁷ Faculty of Computer Science, Higher School of Economics, Kochnovsky pr. 3, 125319, Moscow, Russia

masha306@gmail.com

Abstract. Antisense transcription is widespread among both eukaryotic and prokaryotic genomes. However, we are now only in the beginning on the way to understand their functions. Here, we discuss specific features of bacterial antisense promoters, and possible functions for their products detected from within and upstream of the genes coding for regulatory proteins in *Escherichia coli*. In particular, we found quite strong negative correlation between a gene expression and the presence of antisense transcripts in its regulatory region. Using reporter assays and primer extension, this observation was further confirmed for *dps*, *crp*, *exuR* and *leuO*. In the *leuO* gene, both sense and antisense transcription is repressed by Dps utilizing yet unknown mechanism. Interestingly, lots of short antisense transcripts synthesized from within *uxuR* were found in the fraction of small RNAs secreted by *E. coli* in the extracellular medium, but not in the fraction of intracellular RNAs. Some of them exist in the form of duplexes with complementary sense RNAs.

Keywords: Antisense transcription, Bacteria, ExoRNAs, UxuR, LeuO, Dps.

1 Antisense transcription in pro- and eukaryotes: what do we know?

Antisense transcription is widespread among all kingdoms of life and may affect almost all stages of gene expression, from transcription and translation to RNA degradation. The genomic arrangement of antisense RNAs indicates that they might be part of self-regulatory circuits that allow genes to regulate their own expression. For example, it was recently found that antisense and divergent transcripts can transmit regulatory signal to neighboring promoters [1]. Upon galactose induction in budding yeast, the same transcription factor induces expression of sense and antisense transcript. Antisense transcript runs to the upstream gene and inhibits its activity. The presence of antisense transcripts can also induce an on–off switch on sense-gene regulation, and antisense regulated genes had slightly lower than expected noise levels [2].

At the same time, according to a conventional point of view, in prokaryotic genomes antisense transcripts are just the by-products of noise without any significant functions [3]. Comparative genomics study demonstrated that in contrast to intragenic promoters aimed to initiate synthesis of shortened products in sense direction, antisense promoters has low conservation level even among closely related bacteria and function as "promoters" only due to high A/T content of their upstream regions that allow RNA polymerase binding [4,5]. Recently, it was thought that activity of such intragenic promoters is blocked by H-NS [5] but currently this is not obvious for antisense transcripts [6]. The aim of this study was to compare activity of potential promoters for antisense transcription located within genes coding for regulatory proteins in *E. coli in silico, in vitro* and *in vivo*, and to test their regulatory potential on selected genes.

2 Antisense transcription in bacteria: *in silico*, *in vitro* and *in vivo*.

2.1 PlatProm promoter search algorithm as a tool to find antisense transcription *in silico*.

The PlatPtrom scoring system is based on position-dependent weight matrices. Besides conservative base pairs recognized by the major σ -subunit of RNA polymerase it takes into account several other features of the promoter DNA considering it as a common platform for interaction with both RNA polymerase and transcription regulators. Normally, promoter-like signals form cluster around true starts due to the allowed variation in the spacer between conservative element -10 and the transcription start. Sensitivity of our program is 85,5% at the cut off level ensuring absence of false positives in negative sets [7].

2.2 Distribution of potential intragenic promoters in the *E. coli* genome and their ability to bind RNA polymerase.

Besides expected promoters located in front of coding sequences, PlatProm predicted rather large amount of unexpected promoters within coding sequences. If active, they may drive synthesis of either shortened mRNAs or antisense RNA-products. However some of them may be false positives or interact with RNA polymerase without transcription initiation. To evaluate their ability to bind RNA polymerase, we used two datasets: ChIP-on-chip data reflecting genome-wide distribution of RNA polymerase binding sites *in vivo* and the results of gel-retardation assays reflecting enzyme-binding ability *in vitro*. According to ChIP-on chip data, 77% of

2

co-directed promoters are associated with the registered sites of RNA polymerase binding, indicating rather good correspondence between *in silico* and *in vivo* data. However, only 54% of putative antisense promoters were able to bind RNAPol. This difference was quit unexpected assuming orientation-dependent interference with elongating RNA polymerase.

Then, enzyme binding capacity was tested *in vitro* for 35 DNA fragments, containing intragenic promoters, of which 21 were antisense. Their ability to interact with polymerase appeared to be very similar with known and co-directed promoters both in terms of complex formation by itself and overall binding percentage. Only two intragenic promoters failed to interact with the enzyme of which one was antisense. Thus, both types of unusual promoters can bind RNA polymerase, even though only 23 of them were captured by RNAPol *in vivo*. I.e. ability of enzyme binding may be unrealized in the bacterial cell. The percentage of DNA bound by the enzyme *in vitro* correlated with PlatProm scores assuming that our promoter model predicts the strength of binary complex formation to certain extent. Surprisingly we observed, that the value of correlation coefficient for DNA fragments containing antisense promoters is higher than for the whole set of tested templates (Fig.1A).



Fig. 1. Correlation between promoter strength *in silico* (PlatProm score), and their ability to bind RNA polymerase *in vitro* (A) and *in vivo* (B, C, and D). ChIP-chip data are from [8].

Taking this into account, we estimated correlation between values of Plat-Prom scores and efficiency of RNAPol binding *in vivo* (Fig. 1B-D). No correlation was observed for both "true" promoters and intragenic co-directed ones. This might reflect the presence of additional levels of regulation in the cell. However, the correlation for the set of antisense promoters was statistically significant. This may be explained by the observation that, antisense promoters usually lack additional promoterspecific elements, but possess highly conservative hexanucleotides, that mainly account for RNA polymerase binding [4].

3 Antisense transcription may regulate gene expression.

3.1 Antisense transcription changes transcription efficiency in the sense direction.

We compared the levels of sense and antisense transcription within genes coding for transcription factors in *E. coli*. We chose these genes because transcription factors are involved in multiple regulatory networks, and should be also subjected to a more complex regulation. Overlapping and intraoperonic genes were excluded from the analyses, and in total 160 profiles were analysed, representing 76% of all annotated regulators. Antisense transcription was registered within most of the studied loci, and was absent only in 18 genes (11%). First, we compared profiles of sense and antisense transcription for the exponentially and stationary growing cells. Expression of about 28% of genes was stable, while for 72% it was changed: increased 7%, and logically decreased in about 65%. The same tendency was observed for antisense transcription that was affected in 74% of cases.

3.2 Level of antisense transcription is affected by Dps and growth conditions.

It has been previously observed [5], that potential promoters for antisense transcription are often occupied by H-NS, nucleoid protein of exponential phase and a global transcriptional silencer. However, no evidence currently exists for H-NS influence on the efficiency of antisense transcription *in vivo*. We tested what will happen to the transcription profiles in the absence of Dps, the main nucleoid protein of the stat-phase cells. Despite the point of view, that Dps does not affect transcription specifically, its deletion increased the *leuO* expression. The same was observed for antisense direction. 62% of genes were not responding to the *dps* deletion, but 38 % were changed, which is a lot for a protein without anticipated regulatory function. According to the data obtained in our lab, Dps is involved in regulation of carbon metabolism and belongs to almost a half of transcription profiles during cell growth on D-glucose and D-glucuronate as carbon sources.

In line with the previous observations about 54% of genes were carbon-dependent, while antisense transcription remained unchanged in less than 27%.

3.3 Antisense transcription may interfere with binding sites for regulatory proteins.

We then studied a possibility of antisense modulation for the genes coding for two metabolic regulators: cAMP-CRP, acting mainly during growth on glucose, and ExuR, that is activated during growth with D-glucuronate. While changing carbon source from glucose to a less preferred glucuronate, expression level of *crp* was expectedly decreased. This decrease is accompanied by appearance of divergent transcripts nearby its starts. Vice versa, for ExuR, no antisense transcription was detected in these conditions, but they appeared during growth on glucose when ExuR is inhibited.

4

Both regions where antisense transcripts were detected overlap with binding sites for ExuR itself, suggesting a tight regulatory relationship between transcription factors and antisense transcripts.

4 Short antisense RNAs synthesized from within genes coding for regulatory proteins are widespread among RNAs secreted by *E. coli* in the extracellular medium.

Using whole-genome approaches, we compared the fractions of extremely short (12-22 n) RNAs secreted by Escherichia coli growing in pure culture and jointly with bacteria of the Paenibacillus genus. We found that the set of secreted oligonucleotides (exoRNAs) predominantly consists of mRNA and aRNA fragments usually processed from the regions with stable secondary structures, and differs in composition from the intracellular fraction. One of the striking examples is the *uxuR* gene coding for hexuronic metabolism regulator, with lots of exoRNAs being synthesized from within its coding region in both directions that are almost absent in the intracellular fraction. Precise mapping of oligonucleotides with a particular length allowed revealing many complementary duplexes similar to eukaryotic microRNAs. Interestingly, most exoRNAs are synthesized from within genes coding for regulatory proteins that are not very well expressed themselves. Another intriguing observation is that the amount of these RNAs changes with changing growth conditions - like in the dps gene. In the other genes, the spectrum of short external RNAs changes in the presence of a competing bacterium indicating a population dependent control over their composition. Some of the examples are the *bglG-bglF* operon coding for transcriptional terminatorantiterminator and beta-glucoside transporter, with hundreds of aRNAs being synthesized and extreted by E. coli in the presence of Paenibacillus, and dps, where exoRNAs are synthesized in both directions forming duplexes like in uxuR,

5 Conclusion.

Being in line with a concept of widespread antisense transcription, our findings do not support the idea that bacterial asRNAs are the products of transcriptional noise. Short aRNAs that are significantly overpresented in the fraction of secreted extracellular RNAs may interfere with other regulatory events in bacterial cells and form a novel class of regulatory RNAs with function yet to be understood.

The work was supported by RSF grants 18-14-00358 and 18-14-00348.

References

 Pelechano, V., Steinmetz, L.: Gene regulation by antisense transcription. Nature Reviews Genetics 14, 880–893 (2013).

- Huber, F., Bunina, D., Gupta, I., Khmelinskii, A., Meurer, M., Theer, P., Steinmetz, L.M., Knop, M.: Protein abundance control by non-coding antisense transcription. Cell Reports 15(12), 2625-2636 (2016).
- Llorens-Rico, V., Cano, J., Kamminga, T., Gil, R., Lattore, A., Chen, W.H., Bork, P., Glass, J.I., Serrano, L., Lluch-Senar, M.: Bacterial antisense RNAs are mainly the product of transcriptional noise. Science Advances 2(3), e1501363 (2016).
- Shao, W., Price, M.N., Deutschbauer, A.M., Romine, M.F., Arkin, A.P.: Conservation of transcription start sites within genes across a bacterial genus. MBio 5(4), e01398-14 (2014).
- Singh, S.S., Singh, N., Bonocora, R.P., Fitzgerald, D.M., Wade, J.T., Grainger, D.C.: Widespread suppression of intragenic transcription initiation by H-NS. Genes and Development 28(3), 214-219 (2014).
- Landick, R., Boudreu, E., Jessen, E., Harwig, A.: Antisense transcription modulates gene expression but not H-NS distribution in the *E. coli bgl* operon. In: Proceedings of The Biology and Physics of bacterial Chromosome Organisation 2018, OC015, Biochemical Society, Leiden (2018).
- Shavkunov, K.S., Masulis, I.S., Tutukina, M.N., Deev, A.A., Ozoline, O.N.: Gains and unexpected lessons from genome-scale promoter mapping. Nucleic Acids Research 37(15), 4919-4931 (2009).
- Reppas, N.B., Wade, J.T., Church, G.M., Struhl, K.: The transition between transcriptional initiation and elongation in *E. coli* is highly variable and often rate limiting. Molecular Cell. 24(5), 747-757 (2006).

6

Анализ данных Hi-C одиночных клеток Drosophila melanogaster

Александра А. Галицына^{1,2,3}, Влада В. Захарова², Екатерина Е. Храмеева^{1,3}, Сергей В. Разин^{2,4}, Михаил С. Гельфанд^{1,3,5,6} и Сергей В. Ульянов^{2,4}

¹ Сколковский Институт Наук и Технологий
 ² Институт Биологии Гена РАН
 ³ Институт Проблем Передачи Информации им. Харкевича РАН
 ⁴ Биологический Факультет МГУ им. Ломоносова
 ⁵ Факультет Биоинженерии и Биоинформатики МГУ им. Ломоносова
 ⁶ Факультет Компьютерных Наук Высшей Школы Экономики

agalitzina@gmail.com

Аннотация. В данной работе мы проводим анализ структуры хроматина одиночных клеток *Drosophila melanogaster* с помощью высокопроизводительного метода фиксации конформации хромосом Hi-C. Мы провели сравнение данных популяционных экспериментов с одиночными клетками и обнаружили стохастичность формирования структуры хроматина. Мы обнаружили в одиночных клетках структуры, подобные топологически ассоциированным доменам (ТАДам) популяционных данных. Стабильные ТАДы и их границы обладают отличительными эпигенетическими свойствами.

Ключевые слова: Hi-C, секвенирование одиночных клеток, пространственная организация хроматина

Исследование закономерностей работы генома долгое время было ограничено двумя типами методов исследования: узко специализированных для одиночных клеток (таких, как FISH) или же методов масштабного секвенирования популяции (или ансамбля) большого количества клеток (такие, как RNA-Seq, ChIP-Seq, Hi-C). Появление методов высокопроизводительного секвенирования одиночных клеток (single-cell sequencing) позволило наблюдать за свойствами тотальной РНК, ДНК и белков на уровне единичной клетки. [1]

Так, последние работы по анализу single-cell транскриптомов дрозофилы позволяют наблюдать за развитием целых эмбрионов [2] и изменениями мозга при старении [3].

Одно из последних достижений в этой области – разработка методов высокопроизводительной фиксации конформации хромосомы для одиночных клеток и ядер. Хотя впервые метод появился в 2013 году [4], он был повторен и развит лишь в 2017 году [5,6,7,8]. В этих работах проводится анализ ДНК-ДНК контактов в клетках млекопитающих (мыши и человека), обнаруживается существование компартментов и топологически ассоциированных доменов (ТАДов, также называемых контактными доменами) на уровне одиночных ядер. Статистика по некоторым работам приведена в табл. 1. Видна тенденция к возрастанию количества исследуемых клеток и количества контактов, полученных из одной клетки.

	Количество прочтений на эксперимент	Количество клеток	Количество контактов на клетку		
Nagano 2013 [4]	5 - 15 млн	10 клеток с >1000 контактами	11,000-30,000		
Stevens, 2017 [6]	1.5 - 4.8 млн	8	37,000-122,000		
Flyamer, 2017 [5]	~15-83 млн	36 клеток с >30,000 контактами 219 клеток с > 1000 контактами	до 1,906,000		
Ramani, 2017 [7]	20-500 млн	10,696 клеток с > 1,000 контактами	59,000-72,000		

Табл. 1. Статистика некоторых работ по Ні-С одиночных клеток

В хроматине дрозофилы на уровне популяции клеток были также обнаружены ТАДы и компартменты [9], однако до сих пор не проводился масштабный анализ с помощью single-cell Hi-C. В данной работе мы поводим анализ контактов хроматина 20 клеток D. melanogaster (клеточная линия BG3), полученных по ранее опубликованному протоколу snHi-C [5].

Мы проводим картирование и анализ данных Hi-C с помощью инструментов cooler и pairsamtools [10, 11], а также вводим дополнительный этап фильтрации контактов – обязательное прохождение прочтения секвенирования через точку лигирования двух сайтов рестрикции. Мы также проводим повторный анализ опубликованных данных [5] с помощью предложенного метода. Фильтрация контактов улучшает качество single-cell Hi-C данных по нескольких критериям (например, количество контактов на один рестриктный фрагмент). В результате обработки мы получили до 107823 контактов на клетку дрозофилы.

Мы обнаружили высокую корреляцию данных объединенных данных singlecell Hi-C и популяционных данных дрозофилы. В обоих типах данных присутсвуют выраженные ТАДы. Структуры, подобные ТАДам, обнаруживаются также и в одиночных клетках, их характерный размер около 80-100 т.п.н.

Позиции ТАДов в одиночных клетках значимо коррелируют с ТАДами в популяционных данных Hi-C. Стабильные ТАДы и их границы характеризуются специфическим эпигенетическими свойствами хроматина.

Результаты работы свидетельствуют о том, что хотя укладка хроматина дрозофилы является в значительной мере стохастичной, существуют факторы, определяющие положение границ ТАДов в единичных клетках преимущественно в тех же местах генома, которые являются границами ТАДов в популяционных данных Hi-C.

Данная работа была поддержана грантом РФФИ 17-00-00180, грантом РНФ 16-14-10081, а также грантом "Systems biology Fellowship Program".

Ссылки

- Gawad C., Koh W. and Quake S.R. Single-cell genome sequencing: current state of the science. Nature Reviews Genetics 17, 175–188 (2016)
- Karaiskos N. et al. The *Drosophila* embryo at single-cell transcriptome resolution. Science, 358(6360):194-199 (2017)
- 3. Kristofer D. at al. A Single-Cell Transcriptome Atlas of the Aging Drosophila Brain. Cell, online publication (2018)
- Nagano T. at el. Single cell Hi-C reveals cell-to-cell variability in chromosome structure. Nature, 502(7469) (2013)
- 5. Flyamer I.M. et al. Single-nucleus Hi-C reveals unique chromatin reorganization at oocyte-to-zygote transition. *Nature* 544, 110–114 (2017)
- Stevens T.J. et al. 3D structures of individual mammalian genomes studied by single-cell Hi-C. Nature, 544, 59–64 (2017)
- Ramani V. et al. Massively multiplex single-cell Hi-C. Nature Methods 14, 263– 266 (2017)
- Nagano T. et al. Cell-cycle dynamics of chromosomal organization at single-cell resolution. *Nature*, 547, 61–67 (2017)
- Szabo Q. et al. TADs are 3D structural units of higher-order chromosome organization in Drosophila. Sci. Adv. 4:eaar8082 (2018)
- 10. Cooler tools, https://github.com/mirnylab/cooler, version from 08.05.2018
- 11. Pairsamools tools, https://github.com/mirnylab/pairsamtools, version from 21.05.2018

Система позиционирования внутри зданий мобильной робототехнической платформы на основе детекции краев

Абрамов М.П.^{1,2}, Шипитько О.С.¹, Лукоянов А.С.^{1,2}, Панфилова Е.И.^{1,2}, Кунина И.А.^{1,2}, Григорьев А.С.¹

¹Институт проблем передачи информации им. А.А. Харкевича Российской академии наук — ИППИ РАН, Большой Каретный переулок, д.19 стр. 1, Москва, Россия, 127051 ²Московский физико-технический институт (государственный университет) — МФТИ (ГУ), Институтский переулок, д.9, г. Долгопрудный, Московская область, Россия, 141701

abramov@visillect.com

Ключевые слова: локализация, позиционирование, детектирование краев, модель измерений, фильтр частиц

Аннотация. Необходимой предпосылкой для решения задачи навигации является точное знание собственного положения в пространстве. В работе предлагается метод высокоточного позиционирования мобильной робототехнической платформы внутри зданий. Для предсказания положения используются данные от бортовых датчиков движения (датчик скорости вращения колес и датчик угловых скоростей платформы). В качестве априорной информации используется план здания, хранимый в виде цифровой многоканальной растровой карты. Оценка правдоподобия положения вычисляется путем сопоставления краев, задетектированных на изображениях, получаемых от монокулярной камеры видимого диапазона, с априорно известной картой. Таким образом, предложенный метод использует особенности, присущие любым помещениям — стыки между стенами и полом — и не требует дополнительной подготовки рабочей среды. Для объединения разнородных наблюдений используется фильтр частиц, учитывающий мультимодальное распределение вероятностей оценки собственного положения. Так как для оценки правдоподобия положения фильтром частиц используется вероятностная модель измерений сенсоров, точное моделирование ошибок измерений является ключом к улучшению качества позиционирования. В работе предложена модель измерений детектора краев, учитывающая геометрические ошибки детекции и ложные срабатывания детектора. Предложенная система позиционирования была экспериментально протестирована на роботе-макете легкового автомобиля. Результаты экспериментов демонстрируют, предложенный метод позволяет определять собственное положение что робототехнического комплекса с ошибкой, в среднем не превышающей 0.1 м на каждом из 100 тестовых запусков.

Исследование выполнено за счёт средств гранта гранта Российского научного фонда (проект №14-50-00150).

Оптимальная аффинная аппроксимация проективного преобразования изображений

Коноваленко И.А., Кохан В.В., Николаев Д.П.

Институт проблем передачи информации им. А.А. Харкевича Российской академии наук — ИППИ РАН, Большой Каретный переулок, д.19 стр. 1, Москва, Россия, 127051 konovalenko@iitp.ru, kohan@visillect.com, dimonstr@iitp.ru

Ключевые слова: проективное преобразование, точность проективного преобразования, аффинная аппроксимация преобразований, линеаризация, выпуклый анализ

Аннотация. Замена проективного преобразования существенно более простым аффинным преобразованием встречается во многих областях технического зрения. В точности аффинной аппроксимации то же время, понятие проективного преобразования в литературе не формализуется. Это, в свою очередь, приводит к отсутствию постановок задач и теоретически обоснованных методов аффинной аппроксимации проективного преобразования. Настоящая работа направлена на заполнение этого пробела. Авторами предложено в качестве критериев точности аффинной аппроксимации проективного преобразования изображения использовать поточечную среднеквадратичную И максимальную невязку в координатах преобразованного изображения. На основе этих критериев сформулированы задачи поиска оптимальных аффинных аппроксимаций. Доказана выпуклость полученных задач оптимизации. Предложен метод использования оптимальных аффинных аппроксимаций для экономии вычислительных ресурсов при преобразовании изображений.

Исследование выполнено при частичной финансовой поддержке РФФИ в рамках научных проектов №16-07-00616 и №17-29-03370.

Робастный критерий поиска точки схода проекций прямолинейных траекторий движения детектированных в видеопотоке транспортных средств

Бочаров Д.А.^{1,2}, Аксенов К.А.³, Шемякина Ю.А.^{4,5}, Коноваленко И.А.^{1,2,4,6}

¹Институт проблем передачи информации им. А.А. Харкевича Российской академии наук — ИППИ РАН, Большой Каретный переулок, д.19 стр. 1, Москва, Россия, 127051

²Visillect, Большой Каретный переулок, д.19 стр. 1, Москва, Россия, 127051
³Национальный исследовательский университет Высшая школа экономики — НИУ ВШЭ, ул. Мясницкая, д. 20, г. Москва, Россия, 101000
⁴Smart Engines Ltd., проспект 60-летия Октября, д.9, Москва, Россия, 117312
⁵Институт системного анализа Федерального исследовательского центра «Информатика и управление» Российской академии наук — ИСА ФИЦ ИУ РАН, проспект 60-летия Октября, д.9, Москва, Россия, 117312
⁶Московский физико-технический институт (государственный университет) — МФТИ (ГУ), Институтский переулок, д.9, г. Долгопрудный, Московская область, Россия, 141701
bocharov@visillect.com, aksenov@visillect.com, jshemiakina@smartengines.biz, konovalenko@iitp.ru

Ключевые слова: точка схода, критерий поиска точки схода, алгоритм оценки точки схода, траектории движения, RANSAC.

Аннотация. Одним из подходов к самокалибровке камер для транспортной видеоаналитики является анализ траекторий наблюдаемого движения. В данной работе рассматривается задача поиска точки схода траекторий движения автомобилей на изображении в предположении о прямолинейности и параллельности их движения. Алгоритм оценки точки схода по наблюдаемым трекам движения, предложенный Tuan Hue Thi, устойчив к наличию посторонних траекторий, однако не устойчив к выбросовому шуму точек на траекториях. В данной работе обсуждаются возможные причины возникновения такого шума и предлагается устойчивый к нему алгоритм. На данных, созданных при помощи имитационного моделирования, численно оценивается зависимость точностей референтного и предлагаемого алгоритмов в зависимости от уровня шума. Приведены примеры работы предлагаемого алгоритма на реальных данных, полученных с использованием системы распознавания автомобильных номеров в видеопотоке.

Обзор литературы, создание ПО для проведения вычислительных экспериментов, а также разработка и исследование предложенного в работе алгоритма выполнены за счет средств гранта РНФ (Проект №14-50-00150).

Повышение вычислительной эффективности проективного преобразования изображений на SIMD-архитектурах

Антон Трусов¹, Елена Лимонова^{2,3}, Алмаз Миргасимов⁴

¹Московский физико-технический институт (государственный университет), г. Долгопрудный ²Федеральный исследовательский центр "Информатика и управление"Российской Академии Наук, г. Москва <u>elena.e.limonova@gmail.com</u> ³ООО "Смарт Энджинс Сервис г. Москва ⁴Rock flow dynamics, г. Москва <u>mirgasimov.almaz@gmail.com</u>

Ключевые слова: проективное преобразование изображений, вычислительная эффективность, SIMD-архитектура, обработка изображений

Аннотация. Проективное преобразование изображений широко используется при решении задач технического зрения. В работе предложен метод повышения его вычислительной эффективности, позволяющий сократить число проверок граничных проверки преобразования. Эти возникают, поскольку вследствие условий проективного преобразования на результирующем изображении могут быть точки, выходящим за пределы исходного изображения. соответствующие точкам, Предложенный метод позволяет эффективно задействовать SIMD-расширения современных вычислителей, позволяющие единообразно обрабатывать несколько элементов данных одновременно. Эксперименты показали, что таким образом можно ускорить алгоритм от 1.35 до 2 раз в зависимости от числа точек, попадающих за пределы исходного изображения.

Работа выполнена при частичной финансовой поддержке грантов РФФИ №17-29-03297 и №16-29-09508.

Влияние социального статуса на успех ухаживания самцов сверчка Gryllus bimaculatus

Шестаков Л.С, Веденина В.Ю.

Федеральное государственное бюджетное учреждение науки Институт проблем передачи информации им. А.А. Харкевича Российской академии наук (ИППИ РАН) zicrona@yandex.ru

Ключевые слова: сверчки, Gryllus bimaculatus, мультимодальная коммуникация, сигнал ухаживания, доминантный статус, агрессия, предпочтения самок

Аннотация. Растущее число исследований на представителях многих таксонов животных демонстрирует отсутствие корреляции между социальным статусом самцов и их привлекательностью для самок. Мы проанализировали 19 элементов поведения ухаживания у самцов и самок сверчков Gryllus bimaculatus. Показано, что самки в равной степени копулируют с доминантными и субординантными самцами, но латентный период от начала антеннального контакта до ухаживания оказался короче ٧ субординатных самцов. Bo время ухаживания доминирующие самцы демонстрировали один из элементов агрессии - «тряску», чаще и дольше, чем подчиненные самцы. Обнаружена отрицательная корреляция этого эле- мента с пением у доминантных самцов, но положительно коррелировал с пением у самцов, проигравших в драке. Параметры песни самцов были слабо связаны с успехом у самок. Доминантные самцы, как правило, из- дают большее количество агрессивных сигналов перед самкой, чем проигравшие самцы. Это может объясняться неспособностью доминантов быстро перейти от агрессии к ухаживанию. Полученные результаты свидетельствуют о том, что усиление агрессии, вероятно, будет препятствовать успешному ухаживанию.

Сравнение распознающей и метрической свёрточных сетей на примере распознавания поля "пол" паспорта гражданина РФ

Чирвоная А.Н.¹, Лынченко А.Е.², Чернышова Ю.С.^{2,3}, Шешкус А.В.^{2,3}

¹Национальный исследовательский технологический университет «МИСиС», Москва, Россия

²ООО «Смарт Энджинс Сервис», Москва, Россия,

³Федеральное государственное учреждение "Федеральный исследовательский центр "Информатика и управление" Российской академии наук", г. Москва, Россия <u>{nastyachirvonaya, alynchenko, chernyshova, asheshkus}@smartengines.biz</u>

Ключевые слова: распознавание текста, свёрточные нейронные сети, глубокое обучение, сиамские нейронные сети, обучение метрики

Аннотация. В работе рассматривается использование метрических нейронных сетей в задаче распознавания изображений слов. Подход к распознаванию слов, основанный на распознавании отдельных букв, хорошо изучен, но плохо применим к некоторым видам текста. Действительно, рукописные, написанные арабским языком или имеющие лигатуры тексты трудно сегментируются на буквы. Кроме этого, в тексте могут появляться слипшиеся символы, если изображения сильно зашумлены и/или искажены из-за несовершенства камеры. Все эти проблемы возникают в системах распознавания текста с заданным шаблоном, где набор слов может быть ограничен. В таких случаях разумно распознавать слова целиком, хотя словарь ответов может быть большим и не обязательно известным на этапе обучения. Для решения задачи распознавания изображений слов мы предлагаем использовать нейронную сеть. В работе приводится сравнение метрическую качества распознавания метрической нейронной сети со стандартной классифицирующей на словах, собранных с поля "пол" паспорта гражданина РФ. Параметры всех слоев кроме последнего у метрической и классифицирующей сетей были сделаны одинаковыми для обеспечения чистоты эксперимента. Результаты показывают пригодность метрических нейронных сетей для решения задачи распознавания слов. преимуществами предлагаемого метода являются Основными возможность расширения алфавита сети уже после обучения и отсутствие необходимости сегментировать слово на символы.

Работа выполнена при финансовой поддержке РФФИ в рамках научных проектов №17-29-07092 и №17-29-07093.

РАСПОЗНАВАНИЕ ПРОЕКТИВНО ПРЕОБРАЗОВАННЫХ ПЛОСКИХ ФИГУР. XII.

О новых методах проективно инвариантного описания овалов.

Николаев П.П.

Институт проблем передачи информации им. А.А. Харкевича Российской академии наук — ИППИ РАН, Большой Каретный переулок, д.19 стр. 1, Москва, Россия, 127051 nikol@iitp.ru

Ключевые слова: линейный элемент плоскости, эллиптические точки, вурф, проективно инвариантное отображение, критерии выбора базисных точек

Аннотация. Для сцен, где входным объектом является композиция овала общего вида в сочетании с линейным элементом (ЛЭ. – точкой Р либо прямой L) инцидентной ему плоскости в произвольном их взаимном расположении (внешний, внутренний ЛЭ либо принадлежащий контуру овала) на материале модельных численных реализаций рассмотрены процедурные подходы в задаче проективно инвариантного описания подобной композиции. Предложенные алгоритмы обработки используют полученные ранее теоретические утверждения (теоремы), для овала с выделенной внутренней точкой intP гарантирующие наличие триады внешних проективно инвариантных эллиптичеких точек (ЭТ) E1, E2, E3, а для овала в сочетании с прямой extL (внешнего положения) – наличие двух пар инвариантых точек, порождающих пару стабильных внутренних точек С1, С2. Показано, как из произвольно организованных композиций «овал + ЛЭ» можно сформировать композицию «овал + intP + extL + T-polar». пригодную для вычисления ее проективно инвариантного отображения, не опирающегося исключительно на фиксированный набор инвариантых точек контура, а представляющих интегральное описание исходной композиции. Этот единый алгоритмический подход реализуем в итоге привлечения критериев детерминированного выбора пары ЭТ для задания extL и единственной intP из пары C1, C2 – для задания внешней инвариантной кривой T-polar.

Работа выполнена при частичной финансовой поддержке гранта РФФИ в рамках научного проекта №16-07-00836.

Связь степени возбудимости нейронов моторной коры и сегментарных мотонейронов с возможностью активации непроизвольных ритмических движений верхних конечностей у человека

Гареева Р. Р.^{1,2}, Блинов Е. О.^{1,2}, Солопова И. А.¹, Селионов В.А.¹, Жванский Д.С.¹

¹Институт проблем передачи информации им. А.А.Харкевича, РАН, Москва, Россия ²Московский физико-технический институт (ГУ), Долгопрудный, Московская область, Россия

d.zhvansky@gmail.com

Ключевые слова: Н-рефлекс, ТМС, генератор ритмических движений

Аннотация. Известно, что в условиях разгрузки конечностей генератор ритмических движений конечностей у человека можно активировать различными методами: ЧССМ, электрической стимуляцией нервов, ТМС и др. вибрацией мышц, Эффективность этих методов различная, однако непроизвольные движения никогда не удавалось вызвать у 100% испытуемых. Целью настоящей работы было исследование различий в возбудимости спинальных нейронных сетей шейного отдела спинного мозга и нейронов моторной коры у испытуемых, у которых мышечная вибрация активирует непроизвольные ритмические движения рук и у тех, у которых не активирует. Различия исследовали при выполнении активных и пассивных движений руки в условиях ее вывески. Возбудимость мотонейронов шейного отдела спинного мозга оценивали по величине Н-рефлекса мышцы- сгибателя кисти (FCR). Возбудимость нейронов моторной коры оценивали по величине моторного ответа на транскраниальную магнитную стимуляцию в передней задней головках И дельтовидной мышцы. Показано, что при обоих типах движений испытуемые, у которых удавалось активиро- вать непроизвольные движения руки, имели более высокий уровень Н- рефлекторной возбудимости и меньшие моторные ответы на ТМС, чем те, у которых вибрация была неэффективна для вызова непроизвольной ритмики. Это может свидетельствовать о различиях в состоянии нейронного аппарата шейного отдела спинного мозга и различных кортикоспинальных влияниях у этих двух групп испытуемых, а также, по-видимому, о большей возбудимости мотонейронов и/или меньшем пресинаптическом торможении у первой группы испытуемых.

Синергическое влияние периферической и центральной стимуляции на вызванные ритмические движения рук

Дмитрий Жванский, Ирина Солопова и Виктор Селионов

Федеральное государственное бюджетное учреждение науки Институт проблем передачи информации им. А.А. Харкевича Российской академии наук, 127051, г. Москва, Большой Каретный переулок, д.19 стр. 1., Россия <u>d.zhvansky@gmail.com</u>

Ключевые слова: центральные генераторы ритмики, сенсорная стимуляция, транскраниальная магнитная стимуляция, межконечностные взаимодействия.

Аннотация. Ритмическая активность мышц рук вносит вклад во многие локомоторные задачи, поэтому нейромодуляция цервикальных спинальных сетей представляет интерес для изучения ритмогенеза в спинном мозге человека и исследования взаимодействия между шейными и пояснично-крестцовыми нейронными цепями. Ранее мы показали, что вибростимуляция мышц рук может вызывать непроизвольные колебания верхних конечностей в условиях разгрузки тела. В данной работе мы исследовали возбудимость цервикальных спинальных цепей у здоровых испытуемых, различные приемы центральной и периферической применяя стимуляции. Обоснование применения комбинированной стимуляции заключается в том, что эффективность моностимуляции, как правило, ограничена. Мы обнаружили, что низкоинтенсивная электрическая срединного стимуляция (OC) нерва, иннервирующего мышцы предплечья, может вызывать ритмические движения рук. Кроме того, движения усиливались дополнительной периферической стимуляцией (например, вибрацией мышц рук, поворотом головы или пассивными ритмическими движениями ног). Наконец, ритмогенез значительно облегчался низкочастотной транскраниальной магнитной стимуляцией моторной коры. Результаты исследования обсуждаются в парадигме интерфейса головной-спинной мозг для разработки методов терапии, адаптивно модулирующих центральную генерацию ритмики.

Исследование поддержано грантом РФФИ №16-29-08181-офи_м.

Полисомнографическое исследование центрального апноэ сна у здоровых кошек

Лиманская Александра Васильена

¹ИППИ РАН им. А.А. Харкевича ²МГУ им. М.В. Ломоносова, биологический факультет <u>limasha16@yandex.ru</u>

Ключевые слова: сон, висцеральная теория сна, полисомнография, расстройства дыхания во сне, центральное апноэ сна.

Аннотация. Синдром центрального апноэ - явление полной или частичной остановки дыхания во сне, возникающее по команде из нервного центра на дыхательную мускулатуру. Значительную долю внимания в сомнологии уделяют обструктивному апноэ, прекращению дыхания в результате механической обструкции дыхательных путей, признанного серьезным патологическим явлением, связанным с состоянием сна. Однако, до сих пор остается открытым вопрос, является ли центральное апноэ патологическим состоянием, с которым следует бороться, или же это адаптивная и физиологичная реакция организма. В процессе экспериментальной вполне разработки висцеральной теории сна мы обратили внимание на частые эпизоды центрального апноэ у здоровых кошек. Наличие нестандартных для классической полисомнографии каналов регистрации, включающих расширенный комплекс вегетативных показателей, и возможность проводить исследование на одном животном в течение длительного времени (до нескольких лет), позволило рассматривать и описывать явление центрального апноэ с новой точки зрения, обеспечивающей оригинальный подход к его изучению. Наши наблюдения свидетельствуют в пользу того, что центральное апноэ сна является элементом комплекса вегетативных реакций организма, направленных на восстановление его жизнедеятельности.

Отражение сердечной деятельности в электрической активности корковых нейронов в цикле сон-бодрствование

Лаврова В.Д.

Институт проблем передачи информации им. А.А. Харкевича РАН, Москва

Ключевые слова: сон, кора мозга, импульсная активность нейронов, сердце, электрокардиограмма, висцеральная теория сна.

Аннотация. Как предполагает висцеральная теория сна (Пигарев, 2013), кора головного мозга во время сна переключается на анализ интероцептивной информации, приходящей от внутренних органов. Впервые эта связь была показана нами в экспериментах на желудочно-кишечном тракте, когда в зрительных корковых зонах были обнаружены вызванные ответы, синхронизованные с его активностью. В предыдущем исследовании мы также описали связанные с сердцебиением вызванные потенциалы, усредняемые в электроэнцефалограмме (ЭЭГ) и локальных медленных потенциалах (LFP) во время сна в лобной коре мозга у кошек. Целью данной работы было исследование связанной с сердечной деятельностью электрической активности отдельных нейронов в коре мозга в цикле сон-бодрствование. Методы исследования. У двух взрослых кошек при помощи транскраниальных биполярных микроэлектродов регистрировали локальные медленные потенциалы и нейронную активность в лобной и инсулярной коре обоих полушарий. Зоны отведений были выбраны согласно предшествовавшим исследованиям о местонахождении областей, связанных с окончаниями блуждающего нерва в коре (Черниговский, Мусящикова, 1973; Быков, 1947). ЭКГ записывали от двух электродов, один из которых находился в желудке, а другой – на голове кошки. Кроме того, каждая запись содержала общую ЭЭГ (супрадурально), ритм дыхания, активность желудка и движения глаз кошки для выделения фаз сна. Записи длиной 2-5 часов, содержавшие эпизоды медленного сна, быстрого сна и бодрствования, были проанализированы в программе Spike 2 CED при помощи скриптов, написанных сотрудником лаборатории специально для этой цели. Результаты. В 20 записях были выделены 142 нейрона. Изменения импульсной активности, связанные с сердцебиениями, были обнаружены у трети из них (46 нейронов, 32,4%) как в лобной, так и в инсулярной коре обоих полушарий мозга. У части нейронов частота импульсации оказалась достоверно связана с частотой сердечных сокращений. Этот феномен наблюдался преимущественно во время сна и практически исчезал во время бодрствования. Выводы. В данном исследовании показано, что информация от сердца во время сна действительно достигает коры головного мозга. Наши результаты подтверждают гипотезу, что корковые зоны во сне переключаются на обработку висцеральных сигналов, а их активность, скорее всего, направлена на диагностику и восстановление функций органов тела.

Работа поддержана грантом РФФИ №16-04-00413.

Безопасное управление скоростью наземного беспилотного транспортного средства в условиях неопределенности собственного положения

Кибалов В.И.¹, Шипитько О.С.¹, Коробов Н.С.^{1,2}, Григорьев А.С.¹

¹Институт проблем передачи информации им. А.А. Харкевича Российской академии наук — ИППИ РАН, Большой Каретный переулок, д.19 стр. 1, Москва, Россия, 127051 ²Московский физико-технический институт (государственный университет) — МФТИ (ГУ), Институтский переулок, д.9, г. Долгопрудный, Московская область, Россия, 141701 {kibalov, shipitko, korobov, grigoryey}@visillect.com

Ключевые слова: безопасность движения, управление скоростью, беспилотное транспортное средство, динамическая модель, фильтр частиц

Аннотация. Безопасность движения является ключевой проблемой коммерческой эксплуатации беспилотных транспортных средств (БПТС). В данной работе предложена система безопасного управления скоростью движения наземного БПТС, основанная на математической модели БПТС, системе траекторного управления движением и фильтре частиц. Планируемая траектория движения преобразуется системой управления в управляющие сигналы, которые поступают на вход математической модели. Выход модели - положение транспортного средства в системе координат, связанной с картой. Таким образом в режиме реального времени на основе модели рассчитывается опережающая траектория движения, которая затем применяется к каждой частице - гипотезе системы позиционирования о текущем транспортного средства. Такой подход позволяет производить положении вероятностную оценку безопасности текущей траектории движения в условиях неопределенности собственного положения и ограничивать скорость движения до безопасного значения. Предлагаемый алгоритм был протестирован на данных реального БПТС, а также в составе системы управления движением робота-макета легкого автомобиля. Результаты экспериментов демонстрируют, что разработанная динамическая модель позволяет точно прогнозировать траекторию движения БПТС, а система управления скоростью движения позволяет снижать скорость движения БПТС до безопасного значения при выполнении маневров и проезде узких проемов.

Исследование выполнено за счёт гранта Российского научного фонда (проект №14-50-00150).

Математическая модель

габаритно-весо-энергетическо-ценовых параметров бортовых вычислителей беспилотных транспортных средств

Коробов Н.С.^{1,2}, Шипитько О.С.¹, Коноваленко И.А.^{1,2}, Григорьев А.С.¹, Чукалина М.В.¹

¹Институт проблем передачи информации им. А.А. Харкевича Российской академии наук — ИППИ РАН, Большой Каретный переулок, д.19 стр. 1, Москва, Россия, 127051 ²Московский физико-технический институт (государственный университет) — МФТИ (ГУ), Институтский переулок, д.9, г. Долгопрудный, Московская область, Россия, 141701

{korobov, shipitko, konovalenko, grigorvev}@visillect.com

Ключевые слова: математическая модель, бортовой вычислитель, беспилотное транспортное средство, SWaP-С модель, парето-оптимизация

Аннотация. При конструировании автономных аппаратов (АА), как правило, на стоимость, массу, габариты, а также энергопотребление бортового вычислителя накладываются строгие ограничения. В то же время современные алгоритмы локализации, компьютерного зрения и картирования местности, которые должны быть запущены на АА для обеспечения автономности, могут иметь большую вычислительную сложность. Задачей настоящего исследования является построение математической модели габаритно-весо-энергетическо-производительно-ценовых параметров современных вычислителей для установления ограничений на пространство допустимых вычислителей. Для решения этой задачи построена модель, позволяющая найти пространство парето-оптимальных бортовых вычислителей, что позволяет сузить множество вычислителей рассматриваемых при конструировании AA.

Исследование выполнено при частичной финансовой поддержке РФФИ в рамках научных проектов №17-29-03514 и №17-29-03236.

Исследование функционального состояния двигательной системы в течение первых месяцев жизни у недоношенных детей и детей, родившихся в срок

Долинская И.Ю^{1,2}, Солопова И.А.², Жванский Д.С.², Селионов В.А.², Кешишян Е.С.³

¹Московский физико-технический институт (ГУ), г. Долгопрудный, Московская область, Россия

²Институт проблем передачи информации им. А.А.Харкевича, РАН, Москва, Россия 3 Научно-исследовательский клинический институт педиатрии имени академика Ю.Е.Вельтищева ФГБОУ ВО РНИМУ им. Н.И. Пирогова Минздрава России

Ключевые слова: спонтанные движения, недоношенные дети, электромиографическая активность.

Аннотация. Развитие двигательной системы человека в раннем онтогенезе остается актуальным направлением исследований современной физиологии движений и клинической практики. Особенно важно оценивать состояние мышц и формирование правильных навыков движений у недоношенных детей на ранних этапах их жизни, поскольку именно в этот период развития происходит развитие многих отклонений, которые в будущем могут привести к патологии двигательных паттернов. В данной работе исследовали нейромышечный статус недоношенных и доношенных детей первых месяцев жизни на основе параметров электромиографической активности мышц рук и ног при их спонтанных движениях. Детей исследовали в двух возрастных периодах: 0-2 месяца и 2-4 месяца. Как у доношенных, так и у недоношенных детей наблюдалась сходная динамика изменений ЭМГ параметров, выраженная в уменьшении тонической активности мышц рук и ног (кроме мышц голени), уменьшении амплитуды вспышек ЭМГ активности, увеличении их глубины модуляции. Тем не менее, в возрасте 2-4 месяца выявлены отличия параметров ЭМГ активности у недоношенных детей от активности здоровых доношенных младенцев. Наблюдаемые различия в параметрах активности мышц у доношенных и недоношенных младенцев обсуждаются в терминах созревания нейромышечной системы у детей. Полученные данные при сравнении их с данными младенцев того же возраста с синдромом двигательных нарушений могут помочь выявить критерии для прогноза будущих двигательных нарушений.

Работа поддержана грантом РФФИ № 18-015-00187.

Математическая модель аугментации выборки в задачах классификации изображений

Иванова А.А., Коноваленко И.А.

Институт проблем передачи информации им. А.А. Харкевича Российской академии наук — ИППИ РАН, Большой Каретный переулок, д.19 стр. 1, Москва, Россия, 127051 <u>Ivanova@iitp.ru</u>

Ключевые слова: аугментация обучающей выборки, модель изображения, машинное обучение, классификатор

Аннотация. Одной из актуальных проблем, возникающих в задачах обучения нейросетевых классификаторов изображений, является недостаточное количество примеров в обучающей выборке и её нерепрезентативность. Как правило, в таких случаях применяется т.н. аугментация выборки — автоматическое создание дополнительных обучающих примеров, сгенерированных на основе исходной выборки. В данной работе впервые предложена математическая модель аугментации выборки изображений как функционального оператора в пространстве функций распределения изображений.

Работа выполнена при частичной финансовой поддержке гранта РФФИ в рамках научного проекта №17-29-03514.
Prediction of 3D Chromatin Structure

using Recurrent Neural Networks

Michal B. Rozenwald¹, Ekaterina E. Khrameeva², Grigory V. Sapunov^{1,3}, Mikhail S. Gelfand^{1,2}

michal.rozenwald@gmail.com, ekhrameeva@gmail.com, grigory.sapunov@gmail.codelfm, mikhail.gelfand@gmail.com, grigory.sapunov@gmail.codelfm, mikhail.gelfand@gmail.codelfm, grigory.sapunov@gmail.codelfm, grigory.sapunov@gmail.co

¹National Research University Higher School of Economics, Moscow, Russia;

² Skolkovo Institute of Science and Technology, Moscow, Russia; ³ Intento, Berkeley, US

Abstract. In recent years, applications of Machine Learning have substantially increased. In particular, the utilization of these methods in the field of molecular biology is especially significant.

The development of technologies enables rapid generation of large amounts of epigenetic data. The novel Hi-C technology provides an opportunity to obtain data on the genome interactions that has unraveled many principles of chromosomal folding, including the subdivision of the genome into Topologically Associating Domains (TADs). Several studies have confirmed the correlation between the structure of chromatin and other epigenetic features [4, 5, 7].

My research focuses on using Machine Learning methods to predict the three-dimensional structure of chromatin using ChIP-seq epigenetic data on chromatin markers. This paper considers linear models with three types of regularization and an architecture of Recurrent Neural Networks. As a result, predictive models were trained and evaluated, the best score of weighted MSE was performed by Neural Networks.

Informative epigenetic features were identified that leads to studying their biological significance.

Keywords — 3D chromatin structure; topologically associating domains; Machine Learning; Recurrent Neural Networks.

1. INTRODUCTION

Nowadays, Machine Learning algorithms are used in various disciplines. In particular, the utilization of these methods in molecular biology has brought novel results valuable from both theoretical and practical points of view. Investigating the chromatin structure, i.e. the spatial organization of the genome, is a challenging task. Studying the 3D chromatin architecture is crucial as it can shed light on understanding the mechanism of gene regulation, DNA replication timing, and gain more information about diseases.

DNA-protein complex that forms the eukaryotic chromosomes is called chromatin. Although the chromatin structure is folded 10^4 - 10^5 times it maintains fundamental and vital processes of the cell such as transcription, regulation, replication and recombination. These processes are regulated by transcription factors, architectural proteins, histone modifications and other various factors. Some of the regulation mechanisms were shown to act through the three-dimensional structure formation. By the construction of *in silico* predictive models of chromatin interactions probabilities this complex interconnection between chromatin structure, processes and functions might be studied.



Figure 1. Chromatin structure. On the lowest level DNA exists in the form of double helix coiled around complex histone protein structure. DNA with histones is packed in more complex formations, that are composing the cell nucleus. [9]

To collect large-scale data about the 3D chromatin architecture of multiple organisms and cell types, the novel Hi-C technique was designed. Hi-C experiments identify which fragments of the genome are in contact on a genome-wide basis and have unraveled many principles of chromosomal folding (Figure 2) [1]. Although such high-throughput methods of determining interaction probability and chromatin structure *in vivo* were developed less than 10 years ago, the experiments of high quality started to be published mainly during last several years due to elaborate and expensive protocol.



Figure 2. Overview of Hi-C technique. Cells are cross-linked with formaldehyde, then chromatin is digested with a restriction enzyme and the resulting sticky ends are filled in. DNA is purified and sheared. Biotinylated junctions are isolated with streptavidin beads. Then, paired-end sequencing is performed. Hi-C produces a genome-wide contact matrix. Each pixel represents all interactions between a 1Mb locus and another 1Mb locus; intensity corresponds to the total number of reads. [1].

This 3C-based method has revealed, in particular, that chromosomes are subdivided into a set of self-interacting domains called Topologically Associating Domains (TADs) (Figure 3) [5]. It was proven that TADs correlate with units of replication timing regulation in mammals [14], as well as with either active or repressed epigenetic domains in Drosophila [15].



Figure 3. Hi-C interaction map is typically represented as a heatmap - a genome-wide contact matrix. Bright triangles can be visible across the diagonal. They were interpreted as TADs (topologically associating domains), or compact globules of interacting chromatin [5].

Various factors were shown to contribute to the structure formation, and multiple mechanisms were proposed. The detection of factors importance is typically made by looking for specific chromatin structures based on Hi-C maps and finding the factors that correlate or explain these features the most.

2

At the lowest level of organisation, DNA exists in the form of double helix coiled around complex protein structure called nucleosome. The nucleosome consists of proteins called *histones*. Histones can be chemically modified and affect the structure of DNA. Also, architectural proteins of chromatin (such as CTCF and cohesin) can bind DNA and change the topology.

To assess the binding of modified histones and architectural proteins special biochemical protocols were designed, ChIP-Seq one of the most popular of them. ChIp-Seq or ChIP-sequencing is a high-throughput chromatin immunoprecipitation method which allows to map binding of proteins to the DNA.

The rapid growth of epigenetic data (ChIP-seq) enables us to explore the chromatin structure with more sophisticated and complex methods such as Machine Learning and Artificial Neural Networks. The datasets for various factors such as ChIP-Seq experiments for histone modifications and corresponding datasets of Hi-C experiments on chromatin interaction are available from public databases.

TADs have been studied lately through the chromatin marks [7] via ChIP-seq data (Figure 4) [5]. However, the mechanisms that underlie partitioning of the genome into TADs still remain poorly understood. Moreover, there is no comprehensive work investigating all the factors that are publicly available.



Figure 4. Chromosomes partition into TADs and inter-TADs aligns with the distributions of active and repressed chromatin regions respectively. Heat map of a region of Chromosome 2R for BG3 cells is aligned with RNA polymerase II, H3K27ac, and histone H1 ChIP-Seq modENCODE tracks. Inter-TAD regions and TAD boundaries are shaded in blue [5].

This study focuses on bringing insights into the correlation between the 3D Chromatin Structure and epigenetics data using Machine Learning techniques as a tool. The goal is to identify which of chromatin marks are most significant in predicting information about Topologically Associating Domains (TADs).

To that end, I have exploited 4 types of Linear Regression and constructed Recurrent Neural Networks.

After applying the methods quality metrics were calculated and the informative features were investigated.

The analysis of *Drosophila melanogaster* chromatin was performed to explore the principles of TAD folding and the role of epigenetics in this process. So far, no state-of-art solution for this problem has been proposed.

In addition, it is worth mentioning that the same techniques might be used to explore the 3D chromatin structure of Human beings. Moreover, such reconstruction of the information about Hi-C map might be useful not only for understanding the chromatin structure formation, but also have some practical and medical applications. For example, gliomagenesis and limb malformations in humans were demonstrated to be caused by chromosomal topology disruption [16].

The structure of this paper is the following. Section 2 briefly reviews some of the relevant works. Section 3 presents the data, Section 4 the main methods introduced in this study. The problem statement is described in Section 5. Section 6 outlines the experiments as well as the results. Finally, Section 7 concludes the conducted research.

2. LITERATURE REVIEW

This section briefly overviews previous work on applying Machine Learning to biology and exploring the chromatin structure.

Over the last decade, the volume of produced data has significantly increased and brought the opportunity of applying complex and efficient methods. Both supervised and unsupervised learning algorithms have been successfully applied to biological data.

Tan et al. [2] presented a method to identify and extract complex patterns from genomic data using denoising autoencoders, which employ a learning of unknown biological background based on given data. This research proves that Neural Networks can effectively extract key biological principles from gene expression data and summarize them into constructed features with convenient properties even without prior information.

Another illustration of gathering knowledge from complex biological structures using deep learning was performed by Kelley et al. [3]. In 2016, an open source deep Convolutional Neural Networks (CNNs) package that is used to predict the functional activity of DNA sequences was presented. Kelley et al. trained the model on genomics data and gives information about the noncoding genome. The built tool outperforms all previous solutions due to the used powerful CNN.

Various studies were also performed with a focus on exploring the 3D chromatin architecture. In particular, Fortin et al. in [4] succeeded in extracting the knowledge from ChIP-seq data of histone modification to analyze the Chromatin structure. Fortin et al. constructed a predictive model of the Hi-C that unrevealed the correlation with replication timing.



Figure 5. Hi-C heat map binarization to implement Machine Learning techniques and classification. Taken with modifications from [1].

One of the works that inspired my research was made by Ulianov et al. [5]. They suggested that active chromatin and transcription play a key role in chromosome partitioning into TADs. It was shown that numerous transient interactions between nucleosomes of inactive chromatin lead to the formation of TADs that are potentially highly dynamic self-organized structures. On the other hand, nucleosomes, that tend to interact less often, influence the formation of inter-TADs and TAD boundaries. In this research, the correlation between TADs and chromatin marks is revealed. Ulianov et al. showed that active chromatin marks were preferably present at TAD borders, and repressive histone modifications that reflect nucleosome occupancy were depleted within inter-TADs.



Figure 6. Distribution of the individual chromatin marks and proteins near TAD boundaries. Thick rectangles show TAD boundary bins [5].

Moreover, Schreiber et al. [6] considered nucleotide sequence as input for developing a deep convolutional Neural Network. The objective of this architecture was to estimate the Hi-C contacts. The convolutional Neural Network demonstrated that the predicted outcomes are related to histone modification ChIP-seq data, selected functional elements and replication timing which correlates with theoretical knowledge.



Figure 7. The scheme of Rumbutan Neural Network Model for supervised contacts prediction [6].

This proves the hypothesis of the possibility of extracting information about the Hi-C contacts from epigenetics. The ideological difference from the last described approach to mine is that it is performed on *Human* cell lines and takes Nucleotide Sequence and DNaseI assay signal as input, while I explore the DNA of *Drosophila melanogaster* and ChIP-seq data.

To sum up, various Machine Learning methods are used for structure prediction. For example, linear models were applied to Drosophila genome multiple times producing biologically different results [5, 11]. Deep learning methods such as DNN [12] and CNN [6, 13] were also adopted for the problem solution, but they were limited to the prediction of pairwise genomic interactions, but not TADs properties.

3. DATA

3.1 INPUT DATA

Hi-C data for *Drosophila melanogaster* cells of *S2-DRSC* cell line obtained in the laboratory of Prof. Sergey Razin at IGB was binned at 20-kb resolution similarly to [5]. This leads to the Drosophila DNA division into 6,000 input objects called bins of 20-kb resolution.

Each object (*bin*) can be described by a vast number of biological characteristics. As epigenetic features, ChIP-sec data on chromatin marks was chosen as was described in the Introduction Section. This data was downloaded from the modENCODE database [17] and was also processed identically to [5]. Two subsets of ChIP-Seq data that characterize the input objects were analyzed:

The first small group of 5 biologically significant features:

Chriz, CTCF, Su(Hw), H3K27me3, H3K27a.

The second subset contains 18 biological features:

Chriz, CTCF, Su(Hw), H3K27me3, H3K27a, BEAF-32, CP190, Smc3, GAF, Smc3, H3K36me1, H3K36me3, H3K4me1, H3K9ac, H3K9me1, H3K9me2, H3K9me3, H4K16ac

The input data has been normalized using standardization (each feature was centered by subtraction of its mean, and scaled by a division of this feature by its variance) in order to achieve better performance of the Linear Models and Neural Networks. The distributions of the original first 8 Chip-seq features and the transformed data is shown on Figure 8.



Figure 8. Histograms of (A) original Chip-sec data and (B) normalized data of first 8 Chip-sec features of all input bins.

3.2 TARGET VALUE

Armatus software [7] is used to annotate *Topologically Associating Domains* (TADs) with scaling parameter gamma that determines the average size and the number of TADs. When gamma is fixed, each genomic bin is defined as part of a TAD, inter-TAD or TAD boundary. This means that with different settings of the scaling gamma Armatus annotates different TADs partitions.

Given that active chromatin state and transcription determine the chromosome partitioning into TADs, the propensity of a genomic bin to be identified as a boundary should depend on the proportion of active chromatin and transcription level within this bin.

To test this hypothesis, each bin was assigned to a value of the scaling parameter gamma at which this bin switches from being a part of a TAD to being a part of an inter-TAD or a TAD boundary that is called *transitional gamma*. The higher the gamma value is chosen the smaller the Topologically Associating Domains are in the Armatus annotation.



Figure 9. Annotation of TADs at different gamma parameter values. Higher gamma values correspond to smaller TADs. Transitional gamma is defined as the value of gamma at which this bin switches from being a part of a

TAD to being a part of an inter-TAD or a TAD boundary.

The histogram of the target value (transitional gamma) is presented on Figure 10.



Figure 10. Histogram of Transitional gamma. For bins with a transitional gamma value equal to 10, gamma value at which this bin switched from the TAD to the inter-TAD or TAD boundary was not found (this is influencing the choice of the ML Loss Function).

Whole-genome Hi-C maps of chromatin folding in a set of cultured *Drosophila* of *S2-DRSC* cell line were performed and processed similarly to [5].

4. PROBLEM STATEMENT

To avoid ambiguity and inaccuracies, let us clearly define the task and objects in terms of the classical of Machine Learning problem.

- The objects are "bins" DNA sections of length of 20,000 nucleotides, no intersection - Drosophila melanogaster (see Introduction and Section 3.1 for more details).
- The features are *ChIP-Seq epigenetic* data on chromatin markers (Section 3.1).
- The target value is *transitional gamma* parameter of transformation from TAD to inter-TAD, TAD boundary (Section 3.2).

The task is to predict the characteristics of the 3D structure of Chromatin - transitional gamma.

The aim is to identify which of chromatin marks (ChIP-Seq data) are most significant in predicting information about the Topologically Associating Domains (TADs).

5. Methods

In order to explore the correlation between the 3D Chromatin Structure and epigenetics data I apply the following Machine Learning techniques:

- 1. Linear Regression
- 2. Linear Regression with L1 regularization
- 3. Linear Regression with L2 regularization
- 4. Linear Regression with L1 and L2 regularizations

5. Recurrent Neural Networks LSTM

Each classical Machine Learning approach contains input data points 'x' of dimension KxD and target value 'y' of dimension K:

$$\mathbf{X} = [x_1, x_2, ..., x_k], \mathbf{y} = [y_1, y_2, ..., y_k]$$

Each column of the input data matrix is called data characteristic or *feature*.

The aim is to build a model that learns to predict the target 'y' based on 'x' using mapping function 'f' that minimizes the error between the predicted value 'f(x)' and the true value 'y'.

One of the frequently used minimizing function is mean square error (MSE):

$$\sum_{i=1}^{k} (y_i - f(x_i))^2$$

Machine Learning methods differ from each other mainly by the choice of this mapping function f'.

1. Linear regression (LR).

In Linear Regression, the function 'f' is a set of linear functions with weight 'w':

$$f(x) = w_0 + w_1 x + w_2 x^2 + \dots + w_n x^n$$

$$\mathbf{w} = \arg\min_{\mathbf{w}} \sum_{i=1}^k (y_i - f(x_i))^2$$

2. Linear regression with L1 regularization is called LASSO.

In Linear Regression with L1 regularization, a parameter alpha Ωt is added to prevent the model from overfitting. Lasso coefficients tend to sparse the weights of the model which leads to an easier selection of informative features.

$$\mathbf{w} = \arg\min_{\mathbf{w}} \sum_{i=1}^{k} (y_i - f(x_i))^2 + \alpha \sum_{i=0}^{k} |w_i|$$

 Linear regression with L2 regularization is called Ridge LR.

In Linear Regression with L2 regularization, in comparison to L1, the squared values of weights are taken. This method is used to get a better accuracy of the model on new datasets.

$$\mathbf{w} = \arg\min_{\mathbf{w}} \sum_{i=1}^{k} (y_i - f(x_i))^2 + \alpha \sum_{i=0}^{k} w_i^2$$

4. Linear regression with L1 and L2 regularization is called Elastic-Net.

In Linear regression model trained with L1 and L2 regularization. This combination allows for learning a sparse model where few of the weights are non-zero like Lasso, while still maintaining the regularization properties of Ridge. The parameters alpha (α) and l1_ratio (P) are used to control the combination of L1 and L2.

The minimized objective function is:

$$\min_{w} \frac{1}{2n_{samples}} ||Xw - y||_{2}^{2} + \alpha \rho ||w||_{1} + \frac{\alpha(1 - \rho)}{2} ||w||_{2}^{2}$$

When multiple features are correlated with each other Lasso will most pick one of these at random, however both of them are likely to be picked while using Elastic-net.

5. Recurrent Neural Networks (RNN).

Worth noting that input objects are sequentially ordered in the genome. Thus we selected RNN in order to be able to learn this property of data. The values of the target variable are vastly correlated for neighboring objects (bins) in terms of the physical distance in the DNA.



Figure 11. Scheme of using Recurrent Neural Networks. In this work in the model x₁, ..., x_t are the DNA bins with input window size t. The values y₁, ..., y_t represents the corresponding target values *Transitional gamma* of bins x₁, ..., x_t; h₀, ..., h_t - are the hidden states of the RNN model. The Long short-term memory (LSTM) Units [8] are commonly used as building blocks for layers of RNN. A layer of LSTM Neural Networks is calculated with the following formulas:

$$egin{aligned} f_t &= \sigma_g(W_f x_t + U_f h_{t-1} + b_f) \ i_t &= \sigma_g(W_i x_t + U_i h_{t-1} + b_i) \ o_t &= \sigma_g(W_o x_t + U_o h_{t-1} + b_o) \ c_t &= f_t \circ c_{t-1} + i_t \circ \sigma_c(W_c x_t + U_c h_{t-1} + b_c) \ h_t &= o_t \circ \sigma_h(c_t) \end{aligned}$$

The LSTM unit is built by a so-called *cell* that consists of an *input, output* and *forget gates*.



Figure 12.LSTM Units layout. The illustration was taken from

http://colah.github.io/posts/2015-08-Understanding-LSTMs/

Due to the symmetry in the mutual influence of the objects (DNA bins) of the input sequence, while going through the sequence from left to right and visa-versa, the advanced architecture of Bidirectional LSTM Recurrent Neural Networks should be used [10].



Figure 13. General Structure of Bidirectional Recurrent Neural Networks [10].

The illustration was taken from http://colah.github.io/posts/2015-09-NN-Types-FP/

The main model that I implemented to solve the problem of prediction characteristics of the 3D structure of Chromatin - transitional gamma (Section 4) is returning as output the target value of the middle object of the input sequence (Figure 14).



Figure 14. Scheme of Bidirectional LSTM Recurrent Neural Networks with 1 output that was implemented. x1, ..., xt are the DNA bins with input window size t. The values Yt/2 represents the corresponding target values *Transitional gamma* of bins the middle bin Xt/2. h0, ..., ht - are the hidden states of the RNN model.

5.1 Loss Function

The target object is a continuous value from 0 to 10 (Section 4) which leads to solving a Regression task. The classical optimisation function that is minimized in this type of problems is Mean Square Error (MSE):

$$\sum_{i=1}^{K} (y_{true_i} - y_{pred_i})^2$$

However, the distribution of the target (transitional gamma) is significantly unbalanced (Figure 10). The target value of most of the objects is in the diapason between 0 and 3. Nevertheless, the contribution of the error on objects with a high true value of the target will also be high in the total score using Mean Square Error.

Moreover, the biological nature of objects with a high value of the transitional gamma is different from other objects. For DNA bins with a transitional gamma value equal to 10, gamma value at which this bin passed from the TAD state to the inter-TAD or TAD boundary was not found.

To build a model that accurately predicts the values of the transitional gamma for most objects, it was suggested to introduce our own custom loss function - the *weighted Mean Square Error (wMSE)*:

$$\sum_{i=1}^{K} (y_{true_i} - y_{pred_i})^2 \frac{(11 - y_{true_i})}{11}$$

$$MSE \qquad weight$$

WMSE consists of two main parts: MSE multiplied by the weight (penalty) of the error, depending on the true value of the target variable. Thus, the model is penalized less for errors on objects with a high value of the transitional gamma by using the weighting.

5.2 MOTIVATION OF THE MODELS CHOICE

The choice of the Linear models was made as it is a classical approach that provides the opportunity for intuitive feature importance interpretation and prevents the model from data overfitting.

The implementation of Neural Networks enables us to extract complex dependencies and captures non-linearity of the biological data.

For each of the methods, the wMSE quality metrics were calculated and the selected informative features were investigated.

6. EXPERIMENTS AND RESULTS

After performing data preparation (normalization and selection of data without omissions) the *total number* of DNA bins of 20-kb resolution with ChIP-Seq features and transitional gamma value identified is equal to 5950 objects.

To cope with models overfitting the data was always randomly separated into 2 groups: *Train* dataset 70% of data (4156 bins) and *Test* dataset (1785 bins). And 10% of the Train (415 bins) data was used for *Validation*.

6.1 LINEAR MODELS (LM)

For each of the two subsets of input features with 5 and 18 ChIP-seq characteristics 4 types of Linear Regression were conducted.

In order to get more sustainable and consistent results, I have performed training for each type of the models several times. The results of the learning process of the Regression models may differ between different launches of training the same type of model due to the facts that the coefficients of the models are initialization randomly, the division into train and test sets are also random, as well as the possibility of the model to converge to a local minimum. To tackle this issue, 5 experiments were conducted for the set of 5 input features and 8 trials for the set with 18 characteristics. Next, weighted Mean Square Error (wMSE) that is defined in Section 5 was calculated for each experiment. Finally, the mean value and variance of these experiments on Train and Test datasets was found and is presented on Figure 15.

The best score of the weighted Mean Square Error using Linear Regression with L1 and L2 regularization (Elastic Net model) with parameter alpha (α) equal to 0.2 was performed using grid search.

Weighted MSE of Linear Models

	51	features		18 features			
	Train	Test mean, std		Train	Test mean, std		
Lin Reg	1.19	1.19	0.26	1.13	1.13	0.35	
Lin Reg + L1	1.16	1.16	0.24	1.10	1.10	0.36	
Lin Reg + L2	1.19	1.18	0.26	1.13	1.14	0.35	
Lin Reg + L1 + L2	1.16	1.16	0.23	1.10	1,10	0.31	
Const prediction	1.61	1.62	0.09	1.61	1.62	0.09	

Figure 15. Results of weighted Mean Square Error (wMSE) evaluation of 4 types of Linear models and

Constant prediction by the mean value of the training dataset. There were held 5 experiments for a dataset with 5 features and 8 experiments for a dataset with 18 features and the mean and the standard deviation (std) values on Train and Test dataset are shown.

6.2 FEATURE IMPORTANCE LM

While using Linear Model *feature importance* can be explored by analyzing the weight coefficients of the model. Given that the prediction is made based on multiplication of each weight on the corresponding feature (Section 5) we may say that the bigger the absolute value of the feature is, the higher is the influence of this particular feature on the prediction of the model. This is how we can extract the prioritization in terms of the influence of the features from the model.

After conducting 5 experiments on the first dataset with 5 ChIP-seq characteristics the resulting weights happen to be significantly stable as is shown in the table of feature coefficient of Linear Regression below (Figure 16).

	Chriz	CTCF	Su(Hw)	H3K27me3	H3K27ac
0	-0.478129	-0.305786	0.045819	0.188004	-0.208210
1	-0.458402	-0.309242	0.041717	0.200283	-0.227345
2	-0.434138	-0.322330	0.029390	0.203778	-0.252761
3	-0.452783	-0.306535	0.036500	0.202775	-0.221106
4	-0.475426	-0.309130	0.066029	0.214260	-0.226149

Figure 16. Weights of features of 5 trained Linear Regression models (rows) on the dataset with 5 ChIP-Seq features (columns).

As a result, we obtain that the most valuable in terms of the absolute value of the feature weight is *Chriz*, then *CTCF*, *H3K27ac* and *H3K27me3*, when the weight of Su(Hw) is the smallest.

The same approach was presented on the second dataset with 18 ChIP-seq characteristics with 8 conducted experiments (Figure 17).

	Chriz	CTCF	Su(Hw)	H3K27me3	H3K27ac	BEAF-32	CP190	Smc3	GAF	H3K36me1
0	-0.248587	-0.226543	-0.036198	0.119583	-0.161412	-0.038226	-0.266243	9.0520866-03	0.083495	-0.194714
1	-0.250354	-0.226977	-0.040304	0.130819	-0.160016	-0.027474	-0.265732	-2.515296e+13	0.099725	-0.191472
2	-0.262170	-0.226699	-0.072717	0.128651	-0.159354	-0.040149	-0.251956	6.629252e-03	0.098606	-0.199522
3	-0.273043	-0.234114	-0.039528	0.122537	-0.183555	-0.032864	-0.251123	-2.029188e+12	0.081541	-0.199011
4	-0.252714	-0.222996	-0.035098	0.125873	-0.179882	-0.031258	-0.282027	-1.1725758-03	0.070258	-0.172316

Figure 17. Weights of features of 5 trained Linear Regression models (corresponding to the rows) of the first 10 ChIP-seq characteristics (columns) of the second dataset with 18 features.

With the aim of identifying feature importance this table of feature weights was analyzed. In comparison to the same application on a dataset of 5 features the coefficient order by absolute value of the 18 features is less stable.

To identify which features are detected by the Linear Regression model as more influencing, first of all, I have selected at each of the experiment the top 4 ChIP-seq features that has the biggest absolute value of corresponding coefficients.

Further, the numbers of occurrences of each of the feature indexes in the selected list of most influencing features were calculated. Finally, I have sorted the features based on this frequency number. The result is the following:

Chriz and *H3K9me2* were proved to be influencing most often.

CTCF and *CP190* were identified as the second degree of significant, then *H3K9ac*.

Su(Hw), H3K27me3, H3K27ac, BEAF-32, Smc3, GAF, H3K36me1, H3K36me3, H3K4me1, H3K9me1, H3K9me3, H4K16ac, Smc3.1 were all identified as equally valuable by the described score.

Another result that is worth mentioning is the selection of only one important feature (Chriz) out of all 5 and 18 datasets while using the model of Linear regression with L1 regularization (Figure 18).



Figure 18. Results of the learned coefficients of Linear regression with L1 regularization.

6.3 RNN MODELS

The main Neural Network that I was exploring is bidirectional LSTM. As was described above the sequential relationship of the input objects (DNA bins) in terms of the physical distance in the DNA justifies the usage of Recurrent Neural Networks (RNN).

The variable parameters that I investigated in our LSTM model are:

- 1. A sequence length of:
 - Input RNN objects: The input sequence is a set of consecutive DNA bins with fixed input length (*window size*): 1, ..., 4, 5, 6, ..., 10
 - Output objects: Predicting for each bin from the input sequence its target transitional gamma value.

- The output in this type of model is a sequence of the same length as the input, where each value corresponds to the target variable of the input objects.

- Predict middle value: Predicting one target value of the bit that is in the middle of the sequence, i.e. the bit that has the index that equals the integer value of the input window size divided by 2.

The output is one scalar value of the central target.

- 2. The dimension of feature space of each bin of the input sequence:
 - 5 biologically significant features:

Chriz, CTCF, Su(Hw), H3K27me3, H3K27a

- 18 biological features:

Chriz, CTCF, Su(Hw), H3K27me3, H3K27a,

BEAF-32, CP190, Smc3, GAF, Smc3,

H3K36me1, H3K36me3, H3K4me1, H3K9ac,

H3K9me1, H3K9me2, H3K9me3, H4K16ac.

- 3. Number of LSTM Units: 1, 16, 64, 128, 256
- 4. Number of training Epochs:
 - 1, 16, 64, 128, 256
 - Early Stopping automatic stopping of the training when the loss function stops decreasing after a fixed amount of epochs
- 5. Optimizer: Adam

The target in our task is continuous values from 0 to 10 (see section 4 with the data description) with unbalanced distribution. This is why the main loss function and evaluation score of the predictions is weighted Mean Square Error (wMSE) that is defined in Section 5.

- 6. Loss function:
 - Weighted Mean Square Error (wMSE) our custom evaluation function defined in section 5
 - Mean Square Error (MSE)
- 7. Evaluation metrics:
 - Weighted Mean Square Error (wMSE)
 - Mean Square Error (MSE)
 - Errors counted by true target value RMSE, wMSE.

For each variation of parameters, experiments were conducted and evaluation metrics were calculated (more tables with results can be found in Section 9).

Calculating the error metrics of the trained models and dividing it into components based on the true value of the target value showed that MSE almost linearly increases when increasing transition gamma, while the wMSE for large target values is smaller and higher errors occur on rare gamma values (Figure 19).



Figure 19. Result of calculating RMSE and wMSE of trained models by folds of a true target value.

One of the approaches of evaluation the models is to compare the evaluation metrics of the predictions of transitional gammas by the trained models to the same evaluation score applied to constant prediction. In this work the constant value is equal to the mean value of the training dataset.

In order to implement these models, open source Machine Learning libraries PyTorch and Keras on Python were chosen.



Figure 20. The decreasing wMSE and MSE learning curves during epochs on training and validation datasets prove that the model becomes more accurate.

To explore the dependencies of weighted Mean Square Error on the sizes of *sequence length* Bidirectional RNN models with different input window size and a number of LSTM Units were trained with Cross Validation. The result is shown on Figures 20 and 21 where an optimal sequence size that equals to input window size 6 and 64 LSTM Units was revealed both on the training and testing datasets.



Figure 21. weighted Mean Square Error of trained Bidirectional LSTM with 64 units for different sizes of sequence length. The upper graph presents the results for the training dataset, the bottom one - wMSE counted on the testing objects. The green box is surrounding the minimal score that was achieved using as an input sequence of 6 bins.





Figure 22. Weighted Mean Square Error of trained Bidirectional LSTM with an input sequence of 6 bins for a different number of LSTM Units. The upper graph presents the results for the training dataset, the bottom one - wMSE counted on the testing objects. The green box is surrounding the minimal score that was achieved while using 64 LSTM Units.

6.4 FEATURE IMPORTANCE RNN

In order to explore the importance of each input biological characteristic of DNA bin the following approach was performed.

To check how valuable is feature X from the input space we manually replace the values of the corresponding column of the feature matrix with zeros. Feather we calculate the evaluation metrics and check how significantly different they are from the results obtained while using the complete set of data.

Results of as input all features but one.

missing_feature	n_tries = 4	n_epochs	Wmse_test	Wmse_train
Chriz	mean	21	0.9 ± 0.01	0.87 ± 0.01
CTCF	mean	30	0.89 ± 0.01	0.84 ± 0.01
Su(Hw)	mean	32	0.89 ± 0.02	0.83 ± 0.03
H3K27me3	mean	32	0.89 ± 0.01	0.83 ± 0.01
H3K27ac	mean	32	0.89 ± 0.01	0.85
FULL	mean	27	0.9 ± 0.01	0.85 ± 0.03

Figure 23. Weighted Mean Square Error on Train and Test dataset while dropping out one of the input ChIP-Seq features and using this full dataset. Bidirectional LSTM with 64 units with 6 input bins with Early Stopping was used. The results of wMSE on the test set do not differ dramatically from using full dataset. When we drop out each of the 5 features we get the same score of around 0.89 that is almost equal to using all of them together.

The same described approach was applied to explore the importance of the second dataset with 18 ChIP-seq features. These wMSE scores are presented in Figure 24 as well as the result of training the model on all 18 features.



Figure 24. Weighted Mean Square Error on Train and Test dataset while dropping out one of the input ChIP-Seq

features and wMSE of training using all 18 features (full). Bidirectional LSTM with 64 units with 6 input bins was used.

The difference between the wMSE using all 18 features and omitting each one separately is presented on Figure 25. This provides the opportunity of identifying how valuable a particular biological characteristic is using RNN.



Figure 25. The difference between the weighted Mean Square Error using all 18 features and dropping each one separately. Blue bars corresponds to the wMSE on a testing dataset, red - training.

As a result, the Bidirectional LSTM Recurrent Neural Networks with 64 LSTM Units and sequence of 6 bins taken as input data were trained and achieved better evaluation scores than a constant prediction by the mean value of the training dataset and Linear models:

	Weigh	of all Mo	dels					
	5	features		18	18 features			
	Train	Test mean, std		Train	Test me	an, std		
Lin Reg	1.19	1.19	0.26	1.13	1.13	0.35		
Lin Reg + L1	1.16	1.16	0.24	1.10	1.10	0.36		
Lin Reg + L2	1.19	1.18	0.26	1.13	1.14	0.35		
Lin Reg + L1 + L2	1.16	1.16	0.23	1.10	1.10	0.31		
Const prediction	1.61	1.62	0.09	1.61	1.62	0.09		
biLSTM 64 - 6 bins	0.9	0.85	0.1	0.9	0.9	0.1		

Figure 26. Full results of weighted Mean Square Error (wMSE) evaluation of the best models from the explored classes: 4 types of Linear models,

constant prediction by the mean value of the training dataset and Bidirectional LSTM Recurrent Neural Networks. There were held 5 experiments for dataset with 5 features and 8 experiments for dataset with 18 features and the mean and the standard deviation (std) values on Train and Test dataset are shown.

The smallest weighted Mean Square Error was performed by Recurrent Neural Networks with a size of input sequence equal to 6 bins and 64 LSTM Units.

7. CONCLUSION

The correlation between the ChIP-seq and Hi-C data was confirmed by building and training the predictive models for chromatin feature (gamma transitional) and achieving dignified evaluation metrics. Moreover, the results were proven to be interpretable and biologically relevant.

In particular, evaluation scores were calculated for the following implemented and trained models:

- Linear Regression
- Linear Regression with L1 regularization
- Linear Regression with L2 regularization
- Linear Regression with L1 and L2 regularizations
- Bidirectional LSTM RNN models with custom Loss Function (weighted Mean Square Error) with a variation of the sequence length, number of LSTM Units, number of training epochs.

Best results were obtained while setting the parameters of input window size equal to 6 with 64 LSTM Units using an Early Stopping technique to automatically identify the optimal number of training epochs.

All the models have performed better than constant prediction with the mean value of the training dataset and bidirectional LSTM RNN performed much better than all types of Regression. Further, feature importance analysis of the input ChIP-Seq data was conducted. While exploring the feature weights of the trained linear models an importance order was detected in the set of the selected biologically meaningful ChIP-Seq characteristics.

The most significant in terms of the absolute value of feature weight of the first dataset with 5 Chip-Seq characteristics is *Chriz*, then *CTCF*, *H3K27ac* and

H3K27me3, while the weight of Su(Hw) is the smallest.

The result of exploring the second dataset of 18 ChIp-Seq characteristics is the following:

Chriz and *H3K9me2* were proved to be influencing most often.

CTCF and *CP190* were identified as the second degree of significant, then *H3K9ac*.

Su(Hw), H3K27me3, H3K27ac, BEAF-32, Smc3, GAF, H3K36me1, H3K36me3, H3K4me1, H3K9me1, H3K9me3, H4K16ac, Smc3.1 were all identified as equally valuable by the described score.

Moreover, after training Linear Regression with L1 regularization on both of the datasets only one ChIP-Seq feature (Chriz) was detected as the most influencing.

While applying Neural Network models on the set of 4 ChIP-Seq characteristics out of the first dataset of 5 features the weighted Mean Square Error of all models appeared to be almost equal. The results of applying the same technique while omitting each feature one by one using the second dataset of 18 ChIP-Seq features allowed to evaluate the biological impact of the features.

Besides that, the ideas of future research are to test the transferability of the results between different cell types and species. The same implemented pipeline can be applied to various cell lines and can be easily adapted to any other similar biological dataset of chromatin features. Another direction could be to learn a latent representation of the objects (embedding) and to explore the influence on the predictions. However, this might be done in order to achieve better prediction but might make the interpretability more challenging.

This study provided an opportunity to learn more about the architecture of Linear Models and Recurrent Neural Networks, as well as to practice using open source Python libraries pyTorch and Keras, and to dig into the details of applying Machine Learning to real life data of biological experiments. As a result, the working pipeline was designed and implemented. Interpretable and biologically relevant observations were produced.

8. References

[1] Lieberman-Aiden E., van Berkum N., Williams L et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science, 326(5950), (2009), pp. 289–293.

[2] Tan J., Ung M., Cheng C., Casey S. Greene: Unsupervised Feature Construction and Knowledge Extraction from Genome-Wide Assays of Breast Cancer with Denoising Autoencoders. Pacific Symposium on Biocomputing (2015), pp. 132-143.

[3] Kelley D., Snoek J., and Rinn J. Basset: learning the regulatory code of the accessible genome with deep convolutional Neural Networks. Genome Research 26(7), (2016), pp. 990–999.

[4] Fortin J., Hansen K. Reconstructing A/B compartments as revealed by Hi-C using long-range correlations in epigenetic data. Genome Biology, (2015), pp. 16.

[5] Ulianov S., Khrameeva E.E., Gavrilov A.A., Flyamer I.M., Kos P., Mikhaleva EA, Penin AA, Logacheva MD, Imakaev MV, Chertovich A, Mikhail S. Gelfand et al. Active chromatin and transcription play a key role in chromosome partitioning into topologically associating domains. Genome Research 26, (2016), pp. 70-84.

[6] Schreiber J, Libbrecht M, Bilmes J, Noble W. Nucleotide sequence and DNaseI sensitivity are predictive of 3D chromatin architecture. bioRxiv, (2017), 103614.

[7] Filippova D, Patro R, Duggal G, Kingsford C. Identification of alternative topological domains in chromatin. Algorithms for molecular biology AMB, (2014), pp. 9-14.

[8] Hochreiter S, Schmidhuber J. Long short-term memory. Neural Computation 9 (8), (1997) pp. 1735–1780.

[9] Bonev B., Cavalli G.Organization and function of the 3D genome. Nature Reviews Genetics 17, (2016), p. 772.

[10] Schuste M., Paliwal K.K. Bidirectional Recurrent Neural Networks. IEEE Transactions on Signal Processing archive 45, (1997), 2673-2681

[11] Cubeñas-Potts C. et al. Different enhancer classes in Drosophila bind distinct architectural proteins and mediate unique chromatin interactions and 3D architecture. Nucleic Acids Research, (2017), pp. 1714–1730.

[12] Bkhetan Z.A., Plewczynski D. Three-dimensional Epigenome Statistical Model: Genome-wide Chromatin

Looping Prediction. Scientific Reports 8, (2018) Article number: 5217.

[13] Singh S., Yang Y., Poczos B., Ma J. Predicting enhancer-promoter interaction from a genomic sequence with deep neural networks. bioRxiv, (2018), 85241

[14] Pope B., Ryba T, Dileep V., Yue F. et al. Topologically associating domains are stable units of replication-timing regulation. Nature 515, (2014), pp. 402–405.

[15] Sexton T., Yaffe E., Kenigsberg E. et al. Three-dimensional folding and functional organization principles of the Drosophila genome. Cell 148, (2012), pp. 458–472.

[16] Krijger P., De Laat W. Regulation of disease-associated gene expression in the 3D genome. Nature Reviews Molecular Cell Biology 17, (2016), pp. 771-782.

[17] Celniker S., Dillon L., Gerstein M., Gunsalus K., Henikoff S., Karpen G., et al. Unlocking the secrets of the genome. Nature 459, (2009), pp. 927-30.