

**МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
имени М. В. ЛОМОНОСОВА**

Факультет вычислительной математики и кибернетики

**СБОРНИК ТЕЗИСОВ
ЛУЧШИХ ВЫПУСКНЫХ
КВАЛИФИКАЦИОННЫХ РАБОТ
ФАКУЛЬТЕТА ВМК МГУ 2022 ГОДА**

МОСКВА

2022

Оглавление

Кафедра математической физики	9
<i>М. С. Величкин</i>	
Математическое моделирование структурообразования в модели матричной фурье-фильтрации	9
<i>А. С. Вешкин</i>	
Многомасштабный метод поиска по содержимому для полнослайдовых гистологических изображений	11
<i>А. С. Дунаева</i>	
Обнаружение аномалий на фотографиях поля сельскохозяйственных культур без предварительной разметки	13
<i>В. В. Ефремцев</i>	
Коэффициентная обратная задача для одномерного уравнения теплопроводности	14
<i>В. Б. Ипполитов</i>	
Мнокадровый метод повышения разрешения цветных изображений .	16
<i>Д. М. Кабанов</i>	
Математическая модель рассеяния плоской электромагнитной волны на двух эллипсоидах на основе метода интегрального уравнения в спектральной области	18
<i>А. Е. Колотов</i>	
Исследование метода восстановления волнового фронта по его локальным наклонам с локализованным стабилизатором дробной гладкости	19
<i>Д. Г. Тихвинский</i>	
Математическое моделирование распространения лазерного импульса в нелинейной среде, описываемой уравнением Гинзбурга—Ландау с однородными граничными условиями 1-го рода	21
Кафедра вычислительных технологий и моделирования	23
<i>Б. И. Валиахметов</i>	
Итерационные методы на основе составных пространств Крылова .	23

<i>А. Д. Михайлов</i>	
Оптимальное матричное и тензорное восполнение с дополнительной информацией	25
<i>М. К. Мордвинцев</i>	
Алгоритм дополнения тензора с фиксированными рангами Таккера с использованием оптимизации на римановых многообразиях	26
<i>А. С. Рыгин</i>	
Математическое моделирование динамики новой коронавирусной инфекции	27
Кафедра вычислительных методов	29
<i>М. С. Григорьева</i>	
Численное исследование влияния земли на аэродинамические характеристики жёсткого винта вертолёта	29
<i>В. Ю. Цветинский</i>	
Анализ и прогноз временных рядов на примере рынка криптовалют	31
Кафедра автоматизации научных исследований	33
<i>А. А. Долматов</i>	
Разработка программно-информационной среды для сопровождения экспериментов на установках токамак	33
Кафедра нелинейных динамических систем и процессов управления	35
<i>Г. И. Артамонов</i>	
Моделирование алгоритма управления группой мобильных роботов на основе методов потенциальных полей	35
<i>И. И. Бутиков</i>	
Моделирование распределённых регуляторов для систем с запаздыванием	37
<i>В. Н. Кондратьева</i>	
Построение системы программного управления, реализующей колебательные движения в динамическом объекте с гистерезисом	39
Кафедра системного анализа	41
<i>Ф. А. Федоров</i>	
Управление групповым поведением экономических агентов в приближении игры среднего поля	41
Кафедра математической статистики	42
<i>А. Л. Виляев</i>	
Методы построения торговых стратегий для валютных пар на основе статистических моделей и сетей LSTM	42

<i>С. А. Горбунов</i>	
Статистический анализ данных трафика виртуального мобильного оператора	44
<i>В. К. Клименко</i>	
Критерий проверки биоэквивалентности при наличии пропущенных данных	46
<i>М. Ю. Муляр</i>	
Анализ поведения явных и скрытых потоков тепла в Северной Атлантике с помощью методов математической статистики	48
<i>И. П. Пахненко</i>	
Решение обратной задачи магнитоэнцефалографии с двумя источниками	49
<i>М. В. Селезнёв</i>	
Методы сегментации изображений без учителя	51
<i>Д. П. Хуторной</i>	
Подбор гиперпараметров генеративных состязательных сетей, оптимизация модели	52
Кафедра математических методов прогнозирования	54
<i>С. Ю. Елистратов</i>	
Улучшение автоматической стилизации изображений с помощью преобразующей нейросети	54
<i>А. И. Самбурский</i>	
Нейросетевые методы в задаче оптимизации логических схем	56
<i>В. И. Тыцкий</i>	
Система визуального анализа для активности головного мозга	58
Кафедра математической кибернетики	59
<i>М. И. Миронов</i>	
Некоторые интервалы в решетке замкнутых классов частичных булевых функций	59
Кафедра информационной безопасности	61
<i>Линь Пэйфэн</i>	
Свойства произведения Адамара констациклических линейных кодов	61
<i>Д. Д. Миронов</i>	
Исследование встречаемости небезопасной сериализации программных объектов в клиентском коде веб-приложений	63
<i>Д. И. Назаров</i>	
Поиск информации о принимаемых сервером запросах в закомментированном клиентском коде веб-приложений	65

<i>Н. М. Петрова</i>	Исследование и программная реализация алгоритма шифрования Кукувиноса, основанного на матрицах Адамара с циркулянтным ядром	67
<i>В. А. Тарасова</i>	Разработка архитектуры и прототипа цифрового двойника для умных шахт	69
Кафедра интеллектуальных информационных технологий		70
<i>Ф. Д. Губанов</i>	Полуавтоматическая классификация транзиентов в рентгеновском обзоре неба	71
<i>А. Е. Гуцин</i>	Разработка алгоритма детектирования смен сцен в видео на основе нескольких метрик	73
<i>Д. Т. Сивцов</i>	Исследование и разработка методов автоматической аннотации электронной почты	74
Кафедра автоматизации систем вычислительных комплексов		75
<i>А. В. Абрамов</i>	Построение однопроцессорного расписания с минимизацией пикового потребления ресурса при помощи муравьиного алгоритма	75
<i>Е. А. Ветрова</i>	Разработка прототипа системы масштабирования сетевых функций для MANO (Management and Orchestration) платформы на основе легковесной виртуализации	77
<i>А. Д. Казантаев</i>	Алгоритм балансировки нагрузки в распределенной вычислительной системе с применением мультиагентного обучения с подкреплением	79
<i>А. П. Коваленко</i>	Разработка методов трехмерной визуализации задержек в топологической структуре вычислительных кластеров	81
<i>А. В. Ларин</i>	Исследование метода прогнозирования нагрузки контроллера в программно-конфигурируемых сетях на основе нейронных сетей	83
<i>И. Р. Танкаев</i>	Федеративное обучение с использованием градиентного бустинга	85
<i>Е. А. Чехонина</i>	Совместное планирование вычислений и информационного обмена в системах с разделяемой пропускной способностью сети	86

Кафедра суперкомпьютеров и квантовой информатики	88
<i>И. П. Жердев</i>	
Разработка и исследование параллельных методов кластеризации большого количества текстовых сообщений с использованием алгоритма DBSCAN	88
<i>И. Р. Плужников</i>	
Свойства запутанных состояний многокубитных систем	90
<i>В. В. Федоров</i>	
Применение методов машинного обучения в задачах распознавания русских дореволюционных печатных текстов	92
Кафедра алгоритмических языков	93
<i>А. В. Афракова</i>	
Методы автоматического морфемного разбора на основе машинного обучения	93
<i>Р. О. Бежанян</i>	
Извлечение аргументов из мнений пользователей	94
<i>А. К. Вересов</i>	
Поддержка вычислительной модели Лиспа для языка Си	96
<i>Н. И. Ершов</i>	
Генератор синтаксических анализаторов на основе рекурсивного обхода L-графов	98
<i>Д. А. Рогожин</i>	
Разработка виртуального помощника	100
<i>И. С. Рожков</i>	
Исследование методов автоматического извлечения вложенных именованных сущностей	102
<i>П. А. Степанова</i>	
Автоматическое распознавание терминологических отношений на основе шаблонов и правил	104
<i>Д. В. Хомский</i>	
Методы автоматического обнаружения изменения значений слов	105
<i>Ю. А. Шевернев</i>	
Разработка типового программного решения для публикации документов с использованием языка разметки TEI	107
<i>С. А. Яковлева</i>	
Методы аспектного анализа научно-технических текстов	108
Кафедра системного программирования	110
<i>Е. С. Басков</i>	
Сокращение поверхности атаки на ядро операционной системы Linux	110

<i>Г. Д. Гусев</i>	
Преобразование последовательных Фортран-программ для их распараллеливания на гибридные кластеры	111
<i>Д. О. Мотриченко</i>	
Нейросетевые методы классификации целей цитирования	114
<i>Т. А. Озерова</i>	
Исследование и разработка системы предоставления высокопроизводительных кластеров в облачной среде по запросу с поддержкой видеоускорителей	116
<i>М. С. Черепнина</i>	
Метод увеличения устойчивости моделей распознавания изображений без потери точности на чистых данных	117
Именной указатель	120

Математическое моделирование структурообразования в модели матричной фурье-фильтрации

Величкин Матвей Сергеевич

Кафедра математической физики

e-mail: angry_gnom@mail.ru

Научный руководитель — д.ф.-м.н. проф. Разгулин Александр Витальевич

В обработке сигналов часто используется фурье-фильтрация, то есть изменение фурье-образа сигнала. Оптические системы позволяют легко реализовать фурье-фильтрацию [1], поэтому рассматривалась модель нелинейной оптической системы с обратной связью с пространственным фильтром. Динамика системы описывается уравнением

$$\partial_t u + u - D \partial_{xx}^2 u = F(u, P), \quad (1)$$

где

$$F(u, P) = K |\Phi_{E+P}(A_{in} \exp\{iu\})|^2,$$

$\Phi_{E+P}(f)$ — оператор матричной фурье-фильтрации, линейно преобразующий коэффициенты ряда Фурье функции $f \in L_2[0, 2\pi]$, $D > 0$ — коэффициент диффузии, $E + P$ — матрица фильтра и $A_{in} = A_{in}(x) = const \in \mathbb{C}$ — входная волна.

Общая постановка задачи с матричным фильтром введена в статье [2]. В качестве бифуркационного параметра выбран коэффициент обратной связи K , зафиксировано критическое значение \hat{K} , при котором будут выполнены условия бифуркации, и введён малый параметр μ : $K = K(\mu) = \hat{K} + \mu$. При некоторых ограничениях на матрицу существует пространственно-однородное решение χ , тогда в его окрестности задача может быть сведена к локальной форме и можно исследовать спектр главной линейной части $L = L(\mu)$.

Основная из задач, поставленных в данной работе, — найти способ выбора матрицы фильтра, позволяющий построить решение с наперёд заданными свойствами, при помощи бифуркации Тьюринга для стационарных структур и бифуркации Хопфа для динамических структур.

Также стояла задача исследования устойчивости решений, рождающихся в результате суперкритической бифуркации Тьюринга: по схеме [3] доказано, что при выполнении условий бифуркации Тьюринга стационарное решение существует и асимптотически устойчиво. В [2] показано, что при выполнении условий бифуркации Хопфа периодическое решение существует и орбитально асимптотически устойчиво.

Сначала рассмотрены фильтры, нетривиально воздействующие на одну или две целочисленных частоты. В этом случае, далее называемом «базовым», нетрудно найти в явном виде формулы для элементов фильтра, гарантирующие выполнение условий бифуркации с заданной бифуркационной собственной функцией. Если необходимо больше гармоник в решении, матрица модифицируется так, чтобы собственные значения не изменились, но бифуркационная

собственная функция приняла нужное значение. Для достижения этого фиксируются коэффициенты «базового» случая, 4 для одной частоты или 16 для двух частот, расположенные соответственно в 2 или 4 столбцах, и остальные коэффициенты этих столбцов вычисляются в соответствии с выведенными в работе формулами.

Разработанный метод позволяет получить в линеаризованной задаче в качестве бифуркационной собственной функции любую допустимую линейную комбинацию конечного числа экспонент. То есть для фиксированного $n_0 \in \mathbb{N}$ и малых $\mu > 0$, которым соответствуют малые ε :

для бифуркации Тьюринга

$$u(x, t) = \chi + \varepsilon \sum_{n=1}^{n_0} (a_n \cos(nx) + b_n \sin(nx)) + O(\varepsilon^2),$$

$$\forall a_n, b_n \in \mathbb{R}, \quad \sum_{n=1}^{n_0} (a_n^2 + b_n^2) = 1,$$

для бифуркации Хопфа

$$u(x, t) = \chi + \varepsilon p_{\hat{n}} \cos(\hat{\omega}t + \varphi) - \varepsilon q_{\hat{n}} \sin(\hat{\omega}t + \varphi) + O(\varepsilon^2),$$

$$p_{\hat{n}} = \sum_{n=1}^{n_0} (a_n \cos(nx) + b_n \sin(nx)), \quad q_{\hat{n}} = \sum_{n=1}^{n_0} (c_n \cos(nx) + d_n \sin(nx)),$$

с некоторыми ограничениями на действительные a_n, b_n, c_n и d_n .

Для подтверждения результатов, полученных аналитически для линейной задачи, проведено численное моделирование исходной нелинейной задачи (1), с учётом условий периодичности и с различными начальными условиями. Вычислительные эксперименты показали адекватность метода построения фильтров, а также что в формулах, не определяющих коэффициенты однозначно, правильный выбор коэффициентов позволяет увеличить амплитуду ε , при которой остаточный член $O(\varepsilon^2)$ начинает заметно менять вид решения.

Литература

- [1] Degtiarev E. V., Vorontsov M. A. Spatial filtering in nonlinear two-dimensional feedback systems: phase-distortion suppression // Journal of the Optical Society of America B. 1995. Vol. 12, P. 1238–1248.
- [2] Razgulin A. V., Sazonova S. V. Hopf bifurcation in diffusive model of nonlinear optical system with matrix fourier filtering // Communications in Nonlinear Science and Numerical Simulation. 2019. Vol. 77, P. 288–304.
- [3] Sattinger D. H. Topics in stability and bifurcation theory // Lecture Notes in Mathematics. E. : Springer, 1973. Vol. 309, P. 204.

Многомасштабный метод поиска по содержанию для полнослайдовых гистологических изображений

Вешкин Артемий Сергеевич

Кафедра математической физики

e-mail: artem.veshkin@gmail.com

Научный руководитель — к.ф.-м.н. м.н.с. Хвостиков Александр Владимирович

Целью дипломной работы является разработка метода поиска по содержанию для полнослайдовых гистологических изображений и оценка эффективности его применения в качестве вспомогательного средства в работе врачей гистологов.

В качестве материала для исследования используются полнослайдовые изображения из набора данных PATH-DT-MSU [1, 2]. Изображения имеют размер порядка 100 тыс. × 100 тыс. пикселей и занимают несколько гигабайт. Большая их часть сделана на оптическом увеличении ×40.

Целью метода является возможность для любого выделенного участка изображения найти структурно похожие фрагменты изображений в имеющемся наборе изображений. Это может существенно облегчить и ускорить работу гистологов, так как ручной анализ такого рода снимков крайне затруднителен. Для поиска структурно похожих фрагментов изображений в работе исследуются три способа составления векторов признаков, описывающих гистологические структуры на изображениях.

Первый способ заключается в вычислении текстурных характеристик с использованием матриц смежности и общих яркостных характеристик изображения. Второй способ использует обученную сверточную нейросетевую модель классификации в качестве экстрактора признаков. Для ускорения поиска вектора признаков кодировались при помощи бинарной хэш-функции и сравнивались при помощи модифицированного расстояния Хэмминга.

Третий способ заключается в использовании вариационного автокодировщика [3]. Вариационный автокодировщик - это нейросетевая модель, состоящая из двух основных частей: кодировщика и декодировщика. Кодировщик учится компактно описывать входное изображение, а декодировщик - по этому описанию восстанавливать исходное изображение.

На выходе кодировщика обучаются два вектора: μ_x и σ_x . По ним с использованием репараметризации генерируется вектор $z = \mu_x + \exp(\frac{\sigma_x}{2}) \cdot \varepsilon$, где $\varepsilon \sim \mathcal{N}(0, 1)$. Таким образом, получается, что $z \sim \mathcal{N}(\mu_x, \sigma_x)$. Этот вектор z передается на вход декодировщику и может использоваться в методе поиска по содержанию в качестве вектора признаков. Данный кодировщик выполняет сжатие изображений более чем в 1000 раз.

Разработанный метод ориентирован на работу с многомасштабными полнослайдовыми изображениями и для ускорения поиска используется дополнительная нейросетевая модель, позволяющая автоматически определить масштаб

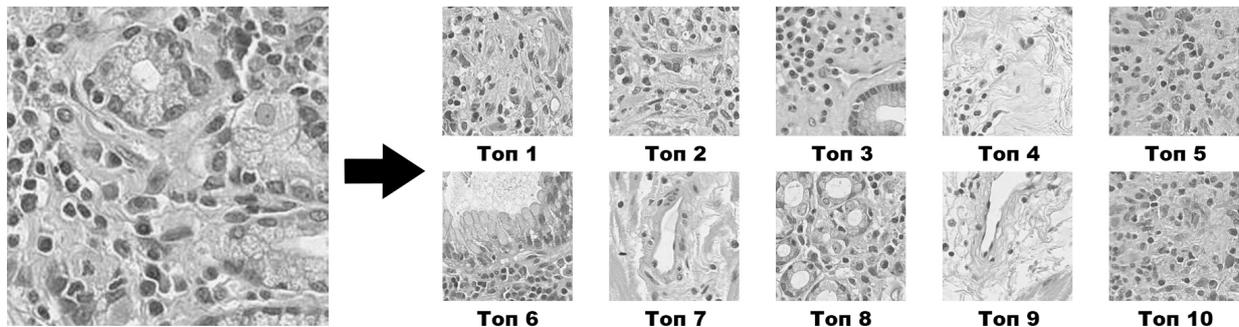
изображения-запроса, что существенно сужает диапазон поиска похожих фрагментов.

Проверка качества работы метода с использованием каждого из способов извлечения признаков проводилась на полнослайдовых гистологических изображениях набора данных PATH-DT-MSU. Для численной оценки качества использовалась мера ранжирования Mean Absolute Precision (*MAP*) для лучших 10 значений:

$$MAP@10 = \frac{\sum_{n=1}^{10} \frac{1}{n} * [class_{target} = class_n]}{\sum_{n=1}^k \frac{1}{n}},$$

где $class_{target}$ - класс изображения, поданного в метод, $class_n$ - класс изображения, находящегося на n -том месте в выдаче метода.

Метод с использованием вариационного автокодировщика в качестве экстрактора признаков оказался эффективнее других вариантов метода и достиг среднего значения $MAP@10$, равного 0.293.



Пример работы метода. Слева - изображение-запрос, справа - топ 10 наиболее похожих на него фрагментов по версии метода.

Метод поиска по содержимому для полнослайдовых гистологических изображений, описанный в этой работе, реализован на языке Python 3.8 в виде отдельного модуля. Весь исходный код доступен по ссылке: <https://github.com/ArtemVeshkin/Content-Based-Image-Retrieval>.

Литература

- [1] Tissue Type Recognition in Whole Slide Histological Images / Khvostikov A., Krylov A., Mikhailov I., Malkov P., and Danilova N. - 2021.
- [2] Development of CNN-based algorithm for automatic recognition of the layers of the wall of the stomach and colon / Mikhailov I., Khvostikov A., Krylov A., Malkov P., Danilova N., and Oleynikova N. // Virchows archiv / Springer one New York plaza, Suite 4600, New York, NY, United States. — 2021. — Vol. 479. — P. S36–S37.
- [3] Kingma D. P., Welling M. Auto-encoding variational bayes // arXiv preprint arXiv:1312.6114. — 2013.

Обнаружение аномалий на фотографиях поля сельскохозяйственных культур без предварительной разметки

Дунаева Анастасия Станиславовна

Кафедра математической физики

e-mail: anastasya.dunaeva@gmail.com

Научный руководитель — к.ф.-м.н. доц. Березин Сергей Борисович

В выпускной квалификационной работе решается задача автоматического обнаружения аномальных областей на фотографиях полей сельскохозяйственной культуры. Причиной появления аномалий может быть влияние болезни или нарушение технологии посадки растений. Своевременное обнаружение аномалий позволяет существенно уменьшить потери урожая и увеличить объем производимой продукции в условиях больших площадей теплиц.

В первой части работы реализован и успешно апробирован на серии изображений из одного из тепличных хозяйств описанный в [1] метод поиска аномальных областей на изображении. Были подобраны гиперпараметры задачи, обеспечивающие достаточно качественное обнаружение аномалий.

Существующий метод требует в качестве входного параметра заранее отобранное из набора изображение, не содержащее аномальных областей, что делает невозможной полную автоматизацию работы алгоритма. Поэтому во второй части выпускной квалификационной работы алгоритм модифицирован мной так, чтобы исключить необходимость выбора эталонного изображения человеком: алгоритму на вход поступает не одно изображение без аномалий, предварительно выбранное человеком из набора данных, а несколько случайных изображений из серии.

Для алгоритмов с предварительной разметкой и без предварительной разметки были получены следующие метрики качества обнаружения: true positive rate (recall), false positive rate, precision, F1. Приведено сравнение алгоритмов по данным метрикам.

Также в алгоритмическую реализацию метода внесены изменения, позволяющие работать с изображениями большого размера и с изображениями, содержащими разные фазы роста сельскохозяйственной культуры.

Полученный алгоритм был апробирован на реальных данных.

Литература

- [1] How to reduce anomaly detection in images to anomaly detection in noise / Thibaud Ehret, Axel Davy, Mauricio Delbracio, Jean-Michel Morel // Image Processing On Line, 9 (2019), pp. 391–412

Коэффициентная обратная задача для одномерного уравнения теплопроводности

Ефремов Всеволод Вячеславович

Кафедра математической физики

e-mail: efremtsev.vsevolod@yandex.ru

Научный руководитель — д.ф.-м.н. проф. Тихонов Иван Владимирович

Научный консультант — д.ф.-м.н. проф. Костин Андрей Борисович

В процессе выполнения ВКР проведено исследование коэффициентной обратной задачи для одномерного уравнения теплопроводности. Общая теория решения подобных задач была разработана в статье [1]. На этой основе составлен численный алгоритм и подготовлена программа на языке Python.

Изучаемая обратная задача имеет вид

$$\begin{cases} u_t = a^2 u_{xx} - c(x)u, & 0 < x < l, \quad 0 < t \leq T, \\ u(0, t) = \mu_0(t), \quad u(l, t) = \mu_1(t), & 0 < t \leq T, \\ u(x, 0) = 0, \quad u(x, T) = \psi(x), & 0 < x < l. \end{cases} \quad (1)$$

Значения параметров a , l , T и функции $\mu_0(t)$, $\mu_1(t)$, $\psi(x)$ предполагаются известными. Требуется найти функции $u(x, t)$, $c(x)$. Помимо финального переопределения $u(x, T) = \psi(x)$ использовалось также переопределение интегрального типа.

Задача (1) рассматривается в рамках

- ограничений $c(x) \geq 0$, $\psi(x) \geq \varepsilon > 0$, $\psi''(x) \geq 0$,
- условий монотонности $\mu_0'(t) \geq 0$, $\mu_1'(t) \geq 0$,
- условий согласования $\mu_0(0) = 0$, $\mu_1(0) = 0$, $\mu_0(T) = \psi(0)$, $\mu_1(T) = \psi(l)$.

Коротко опишем идею алгоритма. Пусть $\tilde{c}(x) = c(x) \cdot \psi(x)$. Введем вспомогательную прямую задачу

$$\begin{cases} u_t = a^2 u_{xx} - \tilde{c}(x)u/\psi(x), & 0 < x < l, \quad 0 < t \leq T, \\ u(0, t) = \mu_0(t), \quad u(l, t) = \mu_1(t), & 0 < t \leq T, \\ u(x, 0) = 0, & 0 < x < l. \end{cases} \quad (2)$$

Определим нелинейный оператор A следующим образом:

$$A\tilde{c} = a^2 \psi_{xx}(x) - u_t(x, T; \tilde{c}),$$

где $u(x, t; \tilde{c})$ — решение вспомогательной задачи (2). Обратная задача (1) эквивалентна операторному уравнению

$$A\tilde{c} = \tilde{c}. \quad (3)$$

Уравнение (3) решается методом итераций. Сходимость метода гарантирована теоремой Тарского об изотонном операторе (см. [1]–[3]). Использовались также стандартные соображения математической физики (см. [4]) и теория неявных разностных схем для параболических уравнений (см. [5]).

Подготовленная компьютерная программа позволяет задавать произвольные параметры задачи, находить решение, проверять его точность и строить графики, визуализирующие сходимость численного алгоритма. На этой основе проведено много вычислительных экспериментов, подтвердивших высокую эффективность метода из работы [1]. Укажем типичные примеры.

Пример 1.

$$\begin{cases} u_t = 0.09u_{xx} - c(x) \cdot u, & 0 < x < 1, 0 < t \leq 2, \\ u(0, t) = t, u(1, t) = 2t, \\ u(x, 0) = 0, u(x, T) = \psi(x). \end{cases}$$

Здесь $\psi(x) = u_{num}(x, T)$ — численное решение соответствующей прямой задачи при непрерывном $c_{exact}(x) = x + 1 - |2x - 1|$. Алгоритм сходится за 18 итераций. Решение восстанавливается с точностью порядка 10^{-7} .

Пример 2. Рассматривается ситуация, аналогичная примеру 1, но при разрывном коэффициенте

$$c_{exact}(x) = \begin{cases} 1, & 0 < x \leq 0.5, \\ 2, & 0.5 < x < 1. \end{cases}$$

Получаем сходимость за 15 итераций с точностью порядка 10^{-6} . Эксперимент подтверждает применимость метода работы [1] для коэффициентов $c \in L^\infty$.

Пример 3. По схеме примера 1 рассматривались также возмущенные ситуации типа $\psi(x) = u_{num}(x, T) - 0.3 \sin(3\pi x)$ с нарушением условия выпуклости для $\psi(x)$. Сходимость алгоритма при этом замедляется. В данном примере ответ получен за 45 итераций с точностью порядка 10^{-6} .

Литература

- [1] Прилепко А. И., Костин А. Б. Об обратных задачах определения коэффициента в параболическом уравнении // Сиб. матем. журнал. 1992. Т. 33, № 3. С. 146–155.
- [2] Tarski A. A lattice-theoretical fixpoint theorem and its applications // Pacific Journal of Mathematics. 1955. Vol. 5, N 2, P. 285-309.
- [3] Люстерник Л. А., Соболев В. И. Краткий курс функционального анализа: Учеб. пособие. М. : Высшая школа, 1982. 271 с.
- [4] Тихонов А. Н., Самарский А. А. Уравнения математической физики. (5-е изд.) М. : Наука, 1977. 735 с.
- [5] Самарский А. А., Гулин А. В. Численные методы. М. : Наука, 1989. 432 с.

Мнокадровый метод повышения разрешения цветных изображений

Ипполитов Владимир Борисович

Кафедра математической физики

e-mail: vovipp2000@gmail.com

Научный руководитель — к.ф.-м.н. с.н.с. Насонов Андрей Владимирович

Съемка на портативную камеру имеет следующую особенность. Необходимо сделать выбор: снимать сцену на короткой выдержке или на длинной. При короткой на изображении появится значительный шум из-за нехватки света, при длинной же выдержке шум будет минимизирован, но в силу неизбежного тремора рук изображение будет размыто. Один из способов решения этой проблемы — это использование серии кадров для получения одного изображения, лишенного вышеперечисленных недостатков. Кроме того, сырые изображения, получаемые с сенсоров камеры, представляют собой цветные мозаики, в которых в каждой точке определен только один цвет: синий, красный или зеленый.

Целью работы является разработка и реализация многокадрового метода повышения разрешения цветных изображений, который принимал бы на вход серию сырых кадров-мозаик, снятых на малой выдержке и с малой задержкой, и формировал бы одно цветное изображение более высокого разрешения.

Для решения поставленной задачи предлагается использовать схему, в которой каждое изображение выравнивается относительно первого, т.е. вычисляется оптический поток — поле смещений, вычисленное в каждой точке, а затем с использованием оптического потока серия кадров объединяется в одно изображение. Оптический поток необходимо оценивать с точностью до субпиксельных смещений, поэтому был выбран широко применимый метод Lucas-Kanade [1], который использует разложение функции в ряд Тейлора и метод наименьших квадратов и который обеспечивает нужную точность.

После вычисления оптического потока необходимо объединить кадры в одно изображение большего разрешения. Для этого все кадры накладываются на общую сетку с использованием оптического потока, и в каждой точке регулярной сетки происходит усреднение всех значений по окрестности с некоторой весовой функцией. Другими словами, используется следующая формула:

$$I(p, c) = \frac{\sum_n \sum_i I_{n,i}(p, c) w_{n,i}}{\sum_n \sum_i w_{n,i}}, \quad (1)$$

где p — пространственная переменная, c — цветовой канал, $I(p, c)$ — итоговое значение в точке x и в канале c , I_n — n -ый кадр в серии, $I_{n,i}(p, c)$ — i -ый пиксель в окрестности точки p в канале c на изображении I_n , $w_{n,i}$ — вес ядра, соответствующий i -ой позиции в окрестности точки на n -ом кадре. Веса $w_{n,i}$ образуют ядра слияния, к которым выдвигаются следующие требования: в результате слияния должны быть сохранены четкие контуры и маленькие детали,

а в областях без контуров и деталей должно производиться шумоподавление. Для этого ядра будут «вытягиваться» вдоль контуров и масштабироваться в зависимости от того, насколько детализированным в окрестности точки является изображение.

Информацию о детализированности изображения и направленности структуры в окрестности точки можно получить, анализируя собственные векторы и собственные значения структурного тензора:

$$S_v(p) = \begin{bmatrix} \int v(r) (I_x(p-r))^2 dr & \int v(r) I_x(p-r) I_y(p-r) dr \\ \int v(r) I_x(p-r) I_y(p-r) dr & \int v(r) (I_y(p-r))^2 dr \end{bmatrix}, \quad (2)$$

где $v(r)$ — весовая функция, I_x и I_y — частные производные изображения по соответствующим координатам, а $S_v(p)$ — структурный тензор изображения I в точке p (в этой формуле изображение приводится к одноканальному, то есть к изображению в оттенках серого). Анализируя спектр структурного тензора, ядра слияния определим, как двумерные гауссовы функции:

$$w_{n,i} = \exp\left(-\frac{1}{2} \mathbf{r}_i \Omega^{-1} \mathbf{r}_i^T\right), \mathbf{r}_i = [x_i - x_0, y_i - y_0], \quad (3)$$

где x_i, y_i — координаты точки, в которой считается вес ядра, x_0, y_0 — координаты точки, для которой конструируется ядро, Ω^{-1} — матрица ковариации гауссовой функции. Матрица ковариации есть функция от собственных векторов и собственных значений структурного тензора. Использование нами структурного тензора в данном контексте вдохновлено работой [2].

Качество реализованного алгоритма оценивается как на синтетических данных, так и на реальных. Для сравнения нашего метода с другими используются метрики PSNR и SSIM, по которым наш метод на синтетических данных ожидаемо превосходит по качеству однокадровые методы. На реальных данных наш метод визуально также показывает себя лучше: производится шумоподавление, минимизируются цветные артефакты, получаемые в ходе демосаикинга, и сохраняется четкость контуров.

Литература

- [1] Lucas B.D., Kanade T. An iterative image registration technique with an application to stereo vision // IJCAI81. 1981. С. 674–679
- [2] Wronski B., Garcia-Dorado I. Handheld Multi-Frame Super-Resolution// ACM Transactions on Graphics (TOG). 2019. Vol. 38, № 2 2019. С. 1–18

Математическая модель рассеяния плоской электромагнитной волны на двух эллипсоидах на основе метода интегрального уравнения в спектральной области

Кабанов Денис Михайлович

Кафедра математической физики

e-mail: s02180016@gse.cs.msu.ru

Научный руководитель — к.ф.-м.н. с.н.с. Лопушенко Владимир Васильевич

Рассматривается задача дифракции, которая сводится к интегральному уравнению (ИУ)

$$\mathbf{E}(M_0) = \mathbf{E}^0(M_0) + \int_V (k^2 - k_0^2) \hat{G}(M, M_0) \mathbf{E}(M) dV_M \quad (1)$$

где k, k_0 — волновое число внутри и вне области V , вектор \mathbf{E}^0 — внешнее возбуждение, а \hat{G} — матрица фундаментальных решений. [1]

В работе предложен метод решения ИУ с переходом в спектральную область, в результате чего уравнение преобразуется к виду

$$\mathcal{F}[\mathbf{U}(M_0)] = \mathcal{A} \hat{\varepsilon}(p, q, s + k_0) + \frac{\hat{\varepsilon}(p, q, s)}{(2\pi)^3} * \mathcal{F} \left[\int_V \hat{G} \mathbf{U}(M) dV_M \right] \quad (2)$$

где функция $\hat{\varepsilon}$ зависит от полуосей a_n, b_n, c_n , смещений центров x_n, y_n, z_n , ориентации частиц, определяемой углами Эйлера, которые задаются в переменных p'_n, q'_n, s'_n . \mathcal{A} задает поляризацию волны. Получим функцию области и функцию внешнего возбуждения:

$$\mathcal{F}[\Delta k] = \hat{\varepsilon}(p, q, s) = \sum_{n=1}^2 4\pi \frac{a_n b_n c_n}{\nu_n'^2} (k_n^2 - k_0^2) \left(\frac{\sin(\nu_n')}{\nu_n'} - \cos(\nu_n') \right) e^{-i[px_n + qy_n + sz_n]} \quad (3)$$

$$\mathcal{F}[\mathbf{U}^0] = \hat{\varepsilon}(p, q, s + k_0) \quad \nu_n' = (p'_n a_n)^2 + (q'_n b_n)^2 + (s'_n c_n)^2 \quad (4)$$

Для решения (2) воспользуемся теоремой отсчётов и, учитывая убывающий характер на бесконечности всех функций, представим функции конечными рядами Котельникова и получим СЛАУ:

$$\mathcal{F}[\mathbf{U}]_{mnl} = \mathcal{A} \hat{\varepsilon}_{mnl}(p, q, s + k_0) + \frac{h_p h_q h_s}{(2\pi)^3} \sum_{i=-N_p}^{N_p} \sum_{j=-N_q}^{N_q} \sum_{k=-N_s}^{N_s} \hat{\varepsilon}_{i-m, j-n, k-l}(p, q, s) (\mathcal{F}[\tilde{G}]_{mnl} \mathcal{F}[\mathbf{U}]_{mnl}) \quad (5)$$

Можно показать, что диаграмма рассеяния $I(\theta, \phi)$ при $R \rightarrow \infty$ выводится непосредственно из решения полученной СЛАУ (5):

$$I(\theta, \phi) = G(\theta, \phi) \mathcal{F}[\mathbf{U}(M)](k_0 \sin(\theta) \cos(\phi), k_0 \sin(\theta) \sin(\phi), k_0 \cos(\theta)) \quad (6)$$

где $G(\theta, \phi)$ — матрица, которая получается из матрицы фундаментальных решений.

В работе исследовались конфигурации эллипсоидов разных размеров и ориентаций, на произвольных расстояниях друг от друга. Во всех рассмотренных случаях исследована сходимость решения с помощью изменения размеров и шагов сетки $N_p, N_q, N_s; h_p, h_q, h_s$.

Заключение. В работе предложен метод решения ИУ в спектральной области с представлением решения рядом Котельникова. В численной схеме удается избежать сингулярности под знаком интеграла. Показано, что для решения задачи дифракции кластером частиц достаточно построить соответствующую функцию области, при этом основная часть численной схемы не изменяется. Решение системы линейных алгебраических уравнений выполнено при помощи высокопроизводительных методов вычисления дискретных сверток итерационным методом.

Литература

- [1] В. И. Дмитриев, У. В. Захаров. Интегральные уравнения в краевых задачах электродинамики. // М.: Изд-во Моск. ун-та, 1987. — 167 с.

Исследование метода восстановления волнового фронта по его локальным наклонам с локализованным стабилизатором дробной гладкости

Колотов Антон Евгеньевич

Кафедра математической физики

e-mail: kolotov.anton07@gmail.com

Научный руководитель — д.ф.-м.н. проф. Разгулин Александр Витальевич

В настоящее время в задачах адаптивной оптики широко применяется датчик Шака-Гартмана [1], позволяющий получить измерения наклонов проходящего через него волнового фронта. Ставится задача — реконструировать измеряемый волновой фронт по данным локальным наклонам.

В данной работе рассматривается семейство методов восстановления волнового фронта, основанных на вариационном подходе [2].

Задачу восстановления волнового фронта $u(x, y)$ по его локальным наклонам $g_1(x, y), g_2(x, y)$ вдоль направлений x и y соответственно, измеренных в квадратной области Ω , формулируется в виде задачи минимизации функционала $J(u)$:

$$J(u) = \iint_{\Omega} \left((u_x - g_1)^2 + (u_y - g_2)^2 \right) dx dy$$

Для построения конечномерной аппроксимации используется базис из кусочно-линейных по каждому направлению сплайнов. Дополнительно в схему вводится стабилизатор дробной гладкости, рассмотренный в работе [3]. Отметим, что работа стабилизатора направлена на коррекцию Фурье-спектра восстанавливаемого волнового фронта. В конечном итоге, задача сводится к решению уравнения для матрицы коэффициентов $u^h = \{u_{kl}^h\}$:

$$B_2 \Lambda_1 u^h + B_1 \Lambda_2 u^h + \bar{\gamma} (\Lambda_1 \Lambda_2)^{\bar{s}} u^h = B_2 G_1 g_1 + B_1 G_2 g_2$$

Здесь, операторы B_1, B_2 задаются с помощью матрицы Грама рассматриваемых сплайнов; Λ_1, Λ_2 — с помощью матрицы Грама градиентов рассматриваемых сплайнов; $\bar{\gamma}, \bar{s}$ — параметры стабилизатора. Решение данного уравнения эффективно осуществляется методом быстрого дискретного преобразования Фурье.

В данной работе был предложен способ локализованного выбора параметров стабилизатора $\bar{\gamma}, \bar{s}$ из условий выравнивания пространственно-частотной характеристики (ПЧХ) метода. Параметры представляются в виде набора $\bar{\gamma} = \{\gamma_{kl}\}, \bar{s} = \{s_{kl}\}$, где каждая пара (γ_{kl}, s_{kl}) направлена на коррекцию восстановления соответствующей гармоники Фурье $\exp\{i(kx + ly)\}$. Минимизируется функционал среднеквадратичного отклонения ПЧХ от идеальной.

Такой подход позволил добиться существенного улучшения ПЧХ метода и, как следствие, лучшего качества реконструкции.

Отдельно в работе рассмотрено применение метода для случая, когда часть измерений наклонов «потеряна». Для этого перед применением непосредственно метода восстановления проводится интерполяция измерений наклонов. В работе предложен способ интерполяции с помощью сверточной нейронной сети архитектуры U-net [4].

Сеть получает на вход пару матриц, содержащих неполные измерения наклонов, интерполирует значения в потерянных точках и возвращает восстановленные матрицы измерений наклонов.

Архитектура сети состоит из четырех сверточных слоев кодирования, содержащих свертки с ядром 3×3 , после которых применяется понижающая пространственную размерность агрегация, и четырех сверточных слоев декодирования, содержащих свертки с ядром 3×3 , перед которыми применяется повышающая пространственную размерность транспонированная свертка. В конце для непосредственно интерполяции применяются свертки с ядром 1×1 . Для стабильности и ускорения сходимости сети используется механизм нормализации. Также, между кодирующими и декодирующими слоями одинаковой размерности существуют связи для сохранения пространственной информации. В качестве оптимизатора используется оптимизатор Адама.

Данная архитектура позволила обеспечить стабильное качество интерполяции наклонов с ростом доли потерянных измерений и характерных размеров «выбитых» областей.

Метод восстановления был реализован на языке программирования Python. Нейронная сеть была построена и обучена при помощи фреймворков глубокого обучения TensorFlow и Keras. Тестирование метода на датасетах реальных волновых фронтов, возникающих в офтальмологических приложениях, показало среднеквадратичное отклонение 5% от исходного волнового фронта в среднем. Также, были проведены тесты восстановления турбулентных волновых фронтов и показана устойчивость метода к малому шуму.

Литература

- [1] Platt B. C., Shack R. History and principles of Shack-Hartmann wavefront sensing // J.Refract. Surg. 2001. Vol. 17. P. 573—577.

- [2] A variational method for wavefront reconstruction from Shack-Hartman sensor measurements / A. V. Razgulin, Y. Z. Kuzhamaliyev, A. S. Goncharov, A. V. Larichev // Atmospheric and Oceanic Optics. 2017. Vol. 30. N 4. P. 399–403.
- [3] A variational method of wavefront reconstruction from local slope measurements using a fractional order of smoothness stabilizer / A. V. Razgulin, Y. Z. Kuzhamaliyev, N. G. Iroshnikov, A. V. Larichev // Computational Mathematics and Modeling. 2019. Vol. 30. N 2. P. 164–176.
- [4] Justin G. Chen J. G., Shah V., Liu L. Performance of a U-Net-based neural network for predictive adaptive optics. // Opt. Lett. 2021. Vol. 46, N 10. P. 2513–2516

Математическое моделирование распространения лазерного импульса в нелинейной среде, описываемой уравнением Гинзбурга—Ландау с однородными граничными условиями 1-го рода

Тихвинский Даниил Геннадьевич

Кафедра математической физики

e-mail: d.tikhvinskii@gmail.com

Научный руководитель — к.ф.-м.н. доц. Федотов Михаил Валентинович

Уравнение Гинзбурга—Ландау часто используется для описания процессов распространения лазерного импульса в оптическом волокне, содержащем усилитель [1, 2]. Несмотря на широкое применение компьютерного моделирования при построении решения подобных задач, его интегралы движения, так называемые инварианты, до настоящего времени отсутствовали даже для линейного случая.

В данной выпускной квалификационной работе были выведены инварианты для трех постановок одномерной задачи: линейной с поглощением, нелинейной без поглощения и нелинейной с поглощением. Рассматривалось следующее нелинейное дифференциальное уравнение в частных производных [3, 4], где функция $A(z, t)$ соответствует медленно изменяющейся комплексной амплитуде оптического излучения:

$$\begin{aligned} \frac{\partial A(z, t)}{\partial z} + (v_r + iv_i) \frac{\partial A(z, t)}{\partial t} + (D_{2\omega} + iD_2) \frac{\partial^2 A(z, t)}{\partial t^2} + \\ + (\delta + i\alpha) |A(z, t)|^2 A(z, t) + \left(\frac{1}{2} \delta_0 + in_0 \right) A(z, t) = 0, \end{aligned} \quad (1)$$

где $0 < z \leq L_z$, $0 < t < L_t$.

Уравнение (1) сопровождалось одним начальным и двумя краевыми условиями, непосредственно вытекающими из физической постановки:

$$A(z, t)|_{z=0} = e^{-(t-\frac{L_t}{2})^2}, \quad (2)$$

$$A(z, t)|_{t=0} = 0, \quad A(z, t)|_{t=L_t} = 0, \quad (3)$$

где $0 \leq t \leq L_t$, $0 \leq z \leq L_z$.

Для упомянутых ранее ключевых случаев поставленной задачи (1)–(3) были сформулированы и доказаны теоремы о соответствующих инвариантах.

Теорема 1. *Задача (1)–(3) в линейном случае с поглощением ($\delta = 0$ и $\alpha = 0$) обладает инвариантом:*

$$P(z, L_t) = e^{-\left(\frac{1}{2}\delta_0 + in_0\right)z} P(0, L_t),$$

где функция $P(z, t)$ определена следующим образом:

$$P(z, t) = \int_0^t A(z, \xi) d\xi.$$

Теорема 2. *Задача (1)–(3) в нелинейном случае без поглощения ($\delta \neq 0$ или $\alpha \neq 0$, $\delta_0 = 0$ и $n_0 = 0$) обладает инвариантом:*

$$P(z, L_t) = P(0, L_t) - (\delta + i\alpha) \int_0^z \left(\int_0^{L_t} |A(\xi, t)|^2 A(\xi, t) dt \right) d\xi.$$

Теорема 3. *Задача (1)–(3) в нелинейном случае с поглощением ($\delta \neq 0$ или $\alpha \neq 0$, $\delta_0 \neq 0$ или $n_0 \neq 0$) обладает инвариантом:*

$$P(z, L_t) = e^{-\left(\frac{1}{2}\delta_0 + in_0\right)z} \left[P(0, L_t) - (\delta + i\alpha) \int_0^z e^{-\left(\frac{1}{2}\delta_0 + in_0\right)\xi} \left(\int_0^{L_t} |A(\xi, t)|^2 A(\xi, t) dt \right) d\xi \right].$$

Также было написано программное обеспечение на языке С, которое позволяет подсчитывать и строить численное решение уравнения (1) и некоторые другие физические характеристики, необходимые для анализа корректности полученных результатов: преобразование Фурье, первые моменты во временном и частотном пространствах и т. д.

Для нахождения численного решения использовалась конечно-разностная схема Кранка–Николсона. Чтобы избавиться от нелинейности в уравнениях, реализовывался итерационный процесс, в котором часть нелинейного слагаемого берется с предыдущей итерации. Полученная система линейных уравнений сводилась к трехдиагональному виду и решалась с помощью метода прогонки. Для построения инвариантов и физических характеристик были сформулированы их разностные аналоги, для аппроксимации интегралов использовалась формула Симпсона.

С помощью той же программы на примере линейной задачи с поглощением детально изучалось поведение интегралов движения. Значения инвариантов на разностной сетке в каждой точке совпали с аналитическими результатами. Фактически это означает, что построенная схема обладает свойством консервативности.

Литература

- [1] Agrawal G. P. Optical pulse propagation in doped fiber amplifiers // Phys. Rev. A 1991, 44, 7493–7501.
- [2] Agrawal G. P. Applications of Nonlinear Fiber Optics // Academic Press: Rochester, NY, USA, 2008.
- [3] Ginzburg V. L. On the theory of superconductivity. In On Superconductivity and Superfluidity // Springer: Berlin/Heidelberg, Germany, 2009.
- [4] Ginzburg V. L. Superconductivity and superfluidity (what is done and what is not done) // Phys.-Uspekhi 1997, 40, 407–432.

Итерационные методы на основе составных пространств Крылова

Валиахметов Булат Ильдарович

Кафедра вычислительных технологий и моделирования

e-mail: valiahmetovbulat@mail.ru

Научный руководитель — акад. РАН, д.ф.-м.н. проф. Тыртышников Евгений Евгеньевич

Научный консультант — м.н.с. ИВМ РАН Желтков Дмитрий Александрович

Решение системы линейных уравнений — классическая задача вычислительной линейной алгебры:

$$Ax = b.$$

Для решения этой задачи разработано множество методов, многие из которых являются итерационными. Существует семейство методов, основанных на пространствах Крылова: $\mathcal{K}_k(b, A) = \mathcal{L}(b, Ab, \dots, A^{k-1}b)$, для решения линейных систем и некоторых других задач.

В неэрмитовом случае естественная идея минимизация нормы невязки $r_k = b - Ax_k$ на k -м подпространстве Крылова приводит к двум подходам. Первый из них — метод сопряженных невязок (GCR) [1] на k -й итерации требует построить и хранить два базиса: ортонормированный базис $V_k \subset A\mathcal{K}_k$ и базис $P_k \subset \mathcal{K}_k$, где $AP_k = V_k$. Одним из его достоинств является естественное обобщение на случай линейных систем с многими правыми частями [3]:

$$AX = B. \quad (1)$$

Причем правые части могут поступать последовательно, в отличие от блочных методов, где они должны быть известны заранее. Решение ищется в *составном пространстве Крылова*:

$$\tilde{\mathcal{K}} = \text{span}(r_0^{(1)}, Ar_0^{(1)}, \dots, A^{n_1-1}r_0^{(1)}, r_{n_1}^{(2)}, Ar_{n_1}^{(2)}, \dots, A^{n_2-1}r_{n_1}^{(2)}, \dots). \quad (2)$$

Второй подход приводит к обобщенному методу минимальных невязок (GMRES) [2], который в точной арифметике эквивалентен GCR. Он требует хранения только *одного* ортонормированного базиса Q_{k+1} пространства \mathcal{K}_{k+1} . Для пересчёта базисов образа и прообраза используется соотношение Арнольди:

$$AQ_k = Q_{k+1}H_{k+1,k} \quad (3)$$

с хессенберговой матрицей $H_{k+1,k}$. В отличие от GCR, GMRES не имеет такого естественного расширения для задачи (1).

В данной работе предлагается метод решения систем с многими правыми частями без использования блочности, основанный на GMRES. Он минимизирует невязки на пространстве (2) и в точной арифметике эквивалентен GCR [3], но требует хранения только *одного* базиса. Его основная идея заключается в определённом выборе базиса P_k пространства поиска решений. На итерации $k + 1$ при правой части m предлагается:

1. Если правая часть сменилась или $k = 0$, добавить в базис Q_k невязку $r_k^{(m)}$, соответствующую проекции $r_0^{(m)}$ на V_k , затем выбрать $p_{k+1} = \frac{r_k^{(m)}}{\|r_k^{(m)}\|}$.
2. Если правая часть не менялась, выбрать

$$p_{k+1} = \frac{r_k^{(m)} - r_{k-1}^{(m)}}{\|r_k^{(m)} - r_{k-1}^{(m)}\|} e^{i\varphi_{k+1}}.$$

Базис P_k векторов направлений является почти ортогональным, и его можно выразить через базис Q_{k+m-1} и получить следующее представление:

$$AQ_{k+m-1}C_{k+m-1,k} = Q_{k+m}H_{k+m,k},$$

аналогичное (3), где m — номер текущей правой части. Матрица $C_{k+m-1,k}$ имеет определённую структуру, позволяющую хранить лишь малое число её столбцов и быстро умножать её на вектор. Матрица $H_{k+m,k}$ является блочно-верхнетреугольной.

Выполнена реализация построенного алгоритма на языке C++ с использованием техники шаблонов, а также MPI и OpenMP параллелизма. Численные эксперименты в сравнении с реализацией GCR [4] показали высокую эффективность предложенного метода. Затрачиваемая память и время расчёта сократились вдвое.

Литература

- [1] Eisenstat S. C., Elman H. C., Schultz M. H. Variational iterative methods for nonsymmetric systems of linear equations // SIAM Journal on Numerical Analysis. – 1983. – Vol. 20. – N 2. – P. 345-357.
- [2] Saad Y., Schultz M. H. GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems // SIAM Journal on scientific and statistical computing. – 1986. – Vol. 7. – N 3. – P. 856-869.
- [3] Lingen F. J. A Generalised Conjugate Residual method for the solution of nonsymmetric systems of equations with multiple right-hand sides // International journal for numerical methods in engineering. – 1999. – Vol. 44. – N 5. – P. 641-656.
- [4] Aparinov A., Setukha A., Stavtsev S. Supercomputer modelling of electromagnetic wave scattering with boundary integral equation method // Russian Supercomputing Days. – Springer, Cham, 2017. – P. 325-336.

Оптимальное матричное и тензорное восполнение с дополнительной информацией

Михайлов Арсений Денисович

Кафедра вычислительных технологий и моделирования

e-mail: mikhailov.arseniy.d@gmail.com

Научный руководитель — к.ф.-м.н, с.н.с. ИВМ РАН Замарашкин Николай Леонидович

Пусть $X \in \mathbb{R}^{n_1 \times n_2}$ — неизвестная малоранговая матрица, и мы будем решать оптимизационную задачу:

$$\min_{\tilde{X}} \text{rank}(\tilde{X}), \mathcal{P}_\Omega(\tilde{X}) = \mathcal{P}_\Omega(X), \text{Im}(\tilde{X}) \subset L_A, \text{Im}(\tilde{X}^T) \subset L_B$$

где $\mathcal{P}_\Omega(X) = \mathbb{1}[(i, j) \in \Omega] \circ X$ — проектор на множество известных элементов, L_A и L_B — линейные пространства, содержащие строки и столбцы X соответственно, которые мы будем называть дополнительной информацией. Пусть $s_1 = \dim L_A$, $s_2 = \dim L_B$ и $s = \max s_i$, $n = \max n_i$.

Если решение этой задачи единственно, то мы восстановили всю матрицу X по известному набору элементов X_{ij} , $(i, j) \in \Omega \subset [n_1] \times [n_2]$ и дополнительной информации. В общем случае решение не единственно, но мы докажем его единственность и построим алгоритм поиска при некоторых предположениях на матрицу X , дополнительную информацию и множество Ω .

Задача восполнения матриц с дополнительной информацией исследовалась в [2]. В этой работе было показано, что решение (в постановке минимизации ядерной нормы) единственно (с вероятностью близкой к 1), при равномерном выборе Ω , так что $|\Omega| = O(rs \ln s \ln n)$. Эта оценка существенно лучше, чем оценки для случая без дополнительной информации, но она не является оптимальной по n .

В этой работе был построен алгоритм SVPWS, для решения задачи матричного восполнения с дополнительной информацией, имеющий полностью доказанную теорию сходимости при выборе $|\Omega| = O(rs^3 \ln s)$, а также теорию сходимости аналогичную SVP ([1]), если $|\Omega| = O(r^2 s \ln s)$. Сложность одной итерации алгоритма: $O(|\Omega| s^2)$. Все характеристики алгоритма не зависят от n , а только от качества дополнительной информации.

В тензорном случае мы будем использовать малоранговость в смысле рангов Таккера, и будем решать аналогичную оптимизационную задачу для неизвестного тензора $X \in \mathbb{R}^{n_1 \times \dots \times n_d}$:

$$\min_{\tilde{X}} \text{trank}(\tilde{X}), \mathcal{P}_\Omega(\tilde{X}) = \mathcal{P}_\Omega(X), X = G \times_1 S_1 \dots \times_d S_d$$

где $\mathcal{P}_\Omega(X) = \mathbb{1}[(i_1, \dots, i_d) \in \Omega] \circ X$ — проектор на множество известных элементов, $\text{Im}(S_i)$ — линейные пространства, содержащие столбцы по i -той размерности X , которые мы будем называть дополнительной информацией, $\text{trank}(X) = \max r_i = r$, где r_i ранги Таккера тензора X .

Алгоритм SVPWS был обобщён для тензорного случая. Было доказано свойство частичной изометрии для тензорного оператора с числом достаточных

элементов $|\Omega| = O((dsr + r^d) \ln s)$. Сложность одной итерации алгоритма также не зависит от размеров исходного тензора.

Литература

- [1] Raghu Meka, Prateek Jain, Inderjit S. Dhillon Guaranteed Rank Minimization via Singular Value Projection 2009
- [2] Miao Xu, Rong Jin, Zhi-Hua Zhou Speedup Matrix Completion with Side Information: Application to Multi-Label Learning 2013

Алгоритм дополнения тензора с фиксированными рангами Таккера с использованием оптимизации на римановых многообразиях

Мордвинцев Михаил Константинович

Кафедра вычислительных технологий и моделирования

e-mail: mordvincevmisha@mail.ru

Научный руководитель — к.ф.-м.н. доц. Замарашкин Николай Леонидович

Задача дополнения тензора — по некоторым известным элементам восстановить тензор, если известно, что ранги Таккера тензора не выше r_1, r_2, \dots, r_d .

Эта задача сводится к минимизации нормы ошибки на известных элементах: вводится оператор дополнения, не меняющий известные элементы, и преобразующий в 0 все остальные. Для такой нормы записывается градиент, выполняется градиентный шаг и происходит проектирование полученного результата на множество тензоров с заданным рангом (проектируется тензор с полным рангом в общем случае). В новом алгоритме градиент сначала проектируется на касательное пространство к многообразию тензоров фиксированного ранга. Это позволяет выполнять операцию проектирования на множество тензоров фиксированного ранга асимптотически эффективно.

Первый алгоритм является аналогией матричного случая [5]. Кроме высокой вычислительной сложности этот подход добавляет новые проблемы, которых не было в матричном случае: из-за того, что проектор на множество тензоров с фиксированным рангом неточный, не удастся доказать сходимость такого алгоритма [3].

Вышеперечисленные проблемы призван решить дополнительный проектор на касательное пространство к многообразию тензоров фиксированного ранга. Теория римановых многообразий дает некоторые оценки на сходимость алгоритма, использующего этот проектор [2]. Так же можно доказать, используя свойства оператора дополнения тензора, что алгоритм с проектором на касательные многообразия сходится [3] при достаточном числе известных элементов. Использование этого проектора позволяет избежать применения дорогостоящей операции HOSVD — проектирования на пространство тензоров с заданными рангами с помощью SVD разложения разверток матрицы — к тензору с исходными размерами.

Использование этого проектора уменьшает сложность алгоритма и помогает доказать его сходимость. Также новый алгоритм обладает некоторыми недостатками. Исходный алгоритм требует меньше элементов для эффективного дополнения, чем алгоритм, в котором применяется проектор.

Тензорная структура позволяет уменьшить число известных элементов, необходимых для восстановления. Теорема, оценивающая их число, а так же сравнение ее с матричным случаем [5] приводится в работе.

Литература

- [1] Dynamical Tensor Approximation. Koch, Othmar and Lubich, Christian. SIAM Journal on Matrix Analysis and Applications. V31 N5 pages 2360-2375.
- [2] Low-Rank Tensor Completion by Riemannian Optimization. Daniel Kressner, Michael Steinlechner, Bart Vandereycken. Bit Numer Math 54, 447–468 (2014). <https://doi.org/10.1007/s10543-013-0455-z>
- [3] Provable Near-Optimal Low-Multilinear-Rank Tensor Recovery. Jian-Feng Cai, Lizhang Miao, Yang Wang, Yin Xian. arXiv:2007.08904
- [4] Матрицы, тензоры и вычисления. Е. Е. Тыртышников
- [5] Guaranteed Rank Minimization via Singular Value Projection. Raghu Meka, Prateek Jain, Inderjit S., Dhillon. arXiv:0909.5457

Математическое моделирование динамики новой коронавирусной инфекции

Рыгин Антон Сергеевич

Кафедра вычислительных технологий и моделирования

e-mail: a.rygin@yandex.ru

Научный руководитель — д.ф.-м.н. проф. Романюха Алексей Алексеевич

Научный консультант — к.ф.-м.н. Санникова Татьяна Евгеньевна

Основным инструментом математической эпидемиологии являются так называемые компартментные модели (или SIR-модели), в рамках которых популяция разбивается на группы по состоянию здоровья и другим признакам. Эти модели хорошо зарекомендовали себя в прошлом, однако во время пандемии COVID-19 выявились значительные расхождения между их предсказаниями и реальной динамикой заболеваемости [1]. Схожая проблема уже поднималась ранее в связи со вспышкой SARS-CoV в 2002-2004 гг. В [2] отмечается, что одной из главных причин является базовое предположение об однородности популяции в данных моделях. Решением могут послужить сетевые методы, в которых популяция представляется в виде графа: вершины соответствуют индивидам, рёбра – их контактам. Топология сети описывается численными характеристиками (средняя степень узла, коэффициент кластеризации и др.), которые могут быть использованы в SIR-модели для более точного описания процесса распространения инфекции. В данной работе проводится сравнение стандартной и сетевой SIR-модели на основе приближения реальных данных о заболеваемости COVID-19 в Москве и Нью-Йорке.

В соответствии с особенностями вируса SARS-CoV-2 и характером его распространения, популяция разбивается на следующие шесть компартментов: восприимчивые к заболеванию S , латентно инфицированные E , невыявленные/выявленные инфицированные I_u/I_d , выздоровевшие R , умершие D . Заболевание распространяется в результате контактов восприимчивых индивидов с невыявленными инфицированными; инфицированные выявляются в ходе ПЦР-тестирования; выздоровевшие имеют полный иммунитет. Математически данная модель описывается системой дифференциальных уравнений.

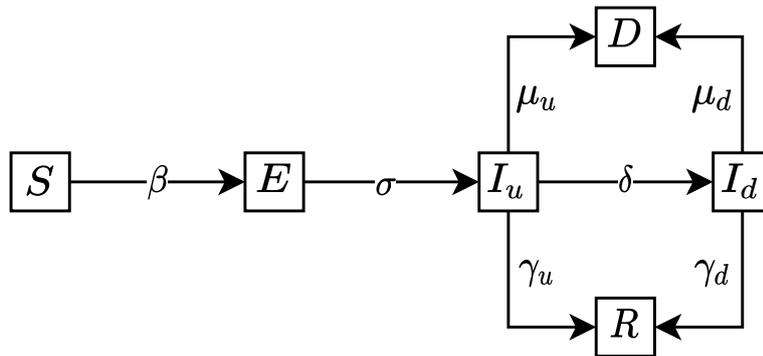


Рис. 1: Схема используемой SIR-модели.

В качестве альтернативы рассматривается предложенная в [3] сетевая SIR-модель, которая использует распределение степеней узлов $P(k)$ для описания сети контактов. Кроме того, все узлы-индивиды разбиваются на классы по их степени k (т.е. по числу контактов), и для каждого класса фактически решается отдельная SIR-система. Основное предположение о распространения инфекции состоит в том, что для восприимчивого узла степени \hat{k} вероятность контакта с невыявленным инфицированным пропорциональна $\hat{k}\Theta(t)$, где

$$\Theta(t) = \frac{\sum_k kP(k)I_{uk}(t)}{\sum_s sP(s)} = \frac{\sum_k kP(k)I_{uk}(t)}{\langle k \rangle}.$$

В статьях, посвящённых сетевым SIR-моделям, сеть контактов представляется преимущественно синтетическими графами, например, безмасштабной сетью Барабаши-Альберт. В данной работе были построены реалистичные модели сетей контактов Москвы и Нью-Йорка на основе таких демографических данных, как распределение населения по полу и возрасту, размеры и структура домохозяйств, среднее число контактов между различными возрастными группами.

Идентификация оптимальных значений параметров обеих моделей проводилась с помощью метода Левенберга-Марквардта. Калибровка происходила на данных по заболеваемости, смертности и выздоровлениям от COVID-19 в Москве и Нью-Йорке в ходе первой волны. Рассматриваемый период разделён на два промежутка – до и после введения основных ограничительных мер. В случае стандартной модели предполагается изменение скорости распространения инфекции β на втором промежутке, в случае сетевой удаляется 75% связей, формирующих сообщества в сети контактов.

В результате численных экспериментов было продемонстрировано, что сетевая компартментная модель превосходит стандартную по точности приближения реальных данных о заболеваемости и смертности. Кроме того, сравнение полученных значений параметров обеих моделей свидетельствует о том, что в Москве скорость распространения инфекции была ниже, чем в Нью-Йорке, скорость выявления инфицированных – выше, что может послужить объяснением различий в динамике заболеваемости в этих городах.

Литература

- [1] Inefficiency of SIR models in forecasting COVID-19 epidemic: a case study of Isfahan / S. Moein, N. Nickaeen, A. Roointan, N. Borhani, Z. Heidary, S. H. Javanmard, J. Ghaisari, Y. Gheisari // Scientific Reports. 2021. Vol. 11, N 1. P. 4725.
- [2] Network theory and SARS: predicting outbreak diversity / L.A. Meyers, B. Pourbohloul, M.E.J. Newman, D. M. Skowronski, R.C. Brunham // Journal of theoretical biology. 2005. Vol. 232, N 1. P. 71–81.
- [3] Y. Moreno, R. Pastor-Satorras, A. Vespignani. Epidemic outbreaks in complex heterogeneous networks // The European Physical Journal B - Condensed Matter and Complex Systems. 2002. Vol. 26, P. 521–529.

Численное исследование влияния земли на аэродинамические характеристики жёсткого винта вертолётa

Григорьева Мария Сергеевна

Кафедра вычислительных методов

e-mail: mashama9919@mail.ru

Научный руководитель — к.ф.-м.н. н.с. Бобков Владимир Георгиевич

Расчёт аэродинамических характеристик при обтекании винта вертолётa является актуальной задачей. Численные исследования обтекания винта необходимы и все шире применяются при разработке и проектировании вертолётov для определения характеристик летательного аппарата на этапе проектирования. Одним из сложных для анализа аэродинамики винта вертолётa является режим работы на предельно низких высотах из-за влияния земли. Целью данной работы являлось численное исследование влияния земли на аэродинамические характеристики жёсткого винта вертолётa при осевом обтекании.

Для описания течения вблизи вращающегося винта использовались модель на основе системы уравнений Навье-Стокса записанных в неинерциальной вращающейся системе координат [1]. Для пространственной дискретизации конвективных потоков использовалась EBR-схема построенная для неструктурированных гибридных расчетных сеток [2]. Интегрирование по времени осуществляется явными методами типа Рунге–Кутта. Данная численная методика реализована в исследовательском программном комплексе NOISEtte [3].

Исследовались аэродинамические характеристики четырёхлопастного жёсткого несущего винта вертолёта радиусом 1.3 метра при осевом обтекании. Все лопасти были идентичные прямоугольные в плане с хордой 0.13 м и углом установки 12° . В силу симметричности геометрии винта и осевого характера обтекания было возможным моделировать сектор с раствором 90° с одной лопастью с периодическим замыканием по азимутальным плоскостям. Для численного исследования была построена последовательность неструктурированных гибридных расчетных сеток с расчетной областью в виде сектора цилиндра с одной лопастью в которых в нижняя плоскость моделировала поверхность земли на разной высоте.

По результатам расчетов была построена зависимость отношения силы тяги винта к силе тяге изолированного винта от относительного расстояния до поверхности земли (рис. 1). Референсные данные приведенные на данном графике представлены в работе [4] и включают в себя данные физических экспериментов, результаты численного моделирования других авторов и аналитические оценки.

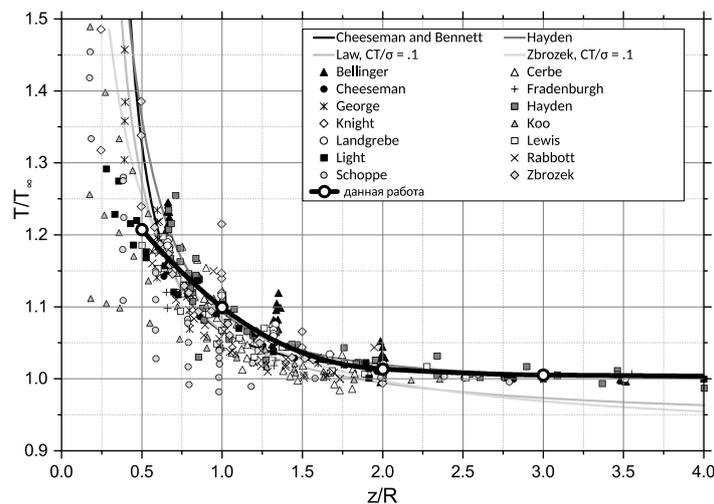


Рис. 1: Зависимость относительной силы тяги винта от высоты над землей

Анализ полученных полей показал, что увеличение относительной тяги с приближением поверхности земли вызвано как увеличением избыточного давления на поверхности лопасти, так и увеличением давления в зоне между винтом и землей.

В ходе данной работы методика численного эксперимента была успешно применена для исследования физического эффекта влияния поверхности земли на аэродинамические характеристики несущего винта вертолета. Проведенный анализ результатов, подтвердил эффект увеличения тяги винта с уменьшением расстояния до земли. Полученный характер роста относительной тяги винта с приближением к поверхности земли качественно и количественно хорошо согласуется с данными физических экспериментов, с численными результатами других авторов и с аналитическими оценками, что подтверждает применимость

используемой методики для исследования характеристик винтов вертолета в подобных режимах.

Литература

- [1] Numerical investigation of the aerodynamic and acoustical properties of a shrouded rotor / Abalakin I. V., Anikin V.A., Bakhvalov P.A., Bobkov V.G., Kozubskaya T.K. // Fluid Dynamics. 2015. Vol. 51(3). P. 419–433. DOI: 10.1134/s0015462816030145.
- [2] Bakhvalov P.A., Kozubskaya T.K. Construction of edge-based 1-exact schemes for solving the Euler equations on hybrid unstructured meshes // Computational Mathematics and Mathematical Physics. 2017. Vol. 57(4). P. 680–697. DOI: 10.1134/s0965542517040030.
- [3] Параллельный программный комплекс NOISETTE для крупномасштабных расчетов задач аэродинамики и аэроакустики / Абалакин И.В., Горобец А.В., Дубень А.П., Козубская Т.К., Бахвалов П.А. // Вычислительные методы и программирование. 2012. № 13. С. 110–125.
- [4] Experimental investigation of rotorcraft outwash in ground effect / Tanner P.E., Overmeyer A.D., Jenkins L.N., Yao C.-S., Bartram S.M. // AHS International Annual Forum & Technology Display. 2015. No. NF1676L-21054.

Анализ и прогноз временных рядов на примере рынка криптовалют

Цветинский Вадим Юрьевич

Кафедра вычислительных методов

e-mail: vadim.tsvet@gmail.com

Научный руководитель — к.ф.-м.н. доц. Терновский Владимир Владимирович

Основная задача работы заключается в выяснении того, подчиняется ли рынок криптовалют законам нелинейной динамики или, другими словами, существует ли закон, описывающий эволюцию подобных систем во времени. При исследовании данной проблемы применяются математические методы, в частности, теорема Флориса Такенса [1]. Актуальность темы обусловлена тем, что на государственном уровне и в бизнес среде непрерывно увеличиваются капиталовложения в криптовалюты.

Теорема Такенса. Теорема утверждает, что возможно получить структуру, топологически эквивалентную исходному аттрактору динамической системы, с помощью вложения в подпространство большей размерности и использования временной задержки.

Теорема Такенса

Пусть $x(t) = v_j(t)$ для некоторого $j = 1, \dots, n$, где $v(t) = (v_1(t), \dots, v_n(t))$ кривая на многообразии Ω . Пусть $v(t)$ плотно в топологии пространства Ω . Тогда существует такое $\tau > 0$, $K \in \mathbb{Z}$, что соответствующие вектора $(x(t), x(t + \tau), \dots, x(t + K\tau))$ на многообразии сохраняют топологию Ω .

Реконструкция аттрактора Лоренца. Сдвиг временных рядов, создаваемых системой дифференциальных уравнений Лоренца, создает вложение более высокой размерности и, согласно теореме Такенса, позволяет реконструировать фазовое пространство аттрактора. Параметры временного сдвига τ и размерности встраиваемого подпространства находятся с помощью взаимной информации [2] и метода ложных ближайших соседей [3] соответственно. Результат применения теоремы: Рис. 1.

Прогнозирование системы Лоренца. Идея здесь заключается в том, что после встраивания в новое пространство у нас есть матрица X со сдвигом нашего исходного временного ряда. Мы можем использовать первый $m - 1$ столбец в этой матрице в качестве признаков для нашего регрессора.

Для системы Лоренца $\tau = 16$, $m = 3$, на Рис. 2 изображен прогноз:

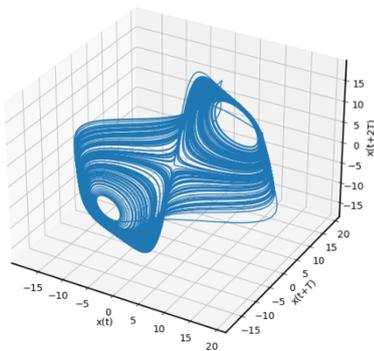


Рис. 1: Восст. Аттрактор

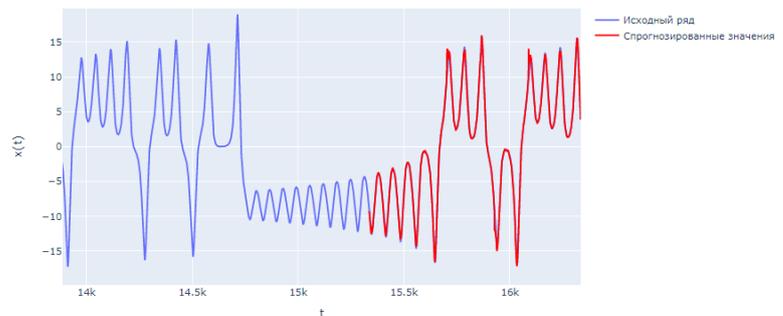


Рис. 2: Прогн. сист. Лоренца

Прогнозирование криптовалютных активов. Прогноз на реальных данных строится в два этапа, сначала на зашумленных данных, затем на данных, очищенных с помощью метода эмпирической модовой декомпозиции. Прогноз построенный на очищенных данных:

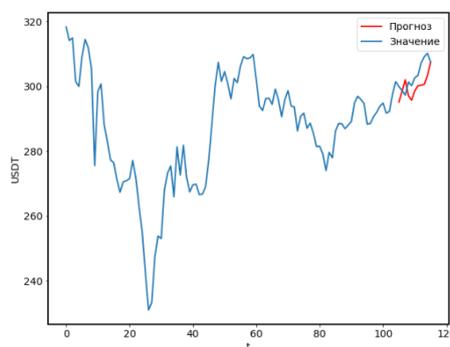


Рис. 3: Прогноз и реальные значения цены актива

Выводы Применена теорема Такенса, реализованы методы нахождения ее параметров, реализован и применен метод эмпирической модовой декомпозиции, построены прогнозы для системы Лоренца и реальных данных.

Проведенный анализ полученных результатов показывает, что подобные децентрализованные системы являются слабопредсказуемыми, но тем не менее поддаются прогнозированию, при условии корректного выбора параметров теоремы Такенса и порога очищения шумов.

Литература

- [1] Takens F. The Takens embedding theorem // International Journal of Bifurcation and Chaos. 1991. С.867–872.
- [2] A. Dionisio, R. Menezes, and D. A. Mendes. Mutual information: a measure of dependency for nonlinear time series // Physica A: Statistical Mechanics and its Applications. 2004. С.326–329.
- [3] Kennel, Matthew B. and Brown, Reggie and Abarbanel, Henry D. I. Determining embedding dimension for phase-space reconstruction using a geometrical construction // Phys. Rev. A. 1992. С.3403–3411.

Разработка программно-информационной среды для сопровождения экспериментов на установках токамак

Долматов Александр Андреевич

Кафедра автоматизации научных исследований

e-mail: zarokdol@yandex.ru

Научный руководитель — д.ф.-м.н. проф. Сычугов Дмитрий Юрьевич

Научный консультант — д.ф.-м.н. проф. Соловьёв Сергей Юрьевич

Аннотация. Работа посвящена интернет-технологиям интегрированного моделирования сложных технических устройств и их применению к установкам токамак и состоит из трёх частей. В первой части приведены результаты, полученные автором при работе в составе научного коллектива по модернизации средств обработки экспериментальных данных установки токамак Т-15МД [1]. Во второй части приводится описание программной среды SIEMNED, предназначенной для моделирования и поддержки экспериментов на сложных технических устройствах [2], и излагаются результаты, полученные автором при разработке её серверной части. В третьей части приведены результаты исследования системы пассивных витков в установке токамак Т-15МД, полученные автором с помощью среды SIEMNED. Часть результатов опубликована в работе [2].

Виртуальный аналог как инструмент исследования. Виртуальным аналогом сложного устройства назовём интегрированную математическую модель, полученную путём объединения в единую среду численных кодов, описывающих все основные стороны функционирования такого устройства. Создание виртуальных аналогов стало возможно в результате развития технологий численного моделирования. Впечатляющие результаты применения виртуальных аналогов в авиационной, автомобильной и военной промышленности можно найти в обзоре [3]. Точно так же эти технологии эффективно используются в

исследованиях по управляемому термоядерному синтезу, в том числе для строящегося международного реактора ITER [4] и российской установки токамак Т-15МД [1].

Результаты модернизации системы обработки и хранения данных установки токамак Т-15МД. Первоначально предполагалось, что имеющаяся база обработки и хранения данных установки Т-10 может без особых изменений перенесена на Т-15МД. Однако оказалось, что программная основа СУБД Т-10 значительно устарела и требует серьёзной модернизации. Автор в составе научного коллектива принимал участие в этой работе и лично получил следующие результаты: реализована новая структура записи данных БД DASSQL, обновлена СУБД и выполнен перенос БД на новый сервер.

Реализация серверной части Единой программной среды SIEMNED. Для корректного моделирования всякого сложного технического устройства необходимо наличие большого числа вычислительных кодов, предназначенных для расчёта той или иной стороны работы данного устройства. Виртуальный аналог приходится создавать путём интеграции данных кодов в единую программную среду, которая могла бы из отдельных частей создать целостную картину. В полной мере это относится к проектированию, строительству и, в частности, к сопровождению экспериментов на установках токамак. Целью проекта является разработка Единой программной среды, позволяющей создавать виртуальные аналоги сложных технических устройств для проведения с такими устройствами численных экспериментов, включая виртуальные испытания и оптимизацию их конструкций. Создание подобной среды, несомненно, будет способствовать ускорению во много раз сроков разработки проектов новых установок токамак, что в высшей степени актуально (например, строящаяся сейчас установка ITER [4] проектировалась примерно 20 лет). В рамках проведения работ по модернизации Единой программной среды автором были выполнены следующие основные задачи: произведено объединение двух составляющих проекта – справочно-информационной и функциональной; осуществлён переход на новую платформу, которая располагает расширенными вычислительными мощностями; выполнен перенос базы данных и обновлён словарь элементов.

Исследование системы пассивных витков в установке токамак Т-15МД. Поскольку большинство современных установок токамак характеризуется вытянутым сечением плазмы, сепаратрисой и наличием полоидального дивертора, то в токамаке Т-15МД (как и в других установках подобного типа) важен контроль вертикальной неустойчивости плазмы на протяжении всего разряда. Данный контроль осуществляется с помощью включения в заданные моменты времени сценария, встроенного в общую систему блоков. В основе модели, применяемой в данном коде, лежит предположение о том, что плазменный шнур движется по вертикали как твёрдое тело и распределение тока в нём и полный ток не меняются. Были проведены расчёты характерного времени затухания токов и его зависимость от конкретного расположения пассивных

элементов в камере установки токамак Т-15МД с помощью входящих в Единую программную среду расчётных кодов.

Литература

- [1] Экспериментальная термоядерная установка токамак Т-15МД / Хвостенко П. П., Анашкин И. О., Бондарчук Э. Н., Инютин Н. В., Крылов В. А., Левин И. В., Минеев А. Б., Соколов М. М. // Вопросы Атомной Науки и Техники. Серия Термоядерный синтез. 2019. Т. 42, № 1. С. 15–38.
- [2] Жилкин А. С., Соловьёв С. Ю., Бурдуковская С. Ю., Сычугов Д. Ю., Высоцкий Л. И., Долматов А. А. / Единая программная среда SIEMNED и её применение к моделированию процессов в плазме. // Сборник тезисов докладов. Научная конференция «Тихоновские чтения» (Москва, 25–30 октября 2021г.). М. : МАКС ПРЕСС, 2021. 31 с.
- [3] Дерюгин Ю. Н. // Суперкомпьютерные технологии в промышленности. Опыт применения и актуальные задачи. (Москва, 23 октября 2013г.).
- [4] ITER Physics Basis. // Nuclear Fusion. 1999. Vol. 39. N. 12.
- [5] Zhilkin A. S., Sychugov D. Yu, Vysotsky L. I., Zotov I. V., Soloviev S. Yu., Sadykov A. D. New open computational resource for plasma processes modeling. Journal of Physics. Conference Series. 2021. Vol. 1730. P. 012–048.

Моделирование алгоритма управления группой мобильных роботов на основе методов потенциальных полей

Артамонов Георгий Игоревич

Кафедра нелинейных динамических систем и процессов управления

e-mail: s02180002@gse.cs.msu.ru

Научный руководитель — д.ф.-м.н. проф. Фомичев Василий Владимирович

Постановка задачи. В настоящее время большую популярность набирают мобильные роботы. Их используют в самых различных отраслях. Особый интерес при этом представляет групповое управление, поскольку использование группы роботов позволяет более эффективно решать практические задачи. В рамках ВКР рассматривалась одна из задач ГУ. Требовалось построить локальный децентрализованный алгоритм управления группой с ведущими и ведомыми роботами, обеспечивающий выполнение следующих условий:

- Ведущие роботы за время T должны посетить не менее τ_{min} целевых точек.
- Ведомые роботы должны следовать за своими ведущими, то есть расстояние между ведомым и ведущим роботами для всех пар в каждый момент времени не должно превышать ρ_{max} .
- Все роботы должны обходить препятствия и избегать столкновений друг с другом.

Математическая модель. Для описания движения мобильного робота используется дискретная система с ограничениями на управление (1).

$$\begin{cases} x_{i+1} = x_i + v_i \cos \varphi_i \\ y_{i+1} = y_i + v_i \sin \varphi_i \\ \varphi_{i+1} = \varphi_i + \Delta\varphi_i, & |\Delta\varphi_i| \leq \delta_1 \\ v_{i+1} = \text{sat}(v_i + \Delta v_i, 0, v_{max}), & |\Delta v_i| \leq \delta_2 \\ (x_0, y_0, \varphi_0, v_0) = (x^0, y^0, 0, 0) \end{cases} \quad (1)$$

Здесь (x_i, y_i) — координаты робота в дискретный момент времени i . (x^0, y^0) — стартовая точка робота. φ_i, v_i — угол поворота и скорость робота в момент времени i соответственно. Управление мобильным роботом осуществляется изменением его скорости и угла поворота. Дополнительно наложены ограничения на управление: $|\Delta\varphi_i| \leq \delta_1, |\Delta v_i| \leq \delta_2$ и на скорость мобильного робота: $v_i \leq v_{max}$. В рассматриваемой модели форма робота аппроксимируется кругом радиуса r с центром в точке (x_i, y_i) . Препятствия задаются в виде произвольных выпуклых многоугольников и эллипсов. Для хранения карты местности используется матрица $C \in \mathbb{R}^{p \times q}$. Обозреваемая роботом локальная окрестность рассматривается в виде центрированного относительно текущего положения робота квадрата, таким образом робот получает информацию о препятствиях в виде квадратной подматрицы $C_a \in \mathbb{R}^{w_s \times w_s}$ матрицы C , соответствующей текущему положению робота на карте.

Алгоритм управления. Для построения алгоритма управления использовались методы потенциальных полей. В частности, применялся метод виртуального силового поля VFF [1]. Существенным недостатком данного метода было затрудненное движение робота в узких проходах, что зачастую приводило к застреванию и полной остановке мобильного робота (проблема локального минимума). Также данному алгоритму управления свойственна излишняя колебательность при движении робота по узким коридорам. Наиболее удовлетворительных результатов с точки зрения поставленных целей управления удалось достичь при использовании метода гистограммы векторного поля VFH [2, 3]. Этот метод стал дальнейшим развитием VFF алгоритма. Основное отличие заключается в двухэтапной обработке данных. На первом этапе строится двумерная декартова гистограмма C_a . На втором этапе по матрице C_a строится одномерная полярная гистограмма $H[k]$. Далее осуществляется поиск на полярной гистограмме долин — последовательно идущих секторов с полярной плотностью препятствий ниже некоторого порогового значения h_{max} . Среди всех найденных долин выбирается наиболее подходящая для движения к цели, и направление движения задается в нее. Выбор скорости робота осуществляется исходя из загруженности препятствиями текущего направления движения. После того как выбраны желаемые направление и скорость движения, вычисляются соответствующие уставки для изменения угла поворота и скорости робота с учетом установленных ограничений. Несмотря на значительно меньшее число застреваний и закливаний роботов по сравнению с VFF

методом, по-прежнему происходили ситуации заторов в узких проходах. Для решения этой проблемы метод VFH был усовершенствован добавлением белого шума к управлению углом поворота, что в серии экспериментов позволило роботам разъехаться.

Полученные результаты. Была разработана среда моделирования для проведения экспериментов с группой мобильных роботов с визуализацией работы системы. Был построен алгоритм управления группой мобильных роботов на основе модификации существующего метода гистограммы векторного поля. Полученный алгоритм управления был промоделирован и были получены оценки его качества в зависимости от соотношения максимальных скоростей ведущих и ведомых роботов.

Литература

- [1] Borenstein J., Koren Y. Real-time obstacle avoidance for fast mobile robots // IEEE Transactions on Systems, Man, and Cybernetics, 1989, Vol. 19, N 5, P. 1179–1187.
- [2] Borenstein J., Koren Y. Real-time obstacle avoidance for fast mobile robots in cluttered environments // Proceedings of IEEE International Conference on Robotics and Automation, Cincinnati, Ohio, 1990, P. 572–577.
- [3] Borenstein J., Koren Y. The Vector Field Histogram – Fast Obstacle-Avoidance for Mobile Robots // IEEE Journal of Robotics and Automation, 1991, Vol. 7, N 3, P. 278–288.

Моделирование распределённых регуляторов для систем с запаздыванием

Бутиков Илья Иванович

Кафедра нелинейных динамических систем и процессов управления

e-mail: ilyaboomer@mail.ru

Научный руководитель — к.ф.-м.н. асс. Атамась Евгений Иванович

Введение. Запаздывание — это свойство физической системы, характеризующееся отложенной реакцией системы на приложенное к ней воздействие.

Системы с запаздыванием появляются во многих науках: в биологии, химии, экономике, экологии, физике. Системы с запаздыванием часто используются для моделирования большого класса инженерных систем, в которых материал, информация или энергия физически перемещаются из одного места в другое.

Наличие запаздываний, особенно длительных, значительно усложняет процесс анализа системы и разработку систем управления. Поэтому стоит избегать запаздываний при разработке систем. Но в силу различных причин на практике не всегда удаётся полностью избежать запаздываний, поэтому построение регуляторов для систем с запаздываниями является востребованной задачей.

Рассматриваемый объект. В работе рассматривается следующий класс систем с запаздыванием:

$$\begin{cases} \dot{x}(t) = Ax(t) + Bu(t - h) \\ y(t) = Cx(t) \end{cases} \quad (1)$$

Где $A \in R^{n \times n}$, $B \in R^{n \times 1}$, $C \in R^{1 \times n}$, $h > 0$ — запаздывание по времени.

В качестве регулятора используется алгоритм назначения конечного спектра (Finite-spectrum assignment, FSA) [1]:

$$\begin{cases} u(t) = Fx_p(t) \\ x_p(t) = e^{Ah}x(t) + v(t) \\ v(t) = \int_0^h e^{A\theta}Bu(t - \theta)d\theta \end{cases} \quad (2)$$

Где $F \in R^{1 \times n}$, $x_p(t)$ — предсказываемое состояние объекта к моменту времени $t + h$.

Основная проблема моделирования связана с реализацией интегрального слагаемого $v(t)$. Ожидается, что при аппроксимации интеграла с помощью стандартных методов (прямоугольника, трапеции) полученный регулятор схемы FSA будет устойчивым даже при небольшой вариации параметров. Однако результаты экспериментов в литературе [2] показывают, что существуют примеры, для которых полученная система управления не является устойчивой независимо от точности стандартных методов аппроксимации.

Это противоречие между теоретическим предположением и практической реализацией представляет интерес данной работы.

Постановка задачи. Целью данной выпускной квалификационной работы является программная реализация интегральных слагаемых распределенного регулятора в среде Matlab/Simulink.

Для её достижения были поставлены и решены следующие задачи:

1. Изучить литературу по теме Finite spectrum assignment (FSA) и распределенных регуляторов.
2. Изучить методы реализации распределённых запаздываний.
3. Изучить способы построения S-функций в Matlab/Simulink.
4. Создать в среде Simulink непрерывный блок-буфер, реализующий одно или несколько запаздываний.
5. Создать в среде Simulink блок для реализации интегрального слагаемого и сравнить полученные реализации.
6. Используя полученные блоки построить регулятор по алгоритму FSA (2) для систем класса (1) и провести численные эксперименты.

Результаты. Одним из результатов работы является написанный на языке C++ непрерывный блок-буфер, поддерживающий реализацию как одного, так и нескольких запаздываний, с возможностью задания начальных условий в виде ломаной функции.

Основная сложность при его разработке была связана с внедрением в систему Matlab/Simulink, что недостаточно хорошо описано в документации.

В данной работе этот блок используется как в реализации запаздывания системы (1), так и в блоке, реализующем интегральные слагаемые вида $v(t)$. Ожидается, что разработанный блок-буфер найдёт применение при моделиро-

вании систем с запаздываниями других классов, не рассмотренных в данной работе.

Так же в ходе работы были написаны скрипты для анализа устойчивости систем с запаздыванием рассматриваемого класса на основе формул из [2].

В ходе проведенных экспериментов получены следующие результаты:

1. Устойчивость системы зависит от качества дискретизации.
2. При увеличении величины запаздывания необходимо уменьшать шаг дискретизации.
3. Найден пример, на котором метод средних прямоугольников не сходится.
4. Использование импульсного регулятора вместо непрерывного может улучшить сходимость систем с запаздыванием вида (1).

Литература

- [1] Zhong Q. -C. Robust Control of Time-Delay Systems. London : Springer-Verlag, 2006.
- [2] Engelborghs K., Dambrine M., Roose D. Limitations of a class of stabilization methods for delay systems // IEEE Transactions on Automatic Control. 2001. Vol. 46, N 2. P. 336–339

Построение системы программного управления, реализующей колебательные движения в динамическом объекте с гистерезисом

Кондратьева Варвара Николаевна

Кафедра нелинейных динамических систем и процессов управления

e-mail: varya.kondratyeva@gmail.com

Научный руководитель — д.ф.-м.н. проф. Фурсов Андрей Серафимович

В работе [1] описана термомеханическая установка, являющаяся устройством сбора энергии, осуществляющим преобразование тепловой энергии в механическую. Такое устройство может быть частью преобразования тепловой энергии в электрическую (сбор тепловой энергии — преобразование в механическую — преобразование в электрическую), что является полезным для сохранения и экономии электроэнергии. Управление системой осуществляется за счет нагрева и охлаждения нити из материала с памятью формы: при нагреве нить укорачивается, при охлаждении удлиняется. Повторяющиеся растяжения и сокращения нити изгибают пьезоэлектрический кантилевер *, который генерирует электрические заряды.

В работе [2] представлена упрощенная математическая модель термомеханической установки в виде замкнутой управляемой нелинейной динамической системы:

*упругий элемент, который перемещается между двумя конечными положениями, соответствующими нагретому и охлажденному состоянию, и генерирует электрические заряды

$$\begin{cases} m\ddot{y}(t) + \beta\dot{y}(t) + ky(t) = \alpha l^{-1}E(T(t))(\Delta - y(t)) - mg \\ \alpha\gamma^{-1}\dot{T}(t) + T(t) = u \\ y(0) = y_0, \quad \dot{y}(0) = 0, \quad T(0) = T_0 \end{cases} \quad (1)$$

Здесь m — масса груза, β — коэффициент затухания колебаний кантилевера, k — коэффициент упругости кантилевера, α — диаметр нити из материала с памятью формы, l — длина нити при отсутствии механического напряжения, T — температура нити, E — модуль Юнга материала с памятью формы, представляющий собой нелинейную характеристику с гистерезисом и насыщением, зависящую от температуры T , Δ — высота конца ненапряженной нити над точкой опоры кантилевера, g — ускорение свободного падения, γ — характеристика скорости тепловых процессов, u — кусочно-постоянное управление, y — высота груза над точкой опоры кантилевера (выходная переменная), y_0 , T_0 — начальные условия.

Доказано [2], что в системе (1) возможен периодический режим, однако нахождение соответствующих параметров установки аналитически представляется трудоемкой задачей, в связи с чем было принято решение попытаться найти параметры установки численными методами. В качестве численного метода был выбран генетический алгоритм.

Было задано множество гармонических сигналов, описывающих искомый периодический режим: $G = \{A \sin(2\pi\nu t) : A \in (0, \Delta), \nu \in (0, +\infty)\}$, A — амплитуда, ν — частота сигнала. Для заданного сигнала $g^*(t) = A^* \sin(2\pi\nu^*t) \in G$ искалось кусочно-постоянное программное управление вида:

$$u = \begin{cases} u_1, & q\theta \leq t < q\theta + t_1, \quad \theta = \frac{1}{\nu^*} \\ u_2, & q\theta + t_1 \leq t < (q+1)\theta, \quad q = 0, 1, 2, \dots, \quad u_1 > u_2 \end{cases} \quad (2)$$

такое, чтобы установившийся режим системы был колебательным с гармоническими характеристиками A^* , ν^* , где θ — период гармонического сигнала.

Кроме управления (2) искались легко интерпретируемые и измеряемые физические параметры системы (1) m , l , Δ , t_1 . Они были выбраны для увеличения размерности пространства поиска для численного метода, что повышает шанс найти требуемое программное управление. Таким образом, вектор параметров для поиска — $(u_1, u_2, t_1, m, l, \Delta)$.

Для сравнения реального выходного сигнала системы (1) $y(t)$ и гармонического колебания с характеристиками A^* , ν^* была предложена оценка качества выходного сигнала:

$$r(y) = q_\nu |\bar{\nu}^* - \nu^*| + q_A |\bar{A}^* - A^*|, \quad (3)$$

где A^* , ν^* — заданные искомые гармонические характеристики, \bar{A}^* , $\bar{\nu}^*$ — характеристики $y(t)$, измеренные с помощью дискретного преобразования Фурье [3], $q_\nu = 1/\max(\nu^*, \bar{\nu}^*)$, $q_A = 1/\max(A^*, \bar{A}^*)$ — коэффициенты масштабирования.

Также были исследованы возможные трудности при использовании дискретного преобразования Фурье и приняты меры по предотвращению возможных проблем.

В ходе работы был разработан программный модуль в Matlab, реализующий генетический алгоритм поиска вектора параметров $(u_1, u_2, t_1, m, l, \Delta)$ замкнутой системы управления (1), при котором минимизируется функционал (3). Система (1) была реализована как схема в Simulink. Затем было проведено тестирование программного модуля, показавшее, что предложенный способ решает поставленную задачу.

Результаты данной работы могут быть использованы в дальнейших исследованиях термомеханической установки [1].

Литература

- [1] Modelling and investigation of a hybrid thermal energy harvester / T. Todorov, N. Nikolov, G. Todorov, Y. Ralev // MATEC Web of Conferences. 2018. Vol. 148. P. 12002.
- [2] О существовании периодического режима в одной нелинейной системе / А. С. Фурсов, Р. П. Митрев, П. А. Крылов, Т. С. Тодоров // Дифференциальные уравнения. 2021. Т. 57, № 8. С. 1104–1115
- [3] Павлейно М. А., Ромаданов В. М. Спектральные преобразования в MATLAB // СПб., 2007. 160 с.

Управление групповым поведением экономических агентов в приближении игры среднего поля

Федоров Федор Алексеевич

Кафедра системного анализа

e-mail: s02180555@gse.cs.msu.ru

Научный руководитель — акад. РАН, д.ф.-м.н. проф. Шананин Александр Алексеевич

Представленная работа берет начало со статьи [2] за авторством Лорелла Фатоне, Франчески Мариани, Марии Кристины Рекчиони, Франческо Зирилли. В статье исследуется модель, описывающая поведение высокочастотных трейдеров на рынке акций.

Модель основна на концепции игр среднего поля и является системой из двух связанных уравнения в частных производных: уравнение Колмогорова-Фоккера-Планка, условия для которого задаются в начальный момент времени, и уравнение Гамильтона-Якоби-Беллмана, условия для которого задаются в конечный момент времени. Подробно подход игр среднего поля описан в курсе лекций Пьера-Луи Лионса, Жана-Мишеля Ласре и Оливера Гейнона [1]. Полученную систему можно решать численно. Один из способов это сделать — подобрать оптимизируемый функционал специального вида, чтобы при применении к нему метода множителей Лагранжа функция цены из уравнения Гамильтона-Якоби-Беллмана представляла из себя сопряженные переменные, а

условие для него в конечный момент времени являлось условием трансверсальности. Однако остается открытым вопрос о существовании и единственности решения системы.

В статье делается предположение, согласно которому решение уравнения Колмогорова-Фоккера-Планка ищется в виде гауссианы, а функция цены уравнения Гамильтона-Якоби-Беллмана в полиномиальном виде. Оно позволяет свести краевую задачу для уравнений в частных производных к многомерной краевой задаче для уравнений Риккати.

Полученную систему можно решать численно, однако вопрос существования и единственности решения остается открытым. В представленной работе этот вопрос был исследован.

Было установлено, что данная краевая задача не порождается никакой вариационной. Следующие результаты были получены методами анализа поведения решений ОДУ и теории Пуанкаре-Бендиксона. Удалось получить условие существования решения системы, доказать, что в случае выполнения этого условия решение единственно, а также получить оценки на начальные данные, при которых краевая задача разрешима.

Литература

- [1] Guéant O., Lasry J.-M., Lions P.-L.: ParisPrinceton Lectures on Mathematical Finance 2010, chapter «Mean field games and applications», Shpringer, 2011. p. 205–266.
- [2] Fatone L., Mariani F., Recchioni M., Zirilli F. A trading execution model based on mean field games and optimal control // Applied Mathematics, 2014, vol. 5, p. 3091–3116.

Методы построения торговых стратегий для валютных пар на основе статистических моделей и сетей LSTM

Виляев Антон Львович

Кафедра математической статистики

e-mail: avilyaev720@gmail.com

Научный руководитель — д.ф.-м.н. доц. Горшенин Андрей Константинович

Использование автоматизированных торговых систем весьма распространено на современных финансовых рынках, поэтому развитие математических моделей и алгоритмов для них представляет собой сложную и интересную задачу. В данной работе рассматривается совместное использование статистических подходов и нейронных сетей для построения таких систем на валютном рынке.

Первостепенной целью данной работы является повышение точности нейросетевого прогнозирования в задачах торговли на валютных рынках за счет использования параметров конечных смешанных вероятностных моделей для нетривиального расширения признакового пространства. Основные задачи включают в себя: выбор подходящей архитектуры для прогнозирования временных рядов, проверку эффективности использования СРС-компонент в качестве

дополнительных признаков, тестирование полученных методов на реальных данных. Модель, получая на вход временной ряд $X_{1,\dots,t}$, должна выдавать сигнал $y(t) \in \{-1, 0, 1\}$, где «-1» будет означать, что значение временного ряда спустя определенный временной промежуток окажется значительно ниже значения в момент времени t (необходимо продавать актив), «1» – значительно выше (необходимо покупать актив), «0» – отклонится не более, чем на заданную величину (рекомендуется не совершать никаких сделок с активом).

Известно (см., например, работы [1-2]), что для решения поставленной задачи наибольший интерес представляют LSTM-сети. При тестировании различных архитектур (см. работы [3-5]) было выявлено, что разделение исходной задачи тернарной классификации на последовательное решение задачи бинарной классификации и тернарной классификации улучшает качество предсказания. Таким образом, в качестве архитектуры для предсказания выбран ансамбль из двух LSTM-сетей.

Повышение качества нейросетевых алгоритмов зачастую требует привлечения дополнительных наблюдений, которые для реальных данных могут быть недоступны в силу различных обстоятельств. Выходом в такой ситуации становится использование математических моделей для данных, позволяющих эффективно расширять признаковое пространство. Расширение признакового пространства подразумевает передачу дополнительных статистик, построенных по ряду, в модель. При этом их выбор зачастую обусловлен известными свойствами анализируемого процесса. Так, для моделирования колебаний цены в финансовой математике используют стохастическое дифференциальное уравнение вида: $dX(t) = a(t)d(t) + b(t)dW(t)$ - здесь $W(t)$ - винеровский процесс, коэффициенты $a(t)$ и $b(t)$ случайны. Распределение приращений $P(X(t_i) - X(t_{i-1}) < x)$ можно аппроксимировать смесями нормальных распределений вида $E\Phi(\frac{x-A_i}{B_i})$, где $\Phi(x)$ - стандартная нормальная функция распределения, $A_i \in R$, $B_i > 0$ - случайные величины. В свою очередь, для распределения случайных величин A_i и B_i , можно использовать дискретную аппроксимацию. Тогда справедливо приближение вида:

$P(X(t_i) - X(t_{i-1}) < x) \approx \sum_{k=1}^K p_k \Phi(\frac{x-a_k}{b_k})$, где $K \in N$, $p_k \geq 0$, $\sum_k p_k = 1$. Оценивая параметры этого распределения в режиме скользящего окна, получаем дополнительные признаки (так называемые CPC-компоненты) для обучения нейронных сетей.

Построенный алгоритм был протестирован на трех валютных парах (EURGBP, USDGBP, USDJPY), тренировочный набор для каждой из пар составлял 56,106 наблюдений, тестовый – 10,858 наблюдений, для сравнения использовалась метрика доходности в процентах годовых. По результатам прирост годовой доходности торговой стратегии с использованием именно ансамбля LSTM с нетривиальным расширением признакового пространства CPC-компонентами по сравнению с LSTM без них составил до 18.9%

Литература

- [1] Omer Berat Sezer, M. Ugur Gudelek, Ahmet Murat Ozbayoglu. Financial Time Series Forecasting with Deep Learning : A Systematic Literature Review // Department of Computer Engineering, TOBB University of Economics and Technology, Ankara, Turkey, 2019.
- [2] Avraam Tsantekidis, Anastasios Tefas. Transferring trading strategy knowledge to deep learning models // Knowledge and Information Systems, 2021. Volume 63. pp. 87–104.
- [3] Виляев А.Л., Горшенин А.К. Использование рекуррентных нейросетевых архитектур и статистических подходов для построения торговых стратегий // Тихоновские чтения: научная конференция, Москва, МГУ имени М. В. Ломоносова, 25–30 октября 2021 г.: тезисы докладов. – М.: МАКС Пресс 2021 г. С. 121.
- [4] Виляев А.Л., Горшенин А.К. О моделировании торговых стратегий для валютных пар с использованием глубоких нейронных сетей и метода скользящего разделения смесей // Интеллектуальные Системы. Теория и приложения, 2021. Т. 25. Вып. 4. С. 92–96.
- [5] Горшенин А.К., Виляев А.Л. Программа моделирования торговых стратегий для валютных пар с использованием глубоких нейронных сетей. Свидетельство о государственной регистрации программы для ЭВМ №2021668631 от 25.11.2021.

Статистический анализ данных трафика виртуального мобильного оператора

Горбунов Сергей Алексеевич

Кафедра математической статистики

e-mail: s.gorbunov.smc@gmail.com

Научный руководитель — д.ф.-м.н. доц. Горшенин Андрей Константинович

Целью работы является статистическое исследование трафика мобильного оператора с решением задач прогнозирования и выявления аномальных объёмов трафика. Принципиальная новизна данной работы заключается в использовании моделей на основе обобщённого гамма-распределения [1], которые ранее в таких целях в литературе не применялись. Актуальность исследования обусловлена, во-первых, необходимостью эффективного управления сетями мобильной связи, во-вторых, необходимостью совершенствования технологий обеспечения качества обслуживания.

Особенностью анализируемых 44 млн реальных данных является их агрегированность по часам: каждый час по каждому активному пользователю суммируется число отправленных и полученных бит, каждое наблюдение маркируется меткой типа приложения, инициировавшего обмен данными. Таким образом, за каждый час имеется выборка просуммированных объёмов трафика по различным пользователям и типам приложений. В работе предлагается

использование обобщённого гамма-распределения для аппроксимации эмпирических распределений объёмов трафика, его плотность задаётся формулой (1).

$$f(x; \mu, \sigma, \nu) = \frac{|\nu| \theta^\theta z^\theta e^{-\theta z}}{\Gamma(\theta) x}, \quad x > 0, \quad (1)$$

$$\mu > 0, \quad \sigma > 0, \quad -\infty < \nu < \infty, \quad \nu \neq 0, \quad z = \left(\frac{x}{\mu}\right)^\nu, \quad \theta = \frac{1}{\sigma^2 \nu^2}.$$

Благодаря подходящему выбору широкого класса вероятностных распределений было получено высокое качество статистической аппроксимации на многих поразному сформированных выборках: варьировалась ширина окна по времени, способ группировки по типам приложений и типам трафика.

Для каждого из двух типов трафика сформируем семь временных рядов с часовым шагом по времени: количество наблюдений, суммарный трафик, количество уникальных пользователей, средний трафик, приходящийся на уникального пользователя, и три временных ряда параметров обобщённого гамма-распределения. Для прогнозирования временных рядов были выбраны следующие три модели: интегрированная модель авторегрессии – скользящего среднего с сезонностью – SARIMA(p, d, q)(P, D, Q)[s], регрессия на члены тригонометрического ряда Фурье до порядка n с SARIMA(p, d, q)(P, D, Q)[s] ошибками – Fourier(n)(p, d, q)(P, D, Q)[s], векторная авторегрессия порядка p с сезонными фиктивными переменными – VAR(p)[s]. Наиболее точными по метрике RMSE на нормированной тестовой выборке оказались модели Fourier и VAR: величина ошибки не превышает 0.08–0.16 для всех четырнадцати временных рядов. Также стоит отметить быстрое действие модели VAR – на стандартном персональном компьютере она обрабатывает за доли секунды. Полученные точечные прогнозы и доверительные интервалы могут быть использованы для прогнозирования загруженности сети мобильной связи.

Использование обобщённого гамма-распределения открывает возможность выявления аномальных наблюдений с помощью расширения статистической процедуры, описанной в статье [2]. Воспользуемся другой параметризацией распределения (1) на основе следующих замен:

$$r = \frac{1}{\sigma^2 \nu^2}, \quad \gamma = \nu, \quad \tilde{\mu} = \frac{1}{\sigma^2 \nu^2 \mu^\nu}.$$

Пусть V_1, \dots, V_m независимая выборка из обобщённого гамма-распределения с некоторыми параметрами $r > 0, \gamma > 0, \tilde{\mu} > 0, V_1 \geq V_j, \forall j \geq 2$,

$$\mathcal{R} = \frac{(m-1)V_1^\gamma}{V_2^\gamma + \dots + V_m^\gamma}. \quad (2)$$

Тогда при условии, что верна гипотеза H_0 : «значение V_1 не является аномально большим», статистика (2) имеет F-распределение Снедекора с параметрами r и $(m-1)r$. В упомянутой статье в силу специфики предметной области (анализируются данные об осадках) рассматривается случай положительного параметра γ . Однако в задачах анализа мобильного трафика выборки зачастую

имеют обобщённое гамма-распределение с отрицательным параметром γ , поэтому в данной работе предлагается использование обобщённой статистики (3). Новый метод выявления аномалий был протестирован на многих выборках реальных данных и показал адекватные результаты.

$$\hat{\mathcal{R}} = \left(\frac{(m-1)V_1^\gamma}{V_2^\gamma + \dots + V_m^\gamma} \right)^{\text{sgn}(\gamma)}, \quad \hat{\mathcal{R}}|_{H_0} \sim \begin{cases} F(r, (m-1)r) & \text{при } \gamma > 0 \\ F((m-1)r, r) & \text{при } \gamma < 0 \end{cases}. \quad (3)$$

Методы анализа мобильного трафика, описанные в работе, могут быть использованы для тонкой настройки объёмов выделяемых ресурсов при сегментации мобильной сети. В дальнейшем модели на основе обобщённого гамма-распределения могут использоваться в качестве базиса при решении задач анализа различных видов телекоммуникационного трафика, в том числе с использованием нейронных сетей.

Литература

- [1] Горбунов С. А., Горшенин А. К. Об обобщённом гамма-распределении в задачах анализа мобильного трафика // Научная конференция «Ломоносовские чтения»: тезисы докладов (Москва, 14–22 апреля 2022 г.). М. : МАКС Пресс, 2022. С. 166-167.
- [2] Korolev V. Yu., Gorshenin A. K. Probability models and statistical tests for extreme precipitation based on generalized negative binomial distributions // Mathematics. 2020. Vol. 8, Iss. 4. Art. No. 604.

Критерий проверки биоэквивалентности при наличии пропущенных данных

Клименко Виктор Кириллович

Кафедра математической статистики

e-mail: Rockthebrony@gmail.com

Научный руководитель — к.ф.-м.н. доц. Захарова Татьяна Валерьевна

В работе рассмотрена проблема проверки гипотезы биоэквивалентности при наличии пропущенных данных и предложен метод её решения. При перекрёстных дизайнах медицинских исследований потеря одного наблюдения, при традиционном подходе, вынуждает исследователя потерять также и другие собранные наблюдения, из-за чего объём обрабатываемой информации уменьшается. Альтернативно, для сохранения информации потерянные наблюдения можно заменять средними значениями, однако это приводит к неконтролируемому повышению вероятности ошибки первого рода, то есть риска потребителя. Метод, предложенный в данной работе, позволяет использовать все доступные данные и при этом контролировать вероятность ошибки первого рода.

Для исследования был выбран дизайн эксперимента, применяющийся на практике в клинических исследованиях. Предложенный метод основан на традиционном, доказавшем свою эффективность и корректность, критерии Шуирманна, также известном как процедура двух односторонних тестов Шуирманна.

Критерий Шуирманна требует полные пары данных, то есть, при потере одного наблюдения для испытуемого, приходится отказаться и от другого, неутраченного наблюдения. В работе доказано утверждение, позволяющее использовать критерий Шуирманна, не отказываясь от собранных наблюдений, при условии, что существуют такие функции неутраченных данных, которые ограничивают, хотя бы с некоторой уверенностью, значения статистик критерия.

В работе предложено два способа построения таких функций, доказана корректность одного из них (корректность другого очевидна), выбран более эффективный.

Для реализации метода был предложен способ построения доверительных интервалов, оценивающих значения пропущенных наблюдений. При помощи сгенерированных данных было показано, что эмпирический уровень доверия этих интервалов очень близок к теоретическому, доказывая правильность теоретического результата. Таким же образом было показано, что закругление, происходящее при взятии максимумов и минимумов мажорирующих статистики критерия функций по доверительной области для пропущенных наблюдений, незначительно.

При помощи сгенерированных данных при разных параметрах модели были вычислены эмпирические вероятности ошибок первого и второго рода. Теоретический результат был подтверждён, так как предложенный метод действительно контролировал значение вероятности ошибки первого рода, даже слишком хорошо: эмпирическая ошибка была намного меньше заданного 0.05. Однако, ошибка второго рода у предложенного метода оказалась сильно завышенной в сравнении с традиционным методом, в котором беспарные наблюдения просто игнорируются.

Для решения проблемы были предложены возможные модификации, теоретически необоснованные, однако те же симуляции показали, что они либо не контролируют вероятность ошибки первого рода, либо всё равно уступают традиционному методу.

Предложенный метод был применён к реальным данным, в ходе чего снова была подтверждена его корректность, а также эффективность предложенных доверительных интервалов.

Литература

- [1] Т. В. Захарова, А. А. Тархов. Оценка уровня значимости критерия шуирманна для проверки гипотезы биоэквивалентности при наличии пропущенных данных // Информатика и её применения. 2019. Т.12. Вып. 3.
- [2] М. А. Драницына, Т. В. Захарова, Р. Р. Ниязов. Свойства процедуры двух односторонних тестов для признания биоэквивалентности лекарственных препаратов // Журнал "Ремедиум". №3, 2019.
- [3] John N. S. Matthews, Robin Henderson. Two-period, two-treatment crossover designs subject to non-ignorable missing data // Biostatistics. 2013. Volume 14, Issue 4.

- [4] Roger L. Berger, Jason C. Hsu. Bioequivalence Trials, Intersection–Union Tests and Equivalence Confidence Sets // *Statistical Science* 1996. Vol. 11, No. 4. pp. 283–319.
- [5] Sauter R, Steinijans VW, Diletti E, Böhm E, Schulz H-U. Presentation of results from bioequivalence studies // *Int J Clin Pharmacol Ther Toxicol*. 1992. 30:S7–30.

Анализ поведения явных и скрытых потоков тепла в Северной Атлантике с помощью методов математической статистики

Муляр Миллиард Юрич

Кафедра математической статистики

e-mail: my.mulyar@gmail.com

Научный руководитель — д.ф.-м.н. проф. Королев Виктор Юрьевич

Задача изучения климатических изменений в океане привлекает большое количество исследований, так как она является критически важной для понимания глобальных процессов развития климата. Именно статистический взгляд на данные помогает отслеживать изменения, находить структуру и зависимости, которые впоследствии могут быть объяснены с физической точки зрения.

Некоторыми группами исследователей рассматривалось суточное и месячное осреднение для разных промежутков времени и строилась карта мирового океана [1,2]. Наиболее подробные карты с шагом измерения в 3 часа и шагом географической сетки в 0.75 градуса с помощью спутниковых данных и вариационных методов были построены в работах [3,4]. Также было рассмотрено несколько математических методов анализа самих потоков: так, например, в работе [5] были аппроксимированы и проанализированы распределения потоков в конкретных географических точках с помощью метода максимального правдоподобия, в работе [6] временные ряды изучались с помощью регрессионного анализа и метода скользящего среднего.

В данной работе в каждой географической точке акватории фиксировались ежемесячные средние значения потоков, а также максимальные и минимальные их значения за каждый месяц в течение сорока лет. В рассматриваемой работе показано наличие долговременных трендов в эволюции указанных характеристик. Впервые рассмотрено совместное поведение этих характеристик для явного и скрытого потоков тепла. При условии справедливости общепринятой гипотезы о нормальности совместного распределения указанных характеристик выделены зоны в Атлантическом океане, в которых преобладает каждый из рассматриваемых потоков — явный и скрытый — и предложена реалистичная интерпретация обнаруженных закономерностей.

Кроме того, для изучения распределений экстремальных значений был применен EM-алгоритм кластеризации, который изучил природу изменения максимальных значений и показал, как именно меняется система за весь заданный период. Полученные результаты были переданы ученым, занимающимся

изучением океанов, и были даны объяснения полученных результатов с океанологической точной зрения. Также важно отметить, что анализ совместного распределения потоков тепла, который не производился ранее в аналогичных исследованиях, открыл ученым новые возможности для интерпретации и понимания процессов, протекающих в океане.

Безусловно полученные результаты являются хорошей базой для проведения новых исследований, в следствии чего изучение данной задачи будет продолжаться.

Литература

- [1] Josey S. A., Kent E. C., Taylor P. K. New insights into the ocean heat budget closure problem from analysis of the SOC air–sea flux climatology //Journal of Climate. – 1999. – Т. 12. – №. 9. – С. 2856-2880.d
- [2] Grist J. P., Josey S. A. Inverse analysis adjustment of the SOC air–sea flux climatology using ocean heat transport constraints //Journal of Climate. – 2003. – Т. 16. – №. 20. – С. 3274-3295.
- [3] Grodsky S. A. et al. Intraseasonal latent heat flux based on satellite observations //Journal of Climate. – 2009. – Т. 22. – №. 17. – С. 4539-4556.
- [4] Andersson A. et al. Evaluation of HOAPS-3 ocean surface freshwater flux components //Journal of Applied Meteorology and Climatology. – 2011. – Т. 50. – №. 2. – С. 379-398.
- [5] Gulev S. K. et al. North Atlantic Ocean control on surface heat flux on multidecadal timescales //Nature. – 2013. – Т. 499. – №. 7459. – С. 464-467.
- [6] Kumar M. et al. Surface flux modelling using ARIMA technique in humid subtropical monsoon area //Journal of atmospheric and solar-terrestrial physics. – 2009. – Т. 71. – №. 12. – С. 1293-1298.

Решение обратной задачи магнитоэнцефалографии с двумя источниками

Пахненко Иван Петрович

Кафедра математической статистики

e-mail: vrrmob@gmail.com

Научный руководитель — к.ф.-м.н. доц. Захарова Татьяна Валерьевна

Настоящая работа посвящена сложной задаче функционального картирования головного мозга человека, которую предпочтительно решать современными методами неинвазивного исследования головного мозга. Одним из таких методов является магнитоэнцефалография (МЭГ).

Магнитоэнцефалография дает уникальные возможности для неинвазивного изучения процессов в нейронах головного мозга [1, 2]. МЭГ позволяет выявить самые разнообразные активируемые в экспериментальных условиях функциональные части мозга.

Зарегистрированные процедурой магнитоэнцефалографии сигналы содержат информацию о различных электрических активностях, наблюдаемых внутри

головного мозга. Одна из важнейших задач - точная локализация областей активности нейронов [3], заключающаяся в воссоздании положения диполей, которые являются источниками сигналов магнитоэнцефалографии. На данный момент отсутствует унифицированный способ решения подобной проблемы с заданной точностью. Однако в однодипольном случае обратная задача локализации решена авторами в работах [4, 5], что послужило базисом для исследования и нахождения решения в более общих моделях.

Представленный в работе метод реконструкции положения источников разработан для модели с двумя фиксированными диполями, каждый из которых расположен близко к поверхности головы, а также имеет отношение к отличающимся функциональным областям. В этом случае исследуемая проблема при заданных условиях имеет решение в виде мультидипольного приближения.

В текущей работе предложены и исследованы новые ранее неприменяемые методы для решения обратной задачи магнитоэнцефалографии с двумя источниками с целью получения достаточной точности в локализации положения источников для использования в реальных условиях.

Предлагаемый алгоритм решения обратной задачи с двумя диполями включает несколько шагов. Первый шаг представляет собой приблизительное вычисление координат диполей путем применения анализа независимых компонент, метода ансамбля решающих деревьев, EM-алгоритма к первичным сигналам МЭГ и использование аналитического решения для обратной задачи магнитоэнцефалографии в однодипольном случае; на втором шаге проблема нахождения источников активности решается методами оптимизации при нахождении экстремума целевой функции с использованием в качестве начального приближения результатов, полученных на первом шаге.

Методы локализации, представленные в данной работе, имеют несомненно важное значение в реальной клинической практике. Так, в ходе нейрохирургических вмешательств могут быть повреждены различные зоны головного мозга, в том числе и невосполнимые. Так как расположение функциональных зон в мозге человека индивидуально, врачу необходимо иметь возможность локализовать эти области в предоперационный период с высокой точностью. Методы, разработанные в данной работе, и служат для решения такой важной задачи.

В дальнейшем предполагается усложнить модель исследования и решить задачу с помощью суперкомпьютерных технологий.

Литература

- [1] Baillet S., Mosher J. C., Leahy R. M. Electromagnetic Brain Mapping // IEEE Signal Processing Magazine. 2001. P. 14–30.
- [2] Sarvas J. Basic mathematical and electromagnetic concepts of the biomagnetic inverse problem // Physics in Medicine and Biology. 1987. Vol. 32. P. 11–22.
- [3] Решение обратной задачи в многодипольной модели источников магнитоэнцефалограмм методом независимых компонент / В. Е. Бенинг, М. А. Дра-

ницына, Т. В., Захарова, П. И. Карпов // Информатика и ее применения. 2014. Т. 8, № 2. С. 79–87.

[4] Zakharova T. V., Karpov P. I., Bugaevskii V. M. Localization of the activity source in the inverse problem of magnetoencephalography // Computational Mathematics and Modeling. 2017. Vol. 28, № 2. P. 148–157.

[5] Karpov P. I., Zakharova T. V. Magnetoencephalography inverse problem in the spheroid geometry // Journal of Inverse and Ill-Posed Problems. 2019. Vol. 27, № 2. P. 159–169.

Методы сегментации изображений без учителя

Селезнёв Михаил Васильевич

Кафедра математической статистики

e-mail: simplymike8@gmail.com

Научный руководитель — д.ф.-м.н. проф. Ульянов Владимир Васильевич

Сегментация — это задача разбиения изображения на области. В работе рассматривается семантическая сегментация: в таком случае каждая область соответствует одному из заранее предопределённых классов. Например, для приложений сегментации в беспилотных автомобилях можно ввести классы «человек», «дорога», «задний фон».

Потребность в сегментации изображений возникает в системах управления беспилотными автомобилями, в медицине, в видеонаблюдении, выделении объектов на спутниковых снимках, распознавании отпечатков пальцев, распознавании лиц и многих других областях.

При наличии обучающей выборки — набора изображений с известным правильным разбиением на области — можно достичь высокого качества, используя методы глубинного обучения с учителем. Однако, создание обучающей выборки в задаче сегментации крайне трудоёмко: каждому пикселю изображения требуется сопоставить класс, на одном изображении сотни тысяч пикселей, и типичный объём обучающей выборки, необходимый для хорошей работы алгоритмов — тысячи изображений. Поэтому возникает желание исследовать методы, не требующие вручную размеченных данных — так называемые, методы обучения без учителя.

В рамках ВКР выполнено несколько подзадач.

1. Проведён подробный обзор существующих методов глубинного обучения, решающих задачу семантической сегментации без учителя.
2. Впервые решена задача семантической сегментации без учителя на сложном наборе данных ADE20k [1], отличающимся большим числом классов (150 против 27 в предшествующих работах).
3. Предложен достаточно общий метод автоматической разметки данных, основанный на использовании промежуточных представлений в нейронных сетях, дающий, в том числе, прирост в качестве на смежной задаче детекции объектов.

Представляется перспективной дальнейшая работа по улучшению качества сегментации без учителя на ADE20k, связанная с разработкой методов, устойчивых к большому числу классов.

Литература

- [1] Semantic understanding of scenes through the ade20k dataset / B. Zhou [et al.] // International Journal of Computer Vision. 2019. Vol. 127, N 3. P. 302–321

Подбор гиперпараметров генеративных состязательных сетей, оптимизация модели

Хуторной Дмитрий Павлович

Кафедра математической статистики

e-mail: dimon.khutornoy@mail.ru

Научный руководитель — д.ф.-м.н. проф. Ульянов Владимир Васильевич

Научный консультант — к.ф.-м.н. доц. Орлов Михаил Владимирович

Описание проблемы. Данная работа посвящена проблеме подбора гиперпараметров при работе с алгоритмами машинного обучения. Гиперпараметры позволяют настраивать процесс обучения того или иного алгоритма и напрямую влияют на его производительность и итоговую эффективность.

Постановка задачи. Задача ВКР - исследовать проблему обучения генеративных состязательных сетей и сравнить методы функциональной оптимизации для подбора гиперпараметров.

В процессе работы было необходимо решить следующие подзадачи.

1. Проанализировать ряд алгоритмов функциональной оптимизации.
2. Подобрать архитектуру модели и основные гиперпараметры для оптимизации
3. Проанализировать характеристики качества алгоритмов машинного обучения и разработать на их основе функции приспособленности.
4. Сравнить основные характеристики алгоритмов подбора гиперпараметров.

Алгоритмы функциональной оптимизации, используемые в работе

1. Алгоритм дифференциальной эволюции. Основывается на смешивании и замене признаков векторов одной эпохи для перехода на новую эпоху [1]. Эпоху образует множество векторов.
2. Алгоритм «1+1». Признаки одного вектора заменяются на новые с учетом вероятности мутации [2].
3. Алгоритм случайного поиска. Основывается на задании точке случайного ускорения, которое «принимается», если качество было улучшено.
4. Алгоритм роя частиц. Также, как и дифференциальная эволюция, основан на взаимодействии между объектами, каждый из которых меняет свое положение в зависимости от точки наилучшего качества своего и общего (по всем объектам) [4].

5. Байесовская оптимизация. Строится на основе вероятностной модели. На каждой итерации получает новое значение оцениваемой функции и на основе этих данных улучшает аппроксимацию.

Архитектура и гиперпараметры. Для исследования была выбрана архитектура условной генеративной состязательной сети с двухслойными генератором и дискриминатором. Для оптимизации были выбраны 20 гиперпараметров, описывающие гиперпараметры генератора и дискриминатора, а также взаимодействия между ними. Такое небольшое число гиперпараметров обуславливается наличием ограничений на размерности байесовской оптимизации [3].

Функции приспособленности. В работе разработаны три вида функций.

1. Эстиматорные. Функции, основанные на построении дополнительной модели, оценивающей качество генератора при помощи метрик.
2. Примитивные. Функции, основанные на интуитивном понимании о том, что метрика должна быть больше, а потери и разность потерь составных сетей модели должны быть ниже.
3. Усредненное расхождение. Усредненная по каждому из признаков дивергенция Кульбака-Лейблера [5].

Порядок проведения эксперимента. Было проведено по 10 запусков каждой модели с разными начальными условиями. Посчитано 500-1000 итераций для алгоритмов «1+1», байесовской оптимизации и случайного поиска и 100-200 эпох для дифференциальной эволюции и метода роя частиц.

Результаты.

- Обнаружена, исправлена проблема работы «1+1» и добавлен в исследование модифицированный вариант алгоритма.
- Рассчитаны оптимальные наборы гиперпараметров каждого алгоритма и проведено сравнение.
- Проведено построение и сравнение различных функции приспособленности.
- Определены наиболее устойчивые алгоритмы.
- Определены алгоритмы, показывающие наилучшее качество
- Проведено визуальное сравнение распределений генераторов оптимальных моделей

Литература

- [1] Baiocchi M., Di Bari G., Poggioni V., Traccolli M. Differential Evolution for Learning Large Neural Networks // Machine Learning, Optimization, and Data Science. 2017. vol. 10710, pp. 401-413.
- [2] Droste S., Jansen Th., Wegener I. A rigorous complexity analysis of the (1 + 1) Evolutionary Algorithm for linear functions with Boolean inputs. In: Proc. IEEE Internat. Conf. on Evolutionary Computation ICEC '98. 1998. Pp. 499-504.

- [3] Eriksson D., Jankowiak M. High-Dimensional Bayesian Optimization with Sparse Axis- Aligned Subspaces. In: Proceedings of the Thirty-Seventh Conference on Uncertainty in Artificial Intelligence, PMLR. Vol. 161. Pp. 568-577.
- [4] Kennedy J., Eberhart R. C. Particle swarm optimization. In: ICNN'95 - International Conference on Neural Networks. 2002. Pp. 1942–1948.
- [5] Kullback S., Leibler R.A. On information and sufficiency // Annals of Mathematical Statistics. 1951. Vol. 22. Pp. 79-86.

Улучшение автоматической стилизации изображений с помощью преобразующей нейросети

Елистратов Семен Юрьевич

Кафедра математических методов прогнозирования

e-mail: semenelist@gmail.com

Научный руководитель — к.ф.-м.н. доц. Китов Виктор Владимирович

Целью данной исследовательской работы являлось увеличение гибкости существующей модели автоматического переноса стиля [1]. Для этого были поставлены следующие задачи: упростить модель, с целью добиться обобщения на произвольный стиль при обучении на меньшем объеме данных, чем предлагается авторами статьи; реализовать плавный контроль силы стилизации, плавный контроль сохранения исходной цветовой гаммы контента и дополнительные параметры контроля стилизации, отвечающие за перенос паттернов с нескольких изображений стиля.

Задача автоматического переноса стиля [2] заключается в том, что на вход некоторой модели подается два изображения: изображение контента и изображение стиля. Изображение контента обычно некоторая реальная фотография, а изображение стиля - картина, рисунок или текстура. Необходимо получить изображение контента, стилизованное заданным стилем.

Трансформационный подход [3] к задаче стилизации заключается в обучении на выборке стилей конечного размера сети преобразователя, которая получает на вход изображение контента и выдает результат стилизации. Недостатком данного подхода является необходимость для каждого изображения стиля обучения отдельной модели, что является крайне затратным по времени. Поэтому была предложена модель [1], имеющая сеть кодировщик стилей, которая для любого изображения стиля получает его векторное представление фиксированного размера. На основе этого векторного представления вычисляются параметры сети преобразователя, отвечающие за перенос стиля для конкретного изображения стиля. Такой подход позволяет при обучении модели на большой выборке стилей производить стилизацию для произвольных изображений стиля и контента за один проход по сети.

Авторы статьи [1] утверждают, что для обобщения модели для работы с произвольным стилем необходимо обучать ее более, чем на 16000 изображений

стиля, что является крайне затруднительным при отсутствии доступа к мощным вычислительным ресурсам. Поэтому было предложено упростить модель добавив регуляризацию, а именно слой Dropout [4] в конец кодировщика стилей, а также добавлены тождественные связи из кодировщика в декодировщик для ускорения настройки сети. Кроме того, были сформированы три выборки стилей: базовые, дополнительные и тестовые, и предложена двухэтапная схема обучения: сначала модель обучалась базовых стилей, и после того как она начала показывать качественные результаты модель дообучалась на большой выборке дополнительных стилей, причем на каждой итерации с некоторой вероятностью подмешивались стили из базовой выборки. Такой подход позволил модели показывать качественную стилизацию на базовых стилях, но и в тоже время обобщиться на работу с произвольным стилем.

Была решена проблема изменения изображения контента при малых значениях параметра силы стилизации. Контроль силы стилизации производился с помощью линейной интерполяции между векторными представлениями стиля и контента, полученными с помощью кодировщика стилей, а также между связями из кодировщика и остаточных слоев. Кроме того, был реализован пост-процессинг при малых значениях параметра, а именно линейная интерполяция между исходным изображением контента и результатом стилизации.

Был предложен плавный контроль сохранения исходной цветовой гаммы контента путем перекраски результата стилизации в цвета изображения контента методом приведения гистограмм. Метод работает следующим образом: функция распределения интенсивности пикселей относительно каждого канала одного изображения приводится к другой функции распределения, являющийся линейной интерполяцией функций распределений двух изображений. Варьируя параметр линейной интерполяции, можно контролировать сохранение исходной цветовой палитры изображения контента.

Кроме того, были выбраны несколько характерных изображений стиля и вычислены их векторные представления. Варьируя веса между этими векторными представлениями и вычисляя взвешенную сумму можно контролировать перенос паттернов стиля с каждого изображения, что сильно расширяет гибкость модели и позволяет пользователю использовать сразу несколько изображений стиля для стилизации.

Таким образом, в данной работе были предложены различные подходы, позволяющие значительно расширить возможности пользовательского контроля при использовании модели автоматической стилизации изображений.

Литература

- [1] Ghiasi G. et al. Exploring the structure of a real-time, arbitrary neural artistic stylization network // arXiv preprint arXiv:1705.06830. – 2017.
- [2] Gatys L. A., Ecker A. S., Bethge M. A neural algorithm of artistic style // arXiv preprint arXiv:1508.06576. – 2015.
- [3] Dumoulin V., Shlens J., Kudlur M. A learned representation for artistic style // arXiv preprint arXiv:1610.07629. – 2016.

- [4] Srivastava N. et al. Dropout: a simple way to prevent neural networks from overfitting // The journal of machine learning research. – 2014. – Vol. 15. – N. 1. – P. 1929-1958.

Нейросетевые методы в задаче оптимизации логических схем

Самбурский Александр Ильич

Кафедра математических методов прогнозирования

e-mail: sashasamb@yandex.ru

Научный руководитель — к.ф.-м.н. доц. Гуров Сергей Исаевич

В области электротехники одним из важных направлений является моделирование функций алгебры логики (ФАЛ) в классе функциональных схем. При конструировании схем осуществляются попытки представить их в наиболее компактном виде, исходя из соображений вычислительной эффективности. Данная задача сводится к процессу трудоёмкой комбинаторной оптимизации, для её эффективного решения требуется применять эвристики, позволяющие упростить исходную постановку задачи и находить приближённые решения. Среди существующих алгоритмов представляют интерес в том числе и методы, основанные на применении нейросетей и обучения с подкреплением [1, 2]. В представленной работе предложен, описан и исследован новый нейросетевой вероятностный подход для обработки логических схем.

Исследуемая проблема описывается следующей постановкой: для ФАЛ, заданной своей таблицей истинности на существенных наборах, требуется получить моделирующий её работу функциональный граф в зафиксированном логическом базисе. При этом алгоритм должен быть предрасположен к моделированию безызбыточных по структуре схем за контролируемое время.

Базовая архитектура. Для решения поставленной задачи предлагается перейти от рассмотрения детерминированных логических схем к неопределённым вычислительным графам: в данных графах не установлен порядок подачи переменных или констант на входы схемы, а также неизвестна расстановка функциональных элементов. Эту неопределённость можно формализовать в рамках вероятностной модели, определяющей поведение вычислительного графа. В качестве параметров данной модели выступают вероятности реализаций дискретно распределённых случайных величин: номеров переменных, подаваемых на конкретные входы сети, а также типов функциональных элементов, реализуемых в конкретных фрагментах схемы. Наиболее подходящие параметры θ данной модели должны быть определены в ходе оптимизации функционала отклонения прогнозируемых значений y_{prob} от истинных откликов ФАЛ y_{true} на задающих её наборах x при прогонке этих наборов через параметризованную схему:

$$\mathcal{L}(x, \theta) = \|y_{prob}(x, \theta) - y_{true}(x)\| \rightarrow \min_{\theta}$$

Данный метод предоставляет возможность обрабатывать логические схемы нейросетевыми архитектурами. При проектировании подхода необходимо контролировать уровень используемой непрерывной релаксации: в течение обучения алгоритм должен выдавать всё более однозначно определенные вычислительные графы. Это свойство позволит после настройки параметров на этапе перехода к детерминированному вычислительному графу избежать существенное изменение функциональности логической схемы.

Среди полезных особенностей предлагаемого подхода можно выделить его способность разреживать логическую схему. С помощью добавки аддитивного регуляризатора к оптимизируемому функционалу можно поощрять модель подавать на входы сети константы, благодаря чему большая часть элементов схемы будет вырожденной, что позволит удалить их из рассмотрения.

Усиление архитектуры. Для дополнительного повышения гибкости и эффективности модели можно воспользоваться несколькими модификациями базовой архитектуры. С помощью методов ансамблирования схем можно решать задачу моделирования поэтапно: на каждой итерации добавлять к текущей схеме очередной функциональный блок, исправляющий ошибки предыдущих, но при этом моделирующий схему для более простой функции.

Для повышения эффективности и вычислительной устойчивости процедуры обучения можно использовать метод REBAR [3]. Подход позволяет производить несмещённую оценку градиента оптимизируемого функционала в моделях, оперирующих с дискретными случайными величинами.

Для обеспечения глобальности анализа схемы можно пробросить связи между её различными фрагментами. К примеру, передавать агрегацию всех параметров одного слоя схемы на следующий с помощью дополнительных надстроек над архитектурой. Это позволит различным вершинам графа обмениваться более комплексным потоком информации на этапе обучения при обратном распространении ошибки, что повысит гибкость модели.

Результаты. Предлагаемые подходы были применены к ФАЛ различной сложности, в том числе зависящим от нескольких десятков переменных и заданным на нескольких сотнях наборов. Предложенная архитектура моделировала схемы, более компактные, чем стандартные методы синтеза примерно в 23.2% случаях и позволяла получить сопоставимые по сложности схемы, но за меньшее время для 30.9% ФАЛ. Методы показывают неплохие результаты по времени и качеству работы и предоставляют инструменты для регулирования этих параметров, что демонстрирует гибкость предлагаемой модели.

Литература

- [1] Gurov S. I., Zolotarev D. V., Samburskii A. I. Artificial neural networks and logic circuit synthesis // Computational Mathematics and Modeling. 2021. Т. 32. № 4. С. 490–499.
- [2] Гуров С. И., Золоторев Д. В., Самбурский А. И. Обучение с подкреплением в задаче синтеза мажоритарных схем // Современные информационные технологии и ИТ-образование. 2021. Т. 17. № 2. С. 295–307.

- [3] Tucker G. et al. Rebar: Low-variance, unbiased gradient estimates for discrete latent variable models // *Advances in Neural Information Processing Systems*. 2017. T. 30.

Система визуального анализа для активности

ГОЛОВНОГО МОЗГА

Тыцкий Владислав Игоревич

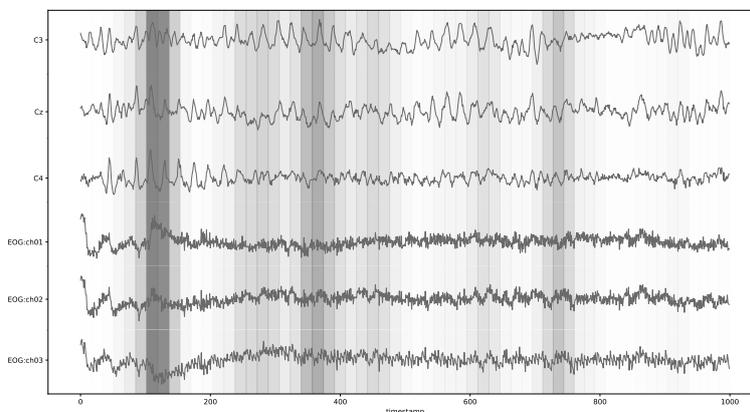
Кафедра математических методов прогнозирования

e-mail: mmp@cs.msu.ru

Научный руководитель — к.ф.-м.н. доц. Майсурадзе Арчил Ивериевич

В рамках данного исследования была поставлена цель разработать методы интерпретации моделей машинного обучения для временных рядов, которые в дальнейшем можно было бы применить для медицинских данных, в частности, для электроэнцефалограммы головного мозга.

В задачах машинного обучения от модели часто требуется не только хорошее качество предсказания, но и ее «интерпретируемость», т. е. «понятность» для человека. Иногда это желание вызвано серьезным риском при принятии решения, иногда интерпретация результата может помочь увидеть новые, нетривиальные зависимости. За последние годы было разработано множество методов интерпретации для моделей, работающих с различными данными, но временные ряды не входят в их число.



В данной работе были подробно изучены основные методы интерпретации, такие как SHAP [2], LIME [3]. В первую очередь они подходят для табличных данных, поэтому при рассмотрении наиболее простого случая, одномерного временного ряда, удобно воспринимать его, как вектор фиксированной длины. В таком случае, мы можем применить подходы SHAP и LIME к такому временному ряду, что было сделано в статье [1]. Такой подход имеет ряд недостатков: плохая масштабируемость, ограничение на размерность ряда, игнорирование специфики данных. В работе предлагаются методы, лишённые этих недостатков. Основная идея основана на разделении ряда на небольшие фрагменты, состоящие из нескольких подряд идущих измерений (*timestamp*). Каждый такой фрагмент мы считаем неделимым признаком и, с некоторыми, модификациями применяем вышеупомянутые методы, LIME или SHAP.

В экспериментах для медицинских данных использовался датасет Motor Movement/Imagery Dataset [4]. Датасет содержит электроэнцефалограммы относящиеся к двум классам: человек *думал* о движении левой рукой, человек *думал* о движении правой рукой. Были обучены три различных алгоритма машинного обучения для решения задачи классификации: логистическая регрессия, сверточная нейронная сеть и градиентный бустинг. Для каждой из них был применен предлагаемый метод. На рисунке визуализирован результат его работы. Помимо хорошей масштабируемости, алгоритм ведет себя более робастно, чем в статье [1], и указывает на те участки ряда, которые действительно важны в определении класса. Таким образом, предлагаемый алгоритм может помочь ученым ставить новые гипотезы о мозговой активности, «подсвечивать» уже известные паттерны в ЭЭГ, чтобы врачам было проще ставить диагноз, а также помочь понять адекватность модели в каждом конкретном случае.

Литература

- [1] Schlegel U. et al. Towards a rigorous evaluation of xai methods on time series // 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW). – IEEE, 2019. – С. 4197-4201.
- [2] Ribeiro M. T., Singh S., Guestrin C. «Why should i trust you?» Explaining the predictions of any classifier // Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. – 2016. – С. 1135-1144.
- [3] Lundberg S. M., Lee S. I. A unified approach to interpreting model predictions // Advances in neural information processing systems. – 2017. – Т. 30.
- [4] Schalk G. et al. BCI2000: a general-purpose brain-computer interface (BCI) system // IEEE Transactions on biomedical engineering. – 2004. – Т. 51. – №. 6. – С. 1034-1043.

Некоторые интервалы в решетке замкнутых классов частичных булевых функций

Миронов Михаил Ильич

*Кафедра математической кибернетики
e-mail: m.mironov_244224@mail.ru*

Научный руководитель – д.ф.-м.н. проф. Алексеев Валерий Борисович

Пусть $E_k = \{0, 1, \dots, k-1\}$, $E_k^n = \{(a_1, a_2, \dots, a_n) \mid \forall i (a_i \in E_k)\}$ и ∞ трактуется как неопределенность. Множество всех функций $f : E_k^n \rightarrow E_k$ ($f : E_k^n \rightarrow E_k \cup \{\infty\}$), ($n = 1, 2, \dots$) с операцией суперпозиции называют k -значной логикой P_k (соответственно, частичной k -значной логикой P_k^*).

Пусть A — произвольный замкнутый класс в P_k . Через $Int(A)$ будем обозначать множество всех замкнутых классов в P_k^* , содержащих A и состоящих только из функций, доопределимых до какой-нибудь функции из A [1].

Ранее многими авторами изучался более широкий интервал $\mathcal{I}(A)$, который состоит из всех замкнутых классов в P_k^* , пересечение которых с P_k есть

в точности A . После долгих исследований в 2017 году была установлена мощность интервалов $\mathcal{I}(A)$ для всех замкнутых классов A булевых функций [2].

Пусть $Pr(k)$ — множество всех предикатов на E_k (то есть функций, принимающих только значения 0 и 1) от любого числа переменных. Пусть $A \subseteq P_k$, $K \subseteq Pr(k)$. Через $[K]_A$ будем обозначать замыкание множества предикатов K относительно следующих операций над предикатами: 1) произвольное переименование переменных; 2) добавление и изъятие фиктивных переменных; 3) конъюнкция предикатов; 4) подстановка в предикат функций из A вместо некоторых переменных.

Множество всех предикатов $1(x_1, \dots, x_n) \equiv 1$ от любого числа переменных будем обозначать $\{1\}$. Через $Z(A)$ будем обозначать семейство всех подмножеств K в $Pr(k)$, содержащих $\{1\}$ и таких, что $[K]_A = K$.

Пусть $f(x_1, \dots, x_n) \in P_k^*$, $R(x_1, \dots, x_n) \in Pr(k)$. Тогда через f/R мы будем обозначать следующую функцию $h(x_1, \dots, x_n) \in P_k^*$:

$$h(x_1, \dots, x_n) = \begin{cases} f(x_1, \dots, x_n), & \text{если } R(x_1, \dots, x_n) = 1, \\ \infty, & \text{если } R(x_1, \dots, x_n) = 0. \end{cases}$$

Пусть $A \subseteq P_k$, $K \subseteq Pr(k)$.

Тогда положим $A/K = \{g \in P_k^* \mid \exists f \in A, \exists R \in K (g = f/R)\}$.

В [3] доказано следующее утверждение, которое сводит изучение интервала $Int(A)$ к изучению $Z(A)$.

Теорема 1. Если A — клон в P_k , то $Int(A)$ состоит из всех классов вида A/K , где K пробегает все классы предикатов из $Z(A)$. При этом $A/K_1 \subset A/K_2$ тогда и только тогда, когда $K_1 \subset K_2$.

Если $K_1, K_2 \subseteq Pr(k)$, то положим $K_1 \cdot K_2 = \{R_1 \cdot R_2 \mid R_1 \in K_1, R_2 \in K_2\}$.

Утверждение. Если $K_1 \in Z(A)$, $K_2 \in Z(A)$ и $K_1 \subseteq K_2$, то $K_1 \cdot K_2 = K_2$.

Класс K из $Z(A)$ будем называть базисным классом для $Z(A)$, если существует такой предикат $R \in Pr(k)$, что $[R \cup \{1\}]_A = K$. Важную роль для описания $Z(A)$ играют следующие утверждения.

Теорема 2. Пусть $A \subseteq P_k$. Тогда семейство $Z(A)$ состоит в точности из всех произведений базисных классов для $Z(A)$.

Теорема 3. Пусть $A, B \subseteq P_k$. Тогда $[K]_{A \cup B} = K$ в том и только в том случае, если $[K]_A = K$ и $[K]_B = K$. Если $K_1 \in Z(A)$, $K_2 \in Z(B)$, то $K_1 \cap K_2 \in Z(A \cap B)$. Если $K_1, K_2 \in Z(A)$, то $[K_1 \cdot K_2]_A = [K_1]_A \cdot [K_2]_A = [K_1 \cup K_2]_A$.

С помощью Теоремы 2 построим $Z(N)$, где $N = M \cap T_0 \cap T_1$.

Пусть $A \subseteq P_2$, $a, b \in \{0, 1\}$. Введём обозначения:

$$A_{ab} = \{f \in A \mid f(0, \dots, 0) = a, f(1, \dots, 1) = b\}$$

$$C = \{R = const\}, \bar{A} = \{\bar{f} \mid f \in A\}$$

$$Q = \{f(x_1, \dots, x_n) \in P_2 \mid \forall \alpha, \beta, \gamma : \alpha < \beta < \gamma, (f(\alpha) = f(\gamma) = 1) \Rightarrow (f(\beta) = 1)\}.$$

Теорема 4. 1) Для $Z(N)$ существует ровно 9 базисных классов:

$\{1\}, C, \{Q_{00} \cup \{1\}\}, \{P_{00} \cup \{1\}\}, \{N \cup \{1\}\}, \{P_{01} \cup \{1\}\}, \{\overline{N} \cup \{1\}\}, \{P_{10} \cup \{1\}\}, P_{11}$.

2) Семейство классов $Z(N)$ состоит из 33 классов. Кроме базисных классов в нём содержатся следующие 24 класса:

$M, \overline{M}, M \cdot \overline{M}, \{Q_{00} \cup M\}, \{Q_{00} \cup \overline{M}\}, \{P_{00} \cup M\}, \{P_{00} \cup \overline{M}\}, \{P_{01} \cup C\},$
 $\{P_{10} \cup C\}, \{P_{11} \cup C\}, \{P_{00} \cup P_{11}\}, \{T_0 \cup \{1\}\}, \overline{T_0}, T_1, \{\overline{T_1} \cup \{1\}\}, \{T_0 \cup \overline{M}\}, \{\overline{T_1} \cup M\},$
 $\{T_0 \cup P_{10} \cup \{1\}\}, \{T_1 \cup C\}, \{\overline{T_0} \cup C\}, \{T_0 \cup T_1\}, \{\overline{T_0} \cup \overline{T_1}\}, \{P_{00} \cup M \cdot \overline{M}\}, Pr$.

Теорема 5. $Int(N)$ содержит ровно 33 замкнутых класса. Это классы вида N/K , где K пробегает все классы предикатов из Теоремы 4.

Используя Теоремы 2–4 и то, что $M \cap T_a = N \cup \{a\}$, $M = (M \cap T_0) \cup \{1\}$, можно получить следующее утверждение.

Теорема 6. 1) Для $a \in \{0, 1\}$ $Z(M \cap T_a)$ состоит из 12 классов предикатов:

$\{1\}, C, Q_{\overline{a}a} \cup \{1\}, M, \overline{M}, Q, Q_{00} \cup M, T_a \cup \{1\}, \overline{T_a}, T_a \cup \overline{M}, \overline{T_a} \cup C, Pr(2)$;

2) $Z(M)$ состоит из 6 классов предикатов: $\{1\}, C, M, \overline{M}, Q, Pr(2)$.

Теорема 6 вместе с Теоремой 1 дают описание $Int(M \cap T_a)$ и $Int(M)$.

Литература

- [1] Алексеев В. Б. О замкнутых классах в частичной k -значной логике, содержащих все полиномы // Дискретная математика. — 2021. — Т. 33, вып. 2. — С. 6–19.
- [2] Couceiro M., Haddad L., Schoelzel K., Waldhauser T. A solution to a problem of D. Lau: Complete classification of intervals in the lattice of partial Boolean clones // J. Mult.-Valued Logic Soft Comput. — 2017. — V. 28. — P. 47–58.
- [3] Alekseev V. B. On some intervals of partial clones // J. Mult.-Valued Logic Soft Comput. — 2022. — V. 38. — P. 3–22.

Свойства произведения Адамара констациклических линейных кодов

Линь Пэйфэн

Кафедра информационной безопасности

e-mail: s02180749@gse.cs.msu.ru

Научный руководитель — к.ф.-м.н. доц. Чижев Иван Владимирович

Введение. В работе описывается атака на криптосистему Мак-Элиса, построенную на основе констациклических линейных кодов, с помощью анализа их свойств произведения Адамара.

Криптосистема Мак-Элиса — один из кандидатов в постквантовые криптосистемы с открытым ключом. Оригинальная криптосистема Мак-Элиса строится на основе двоичных кодов Гоппы. Огромные размеры ключа препятствуют её широкому распространению, хотя до сих пор не существует эффективной атаки на эту систему.

Чтобы уменьшить размеры ключей, сохранив стойкости этой криптосистемы, были придуманы разные модификации, построенные на основе других кодов, но многие модификации уже были взломаны.

Криптосистема Мак-Элиса, построенная на основе констациклических линейных кодов. Пусть открытый ключ криптосистемы Мак-Элиса — это пара матриц (G, G') , где G — $(k \times n)$ порождающая матрица кода C , на котором криптосистема строится, и G' — $(k \times n)$ порождающая матрица кода C^σ . Ключевой шаг восстановления секретного ключа по открытому ключу (G, G') — это нахождение перестановки σ' такой, что $\sigma \cdot \sigma' \in \text{Aut}(C)$, где $\text{Aut}(C)$ — группа перестановочных автоморфизмов кода C . В работе предложен алгоритм нахождения перестановки σ' с полиномиальной сложностью для случая, когда криптосистема строится на основе констациклических линейных кодах с непересекающимися носителями. Далее, используя свойства произведения Адамара констациклических линейных кодов из статьи [1, 2], получено сведение случая, когда криптосистема строится на основе констациклического линейного кода C , имеющего полином-образец, отличный от единицы, к случаю, когда криптосистема строится на основе констациклического линейного кода $C^{\langle w(r(C), \ell_\alpha) \rangle}$, имеющего базис с непересекающимися носителями. Сведение выполнено за полиномиальное время с использованием возведения в степень Адамара кода C .

Для криптосистемы, построенной на основе кода $C^{\langle w(r(C), \ell_\alpha) \rangle}$, легко найти перестановку σ' . Описав группу автоморфизмов констациклического линейного кода $C^{\langle w(r(C), \ell_\alpha) \rangle}$, получим множество перестановок $M = (\sigma')^{-1} \cdot \text{Aut}(C^{\langle w(r(C), \ell_\alpha) \rangle})$.

Все эквивалентные секретные ключи криптосистемы, построенной на основе кода C , составляют множество

$$\{(H'', \sigma'') \mid \sigma'' \in M, H''GP_{\sigma''} = G'\}$$

Для восстановления секретного ключа нужно перебрать перестановки $\sigma'' \in M$ и решить систему линейных алгебраических уравнений $H''GP_{\sigma''} = G'$. Если СЛАУ совместима, то пара (H'', σ'') является эквивалентным секретным ключом, иначе продолжаем перебор перестановок. Мощность группы автоморфизмов кода $C^{\langle w(r(C), \ell_\alpha) \rangle}$ зависит от шаблона-образца $p(x)$ кода C , где $p(x) = \sum_{i=0}^{\frac{n}{v}-1} \alpha^i x^{v \cdot i}$. Если порядок ℓ_α элемента α больше чем n/v , то мощность равна $v!$, что в некоторых случаях является полиномом от длины входа. Например, это выполняется, если $v! < n^2$. Тогда для некоторых слабых ключей, удовлетворяющих условиям $\ell_\alpha > n/v$ и $v! < n^2$, можно за полиномиальное время найти все эквивалентные секретные ключи.

Результаты работы. В работе получены следующие результаты:

1. Описана группа автоморфизмов констациклических кодов C с непересекающимися носителями и дуальных к ним кодов C^\perp ;
2. Описано замыкание констациклических кодов C с непересекающимися носителями и дуальных к ним кодов C^\perp относительно операции произведения Адамара и операции взятия дуального кода;

3. Предложен алгоритм восстановления эквивалентного секретного ключа критосистемы Мак-Элиса, построенной на основе констациклических кодов с непересекающимися носителями с полиномиальной сложностью;
4. Описана атака на критосистему Мак-Элиса, построенную на основе констациклических кодов, последовательность размерностей Адамара которого не достигает n . Атака эффективна, если $\ell_\alpha > \frac{n}{v}$ и $v! < n^2$.

Заключение. Результаты исследования могут быть использованы в описании слабых открытых ключей и атаке на критосистему Мак-Элиса, построенную на основе констациклических линейных кодов, последовательность размерностей Адамара которых не достигает n .

В дальнейшем планируется найти слабые ключи, последовательность размерностей Адамара которых достигает n , и используя процедуру получения $RM^\sigma(1, m)$ из $RM^\sigma(r, m)$ из статьи [3], улучшить процесс построения кода \hat{C} с непересекающимися носителями, принадлежащего замыканию кода C .

Литература

- [1] Randriambololona H. On products and powers of linear codes under componentwise multiplication // Contemporary Mathematics. 2015. Т. 637, С. 3–78.
- [2] Falk B. H., Heninger N., Rudow M. Properties of constacyclic codes under the Schur product // Designs, Codes and Cryptography. 2020. Vol. 88, P. 993–1021
- [3] Бородин М. А., Чижов И. В. Эффективная атака на критосистему Мак-Элиса, построенную на основе кодов Рида–Маллера // Дискретная математика. 2014. Т. 26, № 1. С. 10–20.

Исследование встречаемости небезопасной сериализации программных объектов в клиентском коде веб-приложений

Миронов Денис Дмитриевич

*Кафедра информационной безопасности
e-mail: dmiroNov123@gmail.com*

Научный руководитель — м.н.с. Сигалов Даниил Алексеевич

Определение. Сериализация — процесс преобразования структуры данных или объекта (экземпляра класса) в формат, который может быть сохранен (на диске, в базе данных), передан (по сети, между процессами) и восстановлен обратно (десериализован).

Десериализация клиентских данных может быть опасна по причине того, что злоумышленник может отправить *HTTP*-запрос, содержащий такую структуру данных, при десериализации которой выполнится код непредусмотренный веб-приложением. При этом, нахождение таких недостатков автоматизированными средствами — до сих пор нерешенная проблема [1, 2].

В данной работе проводится исследование встречаемости случаев использования небезопасной десериализации при взаимодействии между клиентским

JavaScript-кодом и серверной стороной веб-приложения. Целями данного исследования являются: выявление шаблонов использования сериализованных объектов в клиентском коде веб-приложений, создание классов на основе найденных шаблонов, автоматизация нахождения данных классов, а также выявление «небезопасных» классов веб-приложений.

В ходе работы был разработан алгоритм, определяющий контекст нахождения сериализованного программного объекта внутри веб-ресурса. Инструментальное средство, реализованное с помощью данного алгоритма, было использовано на ресурсах 60000 веб-приложений из выборки *Alexa TOP*, полученных с помощью статического краулера. Было выявлено 12997 случаев нахождения сериализованного объекта внутри *JavaScript*-кода, что составляет 48% от всех найденных.

На следующем этапе исследования были проведены итеративный процесс ручного просмотра найденных примеров и разработка синтаксических шаблонов для кода, в котором были встречены сериализованные объекты. Также был реализован алгоритм классификатора, который находит заданные классы при помощи сопоставления вершин абстрактного синтаксического дерева кода с разработанными синтаксическими шаблонами.

По итогу классификации было выявлено 23 нетривиальных класса (представителями данных классов является более чем одно веб-приложение) и 37 тривиальных классов.

Финальным этапом исследования являлась проверка найденных классов на наличие уязвимостей. Для приложений из программы *Bug Bounty* эксплуатация проводилась непосредственно на самих приложениях. Для остальных приложений проводился поиск *OpenSource*-компонент, и, в случае нахождения таковых, возможность эксплуатации проверялась путем ручного анализа кода и атак на стендовые приложения, сделанные на основе данных компонент.

Были найдены следующие «небезопасные» классы:

- *Settings*—приложения, сделанные на основе *PHP*-фреймворка *OpenCart* и использующие плагин *So Listing Tabs*. Было создано стендовое приложение. В данном приложении был найден недостаток «десериализация недоверенных данных», а с помощью *gadget-chain* [3], реализуемого на основе структур данных поставляемых базовым фреймворком, удалось добиться выполнения произвольного кода. Данная уязвимость не была публично известной, поэтому отчет о ней был отправлен в компанию *Mitre*, и был получен уникальный идентификатор уязвимости *CVE-2022-24108*.
- *SimpleAjaxManagerWordPress*—приложения, сделанные на основе *CMS WordPress* и использующие плагин *Simple Ajax Manager*. Было создано стендовое приложение. В данном приложении был найден недостаток «десериализация недоверенных данных», а также *SQL*-инъекция, эксплуатируемая при помощи полей сериализованного объекта. Найденная *SQL*-инъекция не была публично известной, поэтому отчет о ней был отправлен в компанию *WPScan*, данная уязвимость была опубликована.

- *LoadmoreWordPress*—приложения, сделанные на основе *CMS WordPress*. Приложения данного класса были найдены в программе *Bug Bounty*. На всех найденных приложениях при помощи недостатка «*десериализация недоверенных данных*» получилось добиться выполнения произвольного кода. Отчет о данных уязвимостях был отправлен разработчикам данных приложений. Уязвимости были подтверждены разработчиками.

Литература

- [1] Сигалов Д. А., Хашаев А. А., Гамаюнов Д. Ю. Обнаружение серверных точек взаимодействия в веб-приложениях на основе анализа клиентского JavaScript-кода // Прикладная дискретная математика. 2021. № 53 С. 32–54
- [2] Раздобаров А. В, Петухов А. А, Гамаюнов Д. Ю. Проблемы обнаружения уязвимостей в современных веб-приложениях // Проблемы информационной безопасности. Компьютерные системы. 2021. № 4 С. 64–69
- [3] Esser S. Utilizing code reuse or return oriented programming in PHP applications // BlackHat USA. 2010. Vol. 69

Поиск информации о принимаемых сервером запросах в закомментированном клиентском коде веб-приложений

Назаров Дмитрий Ильич

Кафедра информационной безопасности

e-mail: nazarovi2011@gmail.com

Научный руководитель — м.н.с. Сигалов Даниил Алексеевич

Функциональность большинства веб-приложений за свой жизненный цикл претерпевает множество изменений. Неиспользуемая или устаревшая со временем часть кода может быть закомментирована вместо удаления. И, поскольку клиентский и серверный код исполняются изолированно друг от друга, то возможна такая ситуация, что часть клиентского кода окажется закомментированной, а при этом с сервера соответствующая функциональность удалена не будет. Поиску такой “забытой” функциональности и посвящена данная работа. Современные сканеры безопасности не имеют возможности обрабатывать закомментированные вызовы. Возникает вопрос, насколько они распространены на реальных сайтах, а соответственно, насколько будет полезен модуль для их извлечения. Ведь, если для закомментированных запросов найдутся активные соответствующие конечные точки на сервере, то их также можно исследовать на предмет уязвимостей.

В ходе исследования был разработан алгоритм, извлекающий закомментированный код из клиентской части веб-приложений. Созданный алгоритм не является самостоятельным инструментом. Для интеграции был выбран анализатор [1], основной задачей которого является поиск точек взаимодействия с сервером.

После успешного внедрения было проведено два эксперимента более чем на 1.000.000 страниц, относящихся к 24212 наиболее и 26072 наименее популярных реальных веб-приложений из списка Alexa Top 1 Million [2]. Первый эксперимент основан на полной выборке, а второй — только на наименее популярных сайтах, исходя из предположения о том, что они содержат больше самописного кода, а следовательно, повышается вероятность нахождения в них закоментированных вызовов.

Основной целью поиска были запросы, являющиеся уникальными для закоментированного кода. Также для каждого из найденных запросов было проверено наличие соответствующей рабочей конечной точки на сервере.

По результатам экспериментов была получена следующая статистика:

— Для полной выборки:

- Всего проанализировано 580056 страниц из базы данных, относящихся к 50291 веб-приложениям.
- AJAX-вызовы, встречающиеся только в закоментированном коде, оказались на 9948 странице ($\sim 1.72\%$ от общего количества) или в 1396 веб-приложении ($\sim 2.78\%$ от общего количества).
- Всего найдено 14395 закоментированных вызовов. Из них оказалось 3923 уникальных ($\sim 27.25\%$).
- Конечные точки на сервере имели 1536 вызовов ($\sim 39.15\%$ среди уникальных), относящиеся к 1232 ресурсам ($\sim 0.21\%$ от общего количества) или 739 различным веб-приложениям ($\sim 1.47\%$ от общего количества).

— Для наименее популярных сайтов:

- Всего проанализировано 568421 страниц из базы данных, относящихся к 26072 веб-приложениям.
- AJAX-вызовы, встречающиеся только в закоментированном коде, оказались на 7112 странице ($\sim 1.25\%$ от общего количества) или в 477 веб-приложении ($\sim 1.83\%$ от общего количества).
- Всего найдено 9202 закоментированных вызовов. Из них оказалось 1646 уникальных ($\sim 17.89\%$).
- Конечные точки на сервере имели 570 вызовов ($\sim 34.63\%$ среди уникальных), относящиеся к 500 ресурсам ($\sim 0.09\%$ от общего количества) или 282 различным веб-приложениям ($\sim 1.08\%$ от общего количества).

В ходе проделанной работы был создан модуль для извлечения закоментированных вызовов, который может быть использован в составе сканера веб-уязвимостей. А также было выявлено, что такие вызовы действительно встречаются на реальных сайтах. В данном случае были обнаружены на 1.47% среди исследованных веб-приложений.

Литература

- [1] Сигалов Д. А., Хашаев А. А., Гамаюнов Д. Ю. Обнаружение серверных точек взаимодействия в веб-приложениях на основе анализа клиентского JavaScript-кода // Прикладная дискретная математика. 2021. № 53 С. 32-54
- [2] Alexa Top 1 Million. — <http://s3.amazonaws.com/alexa-static/top-1m.csv.zip>

Исследование и программная реализация алгоритма шифрования Кукувиноса, основанного на матрицах Адамара с циркулянтным ядром

Петрова Наталия Михайловна

Кафедра информационной безопасности

e-mail: n.petrova.2607@gmail.com

Научный руководитель — к.ф.-м.н. доц. Применко Эдуард Андреевич

Научный консультант — Борисов Август Валерьевич

Данная работа представляет собой анализ статьи Х. Кукувиноса и Д. Симоса, которая была опубликована в 2013 году [1]. Авторы статьи предлагают алгоритмы шифрования и расшифрования на основе матриц Адамара с циркулянтными ядрами, которые являются классами комбинаторных конструкций:

Алгоритм шифрования

1. Запрос на получение открытого текста msg ;
2. Формирование секретного ключа $k \leftarrow (A, d)$, который представляет собой матрицу A и некоторую константу d ;
3. Передача секретного ключа k по каналу связи;
4. Формирование числового вектора m длины n из открытого текста msg с помощью таблицы ASCII;
5. Вычисление соответствующего шифртекста $\bar{c} = \bar{m}A + d\bar{e}_n$;
6. Передача вектора зашифрованного текста \bar{c} по каналу связи.

Алгоритм расшифрования

1. Запрос на получение вектора зашифрованного текста \bar{c} ;
2. Извлечение секретного ключа $k \rightarrow (A, d)$;
3. Расшифрование числового вектора $\bar{m} = 1/n \cdot (\bar{c} - d\bar{e}_n)A^T$;
4. Обратное преобразование вектора m в исходный открытый текст.

Определение 1. Матрицей Адамара называют квадратную матрицу A размера n , составленную из $+1$ и -1 , столбцы которой попарно ортогональны, и для которой справедливо соотношение

$$AA^T = nI_n,$$

где A^T — транспонированная матрица, а I_n — единичная матрица порядка n .

Трудность использования матриц Адамара произвольной формы в качестве секретного ключа шифрования заключается в поиске методов порождения матриц больших порядков и способов их хранения в памяти, поэтому было предложено использовать матрицы с циркулянтным ядром [2].

Определение 2. Матрица Адамара порядка $p + 1$ называется матрицей с циркулянтным ядром, если она может быть представлена следующим образом,

$$\begin{array}{c|c} 1 & 1 \dots 1 \\ \hline 1 & \\ \vdots & C \\ 1 & \end{array}$$

где $C = (c_{i,j})$ — циркулянтная матрица порядка p , для которой выполнено $c_{i,j} = c_{1,j-i+1(\text{mod } p)}$.

Иными словами, ядро $C = (c_{i,j})$ матрицы Адамара получается последовательным циклическим сдвигом вправо на один элемент каждой ее строки относительно предыдущей, начиная со второй. Именно это и является основным преимуществом использования таких матриц, взятым за основу в [1]: чтобы получить матрицу Адамара с циркулянтным ядром, необходимо и достаточно знать одну строку или один столбец ее циркулянтного ядра.

Анализ устойчивости к атакам. В работе проведен анализ трех криптографических атак, рассмотренных в [1]:

- атака «грубой силой»;
- атака с известным открытым текстом;
- атака с выбором открытого текста.

Было выяснено, что рассматриваемая криптосистема не устойчива к атакам с известным открытым текстом и с выбором открытого текста, приведены алгоритмы ее взлома и вычислена их сложность.

Сложность атаки с известным открытым текстом сопоставима со сложностью решения СЛАУ и равна $\mathcal{O}(n^3)$.

Алгоритмы атаки с выбором открытого текста были предложены для двух случаев:

- вектор открытого текста имеет бинарный вид — состоит из 0 и 1;
- вектор открытого текста представим значениями из диапазона $[0, 255]$.

Оба алгоритма обобщены для произвольных матриц Адамара.

Наибольший практический интерес представляет второй случай: с помощью вектора открытого текста $(1, 2, 4, 8, 16, 32, 64, 128, 0, \dots, 0)$ можно вычислить каждый столбец искомой матрицы, построив дерево значений глубины 8, где

каждый лист — одно из возможных значений шифртекста. Путь от корня к листу позволяет однозначным образом определить элементы столбцов матрицы Адамара. Сложность вычисления — сложность поиска по дереву $O(\log(t))$, где t — число вершин.

Полученные выводы. На основании проведенных исследований была поставлена под сомнение идея использования матриц Адамара (не только с циркулянтным ядром) в качестве секретного ключа в детерминированных симметричных алгоритмах. В работе данная гипотеза подтверждена проверкой системы на IND-CPA стойкость [3].

Результаты работы опровергают некоторые утверждения авторов статьи [1], тем самым показывая невозможность применения предложенных алгоритмов для безопасного шифрования информации.

Литература

- [1] Koukouvinos C., Simos D.E. «Encryption Schemes based on Hadamard Matrices with Circulant Cores» // Journal of Applied Mathematics & Bioinformatics, vol.3, no.1, 2013, 17-41.
- [2] Kotsireas I.S., Koukouvinos C., Seberry J. «Hadamard ideals and Hadamard matrices with circulant core» // J. Combin. Math. Combin. Comput., no.57, 2006, 47–63.
- [3] Katz J., Lindell Y. «Introduction to modern cryptography», second edition, chapter 3, 2014.

Разработка архитектуры и прототипа цифрового двойника для умных шахт

Тарасова Валерия Андреевна

*Кафедра информационной безопасности
e-mail: tarasova.valeria.2000@gmail.com*

Научный руководитель — д.т.н. проф. Сухомлин Владимир Александрович

Научный консультант — аспирант Гапанович Дмитрий Антонович

Для современной экономики, вступающей в эпоху Индустрии 4.0, характерной тенденцией является массовый переход промышленных предприятий к технологиям интеллектуального производства, основанного на многомерном высокоточном моделировании свойств и процессов функционирования активов и использовании интеллектуальных методов принятия решений для оптимизации и расширения производства.

Цифровой двойник. Согласно [1], цифровой двойник - это высококачественная виртуальная копия физического актива с двусторонней связью в реальном времени для целей моделирования и поддержки принятия решений, способствующих улучшению обслуживания актива.

Система стандартов. Стандартизация технологии цифровых двойников, как и других технологий Индустрии 4.0, является фундаментальным аспектом, от которого зависит успешное развитие цифровой экономики, так как только

на этом пути можно достичь функциональной совместимости или интероперабельности множества цифровых и физических решений, которая должна рассматриваться в глобальном масштабе. В ВКР проведён анализ существующей системы стандартов в области цифровых двойников производства и выделена архитектура, на основе которой построена собственная архитектура, предложенная в данной работе.

Интегрированный (полный) цифровой двойник. В исследованиях цифровых двойников вводится понятие интегрированного (или полного) цифрового двойника, в котором объединяются все приложения цифровых двойников в единую архитектуру. Предложенная архитектура образует модульную конструкцию, которая характеризуется способностью к масштабированию и обладает повышенной надёжностью. В работе [2] представлена типовая высокоуровневая архитектура интегрированного цифрового двойника, взятая за основу в данной ВКР.

Архитектура цифрового двойника для умных шахт. В качестве области, для которой построена архитектура и программное решение цифрового двойника в представленной работе, выбрана горнодобывающая отрасль. Шахты активно используют цифровые технологии, в частности, цифровых двойников, как в нашей стране, так и за ее пределами. Но решения, разработанные компаниями, зачастую являются проприетарными, не опирающимися на существующие стандарты. В ВКР определённая выше интегрированная архитектура цифрового двойника расширена уровнем приложений и сервисов, предназначенным для настройки системы на конкретные цели и конечных пользователей. Благодаря интеграции приложений такой цифровой двойник обладает функциональной полнотой.

На рис. 1 представлена архитектура приложений и сервисов, предложенная в работе в качестве расширения архитектуры интегрированного цифрового двойника.

Реализация прототипа цифрового двойника для умных шахт. Для реализации прототипа цифрового двойника с предложенной архитектурой выбран язык C# и среда разработки Visual Studio. В рамках работы был проведён эксперимент, подтверждающий работоспособность прототипа.

Литература

- [1] Lim, K. Y. H., Zheng, P., Chen, C.-H. (2019). A state-of-the-art survey of digital twin: techniques, engineering product lifecycle management and business innovation perspectives [Journal Article]. Journal of Intelligent Manufacturing, 1-25.
- [2] Lu, Q., Parlikad, A., Woodall, P., Don Ranasinghe, G., & Heaton, J. (2019). Developing a dynamic digital twin at a building level using cambridge campus as a case study [Journal Article].

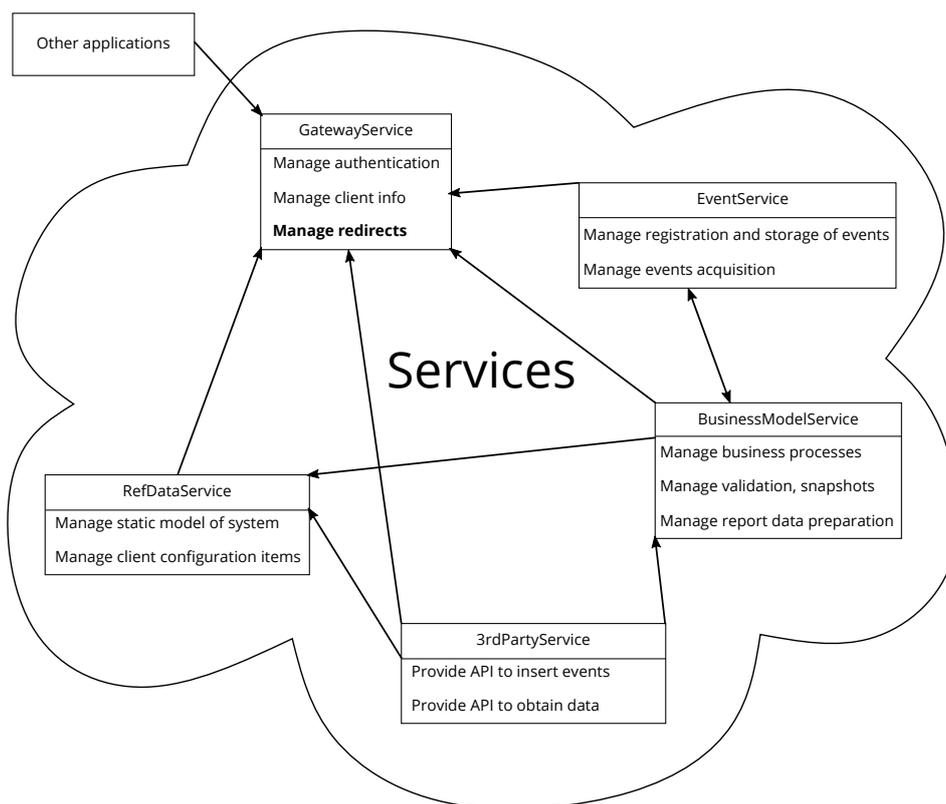


Рис. 1: Архитектура приложений и сервисов цифрового двойника

Полуавтоматическая классификация транзиентов в рентгеновском обзоре неба

Губанов Федор Дмитриевич

Кафедра интеллектуальных информационных технологий

e-mail: fgubanov1533@gmail.com

Научные руководители — к.ф.-м.н. с.н.с. Мещеряков Александр Валерьевич, инженер Герасимов Сергей Валерьевич

С 2019 года в рамках Федеральной космической программы России совместно с Германией проводится уникальный по чувствительности обзор Вселенной в рентгеновском диапазоне, состоящий, на данный момент, из 5 сканов всего неба [1]. Эксперты ИКИ РАН занимаются поиском уникальных объектов среди рентгеновских источников «СРГ/еРОЗИТА» [2], вспыхнувших только в одном из сканов (т.н. транзиентов). В процессе анализа данных выполняется предварительная классификация транзиентов путем ручного сопоставления данных из многоволновых обзоров неба. Ручная разметка транзиентов, в настоящее время, позволяет классифицировать только малую часть (несколько сотен объектов) из десятков тысяч потенциально интересных событий. Целью данной работы было исследование и разработка системы для полуавтоматической классификации рентгеновских транзиентов «СРГ/еРОЗИТА» с использованием многоволновых данных обзоров неба. Данная система позволит значительно

повысить эффективность работы экспертов и увеличить объем анализируемых данных.

Система должна соответствовать определенным требованиям: надежность хранения данных исследования, а также результатов анализа данных экспертами, соблюдение проприетарности данных исследования «СРГ/еРОЗИТА», возможность загрузки данных определенного формата в систему и сохранение взаимосвязей в данных, соответствие интерфейса требованиям экспертов. В результате обзора инструментов разметки данных [3, 4, 5, 6] было принято решение создать собственное решение, которое будет полностью отвечать вышеописанным требованиям. В процессе работы мы столкнулись со сложностями в создании схемы базы данных отражающей связи между анализируемыми источниками, а также в визуализации мультидоменных данных рентгеновских и оптических обзоров. Прототип системы разрабатывался на базе веб-фреймворка Django и СУБД SQLite, в итоговой версии используется более мощная СУБД PostgreSQL. Для автоматизации классификации используются предопределенные критерии, которые отражают знания экспертов о классах транзиентов «СРГ/еРОЗИТА». На основе этих критериев осуществляется предварительная разметка объектов.

Исследована и разработана система для полуавтоматической классификации рентгеновских транзиентов «СРГ/еРОЗИТА», которая отвечает описанным ранее требованиям. Система запущена в ИКИ РАН и протестирована экспертами. На данный момент система используется экспертами, занимающимися исследованием рентгеновских транзиентов «СРГ/еРОЗИТА». В дальнейшем планируется интеграция моделей машинного обучения в целях усовершенствования автоматической разметки.

Литература

- [1] Predehl P. et al. The eROSITA X-ray telescope on SRG //Astronomy Astrophysics. – 2021. – Т. 647. – С. A1.
- [2] Sazonov S. et al. First tidal disruption events discovered by SRG/eROSITA: X-ray/optical properties and X-ray luminosity function at $z < 0.6$ //Monthly Notices of the Royal Astronomical Society. – 2021. – Т. 508. – №. 3. – С. 3820-3847.
- [3] Amazon SageMaker Ground Truth to Label Data [Электронный ресурс]. — Электрон. дан. — [Б. м.] : 2021. — Режим доступа: <https://docs.aws.amazon.com/sagemaker/latest/dg/sms.html>.
- [4] Label Studio/Documentation [Электронный ресурс]. — Электрон. дан. — [Б. м.] : 2021. — Режим доступа: <https://labelstud.io/guide/>.
- [5] Stevens G. et al. AstronomicAL: an interactive dashboard for visualization, integration and classification of data with Active Learning //Journal of Open Source Software. – 2021. – Т. 6. – №. 65. – С. 3635.

- [6] Computer Vision Annotation Tool [Электронный ресурс]. — Электрон. дан. — [Б. м.] : 2021. — Режим доступа: https://en.wikipedia.org/wiki/Computer_Vision_Annotation_Tool.

Разработка алгоритма детектирования смен сцен в видео на основе нескольких метрик

Гуцин Александр Евгеньевич

Кафедра интеллектуальных информационных технологий

e-mail: s02180408@gse.cs.msu.ru

Научный руководитель — к.ф.-м.н. с.н.с. Ватолин Дмитрий Сергеевич

Объем видеоконтента, составляющего большую часть интернет-трафика, стремительно растет. Перед отправкой конечному пользователю видеоконтент проходит целую серию шагов предобработки. Одним из важных этапов обработки видео является разбиение его на сцены – интервалы видео, снятые одной камерой без монтажных склеек. Этот метод обработки также широко применяется для сжатия видео в большинстве современных видеокодеков. В данной работе описан метод детектирования смен сцен в видео.

Предложенный метод состоит из двух основных этапов – оценки близости пары соседних кадров видео и классификация данной пары. На этапе классификации метод определяет наличие между данной парой соседних кадров смены сцен.

Для первого этапа были использованы различные характеристики изображений. В предложенный алгоритм входят 5 методов оценки близости кадров, которые используют границы объектов в изображениях, гистограммы значений в различных цветовых пространствах, насыщенность и яркость. Данные методы используют некоторые техники предобработки изображений, такие как деление на непересекающиеся блоки, применение фильтра Собеля и другие. Также был использован метод Bhattacharyya [1] оценки близости гистограмм.

Для второго этапа (классификации) были применены различные методы машинного обучения, в число которых входят линейная классификация, деревья решений, метод k-ближайших соседей. После серии экспериментов в предложенный алгоритм был включен метод классификации на основе дерева решений.

Для обучения предложенного метода был собран и размечен новый набор данных под названием OS VSD, превосходящий по длине существующие аналоги. Разметка проводилась с помощью платформы Яндекс.Толока [2] с привлечением более 1000 субъектов. Сравнение предложенного набора данных с аналогами показано в таблице 1.

Для экспериментальной оценки предложенный метод был обучен на наборе данных OS VSD и протестирован на наборах данных BBC Planet Earth и RAI. Методом оценки качества служила F-мера (1):

$$F\text{-мера} = \frac{tp}{tp + \frac{1}{2} * (fp + fn)} \quad (1)$$

Набор данных	Длина	Количество смен сцен
RAI	93	985
BBC Planet Earth	539	4844
TRECVID 2007	360	2371
OS VSD (предложенный)	954	6871

Таблица 1: Сравнение наборов данных с аналогами.

tp - количество верно распознанных смен сцен

fp - количество кадров, ложно распознанных как смена сцен

fn - количество пропущенных алгоритмом смен сцен

По результатам экспериментальной оценки предложенный метод превзошел существующие аналоги по F-мере.

Литература

- [1] Bhattacharyya A. On a measure of divergence of two multinomial populations // Sankhyā Indian J Stat. 1945. Vol. 7.
- [2] Яндекс.Толока : [сайт]. URL: <https://toloka.yandex.ru/>

Исследование и разработка методов автоматической аннотации электронной почты

Сивцов Данил Тимофеевич

Кафедра интеллектуальных информационных технологий

e-mail: sivtsovdt@gmail.com

Научный руководитель — к.ф.-м.н. математик Царёв Дмитрий Владимирович

В рамках настоящей работы рассматривалась задача автоматического построения аннотаций переписок электронной почты. Требовалось, чтобы разрабатываемые алгоритмы позволяли отображать лишь краткую аннотацию (длиной в 100 слов) всей переписки, при этом делали это с учетом персональных предпочтений пользователя и поддерживали работу сразу с несколькими языками.

Решение рассматриваемой задачи разбивается на две подзадачи — выбор модели представления предложений текста, исследование и разработка алгоритма ранжирования предложений.

Обычно в задачах аналогичных данной применяются простые модели представления текста, такие как TF—IDF и мешок слов, и модели сопоставления ранга предложению на основе матричных алгоритмов, например SVD и NMF разложений. Такие классические методы ранжирования формируют скрытое пространство тематик, а ранг вычисляется по весам тематик в представлении предложений.

В работе исследована возможность использования нейронных сетей и комбинаций их с классическими алгоритмами с целью улучшения качества аннотирования и персонализации. Для получения модели представления текста переписок сравнивались алгоритмы частотной оценки встречаемости слов и

N-грамм, а также алгоритмы на основе архитектуры трансформера — различных вариаций BERT [1], обученных на мультязычных корпусах текстов. Для ранжирования предложений по важности было предложено два алгоритма. Первый полагается только на свойства модели BERT и в качестве ранга вычисляет обратное значение нормы разности векторов представления предложения и всей переписки. Второй алгоритм использует для ранжирования предложений отдельную модель — автоэнкодер архитектуры Variational autoencoder [2]. Ранг предложения в таком подходе рассчитывается как отрицание ошибки восстановления модели. Данная модель основана на аналогичной с классическими подходами идеи с получением скрытого представления, что происходит за счет снижения размерности в центральных слоях нейронной сети.

В проведённых экспериментальных исследованиях использовались два актуальных на сегодняшний день корпуса текстов: набор аннотированных почтовых переписок ВСЗ, а также набор аннотированных статей CNN/Daily Mail. Качество исследуемых алгоритмов оценивалось по метрикам семейства ROUGE [3], вычисляющих оценку по перекрытиям некоторых объектов (например N-грамм, слов или пар слов) в модельной и полученной алгоритмом аннотациях. Первый предложенный алгоритм на основе разности представлений смог превзойти классические алгоритмы, второй алгоритм на основе BERT и Variational autoencoder показал лучшее качество среди всех сравниваемых алгоритмов.

Литература

- [1] J. Devlin, M.—W. Chang, K. Lee, and K. Toutanova. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. // Association for Computational Linguistics, 2019.
- [2] Kingma D. P., Welling M. Auto-encoding variational bayes // arXiv preprint arXiv:1312.6114.—2013.
- [3] Ch.—Y. Lin. ROUGE: A Package for Automatic Evaluation of Summaries. // Text Summarization Branches Out, pages 74–81, Barcelona, Spain. Association for Computational Linguistics.

Построение однопроцессорного расписания с минимизацией пикового потребления ресурса при помощи муравьиного алгоритма

Абрамов Алексей Владимирович

Кафедра автоматизации систем вычислительных комплексов

e-mail: avabramov@lvk.cs.msu.ru

Научный руководитель — математик Чупахин Андрей Андреевич

В данной работе рассматривается проблема построения оптимального расписания не с точки зрения уменьшения времени его выполнения, а с точки зрения уменьшения пикового потребления памяти. Вычислительная среда (ВС) состоит из одного процессора, способного в каждый момент времени обрабатывать

лишь одно задание, и некоторого объёма памяти. Входными данными является ориентированный ациклический граф заданий, в котором узлы представляют собой задания, а рёбра — зависимости по данным между ними. Результат выполнения каждого задания хранится в вычислительной среде, то есть занимает определённое количество памяти, до тех пор, пока не выполнятся все непосредственные потомки данного задания (то есть, вершины в графе, которые связаны ребром с текущей вершиной). Необходимо построить расписание, минимизирующее пиковое потребление ресурса.

В [1] показано, что в общем случае рассматриваемая в выпускной работе задача является NP-трудной, но алгоритмы, решающие задачу в общем случае, в литературе не рассматривались. Были рассмотрены лишь частные случаи данной задачи, в которых ориентированный ациклический граф заданий имеет определенную структуру: является деревом [1, 2], представляет собой множество параллельных цепей [3], является последовательно-параллельным графом [4]. Для каждого из частных случаев были предложены полиномиальные алгоритмы со сложностями $O(n \log_2 n)$, $O(n \log_2 s)$, $O(n^2)$ соответственно (здесь: n — количество вершин в графе, s — количество параллельных цепей).

В выпускной работе проведён обзор алгоритмов построения расписаний, приведено описание алгоритмов, в числе которых жадный алгоритм, алгоритм имитации отжига, генетический алгоритм и муравьиный алгоритм. Разработан муравьиный алгоритм для решения задачи минимизации пикового потребления памяти, рассмотрены различные модификации алгоритма, в числе которых локальный поиск и подход Ant-Q, основанный на обучении с подкреплением. Показано, что решения, полученные разработанным муравьиным алгоритмом, в среднем лишь в 22% случаев отклоняются от оптимальных решений. Также проведено сравнение разработанного муравьиного алгоритма и его модификаций с алгоритмом имитации отжига, в результате которого выяснилось, что муравьиный алгоритм обладает большей устойчивостью и продолжительностью работы, а также в среднем в 73% случаев находит решения лучше решений, полученных при помощи алгоритма имитации отжига. На рис. 1 изображено сравнение двух разработанных алгоритмов: муравьиного алгоритма и алгоритма имитации отжига, а на рис. 2 сравнение решений муравьиного алгоритма с оптимальными решениями для графов, на которых оптимум достигнут не был. Реализации алгоритмов и результаты проведенных экспериментальных исследований расположены в репозитории по ссылке [5].

Литература

- [1] R. Sethi, Complete Register Allocation Problems // SIAM Journal on Computing. 1975. Vol. 4. № 3. P. 226–248.
- [2] R. R. Redziejowski, On Arithmetic Expressions and Trees // Communications of the ACM. 1969. Vol. 12. № 2. P. 81–84.
- [3] P. G. Hebalkar, Coordinated Sharing of Resources in Asynchronous Systems // Record of the Project MAC conference on concurrent systems and parallel computation, Massachusetts, USA, 1970, P. 151–168.

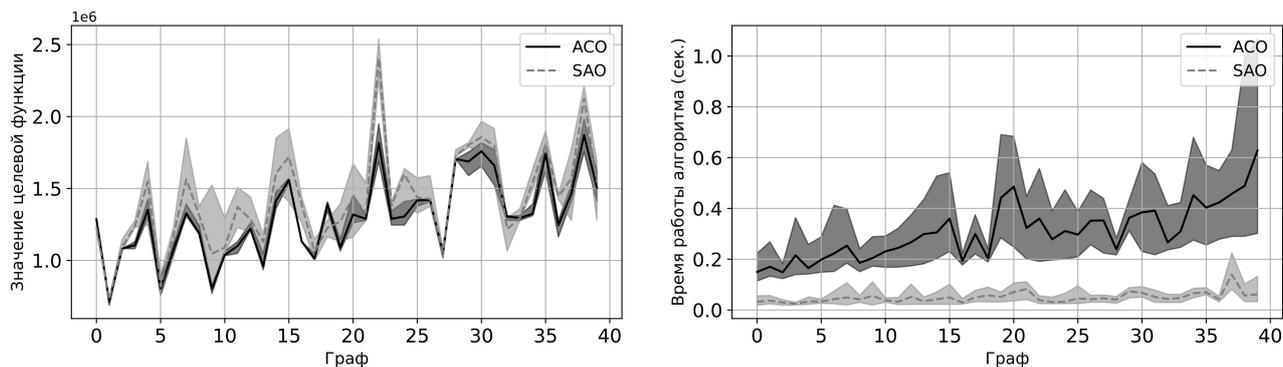


Рис. 1: Слева изображена зависимость целевой функции, а справа зависимость времени работы алгоритма от размерности задачи для алгоритмов имитации отжига (SAO) и муравьиного алгоритма (ACO).

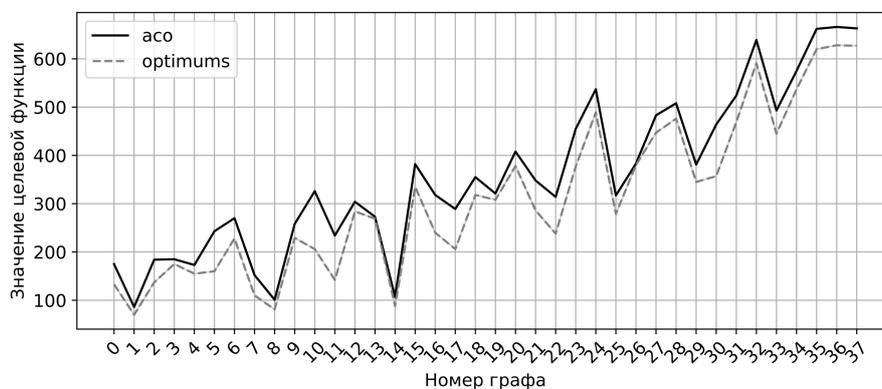


Рис. 2: Сравнение решений АСО с оптимальными решениями.

[4] H. Abdel-wahab and T. Kameda, Scheduling to minimize maximum cumulative cost subject to series-parallel precedence constraints // Operations Research. 1978. Vol. 26. № 1. P. 141–158.

[5] <https://github.com/Crotokot/SchedulingProblem>

Разработка прототипа системы масштабирования сетевых функций для MANO (Management and Orchestration) платформы на основе легковесной виртуализации

Ветрова Екатерина Александровна

Кафедра автоматизации систем вычислительных комплексов

e-mail: katika.belka@yandex.ru

Научный руководитель — к.ф.-м.н. доц. Антоненко Виталий Александрович

Научный консультант — Плакунов Артём Владимирович

Данная работа посвящена проблеме масштабирования сетевых функций, размещённых в кластере легковесной виртуализации.

Сетевая функция — компонент сети, который отвечает за специальную обработку сетевых пакетов. Сетевая функция может действовать на различных уровнях стека протоколов TCP/IP. В качестве примеров сетевых функций мож-

но назвать функции маршрутизации трафика, проверки трафика на наличие вредоносных данных, защиты информации.

Сетевые функции могут быть предоставлены пользователю сети в качестве сервисов. Под сервисом понимают некоторую услугу или комплекс услуг в сфере ИТ, которые предоставлены компанией-поставщиком клиенту.

Чтобы преобразовать трафик нужным образом, могут потребоваться сразу несколько сетевых функций, причём трафик обычно должен проходить через них в заданном порядке. Процесс передачи трафика обеспечивает цепочка сервисных функций (service function chain). При построении цепочки каждая сетевая функция помещается в контейнер, и контейнеры связываются между собой по сети.

В случае, если где-то в цепочке образуется "бутылочное горлышко", может потребоваться увеличить или уменьшить количество ресурсов, отведённых под конкретные сетевые функции, чтобы поддерживать работу сервиса с нужной производительностью. Подход, при котором ресурсы динамически добавляются или удаляются для поддержки производительности сервиса без участия человека, называется автомасштабированием.

Проектирование эффективной системы масштабирования включает в себя решение комплекса проблем, связанных с мониторингом системы, принятием решений на основе полученных от мониторинга данных, выделением и освобождением ресурсов.

Работу сервисов, предоставляемых пользователю, необходимо контролировать. Платформа, которая позволяет создавать сервисы, настраивать их, осуществлять мониторинг и восстановление нормальной работы, называется MANO платформой (Management and Orchestration).

В данной работе был разработан и реализован прототип системы масштабирования сервисов на основе легковесной виртуализации. Данную разработку впоследствии предполагается использовать для организации масштабирования на основе легковесной виртуализации в MANO платформе.

В качестве средства легковесной виртуализации была выбрана система оркестровки контейнеров Kubernetes, так как её встроенные механизмы масштабирования удобны для использования в контексте задачи ВКР.

Для принятия решения об архитектуре будущей системы, были рассмотрены существующие системы масштабирования сервисов на основе легковесной виртуализации [1, 2, 3, 4]. По итогам обзора были определены характеристики системы: масштабирование осуществляется горизонтально; система масштабируется, ориентируясь на сетевые метрики (метрики, свидетельствующие о состоянии предоставляемой клиенту сети), также есть поддержка масштабирования на основании ресурсных метрик (загрузка ЦП, расход памяти); решение о масштабировании принимается по результатам мониторинга, без предсказания заранее.

В качестве основного механизма масштабирования был использован механизм горизонтального масштабирования Kubernetes. Для мониторинга была выбрана система Prometheus для сетевых метрик и cAdvisor для ресурсных.

Было проведено экспериментальное исследование прототипа, которое подтвердило, что система корректно реагирует на изменение нагрузки: автоматически изменяет количество экземпляров сетевой функции до нужного количества. Также было продемонстрировано, что система способна обеспечить равномерную загрузку кластера в течение всей работы системы, не допускает резких колебаний количества экземпляров сетевой функции (т.е. работает устойчиво) и поддерживает стабильное состояние сети при масштабировании сетевых функций.

Литература

- [1] T. Lin, A. Leon-Garcia. Towards a Client-Centric QoS Auto-Scaling System // IEEE/IFIP Network Operations and Management Symposium. 2020. P. 1–9.
- [2] Machine Learning-Based Scaling Management for Kubernetes Edge Clusters / L. Toka, G. Dobreff, B. Fodor and B. Sonkoly // IEEE Transactions on Network and Service Management, Vol. 18, N 1. 2021. P. 958–972.
- [3] D. Balla, C. Simon, M. Maliosz. Adaptive scaling of Kubernetes pods // 2020 IEEE/IFIP Network Operations and Management Symposium. 2020. P. 1–5.
- [4] A. Khaleq, R. Ilkyeun. Intelligent Autoscaling of Microservices in the Cloud for Real-Time Applications // IEEE Access. 2021. P. 1–1.

Алгоритм балансировки нагрузки в распределенной вычислительной системе с применением мультиагентного обучения с подкреплением

Казантаев Алексей Дмитриевич

Кафедра автоматизации систем вычислительных комплексов

e-mail: adkazantaev@lvk.cs.msu.ru

Научный руководитель — математик Чупахин Андрей Андреевич

Балансировка рабочей нагрузки в неоднородных распределенных вычислительных системах (РВС), состоящих из вычислительных устройств (ВУ) различной производительности, является одной из актуальных задач оптимального управления. В неоднородную РВС на ВУ по заранее заданному распределению (Пуассона, равномерное, Эрланга) поступают вычислительные задания с различным временем выполнения и различными требованиями к ресурсам (количество ядер центрального процессора, объем оперативной памяти), в результате ВУ нагружаются неравномерно. Это влечет за собой увеличение времени обработки заданий, и как следствие, снижение производительности всей РВС. Чтобы решить эту проблему, необходимо балансировать входящий поток заданий между устройствами.

В данной работе рассматривается многокритериальная оптимизационная задача балансировки нагрузки в неоднородной РВС. Предполагается, что из-

вестны время выполнения и пересылки заданий между ВУ, распределение времени поступления заданий, а так же подмножество периферийных ВУ, на которые задания поступают извне. Задания независимы по данным и выполняются без прерываний. Требуется разработать алгоритм мультиагентного обучения с подкреплением для равномерного распределения заданий по всем ВУ неоднородной РВС.

Задача балансировки нагрузки в неоднородной вычислительной системе является NP-трудной задачей, что подразумевает использование эвристических алгоритмов для её решения [1-3].

В выпускной работе сначала была сформулирована многокритериальная задача оптимизации: минимизация для каждого типа ресурса (количество ядер центрального процессора, объем оперативной памяти). Далее многокритериальная задача была сведена к однокритериальной и оптимизирующий функционал приобрел вид (1).

$$\mathfrak{F} = \min \sum_1^{|V|} \alpha \frac{\bar{c}_i}{c_i} + \beta \frac{\bar{s}_i}{s_i} + \gamma \left(\left(\frac{\bar{c}_i}{c_i} - \Theta \right)^2 + \left(\frac{\bar{s}_i}{s_i} - \Delta \right)^2 \right) \quad (1)$$

Неоднородная РВС рассматривалась как мультиагентная система, где каждому ВУ сопоставляется обучающий агент. Исходя из текущей загруженности и информации о загруженности других агентов, агенты принимают решение: выполнять задание локально или отправить его другому агенту. Информация о загруженности других агентов передается между агентами периодически, поэтому каждый агент обучается независимо от других агентов.

Была разработана мультиагентная система обучения с подкреплением с использованием алгоритма *PPO (Proximal Policy Optimization)* [4].

Реализация программного модуля на языке *Python 3.8* включает в себя подмодули генерации данных, среды *OpenAI Gym*, подмодули запуска и визуализации результатов экспериментов [5].

Результатом выпускной квалификационной работы является реализованный алгоритм мультиагентного обучения с подкреплением и его программный модуль, решающий поставленную задачу, а также исследование эффективности предложенного алгоритма. Результаты экспериментов (рис. 1) демонстрируют устойчивость алгоритма относительно различных изменений в конфигурации неоднородной РВС (общее число агентов и количество периферийных узлов) и исходных данных (число шагов итерации, распределение заданий).

Литература

- [1] Milan S. T. et al. Nature inspired meta-heuristic algorithms for solving the load-balancing problem in cloud environments //Computers & Operations Research. – 2019. – Т. 110. – С. 159-187.
- [2] Wang Y. et al. Multi-objective workflow scheduling with deep-Q-network-based multi-agent reinforcement learning //IEEE access. – 2019. – Т. 7. – С. 39974-39982.

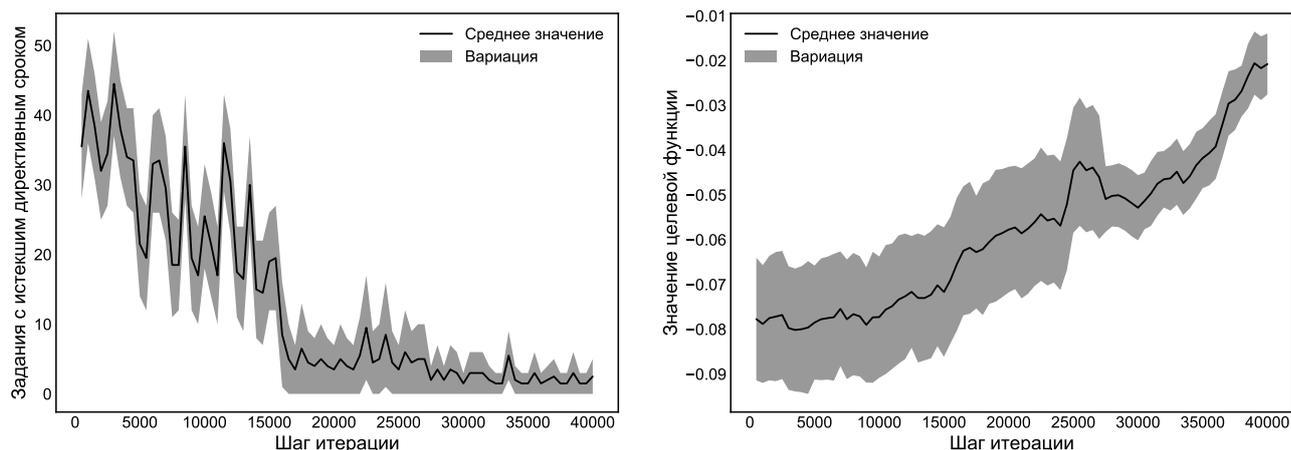


Рис. 1: Графики вариации числа невыполненных в директивный срок заданий и целевой функции при 15 агентах

- [3] Wu J., Xu X. Decentralised grid scheduling approach based on multi-agent reinforcement learning and gossip mechanism //CAAI Transactions on Intelligence Technology. – 2018. – Т. 3. – №. 1. – С. 8-17.
- [4] Schulman J. et al. Proximal policy optimization algorithms //arXiv preprint arXiv:1707.06347. – 2017.
- [5] MARL Load Balancing Environment [Электронный ресурс]. URL: <https://github.com/alexey-kaz/customenv> (дата обращения: 10.06.2022)

Разработка методов трехмерной визуализации задержек в топологической структуре вычислительных кластеров

Коваленко Анастасия Павловна

Кафедра автоматизации систем вычислительных комплексов

e-mail: nastenka29@gmail.com

Научный руководитель — к.ф.-м.н. в.н.с. Сальников Алексей Николаевич

При решении той или иной задачи на вычислительном кластере узлы взаимодействуют через высокоскоростную коммуникационную среду. Однако даже в такой сети присутствуют задержки при передаче данных. Анализ этих задержек крайне важен для более полного понимания процессов, происходящих в кластере, и для выявления возможных ошибок и проблем.

Выпускная квалификационная работа использует в качестве входных данных сведения о задержках в сети, полученные с помощью программы network-tests2, принципы работы которой описаны в статье [1]. Полученные данные можно преобразовать таким образом, чтобы получить матрицу задержек, с помощью которой в приложении Network Viewer 2 производится анализ топологии и визуализация.

Цель выпускной квалификационной работы. Целью работы является усовершенствование метода многомерного шкалирования и его интеграция в приложение Network Viewer 2.

Исследование и построение решения задачи. Одна из основополагающих работ в изучении визуализации графов — статья [2], в которой описывается пружинная модель визуализации ненаправленных графов. Вершины графа рассматриваются как тела, соединенные пружинами. Идея метода — минимизация энергии этой системы тел, в точке минимума функции энергии достигается наилучшая укладка графа. Функция энергии (функция стресса):

$$E = \sum_{i < j}^n \omega_{i,j} (\|x_i - x_j\| - d_{i,j})^2$$

Здесь x_i, x_j — векторы координат вершин i и j соответственно, $\|x_i - x_j\|$ — евклидово расстояние между вершинами i и j , $d_{i,j}$ — «идеальное» заданное расстояние между вершинами. Значение $\omega_{i,j}$ моделирует коэффициент упругости пружины и играет роль нормировочной константы.

Для улучшения пружинного метода был предложен подход, который получил название мажоризация функции стресса (stress majorization). Он заключается в том, что функцию стресса можно оценить сверху выпуклой скалярной функцией $F^Z(X)$ (подробное обоснование приведено в [3]):

$$F^Z(X) = \sum_{i < j}^n \omega_{i,j} d_{i,j}^2 + \text{Tr}(X^T L^\omega X) - 2\text{Tr}(X^T L^Z Z)$$

где Z — произвольная константная матрица размерности $n \times k$, n — число вершин в графе, k — размерность исходного пространства, $\text{tr}(A)$ — след матрицы A , матрица L^ω задана как

$$L_{i,j}^\omega = \begin{cases} -\omega_{i,j}, & i \neq j \\ \sum_{m \neq i}^n \omega_{i,m}, & i = j \end{cases}$$

В итоге для координат вершин, записанных в виде матрицы X , получаем оценку: $\text{stress}(X) \leq F^Z(X)$. Равенство достигается при $Z = X$. Функция $F^Z(X)$ выпуклая, а значит, имеет единственный оптимум и ее минимизация гарантирует быструю сходимость [3].

В работе реализован итерационный алгоритм мажоризации функции стресса, на каждой итерации гарантируется уменьшение значения функции стресса.

Результаты. Результатом выпускной квалификационной работы является усовершенствованное приложение для визуализации топологий коммуникационной среды Network Viewer 2. Это приложение написано на языке C++ с использованием фреймворка Qt и графической библиотеки Qwt.

В рамках работы были рассмотрены существующие методы визуализации графов, а также методы повышения наглядности полученных изображений.

Имеющееся приложение Network Viewer 2 было переписано с устаревшего фреймворка Qt4 на более современную версию Qt5 с использованием библиотеки Qwt6. Пружинный метод многомерного шкалирования был усовершенствован с использованием мажоризации функции стресса. Итоговый

алгоритм работает значительно быстрее исходного (количество необходимых итераций уменьшилось на два порядка) и снижает погрешность построения длин ребер в графе.

Была проведена серия экспериментов с данными задержек, измеренных на суперкомпьютерах JUROPA, Ломоносов, Ломоносов-2. На основании результатов этих экспериментов был проведен сравнительный анализ исходного и полученного алгоритмов.

Итоговое приложение может быть использовано для визуализации задержек в коммуникационной среде кластера, выявления и анализа возможных ошибок и проблем в сети.

Литература

- [1] Банников П. С., Сальников А. Н. Выявление топологии коммуникационной среды вычислительного кластера по результатам нагрузочного тестирования // Вестник Московского университета. Серия 15. Вычислительная математика и кибернетика. 2014. С. 43–52.
- [2] Kamada T., Kawai S. An algorithm for drawing general undirected graphs // Information Processing Letters. 1989. Vol. 31. P. 7–15.
- [3] Gansner E. R., Koren Y., North S. Graph Drawing by Stress Majorization // International Symposium on Graph Drawing. 2004. P. 239–250.

Исследование метода прогнозирования нагрузки контроллера в программно-конфигурируемых сетях на основе нейронных сетей

Ларин Андрей Викторович

Кафедра автоматизации систем вычислительных комплексов

e-mail: s02180464@gse.cs.msu.ru

Научный руководитель — асс. Пашков Василий Николаевич

Ключевым элементом управления в программно-конфигурируемых сетях (ПКС) [1] является контроллер — специальная сетевая операционная система с набором приложений, осуществляющих логически централизованное управление настройками сетевого оборудования, маршрутизацией потоков пользовательских данных на основе глобального состояния сети и в соответствии с установленными сетевыми политиками. Контроллер в режиме реального времени обрабатывает запросы от ПКС коммутаторов и в ответ изменяет их конфигурацию.

При этом контроллер является единой точкой отказа ПКС сети [2]. Резкое увеличение количества запросов к контроллеру от коммутаторов может привести к перегрузке контроллера [3]: значительному увеличению времени отклика на запросы или к отказу в обслуживании этих запросов — что повлечет прерывание или отказ в работе сетевых сервисов для конечных пользователей. Поэтому для внедрения подхода ПКС в реальных сетях актуальной является задача прогнозирования нагрузки на контроллер (количества входящих запросов и

количества исходящих изменений конфигураций коммутаторов). Прогнозирование нагрузки позволит сформировать упреждающие управляющие воздействия для предотвращения перегрузки контроллера.

В работе проводится обзор и сравнительный анализ существующих методов прогнозирования для решения задачи: статистического метода на основе модели ARIMA [3] и методов машинного обучения на основе различных моделей нейронных сетей [4-7]. В работе исследуются решения на базе следующих моделей нейронных сетей: многослойного перцептрона MLP [4]; свёрточной нейронной сети CNN [5]; длинной цепи элементов краткосрочной памяти LSTM [6-7], как наиболее используемой модели рекуррентных нейронных сетей.

Предлагаемое решение предсказывает показатели нагрузки контроллера, соответствующие количеству входящих запросов и исходящих ответов на несколько шагов вперед дискретного времени. При этом пороговые значения (при превышении которых возникает сигнал о перегрузке) могут быть настроены как статически (при помощи правила трёх сигм), так и динамически периодически пересчитываться (с применением скользящего окна и среднеквадратичного отклонения).

В экспериментальной части исследования для выявления наиболее эффективной модели рассматриваемые методы прогнозирования сравниваются по следующим метрикам: по матрице ошибок (включающей ошибки первого и второго рода обнаружения факта перегрузки); по точности и по комбинированной метрике F2-score. Также приводится сравнение данных показателей для каждой модели при использовании статической и динамической формулы задания пороговых значений.

Литература

- [1] Смелянский Р. Л. Программно-конфигурируемые сети // Открытые системы. М.: СУБД, №9, 2012.
- [2] Pashkov V., Smeliansky R. On High Availability Distributed Control Plane for Software-Defined Networks // In Proceedings of the Second International Conference "Modern Network Technologies" (MoNeTec-2018). IEEE. — Moscow, 2018. — pp. 1–10.
- [3] Скобцова Ю. А., Пашков В. Н. Исследование и разработка сетевого приложения для распределенной платформы управления в программно-конфигурируемых сетях // Программные системы и инструменты. – Т. 16 – Москва, МАКС Пресс, 2016.– с. 7–18.
- [4] Volkov A., Proshutinskiy K., Adam A. B. M., Ateya A. A. SDN Load Prediction Algorithm Based on Artificial Intelligence // In Proceedings of the Conference "Distributed Computer and Communication Networks" (DCCN-2019). Communications in Computer and Information Science, 2019, pp. 27–40.
- [5] Borovykh A., Bohte S., Oosterlee C. W. Conditional Time Series Forecasting with Convolutional Neural Networks // In arXiv preprint arXiv:1703.04691. – 2017.

- [6] Lipton Z. C., et al. Learning to Diagnose with LSTM Recurrent Neural Networks // In arXiv preprint arXiv:1511.03677, 2017.
- [7] Malhotra P., et al. Long Short-Term Memory Networks for Anomaly Detection in Time Series // In Presses universitaires de Louvain, 2016.

Федеративное обучение с использованием градиентного бустинга

Танкаев Иван Рустамович

Кафедра автоматизации систем вычислительных комплексов

e-mail: vantank36@gmail.com

Научный руководитель — к.т.н. доц. Костенко Валерий Алексеевич

Технология федеративного обучения актуальна при работе с закрытыми распределенными базами данных, большими распределенными базами данных и для периферийных вычислений. Основной проблемой промышленного применения федеративного обучения является безопасность технологии. Большинство известных алгоритмов обучения не обладают устойчивостью к перехвату обучения на отдельных базах данных. В работе предлагается способ построения алгоритма федеративного обучения с использованием схемы взвешенного голосования и стохастического градиентного бустинга для подбора весов. Приводятся результаты экспериментального исследования его точности в зависимости от числа локальных баз данных на которых перехвачено обучение.

Имеется N различных баз данных (клиентов). Базы данных однородны по хранимой информации, но информация между ними передаваться не может. Каждый объект в базе данных характеризуется набором параметров (характеристик объекта) и принадлежит к одному или нескольким классам объектов. Параметры объекта могут или изменяться со временем или быть постоянными. Классы, к которым принадлежит объект, также изменяются в зависимости от значений параметров, которые могут изменяться. Классификатор после изменения значений параметров объекта должен определять его принадлежность к классам. На каждом клиенте работает свой локальный классификатор. Он обучается на своей выборке данных и отправляется на сервер. Сервер агрегирует полученные локальные классификаторы и на их основании строит глобальный классификатор. Далее глобальный классификатор рассылается клиентам. Для классификации объектов используется полученный глобальный классификатор. По мере обновления своей выборки клиент заново обучает свой локальный классификатор на своей выборке данных и вновь отправляет его на сервер.

Одна из главных проблем использования алгоритмов федеративного обучения - их низкая устойчивость к атакам. В рамках работы было выделено 2 класса атак: атаки, направленные на переобучение локальных классификаторов, и атаки, направленные на нарушение целостности сети, с целью последующей передачи параметров переобученных классификаторов, вместо реальных параметров классификаторов. Основная идея разработанного алгоритма состоит в классификации объектов на каждом клиенте с использованием схемы

взвешенного голосования, где веса голосующих настраиваются с помощью стохастического градиентного бустинга. Схема взвешенного голосования позволяет использовать все классификаторы, обученные на клиентах, а градиентный бустинг позволяет снижать влияние захваченных злоумышленником классификаторов, путём снижения их весов в схеме голосования, и повышать влияние хорошо обученных, незахваченных классификаторов.

Для исследования свойств разработанного алгоритма был реализован тестовый стенд, на котором производилась классификация изображений в рамках задачи федеративного обучения. На этот стенд проводил атаки злоумышленник, пытаясь перехватить процесс классификации на клиентах.

Проведённые эксперименты показали, что алгоритм намного устойчивее других к атакам направленным на переобучение локальных классификаторов или нарушение целостности сети, с целью последующей передачи параметров переобученных классификаторов, вместо реальных параметров классификаторов.

Литература

- [1] V. A. Kostenko, I. R. Tankaev. Federated learning using simple voting scheme // Optical Memory and Neural Networks (Information Optics), 2022, Vol. 31, No. 2. 186-190.
- [2] Танкаев И. Р., Костенко В. А. Исследование устойчивости к перехвату обучения федеративных классификаторов использующих простое голосование // Программные системы и инструменты. Тематический сборник. Т. 21. Москва. 2021. С. 142–153.

Совместное планирование вычислений и информационного обмена в системах с разделяемой пропускной способностью сети

Чехонина Екатерина Андреевна

Кафедра автоматизации систем вычислительных комплексов

e-mail: katerinasheps@gmail.com

Научный руководитель — к.ф.-м.н. с.н.с. Балашов Василий Викторович

При планировании вычислений и обмена данными в ВС реального времени, в т. ч. модульных ВС [1], решаются следующие подзадачи:

1. распределение вычислительной нагрузки по процессорам;
2. построение конфигурации сети (для сетей с виртуальными каналами) или расписания обмена данными через сеть;
3. построение расписания вычислений.

При этом решения подзадач 1, 2, 3 должны быть согласованы между собой. Эта согласованность достигается либо последовательным решением подзадач в порядке 1, 2, 3 с полным решением каждой подзадачи перед переходом к следующей, либо совместным решением подзадач, при котором частичные решения достраиваются «одновременно», оставаясь при этом согласованными.

В рамках курсовой работы [2] был выполнен обзор методов совместного планирования, который показал, что в существующих работах либо используется последовательный подход, либо совместно решаются максимум две из трёх подзадач, а третья рассматривается отдельно.

В ВКР представлен алгоритм совместного решения всех трёх подзадач при планировании вычислений и информационного обмена в многомодульных ВС с общей шиной передачи данных, поддерживающей распределение пропускной способности в виде виртуальных каналов. Алгоритм построен по жадной схеме с элементами ограниченного перебора.

Предлагаемый алгоритм основывается на следующих стратегиях:

- пропускная способность сети конечна, поэтому необходимо ограничивать её выделение потокам данных;
- для попадания в свой директивный интервал выполнения задача должна быть запланирована как можно раньше по её готовности.

В ходе выполнения алгоритма, при движении по оси времени от начального момента к концу интервала планирования выполняется размещение работ на процессорные ядра и выделение пропускной способности канала для потоков данных (периодических сообщений) между задачами, работы которых запланированы на различные модули. Ограниченный перебор выполняется при попадании алгоритма в тупик и заключается в переназначении ранее запланированных работ на другие процессорные ядра и перераспределении пропускной способности канала между сообщениями.

Было проведено экспериментальное исследование успешности и масштабируемости алгоритма относительно числа задач и сообщений. Задача планирования вычислений и обмена данными является NP трудной [3]. Поэтому сначала строилось корректное расписание вычислений и информационного обмена, которое затем преобразовывается в набор входных данных. Для построения корректного расписания рассматривались только задачи с периодом, кратным 2^N . Входные данные были разбиты на классы по количеству ядер в системе и периодам задач для планирования.

При исследовании влияния настроечных параметров на характеристики алгоритма было выявлено, что повышения успешности работы алгоритма можно достичь путём увеличения глубины и числа итераций ограниченного перебора. Однако при этом увеличивается время работы алгоритма.

Предложенный в работе алгоритм успешно работает для систем с количеством ядер, небольшим по сравнению с количеством задач и близкими значениями периодов задач. У этого класса входных данных высокая средняя загрузка процессора. Для этого класса входных данных алгоритм масштабируется относительно увеличения количества задач.

Алгоритм поддерживает уровень успешности при увеличении количества сообщений для входных данных с количеством ядер не более 50% от числа задач. Однако его изначальная успешность на этом классе данных не очень велика.

Программная реализация алгоритма выполнена на языке C++. Объем написанного кода: 2477 строк, 102.7 килобайт.

Представляются перспективными следующие направления развития алгоритма для повышения его масштабируемости на большем количестве классов входных данных:

- увеличение количества настроечных параметров алгоритма и их настройка для каждого класса входных данных;
- добавление схемы ограниченного перебора, ориентированная на изменение пропускной способности, выделенной сообщениям.

Литература

- [1] Парамонов П. П., Жаринов И. О. Интегрированные бортовые вычислительные системы: обзор современного состояния и анализ перспектив развития в авиационном приборостроении // Научно-технический вестник информационных технологий, механики и оптики. 2013. № 2. С. 1–17.
- [2] Чехонина Е. А. Анализ методов совместного планирования вычислений и информационного обмена в вычислительных системах реального времени: курсовая работа. Москва: МГУ имени М.В. Ломоносова, 2021.
- [3] Гэри М., Джонсон Д. Вычислительные машины и труднорешаемые задачи. М.: Мир, 1982. 26 с.

Разработка и исследование параллельных методов кластеризации большого количества текстовых сообщений с использованием алгоритма DBSCAN

Жердев Иван Петрович

*Кафедра суперкомпьютеров и квантовой информатики
e-mail: zherdevivanp@gmail.com*

Научный руководитель — к.ф.-м.н. асс. Жуков Константин Андреевич

Введение. С каждым днем растёт объём текстовых сообщений, анализируемых в вычислительных системах. Это могут быть, например, журналы работы вычислительной системы, сообщения о сбоях или сообщения на естественном языке. Большие объёмы данных требуют автоматизированной параллельной обработки. Кластеризация является одним из методов работы с большим количеством разнообразных текстовых сообщений [1].

Были изучены существующие методы анализа текстовых записей. Программный комплекс ClusterLogs выбран в качестве основы для инструмента анализа сообщений. Выполнено сравнение алгоритмов решения задачи кластеризации, из которых выделен DBSCAN (основанная на плотности пространственная кластеризация для приложений с шумами).

Параллельные версии алгоритма DBSCAN. В работе подробно описан алгоритм кластеризации DBSCAN. Выполнено сравнение доступных парал-

лельных реализаций алгоритма. В ходе первичного исследования отобраны три версии: scikit-learn DBSCAN[†], HPDBSCAN [2], PDSDBSCAN-D [3].

Было выполнено исследование производительности программного комплекса ClusterLogs, а также сравнение производительности рассматриваемых версий алгоритма DBSCAN. Основным недостатком scikit-learn DBSCAN является отсутствие поддержки работы с распределенной памятью. Также были выявлены особенности HPDBSCAN: алгоритм показывает низкую производительность и экспоненциально растущие требования по памяти в зависимости от размерности входных данных. В связи с этим пришлось отказаться от применения HPDBSCAN для решения нашей задачи.

Изучены особенности PDSDBSCAN-D, связанные с обработкой неоднородных данных. Геометрическое перераспределение данных в PDSDBSCAN-D, выполняющееся на начальном этапе работы алгоритма, позволяет распределить точки выборки по параллельным процессам исходя из геометрической близости точек. В результате этой процедуры геометрически близкие точки оказываются в памяти одного процесса. Однако при обработке неоднородных данных, например, текстовых сообщений журналов ошибок, геометрическое перераспределение приводит к неравномерному распределению данных и к потере параллельного ускорения. В работе предложена модификация алгоритма PDSDBSCAN-D, которая позволяет получить более равномерное распределение данных по параллельным процессам и повысить параллельное ускорение. Более подробное сравнение параллельных версий алгоритма DBSCAN опубликовано в статье [4].

Внедрение PDSDBSCAN-D в ClusterLogs. Выполнено программное внедрение модифицированного PDSDBSCAN-D в ClusterLogs. Результат этой работы опубликован в открытом доступе^{‡§}. Вычислительные эксперименты показали эффективность предложенной параллельной реализации. Применен индекс Рэнда [5] для проверки корректности результатов работы полученных алгоритмов.

Заключение. Поддержка PDSDBSCAN-D в ClusterLogs делает весь программный комплекс более гибким, так как PDSDBSCAN-D поддерживает параллельное выполнение с технологией распределенной памяти. В совокупности с алгоритмом кластеризации scikit-learn DBSCAN, параллельно выполняющимся на общей памяти, это дает возможность выбирать подходящий инструмент в зависимости от доступной вычислительной инфраструктуры для наиболее эффективного использования ресурсов. В результате обеспечивается кластеризация текстовых сообщений за более приемлемое для анализа время.

[†] Документация scikit-learn DBSCAN. URL: <https://scikit-learn.org/stable/modules/generated/sklearn.cluster.DBSCAN.html> (дата обращения: 10.06.2022).

[‡] Исходный код и документация модифицированного PDSDBSCAN-D со SWIG-адаптером. URL: <https://github.com/Zherdev/pdsdbscan-mpi-lib> (дата обращения: 10.06.2022).

[§] Исходный код ClusterLogs с поддержкой PDSDBSCAN-D. URL: <https://github.com/Zherdev/ClusterLog/tree/add-pdsdbscan> (дата обращения: 10.06.2022).

Литература

- [1] Grigorieva M. Clustering error messages produced by distributed computing infrastructure during the processing of high energy physics data // International Journal of Modern Physics A. – 2021. – Vol. 36, N. 10, 2150070.
- [2] Götz M., Bodenstern C., Riedel M. HPDBSCAN: Highly Parallel DBSCAN // Proceedings of the Workshop on machine learning in high-performance computing environments. – 2015. – Article N.: 2, P. 1–10.
- [3] A New scalable parallel DBSCAN algorithm using the disjoint-set data structure / M. Patwary, D. Palsetia, A. Agrawal et al. // Proceedings of the International conference on high performance computing, networking, storage and analysis. – 2012. – Article N.: 62, P. 1–11.
- [4] Parallelizing of the DBSCAN algorithm in the ClusterLogs framework / I. Zherdev, K. Zhukov, M. Grigorieva, S. Korobkov. // International Journal of Modern Physics A. – 2022. – Vol. 37, N. 01, 2150247.
- [5] Rand W. M. Objective criteria for the evaluation of clustering methods // Journal of the American Statistical Association. – 1971. – Vol. 66, Article N. 336. – P. 846-850.

Свойства запутанных состояний многокубитных систем

Плужников Иван Русланович

Кафедра суперкомпьютеров и квантовой информатики

e-mail: 02180035@gse.cs.msu.ru

Научный руководитель – д.ф.-м.н. проф. Ожигов Юрий Игоревич

В данной работе рассмотрена задача на распределенные вычисления, в которой использование квантовой нелокальности дает существенный выигрыш в качестве результата.

Так же найден метод моделирования запутанных квантовых состояний ансамблей двухуровневых атомов и однододового поля в резонаторе в рамках модели Тависа-Каммингса в RWA приближении, существенно редуцирующий необходимую память компьютера, а так же скорость вычислений

Установлена возможность квантового превосходства для распределенных вычислений с односторонним управлением - в задаче о синтезе полимерных цепочек.

Мы рассмотрели задачу сравнения индекса склейки во всех разнополюх парах цепочек, полученных путем одновременного синтеза 3 или 4 полимерных цепей в удаленных местах под односторонним контролем. Мы обнаружили, что использование бифотонных ЭПР пар позволяет достичь максимальной разницы в степени склеивания (наложения) между законными и незаконными парами. Этот результат недостижим для классического контроля.

Если цель состоит в том, чтобы увеличить склейку законной пары и в то же время усилить антагонизм для незаконной пары, то мы должны распределить пару EPR между законными участниками и оставить незаконную пару

незапутанной с ними. Разница в индексе склеивания между легальными и нелегальными парами составит тогда 0,35.

Этот вывод сделан только для нашего искусственно построенного примера. Но этот пример напрямую опирается на главную особенность запутанности - нарушение неравенства Белла, следовательно, вывод применим к запутанному управлению для более широкого класса цепочек, собранных в удаленных местах, которые могут представлять интерес для квантовой биологии.

Этот результат является особенно интересным тем, что его можно продемонстрировать на уже имеющихся в России технологиях.

Так же в работе была рассмотрена динамика состояний групп двухуровневых атомов в одномодовом поле в рамках модели Тависа-Каммингса в RWA приближении. Был установлен характер изменения математического ожидания числа возбужденных атомов в одной из двух групп при различных начальных условиях. Так же был предложен и реализован подход отбора состояний, позволивший значительно ускорить моделирование задачи на компьютере, при этом не сильно изменив качество моделирования.

Литература

- [1] Ferreira P., Coelho J.F.J, Almeida J.F. and Gil M.H. Photocrosslinkable Polymers for Biomedical Applications, Biomedical Engineering - Frontiers and Challenges, Prof. Reza Fazel (Ed.), ISBN: 978-953-307-309-5, InTech, Available from: <http://www.intechopen.com/books/biomedical-engineering-frontiers-andchallenges/photocrosslinkable-polymers-for-biomedical-applications>
- [2] Stafford P., Halperin, R., Legutki J.B., Magee D.M.; Galgiani J.; Johnston S.A. (2012-04-01). Physical Characterization of the "Immunosignaturing Effect. Molecular & Cellular Proteomics. 11 (4): M111.011593. doi:10.1074/mcp.M111.011593.
- [3] Rajam C., Roop D.S. Synthesis of Certain Random Copolymers Containing Arylidene Diol Moiety and Studies on Their Antibacterial Activity and Photocrosslinking Efficacy of Blend Nanofibers, International Journal of Science, Engineering and Technology Research (IJSETR), Volume 4, Issue 1, January 2015.
- [4] Pan J., Bouwmeester D., Daniell M., Weinfurter H., Zeilinger A. (2000). "Experimental test of quantum non locality in three-photon GHZ entanglement". Nature. 403 (6769): 515–519.
- [5] Khrennikov A. After Bell. Fortschritte der Physik (Progress in Physics) 65, N 6-8, 1600014 (2017).
- [6] Araújo M, Hirsch f, Quintino M.T. Quantum 4, 353 (2020) DOI:10.22331/q-2020-10-28-353.
- [7] Kysela J., Erhard M, Hochrainer H., Krenn M., Zeilinger A. Experimental High-Dimensional Entanglement by Path Identity, PNAS 117(42), 26118-26122 (2020) DOI:10.1073/pnas.2011405117.

- [8] Erhard M., Malik M., Krenn M., Zeilinger A. Experimental GHZ Entanglement beyond Qubits, *Nature Photonics* 12, 759-764 (2018) DOI:10.1038/s41566-018-0257-6.
- [9] Grover L. A fast quantum mechanical algorithm for database search, *Proceedings, 28th Annual ACM Symposium on the Theory of Computing (STOC)*, May 1996, pages 212-219. *Proceedings, Melville, NY, 2006, vol. 810.*
- [10] Ozhigov Y. Distributed synthesis of chains quth one-way bi-photonic control, *Quantum Information and Computation*, Vol. 18, No. 7&8 (2018) 0592-0598.
- [11] Ю.И. Ожигов *Квантовый компьютер. Место издания: МАКС Пресс Москва. ISBN: 978-5-317-06403-7*

Применение методов машинного обучения в задачах распознавания русских дореволюционных печатных текстов

Федоров Владислав Владимирович

Кафедра суперкомпьютеров и квантовой информатики

e-mail: vlad.fedorov.2001@mail.ru

Научный руководитель — к.ф.-м.н. с.н.с. Ершов Николай Михайлович

Данная работа посвящена возможностям применения технологий машинного обучения для распознавания печатных русских текстов XIX века. Проведен обзор этапов работы OCR-программы и способов их реализации, произведено сравнение существующих программных продуктов – как свободно распространяемых, так и проприетарных – на качество распознавания дореволюционных текстов. Вручную размечены тестовые и обучающие выборки суммарным объемом около 120000 символов. Обучена собственная модель при помощи платформы Tesseract, которая продемонстрировала значительные улучшения в качестве по сравнению с исходной моделью. Создан графический интерфейс для выбора изображения с текстов и наглядного сравнения исходной страницы с полученным распознанным текстом, а также для редактирования текста и сохранения отредактированного текста для дальнейшего дообучения. Наконец, созданы две параллельные реализации обработки многостраничных документов, показавшие свою эффективность. Результаты работы были доложены на конференции Тихоновские чтения 2021 и опубликованы в сборнике [6].

Литература

- [1] Tesseract documentation. Overview of the new neural network system in Tesseract 4.00 [Электронный ресурс]. — URL: <https://tesseract-ocr.github.io/tessdoc/tess4/NeuralNetsInTesseract4.00.html> (дата обр. 05.11.2021).
- [2] Shafait F., Keysers D., Breuel T. Efficient implementation of local adaptive thresholding techniques using integral images // *DRR XV*. — 2008. — Янв. — Т. 6815. — С. 681510.
- [3] Smith R. Hybrid Page Layout Analysis via Tab-Stop Detection // *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR*. — 2009. — Янв. — С. 241–245.

- [4] Smith R. An Overview of the Tesseract OCR Engine // Ninth International Conference on Document Analysis and Recognition (ICDAR 2007). — 2007. — Окт. — Т. 2. — С. 629–633.
- [5] ocr_thesis GitHub repository [Электронный ресурс]. — URL: https://github.com/P34K1N/ocr_thesis (дата обр. 16.05.2022).
- [6] Федоров В. В., Ершов Н. М. Применение методов машинного обучения в задачах распознавания русских дореволюционных печатных текстов // Тихоновские Чтения (2021). — 2021. — Окт. — С. 81.

Методы автоматического морфемного разбора на основе машинного обучения

Афракова Алина Владленовна

Кафедра алгоритмических языков

e-mail: alinaafrakova@yandex.ru

Научный руководитель — к.ф.-м.н. доц. Большакова Елена Игоревна

Задача автоматического морфемного разбора заключается в разбиении слова на морфемы и последующем распознавании их типов (корень, приставка, суффикс, окончание). Например, морфемный разбор для слова *прекрасный*: *пре:PREFIX/крас:ROOT/н:SUFF/ый:END*. В настоящий момент лучшее качество решения задачи морфемного разбора дают методы на основе машинного обучения.

В данной работе проводилось исследование путей улучшения метода морфемного разбора словоформ русского языка на основе высокоточной сверточной нейросетевой CNN-модели, разработанной недавно и обученной на RuMorphs-Words — большом наборе данных, содержащим 2.8 млн словоформ русского языка с их морфемными разборами [1]. Несмотря на высокое качество работы, данная модель имеет несколько недостатков: слова, превосходящие по длине 20 символов, на этапе обучения модели отбрасываются, и в дальнейшем длинные слова разбираются простым способом, не показывающим достаточного качества. Также для работы модели требуется знание части речи разбираемого слова, которая берется из стороннего морфологического словаря или морфоанализатора.

На основе анализа статистики длинных словоформ набора данных RuMorphs-Words и с использованием данной CNN-модели в работе была предложена и реализована процедура морфемного разбора длинных слов через дефис, составляющих подавляющее большинство длинных слов, например, *командно-административный*. Данная процедура разбивает слово по дефису, первая часть слова перед дефисом разбирается моделью как наречие, а вторая часть — как прилагательное, после чего результаты морфемного разбора двух частей склеиваются. Экспериментальная оценка процедуры показывает достаточно высокое качество — значение доли правильных разборов словоформ (*Accuracy_{words}*) достигает 85%, что сопоставимо с качеством для других словоформ.

В работе также были реализованы несколько нейросетевых моделей модифицированной архитектуры с добавлением в исходную сверточную модель рекуррентного LSTM-слоя и слоя внимания Attention. Проведены эксперименты с разработанными моделями для подбора их гиперпараметров (размерности слоя, функции активации, начальных значений матриц весовых коэффициентов) и оценки качества работы. Слой LSTM был реализован на базе готового LSTM-класса из модуля `keras.layers` фреймворка Keras, основанного на открытой библиотеке Tensorflow. Слой Attention был реализован как пользовательский производный класс, наследующий базовый класс `Layer` из модуля `keras.layers` и производящий вычисления по формулам (1), (2), (3), (4).

$$score = \tanh(\bar{x}_t W_t + \bar{x}_t W_x + \bar{b}_h) \cdot W_a + \bar{b}_a. \quad (1)$$

$$\alpha = softmax(score). \quad (2)$$

$$\alpha = \alpha \times mask. \quad (3)$$

$$output = \alpha \odot \bar{x}_t. \quad (4)$$

где W_t, W_x, W_a — матрицы весовых коэффициентов,

\bar{b}_h, \bar{b}_a — вектора смещений,

\bar{x}_t — входной вектор, $output$ — выходной вектор,

$score$ — оценочная функция, α — веса внимания, $mask$ — маска.

По результатам экспериментов показано, что CNN-модель с добавленным слоем Attention размерности 32, без слоя конкатенации и без указания части речи во входных параметрах является оптимальной по соотношению качеству разбора слов ($Accuracy_{words} = 92.66\%$) и производительности (размер модели: 31.57 МБ, количество разбираемых слов в секунду: 365.12).

Литература

- [1] Bolshakova E. I., Sapin A. S. Building Dataset and Morpheme Segmentation Model for Russian Word Forms // Computational Linguistics and Intellectual Technologies: Papers from the Annual International Conference «Dialogue» (Moscow, 2021). 2021. Vol. 20, P. 154–161.

Извлечение аргументов из мнений пользователей

Бежанян Регина Олеговна

Кафедра алгоритмических языков

e-mail: regibezh@gmail.com

Научный руководитель — к.ф.-м.н. д.т.н. Лукашевич Наталья Валентиновна

Введение. В современной компьютерной лингвистике одним из наиболее интересных и перспективных направлений является извлечение аргументации из текстов. Аргументация - это система утверждений, направленная на оправдание или опровержение данной точки зрения.

Поиск мнений пользователей – это популярная задача, которая включает в себя определение цели и полярности мнений, например, в обсуждении какой-либо насущной в настоящее время темы. Полярность этого мнения можно выявить при анализе тональности текста. Тональность – эмоционально окрашенная лексика и эмоциональная оценка, выраженная автором относительно чего-либо [1].

Возможность извлечь такую информацию с обзорных веб-сайтов или из материалов в социальных сетях оказалась интересной. Таким образом, хотелось рассмотреть методы автоматического извлечения аргументации из текстов.

Задачи. В рамках работы требовалось решить следующие задачи:

- Изучить существующие методы автоматической классификации, а именно методы традиционного машинного обучения;
- Выявить модели, которые дают наилучший результат предсказаний для каждой из задач классификации по аргументации и тональности;
- Сравнить результаты с глубокой нейронной сетью BERT;
- Найти способ улучшения решений задач по аргументации и тональности за счет одновременного обучения по разным типам классификации.

Набор данных. При решении поставленной задачи основной рассматриваемой коллекцией предложений были данные соревнований RuArg-2022, собранные из социальных сетей и являющиеся комментариями к постам в социальных медиа по теме "Covid-19". Каждое из предложений было размечено авторами по тональности, позиции автора текста к одному из заданных утверждений по теме вакцинации, карантина и ношению масок, и аргументации, содержанию доводов «за» или «против» по отношению к этим же темам.

Эксперименты. Были рассмотрены различные методы традиционного машинного обучения, например, такие как случайный лес, градиентный бустинг, K ближайших соседей, метод опорных векторов, наивный Байес.

Наилучшие результаты были получены методом опорных векторов и градиентным бустингом. При дальнейшем сравнении и рассмотрении различных библиотек была выбрана реализация бустинга CatBoost [2].

Далее были рассмотрены различные улучшения обучения выбранной модели. Одним из наиболее удачных вариантов оказалось использование дискурсивных маркеров в качестве категориальных признаков. Дискурсивные маркеры – это такой класс слов, который выражает отношение автора текста к ситуации, его намерения, предположения и эмоции.

Также небольшое улучшение метрик и времени обучения дало использование сжатого векторного представления слов на основе FastText [3].

Изучение набора данных показало, что присутствует сильный дисбаланс классов. Поэтому было решено применить методы сэмплирования. Наилучшим образом себя показал метод SMOTE [4]. При этом он дал хороший выигрыш и по времени обучения.

Сравнение с моделями BERT показало, что результат CatBoost с использованием эмбедингов FastText и метода семплирования SMOTE сравним с результатами простой модели RuBert от DeepPavlov. Однако время обучения такой модели BERT оказалось сильно больше. При всем этом методы традиционного машинного обучения лучше интерпретируемые для человека и требуют меньшей предварительной настройки.

Результаты. В рамках выпускной квалификационной работы рассматривалась задача классификации по аргументации и по тональности текстовой коллекции предложений, собранных из комментариев в социальных медиа. По результатам работы:

- Были исследованы основные методы классического машинного обучения;
- Выявлена модель, которая дает наилучший результат предсказаний на совместном обучении по обоим задачам: CatBoost с векторным представлением слов FastText с использованием метода over-семплирования SMOTE.

Также сравнение с моделью ruBERT от DeepPavlov, дообученной на тренировочном датасете, показало, что CatBoost не проигрывает по качеству.

Литература

- [1] Fishcheva I. N., Kotelnikov E. V. Cross-Lingual Argumentation Mining for Russian Texts, Proceedings of the International Conference on Analysis of Images, Social Networks and Texts. AIST-2019, Kazan, pp. 134-144.
- [2] Bentejac C., Csorgo A., Martinez-Muñoz G. A comparative analysis of gradient boosting algorithms // Artificial Intelligence Review. 2021. pp. 1937–1967.
- [3] Athiwaratkun B., Wilson A. G., Anandkumar A. Probabilistic FastText for Multi-Sense Word Embeddings // In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2018. pp 1–11.
- [4] Nitesh V., Kevin W., Lawrence O. H., Kegelmeyer W. P. SMOTE: Synthetic Minority Over-sampling Technique // Journal of Artificial Intelligence Research 16. 2002. pp. 321–357

Поддержка вычислительной модели Лиспа для языка Си

Вересов Алексей Кириллович

Кафедра алгоритмических языков

e-mail: aleksey@veresov.pro

Научный руководитель — к.ф.-м.н. доц. Столяров Андрей Викторович

Проблема многостилевого программирования возникает при потребности использовать в рамках одного проекта несколько стилей, языков или парадигм программирования. Метод непосредственной интеграции является одним из способов решения этой проблемы и состоит в моделировании синтаксиса и семантики альтернативного языка средствами базового языка [1]. Проведённая работа посвящена непосредственной интеграции Лиспа в язык Си. Полученное

в ходе работы решение представлено в виде библиотеки языка Си, не имеющей зависимостей кроме стандартной библиотеки этого языка. Основой решения является сам язык Си, а не язык его препроцессора, активное использование которого ассоциируется с повышенным числом ошибок [2].

Существующие проекты непосредственной интеграции используют в качестве базовых языков объектно-ориентированные языки [1] и функциональные языки [3]. В обоих случаях используются богатые выразительные возможности соответствующих языков, например переопределение операций. Предложенные в существующих работах подходы к непосредственной интеграции неприменимы к языку Си, не обладающему необходимыми выразительными средствами. Поэтому в ходе работы потребовалось предложить новый подход к использованию возможностей языка Си.

Динамическая типизация. Язык Си обладает статической типизацией, в то время как Лисп — динамической. В работе предлагается воспользоваться техникой *fat pointers*, дословно техникой «жирных указателей». При этом объект языка Си доступен в виде указателя на его данные, перед которыми хранится дополнительная информация. Предлагается расширить объект числом, соответствующим типу объекта в языке Лисп, а также флагом, используемым в ходе сборки мусора.

Сборка мусора. Практическое использование Лиспа предполагает наличие в системе сборщика мусора. Предлагается воспользоваться алгоритмом пометок. Исполнение алгоритма пометок при каждом изменении системы вызвало бы ощутимые для пользователя задержки в работе программы. В целях оптимизации предлагается запускать алгоритм только при условии, что занимаемая системой память увеличилась более чем в два раза по сравнению с состоянием системы после прошлой очистки.

Запись списков. Для записи списков предлагается воспользоваться вариативными функциями языка Си. Так как все объекты системы представлены в виде жирных указателей, нулевой указатель заведомо не является объектом системы. Воспользуемся им в качестве признака конца списка. Пусть функция конструкции списка именуется *L*, тогда *L* (можно воспринимать как диграф открывающей скобки Лиспа, а *0*) — как диграф закрывающей. При этом программа на Лиспе будет записываться в виде суперпозиции вызовов этой функции.

Для представления точечных списков и пар введём идентификатор *o*, обозначающий точку Лиспа. Для этого занесём в *o* адрес некоторой статической переменной, нужной лишь для того, что бы получить её уникальный адрес, заведомо не являющимся адресом никакого объекта системы.

Пример использования системы. Пусть *E* — функция вычисления *S*-выражения, *S* — конструктор символов, *I* — конструктор целочисленных объектов. Запишем программу на Лиспе, складывающую 37 и 73, а затем используем полученный результат из Си и выведем его:

```
printf("%d\n", *(int *)E(L(S("+"), I(37), I(73), 0)));
```

Оценка эффективности решения. Оценим эффективность решения в сравнении с MIT Scheme, используя программы, записанные на диалекте, являющемся пересечением Scheme и диалекта Лиспа, который предоставляется разработанной системой. Примерами программ служат самоприменимый транслятор этого диалекта и функция Аккермана. Оценка показала на порядки (примерно в тысячу раз) меньшее потребление памяти разработанной системой. Задачи малой алгоритмической сложности выполняются созданной системой быстрее чем MIT Scheme, что говорит о высокой скорости загрузки. При выполнении задач высокой сложности, созданная система начинает уступать MIT Scheme в скорости, но не более чем в два раза на проверенных примерах. Объясняется это отсутствием оптимизаций кода в предложенной системе — предполагается, что разрыв в скорости можно будет устранить, добавив такие оптимизации.

Заключение. В ходе работы предложен подход к моделированию синтаксиса и семантики Лиспа в рамках языка Си; создана библиотека языка Си, предоставляющая выразительные возможности Лиспа; разработаны интерпретатор и самоприменимый транслятор диалекта Лиспа, показывающий возможность трансляции Лиспа в Си конечным автоматом; проведена оценка эффективности разработанного решения, показавшая приемлемые скорость и потребление памяти.

Литература

- [1] Головин И. Г., Столяров А. В. Объектно-ориентированный подход к мультипарадигмальному программированию // Вестник Московского университета. Серия 15. Вычислительная математика и кибернетика. 2002. Т. 15, № 1. С. 46–50.
- [2] Medeiros F., Ribeiro M., Gheyi R. Investigating Preprocessor-Based Syntax Errors // ACM SIGPLAN Notices. 2014. Vol. 49, N 3. P. 75–84.
- [3] Cheney J., Lindley S., Wadler P. A practical theory of language-integrated query // ACM SIGPLAN Notices. 2013. Vol. 48, N 9. P. 403–416.

Генератор синтаксических анализаторов на основе рекурсивного обхода L-графов

Ершов Никита Игоревич

Кафедра алгоритмических языков

e-mail: e-nik@list.ru

Научный руководитель — к.ф.-м.н. Вылиток Алексей Александрович

Синтаксический анализ – одна из важнейших тем в теории формальных языков. Обычно рассматривается класс контекстно-свободных языков в рамках этой задачи.

Подклассы бесконтекстных грамматик часто рассматриваются на практике, так как для них разработаны более быстрые (в сравнении с универсальными) методы синтаксического анализа. Одним из таких классов является класс LL(1)-грамматик.

Широко известно, что по грамматике, принадлежащей к данному классу, можно построить LL(1)-анализатор, выполняющий нисходящий синтаксический анализ за линейное время от длины входной цепочки символов. В целях универсальности строение такого анализатора является достаточно сложным и требует большого числа дополнительных понятий и конструкций. По этой причине задача построения аналога нисходящего анализатора с более простой внутренней структурой является актуальной. Наглядность графовых представлений открывает перспективу их использования при обучении методам трансляции. Простота использования графического представления формального языка в виде L-графа может в значительной мере ускорить освоение темы формальных языков.

В работе рассматривается подход, основанный на бесконтекстных L-графах [3]. Приведен критерий применимости метода рекурсивного обхода бесконтекстного L-графа, являющийся аналогом критерия применимости рекурсивного спуска при нисходящем анализе контекстно-свободных грамматик. Полученный критерий ограничивает множество L-графов до L-графов, соответствующих классу LL(1)-грамматик. Разработан веб-сервис, позволяющий в интерактивной среде построить L-граф, проверить выполнение необходимых условий для построения по нему синтаксического анализатора, действующего рекурсивным обходом, и в случае выполнения условий сгенерировать и скачать все необходимые файлы синтаксического анализатора по построенному графу. Веб-сервис адаптирован под работу на устройствах с любым разрешением экрана, поддерживает все актуальные на текущий момент браузеры.

Литература

- [1] Вылиток А.А., Ростовский А.В. О подклассах графовых представлений формальных языков // Сборник тезисов посвященных 300-летию М.В. Ломоносова: "Ломоносовские чтения 2011 года. - 2011. - с. 98-99.
- [2] Вылиток А.А., Сутырин П. Г. Характеризация формальных языков графами // Сборник тезисов научной конференции "Тихоновские чтения" (Москва, МГУ имени М.В. Ломоносова, факультет ВМК, 25-29 октября 2010 г. – 2010) М.: Макс-Пресс. 2010. С.81-82.
- [3] Станевичене Л.И. К теории бесконтекстных языков. – М.: Деп. в ВИНТИ РАН 29.05.2000. No 1546B00, 2000. – 155 с
- [4] Хопкрофт, Джон, Э., Мотвани, Раджив, Ульман, Джеффри, Д.. Введение в теорию автоматов, языков и вычислений, 2-е изд.: Пер. с англ. 2008 год
- [5] Ахо, Альфред В. Лам, Моника С., Сети, Рави, Ульман, Джеффри, Д.. Компиляторы: принципы, технологии, инструментарий. 2-е изд.: Пер. с англ. 2008 год.
- [6] Кондратьев Г. Д., Алгоритмы синтаксического анализа на основе L-графов: магистерская диссертация. - Москва: Московский государственный университет имени М. В. Ломоносова, 2018.

Разработка виртуального помощника

Рогожин Денис Андреевич

Кафедра алгоритмических языков

e-mail: den1s.rogozhin@mail.ru

Научный руководитель — к.ф.-м.н. доц. Грацианова Татьяна Юрьевна

Виртуальный помощник (или виртуальный собеседник) — это компьютерная программа, которая ведет диалог с человеком на естественном языке, отвечает на его вопросы, помогает выполнить некоторые действия. В современном мире их используют для обслуживания клиентов, предоставления информации, классификации темы разговора, удержания клиентов на сайте или около телефона и психологической поддержки одиноких людей.

Первые виртуальные собеседники, например, ELIZA и A.L.I.C.E., были реализованы с помощью ключевых слов и шаблонов, а современные программы (Алиса, Siri, Replica, Blenderbot) активно используют новые технологии, такие как нейронные сети и машинное обучение.

Цель данной работы — реализация виртуального помощника, отвечающего на вопросы студентов о кафедре АЯ факультета ВМК. Программа должна уметь классифицировать сообщения пользователя на спам и не спам (то есть отличать вопросы по теме кафедры от остальных сообщений), находить ответы на вопросы по теме кафедры в своей базе знаний, а также поддерживать диалог с пользователем. Далее будут рассмотрены некоторые методы реализации перечисленных умений программы.

Классификация сообщений на спам и не спам. Были исследованы 2 подхода:

1) На основе ключевых слов. В результате экспериментов выяснилось, что в данном случае он обеспечивает высокую полноту ($recall = 0.99$), но крайне низкую точность ($precision = 0.04$).

2) На основе машинного обучения. Были протестированы различные модели машинного обучения, наилучший результат показала модель градиентного бустинга ($precision = 0.99$, а $recall = 0.977$), которая и использована в программе.

Поиск ответа на вопрос в базе знаний. Система имеет базу знаний — набор из четырех собранных вручную таблиц: курсов, преподавателей, вопросов и ответов, синонимов. Для поиска ответа на вопрос в базе знаний происходит ранжирование вопросов в базе знаний на основе их схожести на вопрос пользователя. Для измерения схожести двух вопросов использовались следующие способы:

1) Количество совпадающих слов.

2) Доля совпадающих слов (отношение количества совпадающих слов к длине сообщения).

3) Количество совпадающих биграмм.

4) Косинусное расстояние между эмбедингами вопросов. Эмбединг — это векторное представление текста. Для построения эмбедингов сообщений использовалась русскоязычная версия SentenceBert[1].

5) Модуль разности между координатами эмбедингов вопросов.

6) Похожесть на основе Tf-Idf. $TfIdf(q, d) = \sum_{t \in q} tf(t, d) * idf(t)$

7) Похожесть на основе BM25[2] – модели, используемой во многих поисковых системах. Она является улучшением tf-idf, которое учитывает длины сообщений. $BM25(q, d) = \sum_{t \in q} idf(t) * \frac{(k_1+1)*tf(t,d)}{k_1*(1-b+b*\frac{L_d}{L_{mean}})+tf(t,d)}$

8) Вероятность похожести двух вопросов[2].
 $P(q|d) = \prod_{t \in q} \lambda * \frac{tf(t,d)}{L_d} + (1 - \lambda) * \frac{df(t)}{L_c}$

Обозначения: q – вопрос пользователя, d – вопрос из базы знаний, t – слово, $\lambda \in (0, 1.0]$, $k_1 \in [1.2, 2.0]$, $b = 0.75$, L_c – число слов в базе знаний, L_{mean} – средняя длина вопроса в базе знаний, L_d – число слов в вопросе d, $tf(t,d)$ – частота слова t в вопросе d, $df(t)$ – число вопросов в базе знаний, содержащих слово t, $idf(t) = \lg \frac{N}{df(t)}$, N – число вопросов в базе знаний

Эти методы были реализованы в программе, и проводились эксперименты для выбора лучшего. Для оценки качества методы применялись к множеству из 500 пар <<вопрос – правильный ответ>>, и измерялась точность – доля вопросов, на которые дан правильный ответ. Вышеперечисленные методы использовались как признаки, описывающие пару <<вопрос пользователя – вопрос из базы знаний>> для построения моделей машинного обучения – Логистической регрессии, градиентного бустинга, случайного леса и метода опорных векторов. Наилучший результат был получен с помощью стекинга из этих моделей: точность составила 0.954.

Инициирование диалога. Для поддержания диалога с пользователем программа способна сама предлагать тему для разговора. Она выбирает строку из специальной таблицы и спрашивает пользователя, хочет ли он поговорить на эту тему. Для распознавания ответа (классификации согласен/не согласен) используется метод, основанный на ключевых словах. В случае согласия пользователя программа рассказывает ему о том, что предлагала, после чего снова пытается инициировать диалог.

В рамках работы разработан собственный виртуальный помощник. Он размещен на факультетском сервере и использовался в приемной кампании на кафедру алгоритмических языков.

Программа доступна по ссылке: https://t.me/AL_dep_bot

Литература

- [1] Nils Reimers, Iryna Gurevych. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. [электронный ресурс]. URL: <https://arxiv.org/pdf/1908.10084v1.pdf>
- [2] Christopher D. Manning, Prabhakar Raghavan, Hinrich Schütze. An introduction to information retrieval. Cambridge University Press. Cambridge, England.

Исследование методов автоматического извлечения вложенных именованных сущностей

Рожков Игорь Сергеевич

Кафедра алгоритмических языков

e-mail: fulstocky@gmail.com

Научный руководитель — д.т.н. доц. Лукашевич Наталья Валентиновна

Задача распознавания именованных сущностей (*Named Entity Recognition, NER*) — это задача поиска фрагментов текста, связанных с некоторыми физическими или абстрактными объектами именами, и их классификации в соответствии с их смысловым классом, таким как, например, человек или организация. Такой сущностью можно назвать, например, последовательность *МГУ им. М.В. Ломоносова*, который имеет смысловой класс организации, т. е. *ORGANIZATION*.

Вложенные именованные сущности и ограниченность классов. Традиционно постановка задач и наборы данных извлечения именованных сущностей посвящены извлечению так называемых *плоских* именованных сущностей, где предполагается, что сущность не может содержать другую сущность.

Помимо вложенности существует другая проблема — ограниченность числа некоторых классов сущностей. Она возникает в разных узко специализированных сферах, где создание разметки является большой сложностью [1]. Рассмотрение этой проблемы часто переформулируется в так называемую *few-shot* постановку задачи или задачу с ограниченным количеством данных.

Среди существующих наборов данных для русского языка большинство не рассматривало ни одну из вышеуказанных проблем. Совсем недавно возник набор данных NEREL [2]. Этот набор данных содержит множество вложенных сущностей 29 разных классов. Однако, он не позволял рассмотреть вторую проблему, в связи с чем на его основе был создан набор данных RuNNE [3], содержащий несколько специально ограниченных классов сущностей. Именно на нём проводилось дальнейшее исследование задачи.

Модель машинного чтения. Среди множества подходов к решению задачи была выбрана т. н. *модель машинного чтения (Machine Reading Comprehension, MRC)* [4]. Данная модель позволяет переформулировать задачу извлечения вложенных именованных сущностей как задачу ответа на вопросы по заданной последовательности слов. Выбирая «хорошие» вопросы, можно достичь высокого качества предсказания именованных сущностей. Лучшим типом вопросов в оригинальной работе стали т. н. *определения* классов сущностей, ставшие отправной точкой в исследовании.

Новые подходы. В выпускной работе рассматривались способы улучшения качества предсказания сущности модели MRC на ограниченных классах сущностей. В том числе были предложены новые способы:

- *N* самых частотных сущностей или их компонентов
- *Контекстный*

- *Лексический*
- *Структурный*
- *Полные подходы для всех внешних (внутренних, вложенных)*

В частности, эксперименты показали, что т. н. *полная лексическая подсказка для всех внешних сущностей* оказалась наилучшей из них. Данная подсказка формируется следующим образом: берётся некоторое предложение из обучающего набора данных, в нём находятся все самые внешние сущности заданного класса, и каждая такая сущность заменяется меткой этого самого класса.

Все представленные подходы просты в создании, т. к. могут быть сгенерированы автоматически, в отличие от использования определений, поскольку требуется их тщательный отбор экспертами из разных словарей.

Результаты. Эксперименты на ограниченных классах набора данных *RuNNE* показали, что использование определений достигает качества в 61.21% по F1-мере, в то время как полная лексическая по всем внешним сущностям позволяет улучшить его до 63.88%. Такой результат позволяет сделать вывод о том, что модель машинного чтения получает больше семантической информации из лексических подсказок.

Данная работа была опубликована в рамках международной конференции Диалог [5].

Литература

- [1] A. Beatrice, H. Barry, G. Claire. Recognising nested named entities in biomedical text // Biological, translational, and clinical language processing, 2007. P. 65–72.
- [2] NEREL: a Russian dataset with nested named entities and relations / N. V. Loukachevitch, E. Artemova, T. Batura, P. Braslavski, I. Denisov, V. Ivanov, S. Manandhar, A. Pugachev, E. Tutubalina // International Conference Recent Advances in Natural Language Processing, RANLP 2021, P. 876–886.
- [3] RuNNE-2022 Shared Task: Recognizing Nested Named Entities / E. Artemova, M. Zmeev, N. Loukachevitch, I. Rozhkov, T. Batura, P. Braslavski, V. Ivanov, E. Tutubalina // Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference «Dialog», 2022.
- [4] A unified MRC framework for named entity recognition / X. Li, J. Feng, Y. Meng, Q. Han, F. Wu, J. Li. // In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 2020, P. 5849–5859.
- [5] Rozhkov I., Loukachevitch N. Machine reading comprehension model in RuNNE competition // Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference «Dialog», 2022.

Автоматическое распознавание терминологических отношений на основе шаблонов и правил

Степанова Полина Александровна

Кафедра алгоритмических языков

e-mail: Polina1534@yandex.ru

Научный руководитель — к.ф.-м.н. доц. Большакова Елена Игоревна

Распознавание и извлечение информации из текстов на естественном языке – актуальное направление автоматической обработки текстов. Один из основных подходов к её решению опирается на использование лингвистических шаблонов для формального описания характерных языковых конструкций, содержащих извлекаемую информацию.

В данной работе исследовалась задача автоматического распознавания и извлечения терминологических отношений из текстов на русском языке, в рамках подхода на основе шаблонов и правил. Рассматривались 8 наиболее часто встречающихся в текстах видов отношений (семантических связей) терминов, включая отношения род-вид, часть-целое и причина-следствие. Было разработано 95 лексико-синтаксических шаблонов и правил, формализующих языковые конструкции терминологических отношений и задающих условия их извлечения. Ряд типичных языковых конструкций был взят из работы [1], этот набор затем был существенно расширен. Шаблоны и правила были созданы на языке LSPL, с использованием средств его поддержки [2]. В правых частях разработанных правил указывается вид распознанного терминологического отношения и термины, которые им связываются.

Для применения разработанного набора лексико-синтаксических шаблонов и правил к текстам было реализовано программное средство на языке Python с использованием возможностей библиотеки lxml. Помимо распознавания и извлечения терминологических отношений в его функции входил подсчёт статистики применений шаблонов.

Созданное программное средство было использовано для проведения экспериментов по оценке точности разработанных LSPL-шаблонов на основе русскоязычной коллекции текстов научных статей и учебных пособий из разных предметных областей (1.7 млн. слов). Путём ручного экспертного анализа результатов извлечения терминологических отношений были получены оценки точности всех созданных шаблонов и выделены высокоточные шаблоны (более 80% точности). Также был проведён анализ ошибочных результатов извлечения терминологических отношений, выявлены основные причины неточности и определены пути дальнейшего уточнения шаблонов и правил, а также расширения их набора.

Полученные в данной работе результаты могут быть полезны для автоматизированного построения онтологий предметных областей, реферирования и терминологического анализа специализированных текстов.

Литература

- [1] Никитина С. Е. Тезаурус по теоретической и прикладной лингвистике. М. : Наука, 1978. 375 с.
- [2] Система для извлечения информации из текстов на базе лексико-синтаксических шаблонов / Е. И. Большакова, К. М. Иванов, А. С. Сапин, Г. Ф. Шариков // Пятнадцатая национальная конференция по искусственному интеллекту с международным участием (КИИ-2016). 3–7 окт. 2016 г.: Труды конференции. Т. 1. Смоленск : Универсум, 2016. С. 14–22.

Методы автоматического обнаружения изменения значений слов

Хомский Даниил Владимирович

*Кафедра алгоритмических языков
e-mail: homdani1123@gmail.com*

Научный руководитель — к.ф.-м.н. Арефьев Николай Викторович

Задача обнаружения изменения значений слов (*Lexical Semantic Change Detection, LSCD*). На входе имеется целевое слово и пара текстовых корпусов разных временных периодов, необходимо определить, изменились ли значения слова со временем. Например, слово *тачка* раньше использовалось только в значении *телеги*. Сейчас же данное слово приобрело еще один смысл — *автомобиль*. В задаче имеется несколько постановок: обнаружение изменения значений слов, отранжировав или классифицировав слова. Оценивание решений задачи производилось на русском, испанском, английском, немецком и шведском языках.

Задача выявления значений слов (*Word Sense Induction, WSI*). На входе имеется целевое слово и примеры предложений с данным словом. Необходимо кластеризовать данные предложения по смыслу слова. Оценивание решений задачи производилось на русском и английском языках.

Вспомогательная задача Слово-в-Контексте (*Word-in-Context, WiC*). На входе имеется пара предложений с использованием целевого слова. Необходимо ответить, в одинаковом ли смысле используется целевое слово в двух контекстах. Для того чтобы решить основные задачи можно воспользоваться моделью, которая решает задачу *WiC*. В рамках данной работы использовалась модель [1] с наилучшими параметрами, подобранными в статье [2]. Данный метод принимает на вход пару предложений с целевым словом и возвращает вероятность того, что слово используется в одинаковом контексте в двух предложениях.

Решение задачи обнаружения изменения значений слов. Для решения необходимо найти и семплировать предложения из старого и нового корпусов. Далее нужно создать пары, где одно предложение принадлежит старому корпусу, а другое принадлежит новому корпусу. Данные пары контекстов размечаются моделью Слово-в-Контексте. Полученные предсказания агрегируются

(использовались среднее арифметическое и корреляционная кластеризация) по предложениям для целевого слова.

Решение задачи выявления значений слов. Для решения необходимо составить все пары контекстов для целевого слова. Полученные пары нужно разметить моделью Слово-в-Контексте. То есть, используя данный подход, можно найти матрицу расстояний для всех контекстов для целевого слова. Полученная матрица подается в кластеризатор, который уже кластеризует данные контексты.

Эксперименты В выпускной работе рассматриваются эксперименты, позволяющие определить важность дообучения на разных языках и выборках, длины контекста, количества семплированных пар, а также различные виды преобразования данных для дообучения.

Результаты. В задаче выявления значений слов на русских данных `bts-rnc public test` был достигнут результат 0.647 по метрике ARI, который превосходит наилучшие существующие решения (0.525 ARI). На английской выборке `SemEval 2010` был достигнут результат 56.6 по метрике AVG S10, который также превосходит лучшие существующие решения (54.2 AVG S10).

В задаче обнаружения изменения значений слов на русских данных `RuShiftEval` был достигнут результат 0.85 по метрике корреляции Спирмена, а также было достигнуто 2 место в одноименном научном соревновании. Данный метод был также протестирован на данных `SemEval 2020` с метрикой корреляции Спирмена. На английском языке получен результат 0.740, на немецком 0.871, на шведском 0.750. Данные результаты превзошли результаты участников соревнования: 0.436 (английский), 0.725 (немецкий), 0.604 (шведский). Также данный метод был протестирован на испанских данных, где показал результат 0.719 по корреляции Спирмена, а также занял 2 место в рамках соревнования `Spanish LSCDiscovery 2022`. На испанских данных метод также оценивался на бинарных задачах обнаружения изменения/приобретения/утраты значений слов, где показал результат 0.712 по метрике F1-score в задаче обнаружения изменения, лучшее существующее решение — 0.716. А также был показан наилучший результат в задаче обнаружения приобретенного смысла — 0.591.

По результатам работы опубликованы статьи [2, 3].

Литература

- [1] LIORI at SemEval-2021 Task 8: Ask Transformer for measurements / A. Davletov, D. Gordeev, N. Arefyev, E. Davletov // In Proceedings of the 15th International Workshop on Semantic Evaluation (SemEval-2021), P. 1249–1254, Online. July 2020. Association for Computational Linguistics.
- [2] DeepMistake: Which Senses are Hard to Distinguish for a WordinContext Model / N. Arefyev, M. Fedoseev, V. Protasov, D. Homskiy, A. Davletov, A. Panchenko // In Computational linguistics and intellectual technologies, 20, P. 16 – 30, Russian Federation.

- [3] Homskiy D., Arefyev N. DeepMistake at LSCDiscovery: Can a Multilingual Word-in-Context Model Replace Human Annotators? // In Proceedings of the 3rd Workshop on Computational Approaches to Historical Language Change, P. 173–179, Dublin, Ireland, May 2022. Association for Computational Linguistics.

Разработка типового программного решения для публикации документов с использованием языка разметки TEI

Шевернев Юрий Алексеевич

Кафедра алгоритмических языков

e-mail: shevernev@yandex.ru

Научный руководитель — доц. Леонов Михаил Васильевич

Целями выполнения выпускной квалификационной работы являлось создание применимого на практике типового программного решения для публикации документов с использованием языка разметки TEI – модельной информационной системы, а также исследование и систематизация многообразия современных технологий и инструментов для создания высоконагруженных, эргономичных веб-приложений, приобретение навыков создания полнофункциональных приложений, готовых к развитию и масштабированию – на примере создания модельной информационной системы.

Созданное в рамках выпускной квалификационной работы программное решение позволяет разворачивать в сети Интернет сайты для публикации (размещения в сети Интернет для всеобщего доступа) документов, посвященных какой-либо тематике. Тематика не ограничена какими-либо рамками и выбирается создателем сайта. Под документами подразумеваются текстовые материалы в цифровой форме (книги, статьи, нормативные правовые акты и тому подобное). Материалы в системе размещаются в текстовой и графической формах. Текст материала кодируется с использованием языка разметки TEI. Графические материалы – оцифрованное изображение документа, страницы, иллюстрации к тексту и т.п.

TEI (Text Encoding Initiative) – это формат для кодирования электронных документов, разработанный и поддерживаемый Консорциумом инициативы по кодированию текстов (Text Encoding Initiative Consortium) [1]. Формат TEI начал создаваться в конце 1980-х годов как проект, целью которого было обеспечение целостной и расширяемой системы кодирования текстов любой природы [2]. Наибольшее распространение формат получил в академических кругах. Можно отметить, например, следующие проекты, активно использующие TEI: Библиотека Университета Вирджинии, Британский национальный корпус, Текстовый архив Оксфорда [3] и другие. Документ в формате TEI представляет собой XML-документ, соответствующий определенным требованиям. Требования определяются Руководством TEI (TEI P5: Guidelines for Electronic Text Encoding and Interchange) [4].

Разработанное программное решение позволяет администратору (авторизованному пользователю) редактировать в веб-интерфейсе документы в формате TEI, публиковать их и импортировать (сохранять) на локальный компьютер. Дополнительно создано средство для автономной работы на локальном компьютере (публикацию можно скачать и без отсутствия интернета просмотреть или отредактировать на компьютере). Разработанные в рамках выпускной квалификационной работы программные модули (библиотеки) могут быть переиспользованы в иных проектах, в которых используется формат TEI. Интерес представляют функции конвертации из формата TEI в формат JSON и обратно, импорта публикации в единый ZIP-архив, содержащий собственно текст документа в формате TEI и загруженные на сервер изображения страниц.

Система реализована в трехуровневой архитектуре с использованием таких технологий, как Java, Spring, JavaScript, Vue.js, СУБД PostgreSQL, Keycloak. Архитектура решения предусматривает возможность горизонтального масштабирования, хотя все компоненты могут быть запущены и на одном сервере. Возможность тиражирования (простота развертывания) обеспечивается применением программных средств Docker и Docker-compose. Решение является кроссплатформенным и свободным.

С использованием разработанной системы создан сайт, посвященный работам Ивана Алексеевича Двигубского – российского естествоиспытателя, занимавшего пост ректора МГУ им. М.В. Ломоносова с 1826 по 1833 годы. Сайт доступен в сети Интернет по адресу <https://двигубский.рф>. В разделе «О проекте» размещены исходные коды решения, инструкция по развертыванию, описание функциональных возможностей решения.

Литература

- [1] Text Encoding Initiative [Электронный ресурс] URL: <https://tei-c.org/>.
- [2] Sebastian R., Lou B., Reviewing the TEI ODD system // Proceedings of the 2013 ACM symposium on Document engineering, 2013. P. 193–196.
- [3] Oxford Text Archive [Электронный ресурс] URL: <https://ota.bodleian.ox.ac.uk/>.
- [4] TEI Consortium. TEI P5: Guidelines for Electronic Text Encoding and Interchange. Technical report // TEI Consortium, 2010.

Методы аспектного анализа научно-технических текстов

Яковлева Светлана Андреевна

*Кафедра алгоритмических языков
e-mail: svetych2000@yandex.ru*

Научный руководитель — к.ф.-м.н. ас. Ефремова Наталья Эрнестовна

В настоящее время нам приходится ежедневно обрабатывать, просматривать и изучать огромное количество текстовой информации. Аспектный анализ — один из методов анализа текста, целью которого является нахождение

и выявление определенных аспектов. Эта задача является на данный момент малоизученной для русского языка.

Одним из активно развивающихся направлений является аспектный анализ научно-технических статей, при котором в текстах ищутся аспекты содержания. В рамках работы были выделены пять аспектов:

1. *Рассматриваемая проблема (тема, задача).*
2. *Цель исследования.*
3. *Актуальность + новизна + практическая значимость.*
4. *Идея + метод + особенности решения.*
5. *Результат + выводы.*

Принято выделять два основных подхода к аспектному анализу текстов: инженерный подход, основанный на составлении словарей и правил экспертом, и подход, основанный на машинном обучении. Для применения машинного обучения необходимо наличие размеченных текстовых коллекций значительного объема, которых не существует сейчас для русского языка. Идея, лежащая в основе методов инженерного подхода, заключается в выявлении в тексте аспектных индикаторов — словесных клише, маркеров, указывающих на наличие какого-либо аспекта в тексте [2]. Например, индикатор *в работе рассматриваются методы* говорит о наличии аспекта *Рассматриваемая проблема*.

Цель данной работы — предложить и реализовать методы выявления аспектных индикаторов для их последующего использования в аспектном анализе научно-технических текстов.

В работе [1] был получен список кандидатов в индикаторы из 13 тысяч различных цепочек слов. Было замечено, что словосочетания, являющиеся аспектными индикаторами, имеют схожую лексико-синтаксическую структуру. В них можно выделить главное слово, которое присутствует всегда, и зависимые слова. Отобранные из списка индикаторы были сгруппированы по главному слову. Отдельно в полученных группах выделялись: главная подгруппа, подгруппы из определяющих или характеризующих слов, подгруппа из отрицательной частицы «не», подгруппы из устойчивых фраз. Таким образом были составлены 66 групп индикаторов для рассматриваемых аспектов.

С учетом полученного представления аспектных индикаторов были автоматически созданы два набора шаблонов на языке LSPL: *конкретизированные*, в которых все слова заданы в явном виде, и *обобщенные*, в которых подгруппы определяющих и характеризующих слов заданы своими частями речи.

На основе полученных наборов шаблонов реализованы два метода определения в тексте аспектных индикаторов. Их суть заключается в последовательном применении к предложению шаблонов. При этом определяется:

- длина шаблона (количество подгрупп в нем);
- длина индикатора (количество выделенных с помощью шаблона слов).

На основании этих показателей определяется, насколько вероятно, что предложение содержит искомый аспект.

Предложенное в данной работе решение для проведения аспектного анализа научно-технических текстов было программно реализовано с использованием языка Python 3. Для оценки качества методов рассматривалась задача автоматического реферирования. Была создана коллекция из 64 текстов тезисов ВКР 2020 года, для которых вручную были составлены эталонные рефераты. В результате были получены три набора рефератов:

1. без применения аспектного анализа;
2. с применением аспектного анализа конкретизированными шаблонами;
3. с применением аспектного анализа обобщенными шаблонами.

Экспертный анализ полученных результатов показал, что предложенные методы хорошо справляются с поставленной задачей определения аспектов. Опрошенные респонденты выбирали рефераты, созданные с использованием аспектного анализа, в 75% случаев. Также качество программы проверялось с помощью метрик ROUGE-1, ROUGE-2 и ROUGE-L. Наилучшие результаты показал метод с использованием обобщенных шаблонов.

Литература

- [1] Машигина Н. В. Аспектный анализ научной статьи на основе дискурсивных маркеров: Магистерская диссертация — Москва, 2020.
- [2] Саломатина Н. В., Гусев В. Д. Автоматизация формирования индикаторных словарей и возможности их использования // Труды международной конференции «Диалог 2006». М., 2006. С. 459–463.

Сокращение поверхности атаки на ядро операционной системы Linux

Басков Евгений Сергеевич

Кафедра системного программирования

e-mail: j-basevgser@yandex.ru

Научный руководитель — к.ф.-м.н. доц. Хорошилов Алексей Владимирович

Данная работа посвящена проблеме безопасности ядра операционной системы (ОС) Linux на этапе её развёртывания и инициализации на платформе x86_64. В настоящее время в ОС, использующих ядро Linux, многие техники повышения защищенности кода применяются либо во время работы прошивки, до передачи управления ОС, либо при уже полностью инициализированной и работающей ОС. Поэтому устойчивость к возможным атакам на этапе ранней загрузки в ядре Linux может быть улучшена. Одной из простых, но при этом эффективных техник защиты, является политика *WX* (Write-xor-Execute).

Она заключается в запрете на использование регионов памяти доступных одновременно на запись и исполнение. Её применение не позволяет атакующему записать необходимый код и сразу передать на него управление, усложняя задачу выполнения произвольного кода. Эта политика активно используется в других ОС, а также доступна в Linux, но на более поздних этапах его работы.

Основная сложность реализации W^X заключается в корректном определении регионов памяти и установке требуемых ими атрибутов памяти, так как

- в процессе загрузки механизмы страничной адресации не всегда доступны и не во всех случаях управляются непосредственно ядром;
- память повторно используется для регионов с различными требуемыми атрибутами;
- регионы перемещаются на перекрывающиеся с другими регионами адреса.

Процесс ранней загрузки. Ядро Linux хранится на диске, как файл, содержащий ядро в сжатом виде и небольшой код ранней загрузки, задачей которого является распаковка образа ядра и подготовка окружения для продолжения инициализации (см. Рис. 1). Этот код поддерживает множество интерфейсов загрузки, но все они передают управление на общую для всех интерфейсов часть пути исполнения. Этим вызвана необходимость повторной инициализации состояния процессора и перемещения ядра.

Однако при загрузке с использованием интерфейсов 64-битных UEFI (узлы, выделенные пунктиром на Рис. 1) окружение известно, и, если избежать использования общей для всех путей исполнения части кода и распаковывать ядро напрямую из окружения, предоставляемого UEFI-совместимой прошивкой, можно значительно упростить код загрузки ядра и избежать необходимости копирования образа ядра.

Результаты. В рамках данной работы были разработаны три набора доработок кода ядра Linux. Первый из них исправляет проблемы совместимости кода ядра со стандартом UEFI, что позволяет повысить безопасность процесса загрузки за счёт возможности использовать прошивки, которые устанавливают более строгие атрибуты памяти на регионы, выделяемые прошивкой. Второй повышает безопасность кода общего пути исполнения, реализуя для него политику W^X везде, где это применимо. Третий реализует отдельный путь исполнения при загрузке с использованием 64-битных UEFI-совместимых прошивок, тем самым позволяя избавиться от перемещений ядра и повторного использования памяти, что позволяет реализовать W^X в полном объеме для этого способа загрузки.

Преобразование последовательных Фортран-программ для их распараллеливания на гибридные кластеры

Гусев Григорий Дмитриевич

Кафедра системного программирования

e-mail: gr-gusev@mail.ru

Научный руководитель — д.ф.-м.н. проф. Крюков Виктор Алексеевич

В настоящее время изначально непростая задача создания параллельных программ дополнительно осложняется появлением новых вычислительных устройств и параллельных архитектур. Например, в узлах гибридного кластера

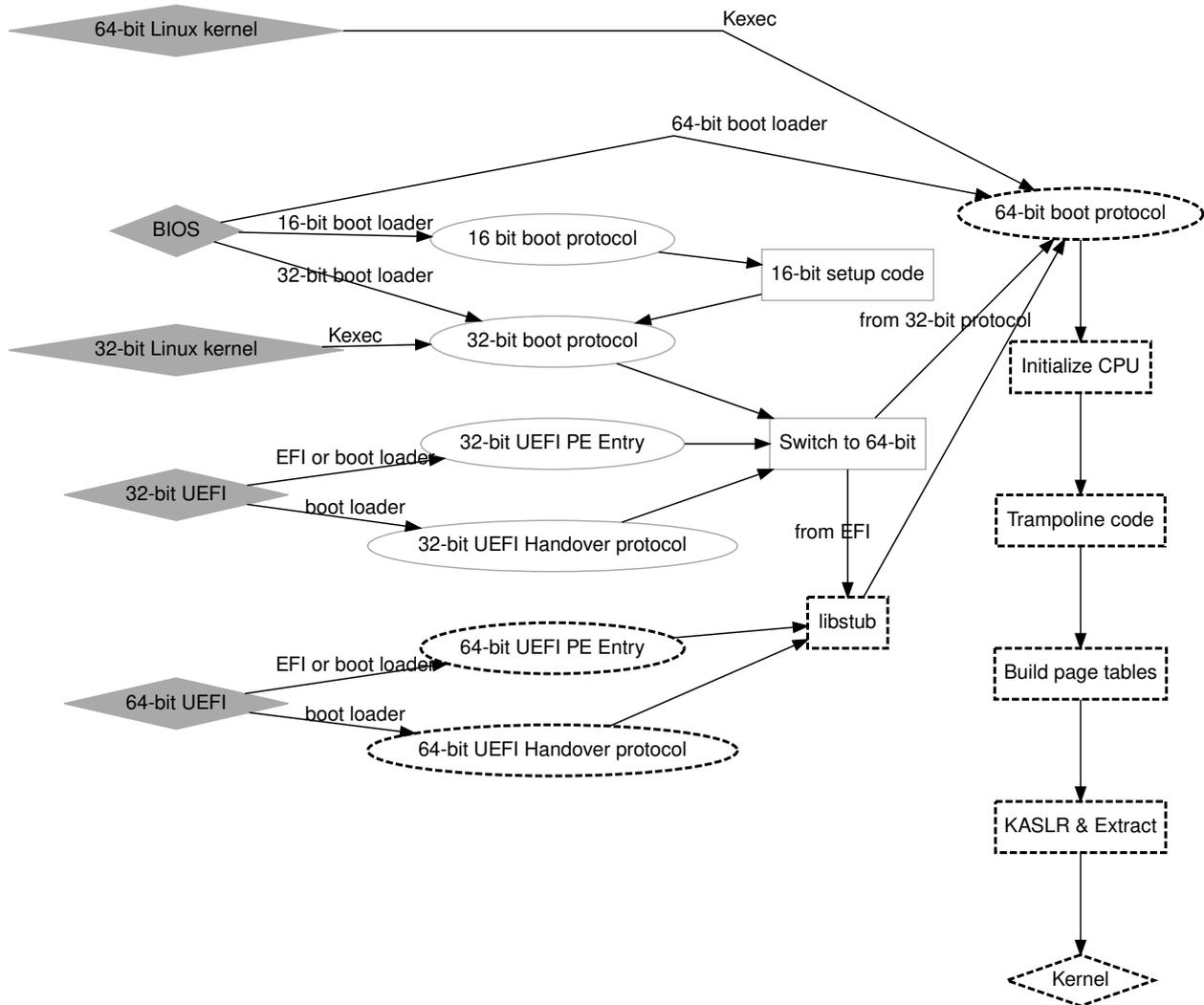


Рис. 1: Схема ранней загрузки

помимо центрального процессора (central processing unit, CPU) используются дополнительные вычислительные устройства — графические ускорители (graphics processing unit, GPU) или многоядерные сопроцессоры. Это приводит к необходимости развития моделей параллельного программирования, позволяющих разрабатывать программы для выполнения на различных архитектурах, и средств автоматизированного создания таких программ.

Примером высокоуровневой технологии параллельного программирования является DVM-система, предназначенная для разработки параллельных программ научно-технических расчётов. Модель DVMH является расширением DVM-системы для гибридных кластеров. Важным дополнением к ней является система автоматизированного распараллеливания Фортран- и Си-программ SAPFOR (System FOR Automated Parallelization), которая анализирует исходный код последовательной программы и на основе полученных результатов осуществляет генерацию параллельной версии в модели DVMH.

Однако полученная автоматизированным распараллеливанием версия программы может оказаться неэффективной из-за структуры последовательной

программы и её оптимизации под последовательное выполнение. Решить возникающие проблемы помогают преобразования последовательных программ, позволяющие изменить структуру программы и её данных.

Целью данной работы являлось расширение возможностей системы SAPFOR по автоматизированному распараллеливанию Фортран-программ через её дополнение несколькими проходами преобразования последовательных программ, а также оценка эффективности реализованных преобразований на примере распараллеливания реальной программы. Были рассмотрены следующие преобразования: объединение/разделение циклов и сужение/расширение приватных переменных циклов.

Объединением двух идущих подряд в программе циклов является их замещение одним циклом с соответствующим переносом и преобразованием операторов из тел объединяемых циклов. Разделением цикла является разбиение его тела на несколько множеств операторов, каждое из которых выносится в отдельный цикл, заголовок которого совпадает с заголовком исходного цикла.

Расширением приватной переменной цикла называется преобразование, при котором к переменной добавляется измерение, размерность которого равна количеству итераций цикла, а индексом является итерационная переменная цикла. Сужением приватной переменной цикла является преобразование, при котором данная переменная заменяется другой переменной, множество измерений которой является строгим подмножеством измерений исходной переменной, т.е. у сужаемой переменной «удаляется» часть измерений.

Алгоритмы данных преобразований были реализованы в виде программных модулей на языке C++ и встроены в систему SAPFOR. Для исследования применимости реализованных преобразований была выбрана программа LU из набора тестов NAS Parallel Benchmarks, которая в исходном виде не поддаётся эффективному распараллеливанию. Данные преобразования были применены в процессе автоматизированного преобразования и распараллеливания программы LU в системе SAPFOR и позволили получить параллельную версию программы. Суммарно реализованными преобразованиями автоматизированным образом было изменено около 27% (1080 строк) кода программы, ручными преобразованиями было затронуто около 1,3% (50 строк) кода. Данные результаты свидетельствуют о существенном уменьшении трудоёмкости получения параллельной версии программы для пользователя системы.

Для оценки эффективности полученной параллельной версии (далее — SAPFOR-версии) были произведены сравнительные запуски на гибридном вычислительном кластере K10 (ИПМ РАН) вместе с распараллеленными вручную версиями программы LU в модели DVMH (DVMH-версия) и в модели MPI (MPI-версия). При выполнении на CPU SAPFOR-версия показала отставание приблизительно в 2 раза во времени выполнения по сравнению с DVMH-версией при любом числе используемых процессов. По сравнению с MPI-версией SAPFOR-версия показала среднее ускорение на 10%.

При выполнении на GPU SAPFOR-версия показала увеличенное потребление памяти по сравнению с DVMH-версией, что объясняется различием в применённых преобразованиях при получении данных версий программы. Однако в лучшей для обеих версий конфигурации запуска отставание SAPFOR-версии от DVMH-версии во времени выполнения составило менее 5%.

Таким образом, автоматизированными преобразованиями была получена параллельная версия программы, достаточно эффективная по сравнению с распараллеленной вручную версией, особенно при выполнении на GPU. Это свидетельствует о том, что реализованные преобразования расширили возможности системы SAPFOR по автоматизированному распараллеливанию последовательных Фортран-программ. На основании полученных результатов был сделан доклад на конференции «Ломоносовские чтения — 2022», статья к докладу опубликована в сборнике [1].

Литература

- [1] Крюков В. А., Колганов А. С., Гусев Г. Д. Преобразование последовательных Фортран-программ для их распараллеливания на гибридные кластеры // Ломоносовские чтения–2022: научная конференция, факультет ВМК МГУ имени М.В.Ломоносова. Тезисы докладов. - с. 101–102

Нейросетевые методы классификации целей цитирования

Мотриченко Дмитрий Олегович

Кафедра системного программирования

e-mail: d.o.motrichenko@gmail.com

Научный руководитель — к.ф.-м.н. доц. Корухова Людмила Сергеевна

Научный консультант — Недумов Ярослав Ростиславович

Мотивация, с которой авторы в своих работах цитируют других авторов может быть различной. Некоторые цитирования указывают на прямое использование метода, в то время как другие просто служат признанием предшествующей работы. Подобная информация о цели цитирования может быть использована в ряде прикладных задач, например, при исследовательском поиске или в наукометрии, но в современных цифровых библиотеках возможность автоматического определения целей цитирований только начинает появляться. Целью данной работы является усовершенствование существующих методов автоматического определения целей цитирований в научных статьях. Подобная классификация цели цитирования производится на основе некоторого контекста цитирования. В данной работе под контекстом цитирования понимается предложение, содержащее цитирование.

Изначально задача автоматической классификации целей цитирования решалась с помощью составленных вручную признаков цитирования, которые подавались на вход классическим алгоритмам машинного обучения. В последние годы начали активно исследоваться приложения методов глубокого обучения к данной задаче. Основными подходами здесь являются использование архитектуры сети, основанной на рекуррентных нейронных сетях, а также

дообучение языковых моделей на задачу классификации. В работе [1] авторы предложили использовать в рамках первой архитектуры обучение модели на несколько задач. В качестве дополнительных задач было предложено использовать задачу классификации секции статьи, в которой находится предложение, содержащее цитирование, а также задачу классификации того, требуется ли в предложении цитирование или нет. В другой работе [2] для решения задачи классификации целей цитирования авторами была использована модель SciBERT, представляющая собой языковую модель BERT, обученную на текстах научных статей, которую затем дообучали на исходную задачу. Для проведения экспериментов и сравнения качества работы моделей были выбраны два набора данных – ACL-ARC, содержащий 6 различных классов целей цитирования, и SciCite, содержащий 3 класса.

В рамках выпускной квалификационной работы был проведен анализ существующих методов автоматической классификации целей цитирования, выбраны два основных набора данных – ACL-ARC и SciCite, используемые для сравнения качества классификации, и разработаны собственные методы, два из которых показали лучшее качество классификации на наборе данных ACL-ARC, чем предыдущие результаты. Первый метод основан на использовании вложений от SciBERT и функции ошибки взвешенной перекрестной энтропии при обучении модели, основанной на использовании рекуррентных нейронных сетей и обучения на несколько задач. Вторым методом представляет собой результат дообучения языковой модели SciBERT на задачу классификации целей цитирования также с использованием обучения на несколько задач и взвешенной перекрестной энтропии в качестве функции ошибки. Качество классификации этих моделей на наборе данных ACL-ARC составило соответственно 0.712 и 0.721 (метрика классификации – макро-F мера), что превосходит лучшее предыдущее значение на этом наборе данных, которое равнялось 0.710. Некоторые результаты исследований приведены в статье [3].

Литература

- [1] Structural Scaffolds for Citation Intent Classification in Scientific Publications / Arman Cohan, Waleed Ammar, Madeleine van Zuylen, Field Cady // Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). — Minneapolis, Minnesota: Association for Computational Linguistics, 2019. — Pp. 3586–3596. <https://aclanthology.org/N19-1361>.
- [2] Beltagy, Iz. SciBERT: A Pretrained Language Model for Scientific Text / Iz Beltagy, Kyle Lo, Arman Cohan // Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). — Hong Kong, China: Association for Computational Linguistics, 2019. — Pp. 3615–3620. <https://aclanthology.org/D19-1371>.

- [3] Motrichenko, Dmitry. Bag of Tricks for Citation Intent Classification via SciBERT / Dmitry Motrichenko, Yaroslav Nedumov, Kirill Skorniakov // 2021 Ivannikov Ispras Open Conference (ISPRAS). — 2021. — Pp. 120–126.

Исследование и разработка системы предоставления высокопроизводительных кластеров в облачной среде по запросу с поддержкой видеоускорителей

Озерова Татьяна Александровна

Кафедра системного программирования

e-mail: tanyaozeroval318@gmail.com

Научный руководитель — к.ф.-м.н. Корухова Людмила Сергеевна

Научный консультант — Аксенова Елена Леонидовна

Решение проблем хранения и обработки больших объемов данных требует использования значительных вычислительных ресурсов и специализированных сервисов, например Slurm [1]. Распределенная система Slurm применяется для решения большого числа задач в разных научных областях: генетика, медицина, биоинформатика, для модуляции физических процессов. Подобные задачи часто выполняются в облачных средах, где можно разворачивать виртуальные вычислительные кластеры с требуемыми характеристиками. В работе исследуется проблема повышения эффективности взаимодействия с облачными платформами на основе предоставления пользователю возможностей по управлению виртуальными ресурсами (создание, конфигурирование) и установке необходимого программного обеспечения. При использовании Slurm в облаках возникает ряд проблем, связанных с настройкой сервиса:

- Для установки сервиса Slurm в ручном режиме необходимо подготовить кластер из виртуальных машин, связанных по сети.
- Настройка Slurm может занимать большое количество времени и требует от пользователя определенного уровня квалификации и знаний.
- Возникает проблема простоя ресурсов.
- Отсутствуют возможности автоматического воспроизведения решения.
- При увеличении числа узлов в кластере существенно возрастает время и сложность разворачивания виртуального Slurm-кластера.

В рамках выпускной квалификационной работы создана система предоставления виртуальных Slurm-кластеров [2] с поддержкой видеоускорителей по запросу. Для реализации этой задачи выбран инструмент управления конфигурациями Ansible. Он позволяет осуществлять удаленный доступ по ssh протоколу к хостам кластера и настраивать их в соответствии с описанными конфигурациями. Написанная роль была дополнена возможностью разворачивания Slurm совместно с CUDA и интегрирована в проект Michman [3]. Michman - система оркестрации, которая взаимодействует с облачной платформой OpenStack и реализует создание и настройку виртуальных кластеров и сервисов по запросу.

Проведено функциональное тестирование полученного решения, проверяющее возможность управления GPU при помощи распределенного сервиса, а также корректность работы расширения CUDA. Путем сравнения производительности виртуального и физического кластеров с установленным сервисом была получена оценка вычислительной мощности облачного Slurm-кластера [4], исследована зависимость производительности bare-metal и виртуального Slurm-кластеров от числа вычислительных узлов.

Разработка системы предоставления высокопроизводительных кластеров в облачной среде позволила расширить конфигурационные возможности Slurm. Существенно возросла вариативность эксплуатации сервиса: от решения задач сплошных сред до запуска расчетов, использующих GPU. Снижился порог входа для пользователя: развертывание сервиса осуществляется при помощи запроса REST API или графического интерфейса. Реализация автоматического развертывания системы Slurm позволила существенно уменьшить (в сравнении с ручным режимом) время установки диспетчера задач для кластера и обеспечить его настройку с указанным числом рабочих узлов.

Литература

- [1] Slurm – workload manager. Version 20.11.Overview. - <https://slurm.schedmd.com/overview.html>
- [2] Озерова Т. А., Аксенова Е. Л., Корухова Л. С. АВТОМАТИЗАЦИЯ РАЗВЕРТЫВАНИЯ СИСТЕМЫ ПЛАНИРОВАНИЯ ЗАДАЧ -slurm В ОБЛАЧНОЙ СРЕДЕ // Ломоносовские чтения-2021: научная конференция, факультет ВМК МГУ имени М.В.Ломоносова. Тезисы докладов. – Т. 2021 из Секция Вычислительной математики и кибернетики. – М.: М., 2021. – С. 117–118.
- [3] Aksenova E. et al. Michman: an Orchestrator to deploy distributed services in cloud environments //2020 Ivannikov Ispras Open Conference (ISPRAS). – IEEE, 2020. – С. 57-63.
- [4] Ozerova T. et al. Slurm PaaS in a cloud environment for development and debug purposes //2021 Ivannikov Ispras Open Conference (ISPRAS). – IEEE, 2021. – С. 127-133.

Метод увеличения устойчивости моделей распознавания изображений без потери точности на чистых данных

Черепнина Мария Сергеевна

Кафедра системного программирования

e-mail: s02170552@stud.cs.msu.ru

Научный руководитель — к.ф.-м.н. доц. Турдаков Денис Юрьевич

На сегодняшний день глубокое обучение широко применяется для решения различных задач компьютерного зрения. В частности, нейронные сети-классификаторы изображений часто используют как основу для моделей классификации и локализации и для моделей обнаружения объектов [1, 2].

Однако классификаторы изображений на основе глубоких нейронных сетей оказались уязвимы для состязательных атак [3, 4], т.е. небольших (часто — незаметных для человеческого глаза) трансформаций входных изображений, из-за которых меняется предсказание модели классификации с корректного $y_{true} = y_{nat}$ на некорректный $y_{true} \neq y_{adv}$, и итоговая точность модели может снижаться практически до нуля:

$$\begin{aligned} y_{true} &= y_{nat} = f(x_{nat}) \\ y_{adv} &= f(x_{adv}) \neq y_{nat} \\ x_{nat} &\approx x_{adv} \end{aligned} \quad (1)$$

Один из распространенных подходов к увеличению устойчивости моделей — это состязательное обучение [5]. Однако такой подход снижает точность модели на неатакованных входных данных x_{nat} на 10–30%.

Альтернативный способ борьбы с состязательными атаками — это обнаружение и отклонение атакованных входных данных во время работы модели [6]. Обнаружить состязательные данные можно благодаря заметным статистическим различиям между x_{nat} и x_{adv} . Т.е. можно определять, подверглись ли входные данные атаке, либо подобрав эвристику на основе заранее заданных статистик, либо обучив сравнительно небольшую нейронную сеть-классификатор. Методы обнаружения состязательных атак достигают практически единичной точности, однако не позволяют классифицировать атакованные входные данные.

Недавно авторы работы Adversarial Examples Improve Image Recognition [7] показали, что добавление в обучающую выборку состязательных данных и использование двух разных наборов обучаемых параметров слоев BatchNorm — одного для исходных входных данных и другого для атакованных — позволяет увеличить точность моделей классификации на исходных данных. Такой подход дал для сети EfficientNet-B7 прирост точности 0.7% на датасете ImageNet и от 4.8% до 7.0% на разных подмножествах датасета ImageNet.

Данная дипломная работа посвящена разработке нового метода защиты от состязательных атак, объединяющего в себе преимущества двух вышеописанных методов защиты: способность модели классифицировать атакованные изображения и сохранение точности модели на исходных данных. Во многом разработанный метод опирается на подход с разделением слоев BatchNorm, описанный в [7].

В рамках работы была выдвинута гипотеза, что для сохранения точности моделей на чистых данных при состязательном обучении достаточно использовать не отдельный слой BatchNorm, а лишь отдельный набор параметров этого слоя, отвечающих за сдвиг. Данная гипотеза подтвердилась экспериментально на датасетах CIFAR-10 и ImageNet.

В работе предлагается построить устойчивую к атакам модель на основе сети ResNet, заменив в ней обычные слои BatchNorm на слои с двойным набором параметров сдвига. Полученная сеть обучается сразу и на натуральных

данных, и на атакованных, при этом x_{nat} и x_{adv} проходят через разные пути в графе обучения модели.

Для того, чтоб в тестовой фазе определить, какой из наборов параметров BatchNorm применять для вычисления итоговых меток, используется классификатор, представляющий из себя сверточную сеть, принимающую на вход внутреннее представление данных из ResNet и возвращающую вероятность того, что данные подверглись атаке.

Эксперименты на датасете CIFAR-10 и ImageNet показали, что разработанная модель на чистых данных не проигрывает классическим моделям и, в то же время, дает высокую точность на атакованных данных.

Таким образом, в рамках работы:

1. На основе объединения двух подходов к защите от состязательных атак, построен новый метод, позволяющий точно классифицировать как натуральные, так и состязательные изображения.
2. Показано, что для пресечения деградации качества классификатора из-за состязательного обучения, достаточно использовать разные параметры сдвига слоя BatchNorm.

Литература

- [1] YOLOv3: An Incremental Improvement / Joseph Redmon, Ali Farhadi // 2018
- [2] SSD: Single Shot MultiBox Detector / Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, Alexander C. Berg // Computer Vision – ECCV 2016
- [3] Explaining and Harnessing Adversarial Examples / Ian J. Goodfellow, Jonathon Shlens, Christian Szegedy // 2015
- [4] Towards Deep Learning Models Resistant to Adversarial Attacks / Aleksander Madry, Aleksandar Makelov, Ludwig Schmidt, Dimitris Tsipras, Adrian Vladu // 2019
- [5] Adversarial Training for Free! / Ali Shafahi, Mahyar Najibi, Amin Ghiasi, Zheng Xu, John Dickerson, Christoph Studer, Larry S. Davis, Gavin Taylor, Tom Goldstein // 2019
- [6] ML-LOO: Detecting Adversarial Examples with Feature Attribution / Puyudi Yang, Jianbo Chen, Cho-Jui Hsieh, Jane-Ling Wang, Michael I. Jordan Proceedings of the AAAI Conference on Artificial Intelligence 2020
- [7] Adversarial Examples Improve Image Recognition Cihang Xie, Mingxing Tan, Boqing Gong, Jiang Wang, Alan Yuille, Quoc V. Le // 2020

Именной указатель

- Абрамов А. В., 75
Артамонов Г. И., 35
Афракова А. В., 93
Басков Е. С., 110
Бежанян Р. О., 94
Бутиков И. И., 37
Валиахметов Б. И., 23
Величкин М. С., 9
Вересов А. К., 96
Ветрова Е. А., 77
Вешкин А. С., 11
Виляев А. Л., 42
Горбунов С. А., 44
Григорьева М. С., 29
Губанов Ф. Д., 71
Гусев Г. Д., 111
Гущин А. Е., 73
Долматов А. А., 33
Дунаева А. С., 13
Елистратов С. Ю., 54
Ершов Н. И., 98
Ефремцев В. В., 14
Жердев И. П., 88
Ипполитов В. Б., 16
Кабанов Д. М., 18
Казантаев А. Д., 79
Клименко В. К., 46
Коваленко А. П., 81
Колотов А. Е., 19
Кондратьева В. Н., 39
Ларин А. В., 83
Линь Пэйфэн, 61
Миронов Д. Д., 63
Миронов М. И., 59
Михайлов А. Д., 25
Мордвинцев М. К., 26
Мотриченко Д. О., 114
Муляр М. Ю., 48
Назаров Д. И., 65
Озерова Т. А., 116
Пахненко И. П., 49
Петрова Н. М., 67
Плужников И. Р., 90
Рогожин Д. А., 100
Рожков И. С., 102
Рыгин А. С., 27
Самбурский А. И., 56
Селезнёв М. В., 51
Сивцов Д. Т., 74
Степанова П. А., 104
Танкаев И. Р., 85
Тарасова В. А., 69
Тихвинский Д. Г., 21
Тыцкий В. И., 58
Федоров В. В., 92
Федоров Ф. А., 41
Хомский Д. В., 105
Хуторной Д. П., 52
Цветинский В. Ю., 31
Черепнина М. С., 117
Чехонина Е. А., 86
Шевернев Ю. А., 107
Яковлева С. А., 108