

МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
имени М.В. ЛОМОНОСОВА

Факультет вычислительной математики и кибернетики

**СБОРНИК ТЕЗИСОВ
ЛУЧШИХ ВЫПУСКНЫХ
КВАЛИФИКАЦИОННЫХ РАБОТ
ФАКУЛЬТЕТА ВМК МГУ
2024 ГОДА**

Москва
2024

Математическое моделирование конкурирующего участия волн второй гармоники в задаче утроения частоты лазерного излучения

Аксенов Павел Николаевич

Кафедра математической физики

e-mail: s02200316@gse.cs.msu.ru

Научный руководитель — к.ф.-м.н. асс. Харитонов Дмитрий Михайлович

В работе исследуется процесс генерации третьей гармоники (ГТГ), который имеет важное значение в современной оптике и лазерных технологиях и применяется в таких областях, как медицинская диагностика, материаловедение и фотоника. Проблема заключается в том, что конкурирующие явления, такие как участие волн второй гармоники, могут существенно влиять на динамику и эффективность процесса ГТГ.

Основной целью работы является оценка влияния слабой второй гармоники на процесс генерации третьей гармоники. В работе рассматривается система нелинейных дифференциальных уравнений, описывающих взаимодействие волн в среде с квадратичной нелинейностью.

Для анализа динамики волновых процессов был применен метод многих масштабов, что позволило свести исходную систему к системе, допускающей аналитического исследования, далее - модифицированная система. В приближении длинных импульсов были разработаны конечно-разностные схемы для численного моделирования исходной и модифицированной систем, которые учитывали выведенные инварианты систем, описаны итерационные процессы для решения систем уравнений. Было показано, что модифицированная система в целом точно аппроксимирует исходную систему. В приближении длинных импульсов также были получены точные решения модифицированной системы уравнений, проведено сравнение с решениями исходной системы, что позволило оценить их применимость. Точные решения оказались верно решающими модифицированную систему. Сравнивая системы со слабой второй гармоникой и без учёта этой волны, для разных кристаллов и разной начальной интенсивности волны первой гармоники получили одинаковый результат: волна второй гармоники хоть и слабая, но оказывает довольно существенное влияние на максимальную амплитуду волны третьей гармоники. Тем самым была показана необходимость учета обеих гармоник второй волны в процессе ГТГ.

В результате проведенного моделирования подтверждены теоретические предположения о значительном влиянии волн второй гармоники на процесс утроения частоты. Полученные результаты показали, что учет обеих волн второй гармоники необходим для правильного описания и моделирования процесса ГТГ в лазерных системах.

Существование точек совпадения у конечного набора многозначных отображений метрических пространств

*Работа удостоена диплома 3 степени
Бугрименко Артём Александрович*

*Кафедра общей математики
e-mail: art.bugrim@gmail.com*

Научный руководитель — д.ф.-м.н. проф. Фоменко Татьяна Николаевна

Введение. Теоремы о неподвижных точках и точках совпадения отображений играют важную роль в различных задачах функционального анализа, дифференциальных и интегральных уравнений и других областей математики. В 2014 году Rakesh Batra, Sachin Vashistha и Kumar Rajesh в работе [1] обобщили понятие F-сжатия из работы [2], определив (F-g)-сжимающие отображения и доказав соответствующую теорему о совпадении двух однозначных отображений.

В работе сформулировано и доказано несколько вариантов обобщений теоремы из работы [1] на случай конечного набора многозначных отображений.

Теоретические сведения. Пусть (X, ρ) — метрическое пространство, тогда $Com(X)$ — совокупность непустых компактных подмножеств X .

Определение 1. Пусть $(X, \rho), (Y, \bar{\rho})$ — метрические пространства, $g_1, g_2, \dots, g_m: X \rightarrow Com(Y), m \geq 2$ — многозначные отображения. Точка $x_* \in X$ такая, что $\bigcap_{i=1}^m g_i(x_*) \neq \emptyset$, называется *точкой совпадения (совпадением)* набора многозначных отображений g_1, g_2, \dots, g_m .

Определение 2. Пусть функция $F: \mathbb{R}_+ \rightarrow \mathbb{R}$ удовлетворяет условиям:

- (F1) F строго возрастает, т.е. $(F(a) < F(b)) \iff (a < b)$.
- (F2) Любая последовательность $\{a_n\}_{n \in \mathbb{N}}$ положительных чисел сходится к 0 тогда и только тогда, когда $\lim_{n \rightarrow \infty} F(a_n) = -\infty$.
- (F3) Существует $k \in (0, 1)$ такое, что $\lim_{a \rightarrow 0+} a^k F(a) = 0$.

Пусть $(X, \rho), (Y, \bar{\rho})$ — метрические пространства, $g: X \rightarrow Com(Y)$ — некоторое многозначное отображение. Тогда многозначное отображение $T: X \rightarrow Com(Y)$ называется *(F-g)-сжатием*, если существует $\tau > 0$ такое, что $\tau + F(D(Tx, Ty)) \leq F(D(g(x), g(y)))$ для всех $x, y \in X$ таких, что $Tx \neq Ty$ и $g(x) \neq g(y)$, где $D(A, B)$ — метрика Хаусдорфа в пространстве $Com(Y)$

Определение 3. Пусть (X, ρ) — метрическое пространство. Тогда *диаметром множества* $A \subset X$ называется величина $diam(A) = \sup_{x, y \in A} (\rho(x, y))$.

Основные результаты. Конечный набор из $m \geq 2$ многозначных отображений g_1, g_2, \dots, g_m компонуется в два новых отображения T и G следующим образом:

$$\begin{cases} T(x) = (g_1(x), \dots, g_1(x)) \subseteq \text{Com}(Y^{m-1}), x \in X; \\ G(x) = (g_2(x), \dots, g_m(x)) \subseteq \text{Com}(Y^{m-1}), x \in X \end{cases} \quad (1)$$

В пространстве Y^{m-1} вводится метрика ρ' :

$$\rho'((u_1, u_2, \dots, u_{m-1}), (v_1, v_2, \dots, v_{m-1})) = \sum_{i=1}^{m-1} d'(u_i, v_i), \quad u_i, v_i \in Y, i = 1, \dots, m-1,$$

где d' — метрика в пространстве Y . Метрика ρ' индуцирует метрику Хаусдорфа в пространстве $\text{Com}(Y^{m-1})$, которую далее будем обозначать D' . Доказано следующее утверждение.

Теорема 1. Пусть (X, ρ) , (Y, d') — метрические пространства, пусть также $g_1, \dots, g_n : X \rightarrow \text{Com}(Y)$ многозначные отображения с компактными образами. Пусть $T, G : X \rightarrow \text{Com}(Y^{m-1})$ — отображения (1), причём T является $(F-G)$ -сжатием для некоторой функции $F : \mathbb{R}_+ \rightarrow \mathbb{R}$, удовлетворяющей условиям $(F1)$ - $(F3)$, а $G(X)$ — полное подпространство в метрическом пространстве $(\text{Com}(Y^{m-1}), D')$. Пусть $T(X) \subseteq G(X)$. Предположим также, что для любых $x, y \in X$ таких, что $T(x) \cap G(y) \neq \emptyset$, найдется такой $z \in X$, что $T(y) \cap G(z) \neq \emptyset$, и существуют константы $L > 0, \alpha > 0$, такие, что выполняются неравенства:

$$D'(T(x), T(y)) \geq D'(G(y), G(z)), \quad \text{diam}(g_1(z)) \leq L(D'(G(y), G(z)))^\alpha$$

Тогда существует точка $x_* \in X$ такая, что $T(x_*) \cap G(x_*) \neq \emptyset$. (т.е. $g_1(x_*) \cap g_2(x_*) \cap \dots \cap g_m(x_*) \neq \emptyset$ (иначе говоря, $g_1(x_*) \in \bigcap_{k=2}^m g_k(x_*)$)).

В работе также предложены следствия теоремы 1, в частности, верны следующие утверждения:

Следствие 1. Пусть выполнены условия теоремы 1, причем g_1 однозначное отображение.

Тогда отображения g_1, \dots, g_m имеют точку совпадения. (т.е. существует точка $x_* \in X$ такая, что $g_1(x) \cap g_2(x_*) \cap \dots \cap g_m(x_*) \neq \emptyset$)

Следствие 2. Пусть $X = Y$, $\rho = d'$, выполнены условия следствия 1, причем отображение g_1 является тождественным, то есть $g_1 = id_X$. Тогда отображения g_2, \dots, g_m имеют общую неподвижную точку, т.е. существует точка $x_* \in X$ такая, что $x_* \in g_2(x_*) \cap \dots \cap g_n(x_*)$.

ЛИТЕРАТУРА

- [1] Rakesh Batra, Sachin Vashistha, Kumar Rajesh. Coincidence Point Theorem for a New Type of Contraction on Metric Spaces // Int. Journal of Math. Analysis. 2014. Vol.8, №27. P. 1315-1320.
- [2] Wardowski D. Fixed points of a new type of contractive mappings in complete metric spaces // Fixed Point Theory and Applications. 2012. №94.

Нейросетевой нелокальный метод для задачи совмещения биомедицинских изображений

Дроздов Никита Александрович

Кафедра математической физики

e-mail: drozdovna@my.msu.ru

Научный руководитель — к.ф.-м.н. с.н.с. Сорокин Дмитрий Васильевич

Совмещение – одна из ключевых задач в области обработки биомедицинских изображений. С распространением глубокого обучения появилось много работ, в которых совмещение производится при помощи нейронных сетей [1]. Однако типичные для медицины и биологии данные (например, микроскопия или МРТ мозга) имеют большое пространственное разрешение, и для обучения моделей на таких данных требуются значительные вычислительные ресурсы. В данной работе предложен нейросетевой нелокальный метод для совмещения биомедицинских изображений. Его главной особенностью является возможность производить совмещение на разных разрешениях без существенной потери точности. Предложенную модель можно обучать на данных в уменьшенном разрешении, что позволяет потреблять меньше ресурсов в ходе обучения.

Нейросетевой метод совмещения на основе нейронного оператора Фурье (Fourier Neural Operator, FNO). Поставим задачу совмещения более формально. Пусть $\mathbf{x} = (x_1, x_2) \in \Omega \subset \mathbb{R}^2$, $f(\mathbf{x})$ – целевое изображение (fixed image), $m(\mathbf{x})$ – совмещаемое изображение (moving image), $m_\phi(\mathbf{x})$ – совмещенное изображение (moved image), $\phi(\mathbf{x}) = (\phi_1(\mathbf{x}), \phi_2(\mathbf{x}))$ – поле деформации. Функционал потерь для непрерывной задачи совмещения имеет вид

$$\hat{L}_{f,m}(\phi) = \int_{\Omega} D_{f,m_\phi} + \lambda \left(\left[\frac{\partial \phi_1}{\partial x_1} \right]^2 + \left[\frac{\partial \phi_1}{\partial x_2} \right]^2 + \left[\frac{\partial \phi_2}{\partial x_1} \right]^2 + \left[\frac{\partial \phi_2}{\partial x_2} \right]^2 \right) dx_1 dx_2,$$

где D_{f,m_ϕ} некоторым образом вычисляет «разницу» между изображениями f и m_ϕ в заданной точке. Для данного функционала система уравнений Эйлера-Лагранжа выглядит следующим образом:

$$\frac{\partial D_{f,m_\phi}}{\partial \phi_1} - 2\lambda \cdot \Delta \phi_1 = 0; \quad \frac{\partial D_{f,m_\phi}}{\partial \phi_2} - 2\lambda \cdot \Delta \phi_2 = 0. \quad (1)$$

Систему (1) можно решать классическими численными методами. Однако в настоящее время существуют нейросетевые модели для решения уравнений в частных производных. Одна из таких моделей называется Fourier Neural Operator [2]. Принцип работы FNO состоит в приближенном нахождении решения операторного уравнения $Lu = f$ на области Ω в виде результата некоторого конечного итерационного процесса. Основной компонент этого процес-

са — сверточный оператор $(Ku_t)(\mathbf{x}) := \int_{\Omega} \kappa_t(\mathbf{x} - \mathbf{y})u_t(\mathbf{y})d\mathbf{y}$, соответствующий сверточному блоку нейронной сети с ядром κ_t . К этому оператору можно применить теорему о свертке и параметризовать в частотной области. Таким образом можно моделировать нелокальные сверточные ядра, благодаря которым модель становится устойчивой к входному разрешению изображений.

В работе была предложена модификация стандартной архитектуры FNO для задачи совмещения, названная FNOReg. Обе модели на основе FNO вместе с классическими сверточными архитектурами для совмещения были протестированы на двумерных и трехмерных снимках МРТ мозга из набора данных OASIS-1 [3]. Для демонстрации устойчивости разных моделей ко входному разрешению все модели обучались на исходном разрешении набора данных и разрешении, пониженном в 2 раза. Тестирование производилось только на оригинальном разрешении данных. Результаты работы различных моделей на трехмерном наборе снимков представлены в таблице 1.

Модель	#параметров	Dice на 1x	Dice на 0.5x	Изменение Dice на 0.5x
Initial (only affine)	-	0.572	0.572	-
Fourier-Net (16 channels)	4.5M	0.814	0.614	-24.57%
Fourier-Net (32 channels)	17.8M	0.822	0.619	-24.70%
VoxelMorph	274K	0.834	0.737	-11.63%
VoxelMorph-Large	15.2M	0.863	0.650	-24.68%
VoxelMorph-Huge	63.3M	0.849	0.722	-14.96%
TransMorph	46.8M	0.867	-	-
FNOReg	98.1M	0.830	0.808	-2.65%

Таблица 1: Результаты работы различных моделей совмещения на тестовой части набора данных 3D OASIS. 1x и 0.5x означают обучение на оригинальном и пониженном разрешении соответственно.

Удалось показать, что потери качества при обучении на изображениях меньшего размера у архитектуры на основе FNO составили менее 3%, а у классических сверточных моделей — около 10–25%. При этом модель FNOReg обучалась более стабильно, чем классическая модель FNO (последняя не включена в таблицу 1 из-за сильной нестабильности процесса обучения на данном наборе снимков).

ЛИТЕРАТУРА

- [1] Balakrishnan G. et al. Voxelmorph: a learning framework for deformable medical image registration //IEEE transactions on medical imaging. – 2019. – Т. 38. – №. 8. – С. 1788-1800.
- [2] Li Z. et al. Fourier neural operator for parametric partial differential equations //arXiv preprint arXiv:2010.08895. – 2020.
- [3] Open Access Series of Imaging Studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults / Marcus D. S., Wang T. H., Parker J., Csernansky J. G., Morris J. C. and Buckner R. L. // J. Cogn. Neurosci. — 2007. — Sep. — Vol. 19, no. 9. — P. 1498–1507

Мозаично-скелетонные аппроксимации и малоранговые структуры для ускорения стохастических и классических алгоритмов численного моделирования агрегации-дробления вещества

Дьяченко Роман Романович

Кафедра вычислительных технологий и моделирования

e-mail: dyachenko.rr@yandex.ru

Научный руководитель — к.ф.-м.н. доц. Матвеев Сергей Александрович

Неупругие соударения хаотически движущихся и агрегирующих частиц лежат в основе самых разнообразных физических явлений [1]. Эволюция концентраций частиц различных размеров определяется балансом процессов агрегации и фрагментации соударяющихся частиц. В случае однородных систем этот процесс описывается системой уравнений Смолуховского [2]:

$$\frac{dn_s}{dt} = \frac{1}{2} \sum_{i+j=s} K_{ij} n_i n_j - \sum_{j=1}^{\infty} K_{sj} n_s n_j,$$

где n_s - концентрация (т.е. количество частиц в единице объема) агрегатов состоящих из s "элементарных" частиц (мономеров).

В работе [3] предложен быстрый численный метод, позволяющий снизить асимптотику алгоритмической сложности шага разностной схемы предиктор-корректор с $O(M^2)$ до $O(M \log M)$ арифметических операций, где M - число уравнений в конечной системе дифференциальных уравнений. Этот метод основывается на применении малоранговых матричных аппроксимаций и быстрых алгоритмов линейной алгебры к ядру агрегации K_{ij} . В работе [6] также предлагается быстрый метод Монте-Карло моделирования для систем, имеющих функцию интенсивности взаимодействия - представляющуюся в малоранговом виде. Сложность этого алгоритма $O(\log M)$ операций для обработки одного инцидента слияния частиц. Более того, этот подход не требует, чтобы само ядро было малоранговым или имело хорошую малоранговую аппроксимацию, вместо этого достаточно ограничить его сверху некоторым малоранговым ядром. Однако на практике существует множество ядер, которые не поддаются прямой аппроксимации малым рангом. Например, имеющие в своей записи модуль разности $K_{ij} = |i - j|$ и некоторые другие [2, 4].

В данной работе мы предлагаем алгоритм для быстрого решения уравнений Смолуховского с ядрами из более широкого класса, чем малоранговые. Применение мозаично-скелетонных аппроксимаций [5] приводит к сложности шага разностной схемы в $O(M \log^2 M)$ арифметических операций. Для стохастического алгоритма к сложности $O(\log^2 M)$ обращений к памяти для обработки одного процесса слияния. Это позволяет решать экстремально большие системы ОДУ порядка $M \equiv 2^{20}$ с широким классом ядер K_{ij} на персональном компьютере.

Найденные решения согласуются с результатами альтернативного подхода, основанного на Монте-Карло сэмплировании с мажорантой. Преимуществом по сравнению с ним является априорно задаваемая более высокая точность.

ЛИТЕРАТУРА

- [1] Галкин В.А. Уравнение Смолуховского. М.: Физматлит, 2001
- [2] Leyvraz F. Scaling theory and exactly solved models in the kinetics of irreversible aggregation // Physics Reports. 2003. 383, N 2–3. 95–212
- [3] Matveev S. A., Smirnov A. P., Tyrtysnikov E. E. A fast numerical method for the Cauchy problem for the Smoluchowski equation // Journal of Computational Physics. – 2015. – Т. 282. – С. 23-32.
- [4] Zagidullin R. R. et al. Aggregation kinetics at sedimentation: the impact of particles diffusion // arXiv preprint arXiv:2112.14584. – 2021.
- [5] Tyrtysnikov E. Mosaic-skeleton approximations // Calcolo. – 1996. – Т. 33. – С. 47-57.
- [6] Osinsky A. I. Low-rank Monte Carlo for Smoluchowski-class equations // Journal of Computational Physics. – 2024. – Т. 506. – С. 112942.

Проекционный метод восстановления негладких волновых фронтов с локализованным выбором параметров стабилизатора дробной гладкости

Работа удостоена диплома 2 степени

Горохов Сергей Александрович

Кафедра математической физики

e-mail: s02200277@gse.cs.msu.ru

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

В настоящее время для коррекции зрения применяются мультифокальные интраокулярные линзы (ИОЛ). Однако использование таких линз может вызвать нежелательные зрительные искажения и затруднить получение чёткого и обширного обзора сетчатки при офтальмоскопии и оптическом исследовании. Изучение оптических свойств мультифокальных ИОЛ требует анализа сложных карт аберраций в многослойной структуре ИОЛ. Для решения этой задачи можно применить проекционный метод восстановления волновых фронтов по их наклонам, исследованный в работах [1],[2]. В работе [2] исследован метод с глобальным выбором параметров стабилизатора дробной гладкости γ и s .

В рамках настоящей работы рассматриваются мультифокальные и спиральные волновые фронты, одна из версий которых взята из работы [3]. Были разработаны модификации проекционного метода с локальным выбором γ и s . Задача состоит в том, чтобы для выбранных классов волновых фронтов получить матрицы значений этих параметров, которые будут давать наилучшее качество восстановления при разных долях зашумления исходных данных.

Поставленная задача решалась с помощью трех разных способов поиска: *Optuna*, генетические алгоритмы и оптимизатор *Adam*. В работе была произведена реализация данных методов применительно к рассмотренной задаче и тестирование полученных значений. Было выполнено исследование применимости этих значений и сделаны выводы о том, какие есть преимущества и недостатки у рассмотренных методов.

Из трех рассмотренных методов поиска матриц значений, оптимизатор *Adam* показал наилучшие результаты в рассмотренных метриках качества восстановления таких, как *MSE* и *DSSIM*. *Adam* обгоняет по качеству версию без стабилизатора при доле зашумления равной 2.5%, в то время, как метод с глобальным выбором параметров, *Optuna* и генетические алгоритмы делают это при 10%, если рассматривать метрику *MSE*. Таким образом, был получен баланс между качественным восстановлением волновых фронтов без зашумления, то есть близких к точным, где метод без стабилизатора дает почти идеальный результат и сильно зашумленным, где уже важно использование стабилизатора дробной гладкости. Полученные таким методом матрицы значения параметров γ и s оказались применимы только в рамках класса волновых фронтов, на котором был обучен метод, то есть значения для мультифокальных плохо работают для спиральных и наоборот. В случае спиральных волновых фронтов, модификации исходных данных такие, как поворот и изменение направления закручивания, не ухудшают качество восстановления.

Работа выполнена при поддержке МНОШ МГУ «Фотонные и квантовые технологии. Цифровая медицина», соглашение № 23-Ш06-20.

ЛИТЕРАТУРА

- [1] Razgulin A. V., Kuzhamaliyev Y. Z., Goncharov A. S., Larichev A. V. A variational method for wavefront reconstruction from Shack-Hartman sensor measurements // *Atmospheric and Oceanic Optics* (2017) 30, № 4, p. 399-403.
- [2] Разгулин А. В, Ирошников Н.Г., Ларичев А.В., Турганбаев С.А., Романенко Т.Е. Оценки точности проекционного метода со стабилизатором дробной гладкости в задаче восстановления волнового фронта по его наклонам // *Дифференциальные уравнения*. 2022. Том 58. № 7. С. 995-1008.
- [3] Kołodziejczyk A, Bará S, Jaroszewicz Z, Sypek M. The Light Sword Optical Element—a New Diffraction Structure with Extended Depth of Focus // *Journal of Modern Optics*. 2007. Vol 37, no. 8, 1283-1286.

Нейросетевой метод нахождения оксидного состава металла по кривой неизотермического газовыделения

Работа удостоена диплома 3 степени

Кнорре Дмитрий Дмитриевич

Кафедра математической физики

e-mail: knorre.dmitry.junior@gmail.com

Научный руководитель — д.ф.-м.н. проф. Крылов Андрей Серджевич

Оксидные включения в металле могут сильно влиять на его рабочие свойства, при этом задача определения оксидного состава металла важна не только для контроля качества металлов, но и для управления процессом плавки. Фракционный Газовый Анализ (ФГА) - один из основных методов анализа состава оксидных включений. Для проведения ФГА образец металла помещается в печь, где плавится при постепенно повышающейся температуре, при этом фиксируется скорость выделения угарного газа из образца. Так как при разложении оксидных включений выделяется CO , и различные оксиды начинают раскладываться при достижении разной температуры, полученные кривые зависимости температуры и скорости выделения CO из образца от времени являются характеристиками оксидного состава металла. В данной работе был разработан алгоритм для интерпретации результатов ФГА, позволяющий определить для каждого типа оксидного включения долю выделения CO , обусловленную разложением оксида данного типа, что позволяет оценить количество содержания включений данного типа в образце.

Используемые модели. Разработанный алгоритм основан на концепции Нейронных Сетей Учитывающих Физическую Информацию ("Physics-Informed Neural Networks PINN) – подходе, в рамках которого нейронная сеть обучается предсказывать значение функции в точках некоторой области так, чтобы она удовлетворяла дифференциальному уравнению. В работе [1] было выведено дифференциальное уравнение (1) моделирующее процесс выделения CO из образца при его неизотермическом плавлении.

$$V'(t) = V(t)(f'(t) - e^{f(t)}). \quad (1)$$

В этом уравнении $V(t)$ - функция зависимости скорости выделения CO из образца от времени, $f(t) = K - \frac{E}{T(t)}$, $T(t)$ - функция зависимости температуры от времени, K и E термодинамические характеристики оксидного включения. K и E значительно отличаются для различных оксидных включений и несколько отличаются для оксидных включений одного типа для каждого образца. Использование PINN позволяет учитывать данную модель при поиске функций составляющих выделение CO из оксидных включений разного типа.

Предложенный алгоритм. Алгоритм позволяет по набору кривых газовыделения определять скорость выделения CO из оксидных включений каждого типа в каждый момент времени. Для этого полносвязная нейронная сеть обучается предсказывать значения функций зависимости скорости

выделения CO из оксидных включений от времени такие, что каждая из предсказанных функций удовлетворяет дифференциальному уравнению (1), а в сумме они соответствуют кривым ФГА. Обучение происходит при помощи алгоритма adams. Кроме весов нейронной сети при помощи оптимизатора Adam [2] для каждого типа оксидных включений содержащегося в образце обучаются параметр E и параметр T^{max} , из которого по формуле (2) считается параметр K .

$$K = \frac{E}{T^{max}} + \log\left(\frac{E}{(T^{max})^2}\right). \quad (2)$$

Алгоритм применяется в двух режимах: сначала для каждого нового состава металла в течение большого числа эпох обучается базовый алгоритм. Затем для каждой новой кривой ФГА для этого типа металла соответствующая базовая модель дополнительно обучается в течение маленького числа эпох.

Результаты. Предложенный алгоритм был реализован на языке Python3 с использованием библиотеки DeepXDE [3]. Далее алгоритм был протестирован на результатах ФГА для металлов двух разных составов. Для имеющихся кривых ФГА была проведена экспертная оценка количества кислорода связанного в некоторых типах оксидных включений. Значения полученные при помощи экспертной оценки в достаточной мере соответствуют значениям предсказанным предложенным алгоритмом.

ЛИТЕРАТУРА

- [1] Григорович К. В., Красовский П. В., Исаков С. А., Горохов А. А., Крылов А. С. Обработка и интерпретация результатов фракционного газового анализа // Заводская лаборатория. Диагностика материалов. 2002. Р. 3-9
- [2] Kingma Diederik P., Ba Jimmy Adam: A Method for Stochastic Optimization // arXiv preprint arXiv:1412.6980
- [3] Lu Lu, Meng Xuhui, Mao Zhiping, Karniadakis George Em DeepXDE: A Deep Learning Library for Solving Differential Equations // SIAM Review. 2021. Р. 208-228

Корректная реализация алгоритма регуляризирующего разложения для рациональных матриц на языке MATLAB

Ковч Николай Сергеевич

Кафедра общей математики

e-mail: nr3ijs@gmail.com

Научный руководитель — д.ф.-м.н. проф. Икрамов Хаким Дододжанович

В современных системах компьютерной алгебры отсутствуют инструменты для проверки конгруэнтности матриц. Данная работа направлена на устранение этого недостатка.

Конгруэнтные преобразования. Мы будем рассматривать квадратные матрицы порядка n над полями \mathbb{Q} и $\mathbb{Q}(i)$. Ограничения на природу элементов связаны с тем, что при работе с действительными числами невозможно избежать вычислительных погрешностей, а для проверки конгруэнтности необходима безошибочная арифметика.

Две произвольные квадратные матрицы A и B порядка n называются *конгруэнтными*, если существует невырожденная матрица S порядка n , такая что

$$A = S^T B S, \quad (A = S^* B S).$$

Сам переход от B к A называется *T-конгруэнцией* (**-конгруэнцией*).

Регуляризующее разложение. В статье [1] предложено следующее промежуточное представление при построении канонической формы относительно конгруэнций:

$$A_R \oplus M, \quad M = J_1^{[m_1-m_2]} \oplus J_2^{[m_2-m_3]} \oplus \dots \oplus J_{q-1}^{[m_{q-1}-m_q]} \oplus J_q^{[m_q]}.$$

Это представление авторы (Сергейчук и Хорн) назвали *регуляризующим разложением*. Оно разбивает матрицу на регулярную составляющую (невырожденная матрица A_R) и сингулярную (прямая сумма нильпотентных жордановых блоков M).

Для того, чтобы проверить конгруэнтность вырожденных матриц A и B необходимо выполнить следующую последовательность действий:

1. Построить регуляризующие разложения для матриц A и B :

$$A_R \oplus M, \quad B_R \oplus N.$$

2. Если $M \neq N$ (с точностью до порядка слагаемых), то A и B не конгруэнтны.
3. Если $M = N$ (с точностью до порядка слагаемых), то для конгруэнтности матриц A и B необходимо и достаточно конгруэнтности регулярных частей A_R и B_R .

Результаты. Найдены и устранены ошибки, допущенные авторами [1] при описании и обосновании алгоритма. Реализован корректный алгоритм регуляризующего разложения для рациональных матриц.

ЛИТЕРАТУРА

- [1] Horn R. A., Sergeichuk V. V. A regularization algorithm for matrices of bilinear and sesquilinear forms // Linear Algebra Appl. 2006. N 412. P. 380–395.
- [2] Икрамов Х. Д. Как проверить конгруэнтность заданных квадратных матриц? // Зап. научн. сем. ПОМИ. 2015. № 439. С. 99–106.

Сравнение алгоритмов декомпозиции динамически адаптивных локально измельчаемых расчётных сеток

Работа удостоена диплома 2 степени

Кузьменко Илья Дмитриевич

Кафедра вычислительных методов

e-mail: ilyehakuzmenko@gmail.com

Научный руководитель — д.ф.-м.н проф. Якововский Михаил Владимирович

В работе представлено сравнение двух алгоритмов декомпозиции динамически адаптивных локально измельчаемых расчетных сеток: послойного алгоритма [1], разработанного в ИПМ имени М. В. Келдыша РАН, в основе которого лежит поэтапное послойное формирование доменов, и алгоритма, использующего библиотеки METIS и ParMETIS [2].

Одним из перспективных подходов к численному моделированию широкого круга задач механики сплошной среды является использование динамически адаптивных расчётных сеток. Эффективность их использования на высокопроизводительных вычислительных системах существенно зависит от качества алгоритмов адаптивной декомпозиции расчетных сеток, обеспечивающих преимущественное сохранение данных на тех процессорах, на которых они обрабатывались на прошлых шагах расчета. Под качеством работы в данном случае имеется в виду не только равномерность распределения данных между процессорами и минимальность объёмов данных, передаваемых между процессорами на каждом шаге расчёта модельного времени. Принимается во внимание и уровень затрат времени на многократное вычисление самих декомпозиций, соответствующих каждому новому виду расчётной сетки, получаемому после каждого акта динамической адаптации.

Для первоначального разбиения расчетной сетки была использована функция METIS_PartGraphRecursive из библиотеки METIS, предназначенная для разбиения нерегулярного графа на заданное количество частей на основе метода многоуровневой рекурсивной бисекции. Для перераспределения существующего разбиения расчетной сетки применяется функция ParMETIS_V3_AdaptiveRepart из библиотеки ParMETIS, предназначенная для адаптивного разбиения.

Для сравнения этих алгоритмов рассмотрены следующие метрики на двух этапах: на первом этапе, на котором производились расчеты и передачи данных процессорам-соседям, рассматривалось время расчета шага модельного времени и время передачи данных интерфейсных ячеек (1), на втором этапе, на котором производилось переразбиение сетки после очередного акта динамической адаптации, рассматривалось время вычисления новой декомпозиции и время на соответствующее перераспределение ячеек расчётной сетки между процессорами (2):

$$Time_calc_i = KNeigh_i \cdot t_{latency} + n_inter_i \cdot 20 \cdot 8 \cdot t_{byte} + n_ec_i \cdot t_{calc}, \quad (1)$$

$$Time_red_i = K_transfer_i \cdot t_{latency} + ec_trans_i \cdot 20 \cdot 8 \cdot t_{byte}, \quad (2)$$

где $Time_calc_i$ – затраты времени i -го процессора на этапе вычислений, $KNeigh_i$ – количество процессоров-соседей у i -го процессора, $t_{latency}$ – латентность, n_inter_i – количество интерфейсных элементарных ячеек i -го процессора, t_{byte} – время передачи одного байта, n_ec_i – количество элементарных ячеек i -го процессора, t_{calc} – время выполнения расчетов в элементарной ячейке, $Time_red_i$ – затраты времени i -го процессора на этапе перераспределения, $K_transfer_i$ – количество доменов, которым необходимо передать данные, ранее расположенные на i -ом процессоре, ec_trans_i – количество передаваемых элементарных ячеек i -го процессора. Считается, что в каждой элементарной ячейке находится 20 величин (температура, плотность, давление и т. д.), каждая из которых занимает 8 байт – отсюда множители во вторых слагаемых.

Рассмотрено два случая: редкого и частого выполнения динамической адаптации. При частом выполнении актов динамической адаптации (при малом числе шагов между ними), конфигурация расчетной сетки преимущественно изменяется мало, что обеспечивает лучше качество работы алгоритмов динамической балансировки.

Выводы. Разработан алгоритм, использующий библиотеки METIS и ParMETIS, и программное обеспечение, обеспечивающее сравнительное тестирование двух методов разбиения динамически адаптивных расчетных сеток, на основе которого проведено сравнение ряда вариантов разбиений. По результатам сравнения выявлено, что алгоритм METIS/ParMETIS и послойный алгоритм являются конкурентоспособными, поскольку результаты затрат времени оказались близки. Следует отметить, что результаты работы этого алгоритма иногда могут иметь дефект, при котором множество вершин, принадлежащих домену, не является связным. В послойном алгоритме такие артефакты появляются реже, причём характер их появления является детерминированным по построению алгоритма. Также следует отметить, что при использовании библиотеки ParMETIS число доменов, между которыми необходимо выполнить перераспределение, должно совпадать с числом процессоров, на которых алгоритм выполняется, что ограничивает возможность выполнения длительных многосеансовых расчётов на различном числе процессоров, а при использовании послойного алгоритма такого ограничения нет. Подготовлено дополнение к документации, поясняющее как именно необходимо хранить и передавать данные при работе с библиотеками METIS/ParMETIS и содержащее инструкцию для работы при произвольном распределении ячеек сетки между доменами.

ЛИТЕРАТУРА

- [1] Корнилина М. А., Якобовский М. В. Оценка накладных расходов при выполнении расчётов на локально измельчаемых сетках. Препринты ИПМ им. М.В.Келдыша. 2022. № 102. 36 с.
- [2] Karypis G. Family of Graph and Hypergraph Partitioning Software. URL: <http://glaros.dtc.umn.edu/gkhome/views/metis/>

Комбинация методов объемных и поверхностных интегральных уравнений в задачах электромагнитного рассеяния на структурах малой толщины

Работа удостоена диплома 1 степени

Масс Илья Александрович

Кафедра вычислительных технологий и моделирования

e-mail: massia@my.msu.ru

Научный руководитель — д.ф.-м.н. проф. Сетуха Алексей Викторович

Постановка задачи. Рассматривается трехмерная задача рассеяния монохроматической электромагнитной волны на системе идеально проводящих поверхностей (замкнутых или разомкнутых), и диэлектриков, имеющих вид плоских слоев малой толщины. Ищутся электрическое и магнитные поля в виде $\mathbf{E}(x, t) = \mathbf{E}(x)e^{-i\omega t}$, $\mathbf{H}(x, t) = \mathbf{H}(x)e^{-i\omega t}$. Решаются уравнение Максвелла в частотной области

$$\begin{cases} \operatorname{rot} \mathbf{E} = i\omega\mu\mathbf{H}, \\ \operatorname{rot} \mathbf{H} = -i\omega\varepsilon\mathbf{E}, \end{cases}$$

В качестве первичного поля рассматривается плоская волна, Полное поле является суперпозицией первичного и отраженного: $\mathbf{E} = \mathbf{E}_{sc} + \mathbf{E}_{inc}$. На поверхности метала выполняется условие отсутствия касательной компоненты электрического поля

$$\mathbf{n} \times \mathbf{E} = 0, \quad (1)$$

вторичное поле удовлетворяет условию Зоммерфельда на бесконечности

$$\mathbf{E}_{sc}(x) = O\left(\frac{1}{|x|}\right), \quad \frac{\partial \mathbf{E}_{sc}(x)}{\partial |x|} - ik\mathbf{E}_{sc}(x) = o\left(\frac{1}{|x|}\right), \quad |x| \rightarrow \infty. \quad (2)$$

Интегральное представление. В рассматриваемой задаче полное электрическое поле \mathbf{E} можно искать в виде [1]:

$$\mathbf{E} = \mathbf{E}_{inc} + \mathcal{K}[\Omega, (\varepsilon' - 1)\mathbf{E}] + \frac{\mathbf{i}}{\omega\varepsilon_0} \mathcal{K}[\Sigma, \mathbf{j}], \quad (3)$$

$$\mathcal{K}[S, \mathbf{g}](x) = \operatorname{rot} \operatorname{rot} \int_S \mathbf{g}(y) F(x - y) d\sigma_y, \quad F(x - y) = \frac{e^{ik|x-y|}}{4\pi|x-y|}. \quad (4)$$

Здесь Σ - поверхность идеального проводника, Ω - объем, занятый диэлектриком. Тогда исходную задачу можно заменить на эквивалентную систему интегральных уравнение относительно неизвестных поверхностных токов \mathbf{j} и объемных токов \mathbf{g} :

$$\mathbf{g} = \mathbf{E}_{inc} + \mathcal{K}[\Omega, (\varepsilon' - 1)\mathbf{g}] + \frac{\mathbf{i}}{\omega\varepsilon_0} \mathcal{K}[\Sigma, \mathbf{j}], \quad x \in \Omega, \quad (5a)$$

$$\mathbf{n} \times \left(\mathbf{E}_{inc} + \mathcal{K}[\Omega, (\varepsilon' - 1)\mathbf{g}] + \frac{\mathbf{i}}{\omega\varepsilon_0} \mathcal{K}[\Sigma, \mathbf{j}] \right) = 0, \quad x \in \Sigma. \quad (5b)$$

Численный метод. Задача решается численно с помощью метода Бубнова-Галеркина. Объем, занятый диэлектриком и металлическая поверхность аппроксимируются призматической и треугольной сеткой соответственно. На поверхности вводятся базисные функции v_i типа РВГ [2], в объеме вводятся константные вертикальные функции a_i , а также горизонтальные функции h_i , являющиеся обобщением функций типа РВГ на случай призм (рис 1).

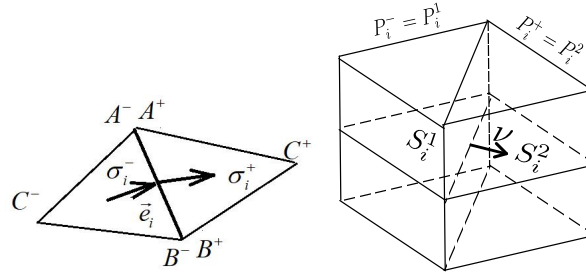


Рис. 1: иллюстрация к базисным функциям

Численный эксперимент. Для тестирования метода была рассмотрена задача о рассеянии плоской волны на металлическом квадрате, покрытом тонким слоем диэлектрика. На основании результатов моделирования строилась диаграмма распределения эффективной площади рассеяния в направлении оптического отражения. Результаты сравнения с методом моментов [3] представлены на рисунке 2. Важно отметить, что сходимость метода как по вертикальному шагу сетки (дробление диэлектрика перпендекулярно плоскости) достигалось при малом числе вертикальных слоев, из-за чего линейная система имела размеры, сопоставимые с количеством поверхностных базисных функций.

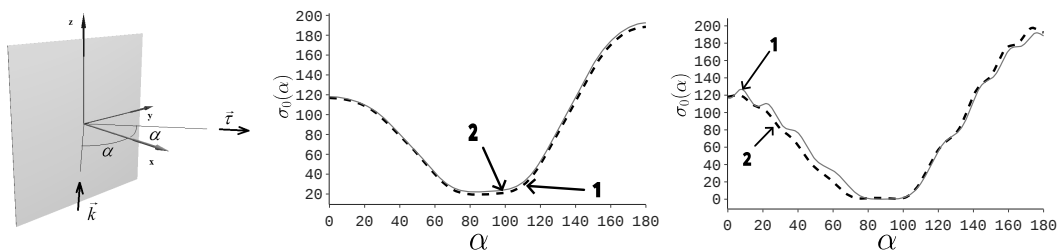


Рис. 2: численный эксперимент. линия 1 - метод из данной работы, 2 - метод моментов.

ЛИТЕРАТУРА

- [1] Samokhin A. B. Volume singular integral equations of electrodynamics (in rus) / AB Samokhin. — Technosphere (in rus), 2021
- [2] Rao, Sadasiva. Electromagnetic scattering by surfaces of arbitrary shape / Sadasiva Rao, Donald Wilton, Allen Glisson // IEEE Transactions on antennas and propagation. — 1982 — Vol. 30, no. 3 — Pp. 409–418.
- [3] Gibson W. C. The Method of Moments in Electromagnetics / W.C. Gibson. — CRC Press, 2021

Об одном методе решения обратной задачи Штурма–Лиувилля по конечному набору собственных значений

Мозговых Василий Владимирович

Кафедра математической физики

e-mail: s02170168@gse.cs.msu.ru

Научный руководитель — д.ф.-м.н. проф. Баев Андрей Владимирович

Данная работа посвящена решению одной из классических спектральных задач — обратной задачи Штурма–Лиувилля. В качестве входных данных обратной задачи использован конечный набор первых собственных значений двух задач Штурма–Лиувилля в постановке Борга–Левинсона. Остальные собственные значения заданы по хорошо известной классической асимптотике. В обратной задаче требуется восстановить непрерывный вещественный потенциал, заданный на отрезке.

Обратная спектральная задача сведена к обратной задаче для телеграфного уравнения с переменным коэффициентом, искомым потенциалом. Эта редукция осуществлена методом обращения преобразования Лапласа по формуле Меллина. В результате вычисления контурного интеграла с помощью вычетов получена явная формула для параметров рассеяния в обратной задаче для телеграфного уравнения. При этом полученная формула представляет собой тригонометрический многочлен, в котором амплитуды и частоты полностью определяются входными данными спектральной задачи.

Обратная задача рассеяния для телеграфного уравнения решена методом обращения разностной схемы путём редукции к задаче Дарбу. Численный алгоритм программно реализован на языке Python и протестирован. Полученная программа позволяет задавать наборы собственных значений и точность вычисления по разностной схеме для задачи рассеяния. Проведена серия вычислительных экспериментов, позволившая исследовать свойства предложенного метода решения и показать его применимость для потенциалов различной гладкости.

ЛИТЕРАТУРА

- [1] Баев А. В. О решении одной обратной задачи для волнового уравнения с помощью регуляризующего алгоритма // ЖВМиМФ. М. : Наука, 1985. С. 140–146.
- [2] Левитан Б. М., Саргсян И. С. Операторы Штурма–Лиувилля и Дирака. М. : Наука, 1988. 350 с.
- [3] Романов В. Г. Обратные задачи математической физики. М. : Наука, 1984. 264 с.

Вычисление управляющих параметров работы электролизной ванны

Работа удостоена диплома 3 степени

Ненахов Нил Денисович

Кафедра вычислительных методов

e-mail: s02200678@смс.msu.ru, nenakhov.neil@mail.ru

Научный руководитель — д.ф.-м.н. вед. науч. сотр. Савенкова Надежда Петровна

Научный консультант — д.ф.-м.н. вед. науч. сотр. Бычков Владимир Львович

Эффективное управление промышленным процессом электролиза алюминия требует внедрения автоматизированных систем управления – АСУТП. Высокие температуры и агрессивность среды, в которой происходит процесс электролиза алюминия, не позволяют провести измерения большинства параметров управления. Наиболее перспективным направлением является разработка новых алгоритмов управления, построенных на понимании и моделировании технологического процесса электролитического получения алюминия.

Для определения выхода по току и потерь выхода по току (η , $\Delta\eta$ соответственно), которые являются важными управляющими параметрами, необходимы замеры межполюсного расстояния, плотности анодного тока и формы поверхности раздела сред металл – электролит, которые на практике определяются весьма приблизительно [1]. В свою очередь, трёхмерная трёхфазная математическая модель процесса электролиза алюминия [2], на основе которой был реализован вычислительный комплекс, позволяет достаточно точно вычислить величины, необходимые для определения управляющих параметров, и учитывает взаимосвязь всех основных динамических процессов, происходящих в электролизной ванне.

В зависимости от режима работы электролизной ванны процесс электролиза может быть МГД – стабилен или МГД – нестабилен. В случае МГД – нестабильной работы ванны расстояние между анодом и катодом может быть критически мало, что ведёт к резкому уменьшению выхода по току. В работе приводятся результаты вычисления величины выхода по току в разные моменты времени при МГД – стабильности и МГД – нестабильности ванны.

На практике в АСУТП применяется эмпирическая формула, разработанная институтом ВАМИ [3] (Всероссийский алюминиево – магниевый институт). При этом значения параметров определяются экспериментально в нескольких отдельных точках рабочего пространства ванны. Поскольку на практике значениям параметров, входящих в формулу, соответствует большая погрешность, то величина выхода по току η по данной формуле вычисляется с значительной погрешностью.

Из проведенных в работе вычислений следует, что выход по току имеет общую тенденцию на уменьшение при искривлении поверхности.

В работе [4] представлена полуэмпирическая формула выхода по току, опирающаяся на распределения МПР по горизонтальному срезу ванны. Однако ее практическая реализация зависит от точности определения поверхности

раздела сред металл – электролит, а также требует достаточно точного вычисления поверхностного интеграла.

При этом величина МПР, как и в эмпирической формуле, определяется очень грубо, поэтому практическое ее использование весьма затруднительно. Более того, как показали проведенные в работе исследования, полуэмпирическая формула является противоречивой. Поэтому ее применение не является целесообразным. В работе предлагается модифицированная формула потерь выхода по току $\Delta\eta$:

$$\Delta\eta = (1 - \eta_0) \cdot \frac{1}{S} \cdot \int_Z \frac{l(x, y) ds}{H(x, y)}, \quad (1)$$

в которой значения величин межполюсного расстояния $l(x, y)$ и глубины жидкого металла $H(x, y)$ в каждый момент времени определяются при помощи трёхмерного математического моделирования.

Для вычисления по модифицированной формуле (1) в работе предложен метод простой триангуляции второго порядка точности. Точность исследована на сгущающихся сетках. Численный расчёт в плоскости электролизной ванны значений потерь выхода по току в случаях МГД – стабильной работы ванны, при выемке анодов и анодном эффекте показал, что наибольшие потери выхода по току соответствуют анодному эффекту, а наименьшие – МГД – стабильному режиму работы ванны. Этот вывод соответствует практическим наблюдениям.

Также с помощью модифицированной формулы (1) получены распределения потерь по току в электролизной ванне. Анализ проведённых численных экспериментов позволяет сделать выводы о качественном соответствии результатов расчетов и результатов лабораторных исследований стабильности режима работы ванны.

ЛИТЕРАТУРА

- [1] Белолипецкий В. М., Пискажова Т. В. Математическое моделирование процесса электролитического получения алюминия. Решение задач управления технологией: монография // (Красноярск, 2012 г.). — Сиб. федер. ун-т. — С. 13–58.
- [2] Калмыков А. В. Математическое моделирование влияния процессов тепломассопереноса на МГД-стабильность алюминиевого электролизёра // Москва: Московский государственный университет имени М. В. Ломоносова. Факультет вычислительной математики и кибернетики. Кафедра вычислительных методов. Диссертация. 2017 г.
- [3] Громова Б. С. Тепловые процессы в электролизерах и миксерах алюминиевого производства. // М.: 1998. С. 322.
- [4] Деркач А. С. Влияние неустойчивости тока серии на технологический режим алюминиевых электролизеров // Цветные металлы. – № 3. – 1967. 39–40 с.

Обратные задачи для уравнения теплопроводности с неизвестными коэффициентом и начальным условием

Работа удостоена диплома 1 степени

Спиридонова Алина Владиславна

Кафедра математической физики

e-mail: s02200488@gse.cs.msu.ru

Научный руководитель — д.ф.-м.н. проф. Денисов Александр Михайлович

Обратные задачи для уравнения теплопроводности представляют собой важный класс обратных задач. Они возникают при исследовании различных процессов в геофизике, теплофизике и других областях науки. Существует большое разнообразие постановок обратных задач. В [1] подробно изучены некоторые из них.

В данной выпускной квалификационной работе рассматриваются три обратных задачи для одномерного уравнения теплопроводности, состоящие в определении неизвестных коэффициента и начального условия.

Обратная задача 1. Функция $u(x, t)$ является решением начально-краевой задачи

$$\begin{aligned} u_t &= ku_{xx}, & 0 < x < \pi, & \quad t > 0, \\ u(0, t) &= u(\pi, t) = 0, & t &\geq 0, \\ u(x, 0) &= \varphi(x), & 0 \leq x \leq \pi, \end{aligned}$$

Требуется определить постоянный коэффициент теплопроводности $k > 0$ и функцию начального распределения температуры $\varphi(x)$ на отрезке $[0, \pi]$, если при $t \geq 0$ заданы функции

$$g(t) = u_x(0, t), \quad h(t) = u_x(\pi, t).$$

Обратная задача 2. Функция $u(x, t)$ является решением начально-краевой задачи

$$\begin{aligned} u_t &= ku_{xx} - qu, & 0 < x < \pi, & \quad t > 0, \\ u(0, t) &= u(\pi, t) = 0, & t &\geq 0, \\ u(x, 0) &= \varphi(x), & 0 \leq x \leq \pi, \end{aligned}$$

Требуется определить постоянный коэффициент теплообмена $q > 0$ и функцию $\varphi(x)$ на отрезке $[0, \pi]$, если $k = 1$ и при $t \geq 0$ заданы функции

$$g(t) = u_x(0, t), \quad h(t) = u_x(\pi, t).$$

В работе приведены примеры, показывающие, что в первой и второй обратных задачах в общем случае нет единственности решения. В виде теорем сформулированы и доказаны достаточные условия единственности:

Теорема 1. *Если функция $\varphi(x)$ такова, что $\int_0^\pi \varphi(\xi) \sin(\xi) d\xi \neq 0$, то обратная задача 1 имеет не более одного решения.*

Теорема 2. Если $\exists N \in \mathbb{N} : \int_0^{\pi} \varphi(\xi) \sin(n\xi) d\xi = 0$ при $n = 1, \dots, N - 1$, а $\int_0^{\pi} \varphi(\xi) \sin(N\xi) d\xi \neq 0$, то обратная задача 1 имеет не более одного решения.

Аналогичные утверждения справедливы и для обратной задачи 2.

Обратная задача 3. Функция $u(x, t)$ является решением начально-краевой задачи

$$\begin{aligned} u_t &= k(x)u_{xx}, & 0 < x < \pi, & \quad t > 0, \\ u(0, t) &= u(\pi, t) = 0, & t &\geq 0, \\ u(x, 0) &= \varphi(x), & 0 \leq x \leq \pi, \end{aligned}$$

где

$$k(x) = \begin{cases} k_1, & 0 < x < x_0, \\ k_2, & x_0 < x < \pi, \end{cases}$$

— кусочно-постоянный коэффициент теплопроводности. Считаем, что постоянные $k_1, k_2 > 0$ известны. Требуется определить точку разрыва коэффициента теплопроводности $x_0 \in (0, \pi)$ и функцию $\varphi(x)$ на отрезке $[0, \pi]$, если при $t \geq 0$ заданы функции

$$g(t) = u_x(0, t), \quad h(t) = u_x(\pi, t).$$

Для поставленных обратных задач были разработаны численные методы решения. Они основаны на редукции исходных задач к вариационным задачам и применении градиентных методов минимизации [2]. Численные методы были реализованы в виде программ на языке Python. Проведенные вычислительные эксперименты показали достаточно высокую точность предложенных методов.

ЛИТЕРАТУРА

- [1] Денисов А. М. Введение в теорию обратных задач: Учеб. пособие. М. : Издательство МГУ, 1994. 208 с.
- [2] Васильев Ф. П. Методы оптимизации. М. : Факториал Пресс, 2002. 824 с.

Блочное уравнивание космических снимков

Строков Вениамин Игоревич

Кафедра математической физики

e-mail: strokov.veniamin@mail.ru

Научный руководитель — к.ф.-м.н. доц. Головина Светлана Георгиевна

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

Метод блочного уравнивания изображений широко используется в сфере космической съёмки [1]. В работе рассматривается применение метода при наличии значительных погрешностей позиционирования снимков в географических координатах.

Метод блочного уравнивания эффективен для коррекции ошибок географической привязки кадров, в которых наблюдается регулярность и незначительные расхождения (обычно порядка десятков пикселей). Однако, если положения соседних кадров в географических координатах сильно различаются, например, отсутствуют географические области перекрытия, поиск связующих точек становится невозможным. Для решения данной проблемы предлагается применить предварительную коррекцию снимков.

Основная идея предварительной коррекции состоит в поиске линии, аппроксимирующей траекторию движения центральной точки обзора съёмочной аппаратуры спутника по поверхности Земли, и выравнивании по ней географических координат снимков, полученных из телеметрии [2],[3]. Для этого находится минимум функционала отклонений точек центров снимков от искомым кривых методом наименьших квадратов

$$\Phi_n(\xi_{n1}, \dots, \xi_{nT}) = \sum_{t=1}^T \xi_{nt}^2 \rightarrow \min, \quad (1)$$

где ξ_{nt} - отклонение точки центра снимка от искомой кривой для каждого из $n = 1..N$ наборов снимков, $t = 1..T$ - количество кадров в каждом наборе.

Из необходимого условия минимума функционала (1) находятся коэффициенты искомым кривых. Далее находятся проекции центров снимков на полученные аппроксимирующие кривые и векторы сдвигов, на которые смещаются все точки соответствующих снимков.

В работе приведены примеры работы алгоритма предварительной географической коррекции, реализованного на языке C++, и проведено сравнение результатов применения метода блочного уравнивания без предварительной коррекции и с предварительной коррекцией.

ЛИТЕРАТУРА

- [1] Егин М.М., Кузнецов А.Е., Пошехонов В.И. Алгоритм блочного уравнивания снимков, привязанных по RPC-модели. Современные проблемы дистанционного зондирования Земли из космоса. 2009. В.6. Т.1. С. 96-104
- [2] Alan O. Sykes, An Introduction to Regression Analysis. Coase-Sandor Institute for Law and Economics Working Paper No. 20, 1993. - 34 с.
- [3] И. В. Митин, В. С. Русаков. Анализ и обработка экспериментальных данных: Учебно-методическое пособие для студентов младших курсов. — М.: Физический Факультет МГУ. — 44 с.

Нейросетевой метод классификации фрагментов полнослайдовых гистологических изображений с использованием частичной разметки

Яковлев Никита Алексеевич

Кафедра математической физики

e-mail: yakovlevnikita0@gmail.com

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

Использование цифровых средств проникает во все сферы медицины, в том числе и в гистологические исследования. Гистологический срез - это снимок тканей человека, полученный с помощью специального микроскопа. Их используют для диагностики состояния пациента. Одним из способов помочь специалисту в исследовании данного препарата является автоматическая классификация фрагментов изображения. Современные методы автоматической классификации требуют большого количества размеченных данных, однако их разметка трудоемка. В силу этого есть потребность в разработке методов классификации требующих малое количество размеченных данных для своей работы.

Для обучения и тестирования использовался набор данных PATH-DT-MSU WSS2-v2 [1] с полигональной разметкой по 5 классам. Так как размечена меньшая часть изображений, разметка является частичной. В силу биологических особенностей есть несбалансированность по классам. Поэтому в качестве метрик использовались точность и сбалансированная точность (среднее значение точности по классам).

Предложенный метод является двухэтапным. В качестве обучаемой модели для классификации фрагментов использовалась свёрточная нейронная сеть EfficientNet-B4.

Первый этап: предобучение весов кодировщика с помощью обучения автокодировщика без использования разметки. Автокодировщик состоит из двух частей — кодировщика и декодировщика. Кодировщик сжимает исходное изображение, декодировщик по сжатому представлению восстанавливает исходное изображение. За счет сжатия автокодировщик учится находить высокоинформативные признаки (текстурные элементы), которые будут полезны при классификации. Кодировщик — EfficientNet-B4, декодировщик — сверточный.

Второй этап: обучение классификатора с использованием подхода аугментаций и псевдо-меток [2]. Используются как размеченные, так и неразмеченные данные. Функция потерь состоит из двух слагаемых:

$$L = \frac{1}{B} \sum_{b=1}^B H(y_b, \text{Model}(\alpha(x_b))) +$$

$$\frac{1}{\mu B} \sum_{b=1}^{\mu B} \mathbf{1}(\text{Model}(\alpha(x_b)) > \tau) H(\tilde{y}_b, \text{Model}(\beta(x_b))),$$

где B — размер пакета размеченных данных, $H(\cdot, \cdot)$ — кросс энтропия, \tilde{y}_b — псевдо-метка класса (т.е. $\text{argmax}(\text{Model}(\alpha(x_b)))$), $\text{Model}(x_b)$ — вектор-вероятности классов сети по объекту x_b , $\tau = 0.95$ — порог псевдо-метки, $\mathbf{1}$ — функция индикатора, $\mu = 7$ — множитель для количества экземпляров в одном пакете для неразмеченной выборки, $\alpha(\cdot), \beta(\cdot)$ — слабые и сильные аугментации соответственно.

Первое слагаемое — это кросс-энтропия на размеченных данных. Второе слагаемое — кросс-энтропия на неразмеченных данных, где предсказанная моделью вероятность какого-то класса превышает определенный порог τ .

В процессе обучения модели использовались различные методы аугментации изображений, которые можно разделить на сильные и слабые. Модель обучалась предсказывать одинаковые классы на слабо и сильно аугментированных изображениях, полученных из неразмеченных данных. За счет использования сильных аугментаций достигается использование текстурной информации на неразмеченных участках.

Была проведена оценка предложенного метода при использовании разного количества размеченных данных по сравнению с базовым методом, который отличается отсутствием использования неразмеченных данных. При использовании предложенного метода с 40 % разметки одного изображения, точность и сбалансированная точность на тестовой выборке составила 0.853 и 0.948 соответственно, что больше по сравнению с базовым методом, который использует 100 % разметки одного изображения (0.808 и 0.947 соответственно). При использовании разметки 100 % с 5 изображений предложенный метод показывает результаты лучше по сравнению с базовым методом: 0.938 и 0.978 против 0.92 и 0.97 соответственно. Откуда следует, что ошибка классификатора уменьшилась на 23 % и 27 % соответственно.

ЛИТЕРАТУРА

- [1] Automatic mucous glands segmentation in histological images / Khvostikov A., Krylov A., Mikhailov I., Kharlova O., Oleynikova N., Mal'kov P. // ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. 2019. T. XLII-2/W12. С. 103-109.
- [2] Fixmatch: Simplifying semi-supervised learning with consistency and confidence / Sohn K., Berthelot D., Carlini N., Zhang Z., Zhang H., Raffel C. A., Cubuk E. D., Kurakin A. Li C.-L. // Advances in neural information processing systems. 2020. Vol. 33. P. 596–608.

Оптимальная стратегия потребления и производства электроэнергии с применением накопителя энергии

Работа удостоена диплома 1 степени

Бабаян Никита Вачеевич

Кафедра исследования операций

e-mail: babayann26@gmail.com

Научный руководитель — д.ф.-м.н. проф. Васин Александр Алексеевич

Введение. Оптовый рынок электроэнергии в России работает по принципу конкурентного рынка, сделки на котором охватывают более 90% производимой электроэнергии [1]. Спрос на электроэнергию меняется в течение суток, цены в различные часы могут отличаться в несколько раз. Использование накопителя энергии позволяет повысить эффективность экономических агентов, потребляющих и производящих электроэнергию, для этого требуется определить оптимальную стратегию управления накопителем. В работе [2] приводятся постановки задач оптимизации прибыли для потребителя, использующего накопитель энергии, а также для накопителя энергии, который действует независимо и извлекает прибыль из покупки электроэнергии по низкой цене и продажи — по более высокой.

Задача оптимизации для накопителя, перепродающего энергию. Рассматривается модель с дискретным временем, определены цены на электроэнергию p^t , $t = 1, \dots, T$, изменяющиеся с периодом T — длиной планового периода. Накопитель характеризуется предельной ёмкостью E , мощностью зарядки и разрядки V , коэффициентами потерь разрядки η_{dis} и зарядки η_{ch} . Стратегия накопителя v_{Bat} определяется объемами приобретенной (v_{ch}^t) и проданной (v_{dis}^t) энергии в каждый момент t . Задача оптимизации

$$\frac{1}{\eta_{dis}} \sum_{t=1}^T v_{dis}^t p^t - \eta_{ch} \sum_{t=1}^T v_{ch}^t p^t \rightarrow \max, \quad (1)$$

$$\sum_{t=1}^T (v_{ch}^t - v_{dis}^t) = 0,$$

$$v_{dis}^t \leq V_{dis}, \quad v_{ch}^t \leq V_{ch}, \quad 0 \leq \sum_{k=1}^t (v_{ch}^k - v_{dis}^k) \leq E, \quad t = \overline{1, T}, \quad (2)$$

$$0 \leq v_{ch}^t, \quad 0 \leq v_{dis}^t, \quad t = \overline{1, T},$$

является задачей линейного программирования (ЗЛП).

Задача оптимизации с кусочно-линейной целевой функцией. В настоящей работе рассматриваются методы решения указанной задачи, также сформулирована постановка задачи определения оптимальных параметров накопителя при проектировании изолированной энергосистемы для минимизации капитальных расходов. Указан метод сведения задачи с линейны-

ми ограничениями и кусочно-линейной целевой функцией к ЗЛП с помощью добавления новых переменных.

Решение с помощью симплекс-метода. Для приведения задачи (1) к каноническому виду в ограничения-неравенства (2) вводятся слабые переменные, тогда ЗЛП будет иметь $4T + 1$ ограничение и $6T$ переменных.

Симплекс-метод требует существенного повышения размерности задачи, не учитывает периодичность цен, то есть для нахождения оптимальной стратегии нужно также указать оптимальное начало периода планирования.

Решение с помощью разработанного алгоритма. В работе [3] был описан алгоритм для решения задачи (1), гарантирующий выполнение необходимых условий оптимальности. Алгоритм использует разбиение планового периода на непересекающиеся связные интервалы.

Определение. *Связным интервалом называется период (t', t'') наибольшей длины, на котором ограничение на текущий заряд накопителя $0 \leq \sum_{k=1}^t (v_{ch}^t - v_{dis}^t) \leq E$ не является активным.*

Сначала строится стратегия для начального значения e^* , определяемого из графика цен, затем индуктивным методом по оптимальной стратегии для заданного значения ёмкости e строится стратегия для $e + 1$, пока не будет получена стратегия для E .

На основе описанного подхода в настоящей работе построен новый алгоритм, обеспечивающий достаточные условия оптимальности с помощью проверки нескольких условий для наборов следующих друг за другом связных интервалов. Применение алгоритма не требует повышения размерности задачи в отличие от симплекс-метода, а также учитывает периодичность цен (начало и конец периода планирования считаются соседними).

Результаты вычислительных экспериментов. Написана программа на языке Python, применяющая алгоритм к случайно сгенерированным векторам цен. Показано, что число элементарных операций, необходимых для расчета стратегии у данного алгоритма меньше, чем у симплекс-метода.

ЛИТЕРАТУРА

- [1] Давидсон М. Р. и др. Математическая модель конкурентного оптового рынка электроэнергии в России // Известия Российской академии наук. Теория и системы управления. – 2004. – №. 3. – С. 72-83.
- [2] Vasin A., Grigoryeva O. Optimal strategies of consumers with energy storages in electricity market // Mathematical Optimization Theory and Operations Research: Recent Trends. MOTOR 2022. Communications in Computer and Information Science. Vol. 1661. Cham: Springer, 2022.
- [3] Vasin A., Grigoryeva O. On the optimal management of energy storage, 10 p. (в печати).

Методы машинного обучения для анализа биологической значимости экспрессии генов

Работа удостоена диплома 1 степени

Боровко Никита Алексеевич

Кафедра математической статистики

e-mail: borovkko.kt@mail.ru

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

Научный консультант — к.б.н. Поверенная Екатерина Владимировна

Кластеризация генов широко используется для решения важных задач в биологических и медицинских исследованиях. Одной из таких задач является выявление групп генов, участвующих в общих биологических процессах — она необходима для понимания молекулярных механизмов, лежащих в основе клеточных процессов и важно для открытия новых мишеней для терапевтического воздействия и разработке более эффективных стратегий лечения различных заболеваний [1].

Данные в исследовании представляют собой результаты анализа экспрессии генов клеточных линий человека, полученные Project Achilles — исследовательским проектом по выявлению генов, важных для выживания и размножения раковых клеток. Они включают клетки из различных тканей и органов человека, всего имеется 17 283 гена и 973 клеточных линии.

Решается задача определения порога биологической значимости экспрессии генов и их кластеризации на группы с общими биологическими свойствами. Алгоритм, предложенный для этой задачи, состоит из нескольких этапов.

Биологическая гипотеза порога биологической важности экспрессии генов заключается в предположении существования определенного уровня экспрессии у каждого гена, ниже которого изменения в экспрессии не имеют существенного биологического эффекта [2]. В рамках работы предложен подход к вычислению этого порога, основанный на нелинейном преобразовании выживаемости генов и использовании метода взвешенных квантилей для каждого гена:

$$\alpha \geq \sum_i I[x_i \leq x_\alpha] \cdot w_i, \quad (1)$$

$$1 - \alpha \leq \sum_i I[x_i \geq x_\alpha] \cdot w_i, \quad (2)$$

где x_i — экспрессия в i -ой линии ($x_1 \leq x_2 \leq \dots \leq x_N$), x_α — значение взвешенного квантиля, w_i — значимость i -ого гена ($\sum_i w_i = 1$).

С помощью знания порога значимости экспрессии вычисляется доля клеточных линий, экспрессируемых для каждого гена — в которых наблюдение экспрессии превышает вычисленный опрог. В соответствии с этой долей происходит группировка всех генов, в зависимости от того, в какой диапазон попала такая доля. Границы диапазонов фиксированные: 0.95, 0.9, 0.75, 0.5,

0.25 и 0.1 наблюдений экспрессии выше порога. Таким образом, гены, экспрессирующиеся схожим образом, окажутся в одной группе.

Для полученных групп происходит сокращение размерности исходной матрицы экспрессий до 3 измерений при помощи t-SNE [3] (t-distributed Stochastic Neighbor Embedding). К каждой проекции экспрессии группы применяется иерархическая агломеративная кластеризация [4], число кластеров для каждой группы подбирается, исходя из количества генов в ней и возможности добиться различий в характеристиках генов в разных кластерах.

По результатам кластеризации проводится функциональная аннотация [5] (согласно системе Gene Ontology). Для полученных кластеров вычисляется доля генов, отвечающих за общие биологические функции и сравнивается с аналогичной долей в исходном наборе генов. От различия этих долей зависит степень достоверности выявления таких функций в кластере. Достоверность результатов считалась надежной при $FDR < 10^{-7}$, нулевая гипотеза отклонялась при $q\text{-value} < 0.05$.

Всего было получено 24 кластера, из них было статистически значимо выделено 8 монокластеров, отражающих единый биологический процесс, такой как эмбриогенез, морфогенез скелета и т.д. и 7 мультикластеров, в которых гены вовлечены в различные биологические процессы, напрямую не связанные между собой.

Полученные в результате исследования данные позволяют предположить существование индивидуального порогового уровня экспрессии генов, участвующих в определенных биологических процессах. Предложенный метод продемонстрировал способность достоверно выделять моно- и мультикластеры генов, отвечающих за общий набор функций.

ЛИТЕРАТУРА

- [1] Jiang D., Tang C., Zhang A. Cluster analysis for gene expression data: a survey // IEEE Transactions on knowledge and data engineering. 2004. Т. 16, № 11. С. 1370–1386.
- [2] Mukherji S. et al. MicroRNAs can generate thresholds in target gene expression // Nature genetics. 2011. Т. 43, № 9. С. 854–859.
- [3] Cai T. T., Ma R. Theoretical foundations of t-sne for visualizing high-dimensional clustered data // Journal of Machine Learning Research. 2022. Т. 23, № 301. С. 1–54.
- [4] Hossen B., Siraj-Ud-Douhah H. A., Hoque A. Methods for evaluating agglomerative hierarchical clustering for gene expression data: a comparative study // Computational Biology and Bioinformatics. 2015. Т. 3, № 6. С. 88–94.
- [5] Eden E. et al. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists // BMC bioinformatics. 2009. Т. 10, С. 1–7.

Ньютоновские методы вблизи критических решений кусочно—гладких нелинейных уравнений

Доровских Дмитрий Игоревич

Кафедра исследования операций

e-mail: deamterr@gmail.com

Научный руководитель — д.ф.-м.н. проф. Измайлов Алексей Феридович

Метод Ньютона является одним из самых известных методов численного решения нелинейных уравнений. Он обладает высокой скоростью сходимости в стандартных предположениях и дает хорошее приближение к решению. В данной работе исследуются кусочные ньютоновские методы, которые представляют интерес для решения кусочно—гладких уравнений с ограничениями и, в частности, нелинейных комплементарных задач в их переформулировке через функцию минимума.

Основными целями данной работы являются глобализация сходимости кусочных ньютоновских методов для кусочно—гладких уравнений с ограничениями, исследование поведения кусочного метода Ньютона вблизи «особых» решений кусочно—гладких уравнений, а также численное сравнение глобализованного кусочного метода Ньютона для переформулировки нелинейных комплементарных задач через функцию минимума с методом Ньютона для гладкой переформулировки нелинейных комплементарных задач.

В работе предложены алгоритмы, реализующие глобализацию сходимости кусочных ньютоновских методов (а именно, методов Ньютона и Гаусса—Ньютона) для кусочно—гладких уравнений с ограничениями. Получены результаты, характеризующие свойства возможных предельных точек последовательностей, генерируемых этими методами, а именно, стационарность всякой такой точки хотя бы для одного активного в ней гладкого кусочного отображения, а также условия асимптотической сверхлинейной скорости сходимости таких последовательностей.

Ключевую роль в анализе глобальной сходимости сыграло требование мажорирования нормы отображения нормами гладких кусочных отображений, которое уже появлялось в [1, 2]. Приведены примеры, демонстрирующие, что при невыполнении этого условия рассматриваемые алгоритмы могут генерировать последовательности, сходящиеся к точкам, не являющимся стационарными ни для одного активного гладкого кусочного отображения.

В [2] было показано, что кусочный метод Ньютона в определённой области вблизи «особых» решений кусочно—гладких уравнений может демонстрировать всего лишь линейную скорость сходимости. В работе этот теоретический результат проиллюстрирован численными примерами.

При численном сравнении глобализованного кусочного метода Ньютона для переформулировки нелинейных комплементарных задач через функцию минимума с методом Ньютона для гладкой переформулировки нелинейных комплементарных задач получен вывод о более высокой скорости сходимости

сти первого метода, что соответствует теоретическим результатам. При этом значимых различий в робастности между этими методами не обнаружено.

Основные результаты работы содержатся в [3].

ЛИТЕРАТУРА

- [1] A globally convergent LP–Newton method / A. Fischer, M. Herrich, A. F. Izmailov, M. V. Solodov // SIAM J. on Optimization. 2016. V. 26. P. 2012–2.
- [2] Fischer A., Izmailov A. F., Jelitte M. Newton–type methods near critical solutions of piecewise smooth nonlinear equations // Computational Optimization and Applications. 2021. V. 80. P. 587–615.
- [3] Доровских Д. И., Измаилов А. Ф., Усков Е. И. Глобализация сходимости кусочных ньютоновских методов // Вестник российских университетов. Математика. 2024. Т. 29. № 146. В печати.

Исследование алгоритмов классификации текста на русском языке

Кадрова Надия Ренатовна

Кафедра математической статистики

e-mail: kaderova.n@mail.ru

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

Задача классификации текста — это процесс анализа текстовых последовательностей и присвоения им меток, а также группирования их в зависимости от их содержания. Это ключевой элемент в области искусственного интеллекта и машинного обучения, особенно в контексте обработки естественного языка.

В классификации текста на русском языке возникают дополнительные трудности из-за морфологической сложности и множества грамматических признаков, а также из-за вариантивности текста и аспектов его восприятия [1].

Целью данной работы является выбор наиболее подходящей модели для эффективной классификации для конкретного набора данных с новостями на русском языке по разным темам. Для исследования различных методов классификации текста на русском языке были взяты данные с новостных порталов, которые перед применением моделей необходимо было предобработать. В ходе работы были выделены следующие методы: наивный байесовский классификатор [2], RandomForest, Word2Vec [3], BERT [4], а также их объединения. В качестве оценки результатов работы моделей была выбрана метрика *accuracy* [5].

Результаты. В ходе применения различных методов к исходному набору данных с новостями на русском языке были получены выводы, описанные ниже.

Наивный байесовский классификатор в данной задаче показал невысокий результат ($accuracy = 0.51$), что можно объяснить недостаточной однородностью фрагментов текстов, выбранных для классификации. При снижении количества уникальных классов (рубрик новостей) в задаче классификации точность сильно улучшилась ($accuracy = 0.96$ - лучший результат у наивного байесовского классификатора при трех узких, различных по смыслу классах).

Word2Vec позволяет захватывать и кодировать семантические отношения между словами. Близкие по смыслу слова имеют близкие векторные представления, Word2Vec позволяет учитывать контекстные особенности слов при их представлении в векторной форме. Однако векторные представления не всегда обладают полным пониманием различных смыслов слова, и Word2Vec учитывает только локальный контекст слова и не учитывает глобальные особенности или связи в предложении или документе. При этом особенность BERT заключается в его способности понимать семантику, контекст и зависимости между словами, что позволяет модели достигать высокой производительности для различных задач обработки естественного языка, поэтому при объединении методов Word2Vec и BERT был получен лучший результат в задаче классификации на исходном наборе данных с множеством классов ($accuracy = 0.74$).

Таким образом, был реализован алгоритм классификации текста на русском языке для новостей с интернет-порталов, показывающий высокие результаты. Данный алгоритм включает в себя предобработку данных и последовательное применение методов классификации. Также был сделан вывод, что для более корректного и точного выбора подходящего метода для конкретного набора данных необходимо учитывать следующие параметры: объем и качество данных (насколько хорошо данные размечены и структурированы), сложность задачи, специфика задачи (необходимость учитывать отдельные фразы или в приоритете учет тональности текста и тд), имеющиеся ресурсы и время.

Данная работа была представлена на XXXI Международной конференции студентов, аспирантов и молодых учёных "Ломоносов".

ЛИТЕРАТУРА

- [1] Bahl L. R. et al. A tree-based statistical language model for natural language speech recognition //IEEE Transactions on Acoustics, Speech, and Signal Processing. – 1989. – Т. 37. – №. 7. – С. 1001-1008.
- [2] Murphy K. P. et al. Naive bayes classifiers //University of British Columbia. – 2006. – Т. 18. – №. 60. – С. 1-8.
- [3] Mikolov T. et al. Distributed representations of words and phrases and their compositionality //Advances in neural information processing systems. – 2013. – Т. 26.

- [4] Vaswani A. et al. Attention is all you need // Advances in neural information processing systems. – 2017. – Т. 30.
- [5] Mathews S. M. Explainable artificial intelligence applications in NLP, biomedical, and malware classification: a literature review // Intelligent Computing: Proceedings of the 2019 Computing Conference, Volume 2. – Springer International Publishing, 2019. – С. 1269-1292.

Методы повышения обобщающей способности линейных ансамблевых процедур

Работа удостоена диплома 3 степени

Кривуля Полина Юрьевна

Кафедра математических методов прогнозирования

e-mail: polina_krivulya@mail.ru

Научный руководитель — д.ф.-м.н. проф. Сенько Олег Валентинович

Бустинг — схема ансамблирования, в которой каждый следующий базовый алгоритм осуществляет поправку предыдущих ответов ансамбля до верных ответов на обучающей выборке. Градиентный бустинг — популярный вид бустинга, использующий деревья решений в качестве базовых алгоритмов [1]. Различные варианты градиентного бустинга получили широкое распространение и позволяют достигать высокой точности решения задач прогнозирования по сравнению с альтернативными методами. Однако градиентный бустинг требует дифференцируемости функции потерь и аппроксимации ее градиента в пространстве прогнозов. Неявная подмена оптимизируемого функционала приводит к снижению качества отдельных базовых алгоритмов и избыточному увеличению их количества [2].

В работе предложено вместо шага с вычислением градиента напрямую строить дерево, минимизирующее функцию потерь, что позволяет обойти необходимость ее дифференцируемости и аппроксимации градиента. Целью работы является экспериментальное исследование схем, повышающих эффективность бустинга.

Предложенный в работе новый вариант бустинга основан на поиске суммы деревьев, минимизирующей потери. При этом на каждом шаге строится дерево, в результате добавления которого в сумму потери оказываются на данном шаге минимальными. Преимуществом схемы является возможность использования новых схем регуляризации. Одна из предложенных схем основана на оптимизации степени корреляции добавляемого дерева с текущим ансамблем.

Определение 1. Будем называть корреляцией между двумя алгоритмами $a_1(x)$ и $a_2(x)$ линейный коэффициент корреляции между их предсказаниями на обучающей выборке и обозначать $\rho(a_1, a_2)$.

Введем следующие обозначения: A_k — бустинг, полученный на k -м шаге, y_{true} — целевые значения, T_{k+1} — решающее дерево, строящейся на $k + 1$ -м

шаге, $\mathcal{L}(y_{true}, A_{k+1})$ — оптимизируемый функционал, β — коэффициент регуляризации. Тогда задача состоит в оптимизации функционала (1).

$$\mathcal{L}(y_{true}, A_{k+1})|\rho(T_{k+1}, A_k)|^\beta \rightarrow \min_{T_{k+1}} \quad (1)$$

Прогнозы дерева в процессе его построения зависят от порядка обработки узлов, что влияет на корреляцию между деревом и предыдущим ансамблем. В связи с этим, в работе рассматриваются два варианта обхода дерева: прямой вертикальный и горизонтальный обход.

Экспериментальное исследование показывает, что точность модифицированного таким образом бустинга возрастает при увеличении корреляции нового дерева с ансамблем. При этом для многих задач эффективность модифицированного метода превосходит эффективность популярного метода CatBoost [3].

Другим важным преимуществом предложенного в работе варианта бустинга является возможность рассматривать не только линейные суммы деревьев, но также и алгебраические комбинации деревьев, при построении которых используются операции сложения и умножения. Проведены исследования такого подхода, в результате которых получено, что на некоторых задачах качество метода превосходит качество других известных схем.

Таким образом, в работе исследованы новые схемы бустинга: метод, основанный на оптимизации степени корреляции добавляемого дерева с текущим ансамблем, и метод с применением для коррекции ансамблей наряду с аддитивными деревьями также деревьев—мультипликаторов. Ряд экспериментов показывает, что оптимизация степени корреляции позволяет достичь на многих задачах качества лучше, чем у других популярных методов. При использовании алгебраических комбинаций деревьев также достигнуты некоторые улучшения качества.

ЛИТЕРАТУРА

- [1] Friedman J. H. Greedy function approximation: a gradient boosting machine // *Annals of statistics*. 2001. P. 1189-1232.
- [2] Воронцов К. В., Каневский Д. Ю. Коэволюционный метод обучения алгоритмических композиций // *Таврический вестник информатики и математики*. 2005. Т. 2. С. 51.
- [3] CatBoost: unbiased boosting with categorical features / Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., Gulin, A. // *Advances in neural information processing systems*. 2018. Vol. 31.

Оптимизация размещения рекламы с учетом эффективного охвата

Работа удостоена диплома 2 степени

Куренкова Дарья Александровна

Кафедра исследования операций

e-mail: s02200661@gse.cs.msu.ru

Научный руководитель — к.ф.-м.н. доц. Поспелова Ирина Игоревна

Введение. Целью данной работы является разработка метода определения оптимального расписания размещения телевизионной рекламы, который позволит обеспечить максимальную эффективность рекламной кампании.

Математическая модель. Необходимо распределить выделенный фирмой бюджет между различными однотипными медиасобытиями [1] таким образом, чтобы максимизировать суммарный эффективный охват при заданных ограничениях:

$$\begin{cases} \sum_{i=1}^N F_{\text{eff}}(k_i) \rightarrow \max_k \\ \sum_{i=1}^N k_i a_i \leq S, \\ 0 \leq k_i \leq n_i, \quad i \in [1, N], \end{cases}$$

где F_{eff} — эффективный охват целевой аудитории, i — номер рассматриваемого медиасобытия, k_i — число размещений на определенном канале в определенное время, N — количество рассматриваемых медиасобытий по всем телеканалам, n_i — максимально возможное число размещений однотипного события в сетке вещания за некоторый промежуток времени, a_i — стоимость одного размещения, S — заданный рекламный бюджет.

Формула эффективного охвата. Эффективный охват представляет собой среднюю долю людей из целевой аудитории, провзаимодействовавших с эфирным событием не менее, чем с эффективной частотой [1]. Эффективный охват выражается через спектр охвата, формула для которого выведена в работе и зависит от рейтинга медиасобытия, а также от предельного охвата целевой аудитории. Таким образом, получена итоговая задача оптимизации:

$$\begin{cases} \sum_{i=1}^N \sum_{\nu=\nu_{\text{eff}}}^{k_i} F_i^{\infty} C_{k_i}^{\nu} \left(\frac{R_i}{F_i^{\infty}} \right)^{\nu} \left(1 - \frac{R_i}{F_i^{\infty}} \right)^{k_i-\nu} \rightarrow \max_k \\ \sum_{i=1}^N k_i a_i \leq S, \\ 0 \leq k_i \leq n_i, \quad k_i \in \mathbb{Z}, \quad i \in [1, N]. \end{cases}$$

Основными трудностями при ее решении являются: наличие биномиальных коэффициентов с искомой переменной, суммирование от величины эффективной частоты до неизвестной переменной внутри целевой функции, нелинейностью целевой функции, наличие ограничений задачи и ограничения целочисленности на искомые переменные. В качестве результата формируется оптимальное расписание размещений, которое включает в себя название телевизионного канала, время размещения рекламного сообщения, а также найденное количество необходимых размещений.

Методы решения. В связи с большим объемом входных данных и особенностями задачи был рассмотрен метод проекции градиента, основная идея которого заключается в том, чтобы совместить принципы градиентных методов для оптимизации без ограничений с процессом проецирования создаваемых приближений на множество, соответствующее всем условиям задачи.

Также был рассмотрен метод динамического программирования. Задача состоит из N этапов, на каждом из которых принимается решение о выборе альтернативы в зависимости от текущего состояния, затем происходит переход в новое состояние. Решение задачи сводится к последовательному вычислению функции Беллмана [2], начиная с последнего шага и заканчивая первым. Находятся оптимальные значения альтернатив на каждом этапе.

Применение метода динамического программирования приводит к лучшим результатам, однако разница в значениях целевой функции не значительна, около 1.4%, таким образом возможно использование обоих методов.

Анализ конкурентной среды. Для анализа конкурентной среды и оценки влияния конкурентных факторов на эффективность рекламной кампании был разработан алгоритм, использующий метод наименьших квадратов и тест Чоу, который с помощью F-статистики проверяет, описываются ли два набора данных одним и тем же уравнением регрессии. Этот алгоритм позволяет автоматически подбирать панель конкурентов для проведения аудита рекламной кампании на телевидении с использованием средств эконометрического анализа.

Заключение. В работе получены следующие результаты: определена формула для вычисления эффективного охвата целевой аудитории, сформулирована задача оптимизации, описаны возможные методы решения оптимизационной задачи с применением метода проекции градиента и метода динамического программирования, проведена практическая часть исследования, включая реализацию методов, их сравнение, анализ результатов, разработан алгоритм автоматического подбора панели конкурентов для анализа конкурентной среды и оценки влияния конкурентных факторов на эффективность рекламной кампании. Важность данной работы заключается в том, что она представляет собой комплексный подход к оптимизации размещения рекламы на телевидении с учетом максимального охвата аудитории и анализа конкурентной среды.

ЛИТЕРАТУРА

- [1] Шматов Г. А. Основы экономико-математической теории медиапланирования : Учебное пособие / Науч. ред. Попов Е. В. Изд. 2-е, испр. и доп. М.: ЛЕНАНД, 2021. 456 с.
- [2] Bellman, Richard. Dynamic Programming // A Rand Corporation Research Study. Princeton University Press, 1957. Sixth Printing, 1972. ISBN 0-691-07951-X.

Методы интеллектуального анализа телекоммуникационных данных

Работа удостоена диплома 2 степени

Навалихин Александр Ильич

Кафедра математической статистики

e-mail: Alex.Navalikhin@yandex.ru

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

Эффективный анализ телекоммуникационных данных может позволить существенно улучшить качество сервисов, сократить издержки и ускорить процесс обработки трафика. Все это делает анализ телекоммуникационного трафика необходимым и важным процессом, который позволяет существенно упростить работу многих сервисов и обеспечивает более эффективное взаимодействие с получаемыми данными.

В работе рассматриваются современные подходы к анализу телекоммуникационных данных, основанные на вероятностном прогнозировании, методах машинного обучения и нейронных сетях. Все методы были протестированы на нескольких наборах данных, один из которых представляет собой реальный набор данных, предоставленный мобильным оператором. В данных содержатся почасовые объемы загруженных и отправленных данных по каждому пользователю и приложению с 10.02 по 22.02 и с 01.03 по 04.03, часовые метки внутри суток, идентификаторы пользователей (более 94 тысяч уникальных пользователей за период наблюдений) и одного из 16 различных типов приложений. Другим набором данных был открытый набор данных мобильного приложения, включающий в себя почасовые наблюдения в течение недели с 22.12.2018 по 29.12.2018 и содержащий в себе 169 наблюдений числа активных пользователей, количества активных сессий, числа сбоев и числа новых пользователей за последний час.

Одной из решаемых задач была реализация методов обнаружения экстремальных значений: вероятностное прогнозирование [1] и метод превышения пороговых значений [2].

Экстремальный процесс представляет собой маркированный точечный процесс $\{(\tau_i, X_i)\}_{i>0}$, где $\{\tau_i\}_{i>0}$ - точечный случайный процесс, а $\{X_i\}_{i>0}$ - случайные величины, отмечающие величину превышения порога в каждый момент времени τ_i . Другими словами, экстремальный процесс описывает события, происходящие в точки времени, когда происходит превышение порога, и величину этого превышения.

Поскольку экстремальный процесс является случайным, точную дату следующей катастрофы предсказать невозможно. Однако можно вычислить или оценить вероятность наступления катастрофы в течение некоторого интервала времени. Именно эту оценку и вычисляет метод вероятностного прогнозирования.

Оба метода и вероятностное прогнозирование, и метод превышения порогового значения основаны на теореме [3], которая показывает, что превышения порогового значения соответствуют обобщенному распределению Парето.

В результате работы методов были получены критические уровни, после которых данные считаются аномальными. Далее производилась разметка данных с помощью модели классификации, что позволило эффективно детектировать аномалии.

Другой решаемой задачей было прогнозирование временного ряда в телекоммуникационном трафике с помощью модели ARIMA, нейронной сети с ячейкой LSTM [4], модели NeuralProphet [5]. Все методы обучались и тестировались на независимых выборках. Для моделей ARIMA и NeuralProphet обучающая и тестовая выборки строились методом скользящего окна, для нейронной сети с ячейкой LSTM выборки брались фиксированного размера. Для оценки качества прогнозов используется среднее абсолютное процентное отклонение (MAPE). Для каждого набора данных алгоритмы дали близкие результаты, наименьшая ошибка прогноза составила 2.4%.

ЛИТЕРАТУРА

- [1] Королев В. Ю., Соколов И. А. Математические модели неоднородных потоков экстремальных событий. 2008.
- [2] Горшенин А. К. О некоторых математических и программных методах построения структурных моделей информационных потоков // Информатика и её применения. 2017. Т. 11. №. 1. С. 58–68.
- [3] Balkema A. A., De Haan L. Residual life time at great age // The Annals of probability. 1974. Т. 2. №. 5. С. 792–804.
- [4] Hochreiter S., Schmidhuber J. Long short-term memory // Neural computation. 1997. Т. 9. №. 8. С. 1735–1780.
- [5] Forecasting of electric load using a hybrid LSTM-neural prophet model / Shohan M. J. A., Faruque M. O., Foo S. Y. // Energies. 2022. Т. 15. №. 6. С. 2158.

Диффузионные модели с адаптивными расписаниями генерации

Оганов Александр Артурович

Кафедра математических методов прогнозирования

e-mail: oganovaa@my.msu.ru

Научный руководитель — ас. Темирчев Павел Георгиевич

*Научный консультант — без степени, к.ф.-м.н. ас., проф. Охотин Андрей Сергеевич,
Ветров Дмитрий Петрович*

Диффузионные модели [1] – это современный подход к решению задачи генеративного моделирования. Модели, основанные на диффузии, включают в себя латентные переменные, через которые оценивается искомая плотность распределения. Современные диффузионные модели требуют большое

количество шагов, то есть применений нейронной сети, для качественной генерации. При этом используемые нейронные сети содержат огромное число параметров, что отражается на времени генерации.

Основной задачей работы является построение диффузионной модели, которая бы требовала меньшее число шагов для получения хороших семплов. Мы считаем, что равномерные шаги генерации являются ограничением современных моделей, поэтому использование адаптивной генерации могло бы позволить повысить качество и уменьшить число шагов.

В работе используется подход генеративных потоковых нейронных сетей [2], с помощью которого мы можем иначе описать диффузионные модели. В нашей работе предложена модель с адаптивным расписанием генерации, которая является генеративной потоковой нейронной сетью. Ядра перехода P_F и P_B определены в выражениях ниже.

$$\begin{aligned} P_B((x_{t_{i-1}}, t_{i-1}) | (x_{t_i}, t_i); \phi) &:= q(x_{t_{i-1}} | x_{t_i}) P_{B_{time}}(t_{i-1} | (x_i, t_i); \phi) \\ P_F((x_{t_i}, t_i) | (x_{t_{i-1}}, t_{i-1}); \psi) &:= p_\theta(x_{t_i} | x_{t_{i-1}}) P_{F_{time}}(t_i | (x_{i-1}, t_{i-1}); \psi) \end{aligned}$$

Основной целью нашей работы является минимизация $D_{KL}(P_B(\cdot; \phi) \| P_F(\cdot; \psi))$ по ϕ и ψ , то есть изучение поведения модели при обучении двух ядер перехода вместе. Так как в нашей модели с адаптивной генерацией минимизация происходит по двум аргументам сразу, мы предполагаем, что найденное решение будет эффективным. Для вычисления градиента по ϕ в работе используется градиент функции $\mathcal{L}_{TVC}(\tau, \tau')$ [3], где аргументы были получены из P_B .

$$\mathcal{L}_{TVC}(\tau, \tau') = \left[\log \frac{P_F(\tau)}{P_B(\tau | s_n = x)} - \log \frac{P_F(\tau')}{P_B(\tau' | s_n = x)} \right]^2$$

Эксперименты проводились с синтетическими данными состоящими из векторов в \mathbb{R}^2 с распределением в виде дерева. Оценка качества проводилась с помощью выборочной оценки дивергенции Йенсена—Шеннона, полученной методом k -ближайших соседей. В табл. 1 продемонстрированы значения дивергенции Йенсена—Шеннона при значениях параметра k равного 1, 2, 3, 4 соответственно.

	JSD_1	JSD_2	JSD_3	JSD_4
Стандартная	0.86 ± 0.017	0.77 ± 0.013	0.71 ± 0.012	0.67 ± 0.012
Адаптивная	0.63 ± 0.024	0.55 ± 0.019	0.51 ± 0.018	0.48 ± 0.017

Табл. 1: Результаты подсчета метрик, усредненные по 10 запускам. На каждом запуске метрики считались по выборкам размером 10^5 .

В работе была предложена новая модель, использующая комбинацию диффузионных моделей и генеративных потоковых нейронных сетей, описан весь

теоретический вывод формул и удобной параметризации получаемых траекторий. При проведении всех экспериментов бралась уже обученная диффузионная модель, что делает предложенную адаптивную генерацию обучаемой модификацией для стандартных диффузионных моделей. В работе показана эффективность предложенной модели по приведенным метрикам.

ЛИТЕРАТУРА

- [1] Jonathan Ho, Ajay Jain, Pieter Abbeel Denoising Diffusion Probabilistic Models // Advances in Neural Information Processing Systems. Curran Associates, Inc. 2020. Vol. 33, P. 6840–6851.
- [2] Flow Network based Generative Models for Non-Iterative Diverse Candidate Generation / Emmanuel Bengio, Moksh Jain, Maksym Korablyov, Doina Precup, Yoshua Bengio // Advances in Neural Information Processing Systems. Curran Associates, Inc. 2021. Vol. 34, P. 27381–27394.
- [3] Unifying Generative Models with GFlowNets and Beyond / Dinghuai Zhang, Ricky T. Q. Chen, Nikolay Malkin, Yoshua Bengio // arXiv preprint 2209.02606. 2023.

Исследование модели крупных закупок с учетом биржевого индекса и инфляции

Племяшова Лидия Геннадьевна

Кафедра исследования операций

e-mail: lplemyashova@mail.ru

Научный руководитель — к.ф.-м.н. доц. Морозов Владимир Викторович

В работе рассматривается проблема крупных закупок. Она заключается в том, что в результате крупных закупок текущая стоимость ценных бумаг резко увеличивается из-за ухода с рынка продавцов с низкими ценами. Поэтому при единовременной покупке большого объема акций инвестор приобретает часть бумаг по цене сильно выше рыночной. Для уменьшения затрат на покупку акций можно разбить общий объем на части и совершать закупки в течение заданного временного интервала.

Проблема крупных закупок подробно рассматривается в работах [1], [4]. В них предлагаются различные модели влияния крупных закупок на цену актива.

В работе рассматривается модель, построенная по аналогии с моделью Обижаевой и Ван [4]. Решается задача, поставленная в работе Бертсима и Ло [1], то есть задача минимизации средней величины затрат, которые требуются для покупки актива в заданном объеме. Также рассматриваются три модификации модели крупных закупок: одна из них учитывает влияние инфляции на цену актива, вторая учитывает влияние биржевого индекса, третья учитывает одновременное влияние биржевого индекса и инфляции на цену актива. Для каждой модели ставится задача минимизации средних затрат и находится оптимальная стратегия инвестора.

Для решения поставленных задач минимизации используются метод Лагранжа и метод динамического программирования. В результате находятся оптимальные стратегии двух видов: статистические и динамические. Для каждой модели реализован алгоритм для численного нахождения оптимальной стратегии при заданных параметрах.

ЛИТЕРАТУРА

- [1] Bertsimas D., Andrew W. Lo. Optimal control of execution costs // Journal of Financial Markets. 1998. N 1. P. 1–50.
- [2] Hausman J., Lo A., MacKinlay C. An ordered probit analysis of transaction stock prices. // Journal of Financial Economics. 1992. Vol. 31, N 3. P. 319–379.
- [3] Keim D., Madhavan A. The anatomy of the trading process. // Journal of Financial Economics. 1995. Vol. 37, N 3. P. 371–398.
- [4] Obizhaeva A., Wang J. Optimal trading strategy and supply/demand dynamics // Journal of Financial Markets. 2013. Vol. 16, N 1. P. 1–32.

Оптимальная стратегия экономического регулирования с учетом импорта

Работа удостоена диплома 3 степени

Сиваченко Артём Дмитриевич

Кафедра исследования операций

e-mail: artem.sivachenko@gmail.com

Научный руководитель — д.ф.-м.н. проф. Васин Александр Алексеевич

Введение. Одной из задач государства является регулирование внутренних отраслевых рынков. Для повышения национального благосостояния государству может быть целесообразно поддерживать отечественных производителей, инвестируя в отрасль и внедряя таможенные пошлины.

При введении таможенных пошлин государство также обязано осуществлять регулирование для предотвращения уклонений от уплаты налогов. Это требует разработки механизмов, позволяющих обеспечивать честное поведение при минимальных затратах.

Задача выбора стратегии государства. Рассматривается рынок однородного товара с двумя производителями в условиях совершенной конкуренции. Производители представляют из себя множества независимых небольших фирм, не способных самостоятельно влиять на равновесную цену. Функция спроса $D(p)$ имеет стандартные свойства. Производители характеризуются функциями полных затрат $C^a(V)$ и $C^b(V)$. Государство может влиять на рынок, устанавливая импортную пошлину t , $t \in [0, +\infty)$.

Отечественный производитель может провести N этапов модернизации технологического процесса, чтобы снизить себестоимость единицы товара. Модернизация проводится последовательно, кроме того, для неё нужны финансовые вложения в размере F_1, F_2, \dots, F_{N-1} . Производитель может полу-

чить финансы, реализуя товар на рынке, либо получая инвестиции от государства.

Утверждение 1. *В случае, если государство может совершать неограниченные инвестиции в отечественного производителя, общественное благосостояние достигает максимума при отсутствии таможенной пошлины. Модернизация до необходимого уровня должна происходить без задержек, недостающие для модернизации средства выделяются государством.*

Алгоритм расчета таможенных пошлин. Для расчета таможенных пошлин был разработан соответствующий алгоритм. Основная идея заключается в последовательном поиске наилучшей возможности раньше провести модернизацию в определённом периоде, повысив таможенные пошлины в предыдущие периоды.

Результаты сравнения стратегий. Показано, что применение таможенных пошлин понижает благосостояние в конкретный период, но может положительно сказаться на благосостоянии в будущем. Инвестиции позволяют провести модернизацию быстрее, однако, необходимо рассматривать альтернативы и учитывать эффект на всю экономику, поэтому инвестирование в конкретный рынок может быть нецелесообразным.

Модель налогового регулирования. В работе исследована модель организации налоговой инспекции с перекрёстными проверками. Далее эта модель сравнивалась с уже исследованной в работах [1, 2, 3] трёхуровневой иерархической моделью с точки зрения минимизации затрат на обеспечение честного поведения.

В иерархической модели плательщик с некоторой вероятностью проверяется инспектором первого уровня, с некоторыми вероятностями могут проводиться ревизии. При выявлении уклонения плательщик и инспектора платят штрафы с заданными коэффициентами f и f_i . Максимальное значение штрафа для инспектора соответствует его увольнению, при котором он переходит на работу с заданной альтернативной зарплатой s_a , при расчете учитывается коэффициент приведения α . Плательщик может войти в коалицию с проверяющими его инспекторами, если это будет взаимовыгодно. На верхнем уровне находится руководитель, который является заведомо честным и никогда не берёт взятки. Задана стоимость его проверки c_w .

В модели с перекрёстными проверками инспекторы, кроме руководителя, находятся на одном уровне. Повторная проверка проводится с некоторой вероятностью инспектором из того же уровня.

Теорема 1. *Оптимальная вероятность повторной проверки инспекторами равна 0. Модель с перекрёстными проверками сводится к двухуровневой иерархической модели.*

Теорема 2. *Для иерархической модели налоговой инспекции оптимальным является выбор трёх уровней организации относительно двух уровней, если*

верно неравенство:

$$\frac{s_a}{f + f_1} + \frac{2\sqrt{c_w}}{\sqrt{\alpha(f + f_1)}} + \frac{f_1}{\alpha f} - \frac{2\sqrt{c_w}}{\sqrt{\alpha f}} < 0,$$

где f_1 — оптимальный штрафной коэффициент первичной проверки, рассчитанный для трёхуровневой организации.

Сравнение затрат. При различных условиях системы может быть выгодно использовать как трёхуровневую, так и двухуровневую модели. Чем дороже стоимость проверки начальником и ниже альтернативная зарплата, тем больше уровней следует использовать.

ЛИТЕРАТУРА

- [1] Васин А. А., Николаев П. В., Огнева М. С. Задача отбора кандидатов на работу в налоговую инспекцию // 2018.
- [2] Sanchez I., Sobel J. Hierarchical design and enforcement of income tax polices // J. of Public Economics. 1985. Vol. 26, P. 1–18.
- [3] Chandler P., Wilde L. A general characterization of optimal income tax enforcement // Review of Economics Studies. 1987. Vol. 65. P. 165–189.

Исследование систем обслуживания с повторными вызовами

Работа удостоена диплома 3 степени

Смирнов Иван Сергеевич

Кафедра математической статистики

e-mail: smirnov_i.s@mail.ru

Научный руководитель — д.ф.-м.н. проф. Ушаков Владимир Георгиевич

Системы массового обслуживания (СМО) со временем становятся сложнее, больше. Количество приборов, обрабатывающих заявки, кратно увеличивается, как и число каналов поступления вызовов, так как растут потребности в обработке различных данных, очередей, построении облачных хранилищ, суперкомпьютеров и вычислительных центров. Именно поэтому появляются новые математические модели, которые помогают в воссоздании новых сложных процессов. Одними из таких нововведений стали, так называемые системы массового обслуживания с орбитами. Такие СМО сложны в аналитическом исследовании.

Заявки, которые поступают в систему в момент занятости прибора и не могут встать в очередь, а потом повторно пытаются попасть на обслуживание через некоторое время, называются повторными. Под повторными заявками могут подразумеваться, например, абоненты, которые не смогли дозвониться на нужный номер с первого раза и повторяют свои попытки пока не добьются желаемого, или пользователи сети интернет, пытающиеся получить доступ к неоткрывающемуся сайту.

Целью данной работы является демонстрация того, что более простые модели вида $M/M/1/0$, $M/M/1/1$, которые могут быть аналитически исследованы и обобщены, при должных параметрах не хуже в смысле некоторых характеристик, чем модель с динамически изменяющимися параметрами, которая плохо поддаётся анализу. В связи с этим была рассмотрена СМО вида $M/G/1$ с орбитой и переменной интенсивностью обращения повторных заявок, исследованы модели $M/M/1/0$, $M/M/1/1$ с орбитами и выведены их характеристики.

В работе были найдены формулы для расчёта среднего количества повторных заявок в системах $M/M/1/0$, $M/M/1/1$, выведены теоремы о производящих функциях для данных моделей с указанием условий существования стационарного распределения случайного процесса количества заявок на орбите, а также найдены формулы для расчёта ряда других характеристик систем, которые позволяют оценить их эффективность.

В ходе сравнения указанных СМО было показано, что при некоторой фиксированной интенсивности обращений повторных заявок в модели Г. И. Фалина в стационарном режиме найдётся такая частота s для системы $M/M/1/0$, при которой длина очереди заявок на орбите будет не больше, чем для первой. Если же изменять интенсивность одинаково на двух моделях, то разность средних длин очередей повторных заявок стремиться к нулю. В случае системы $M/M/1/1$ и вовсе было замечено, что длина очереди в ней почти всегда меньше, а различие может быть на порядок. Для наглядной демонстрации этих фактов в работе были приведены графики, построенные на промоделированных данных.

Таким образом, целесообразность в использовании СМО вида $M/M/1/0$ и $M/M/1/1$ полностью может быть оправдана, так как при проектировании высоконагруженных систем важно учитывать ограниченность ресурсов, которые дополнительно затрачиваются на механизм динамического регулирования системы.

ЛИТЕРАТУРА

- [1] Falin G. I. A survey of retrial queues // Queueing Systems 7, 1990. С. 127—167. DOI: <https://doi.org/10.1007/BF01158472>.
- [2] Jonin G.L., Sedol. J.J. Telephone systems with repeated calls // 1970. С. 435/1—435/5
- [3] Jonin G.L., Sedol J.J. Fully-availability groups with repeated calls and time of advanced service // 1973. С. 137/1—137/4
- [4] Ушаков В. Г., Матвеев В. Ф. Системы массового обслуживания // Издательство Московского университета, 1984.

Генетические алгоритмы корректного понижения значности признаков в задаче классификации

Работа удостоена диплома 2 степени

Сорокин Олег Владиславович

Кафедра математических методов прогнозирования

e-mail: ovs02@mail.ru

Научный руководитель — д.ф.-м.н. доц. Дюкова Елена Всеволодовна

Одним из подходов к задаче классификации по прецедентам является логический подход, основанный на применении аппарата дискретной математики. Преимущество указанного подхода состоит в возможности построения алгоритма классификации по небольшой обучающей выборке и без дополнительных предположений о вероятностных распределениях объектов. При этом особое внимание уделяется корректным алгоритмам, правильно распознающим всю обучающую выборку.

На практике часто встречаются задачи с целочисленными признаками высокой значности, то есть с признаками, имеющими большое число допустимых значений. Этап обучения логического классификатора состоит в поиске фрагментов описаний обучающих объектов, позволяющих различать объекты из разных классов. Наличие признаков высокой значности обычно приводит к повышению временных затрат при поиске искомых фрагментов и к падению качества классификации. В связи с указанной особенностью логических классификаторов актуальной является поставленная академиком РАН Ю.И. Журавлевым задача «корректного» понижения значности данных, т.е. такого преобразования исходных данных, при котором описания объектов из разных классов остаются различимыми. Методика корректного перекодирования данных впоследствии была развита в [1-3]. Показано, что поставленная задача может быть сведена к поиску специальных наборов столбцов булевой матрицы, построенной по обучающей выборке. Каждый такой набор столбцов порождает корректное преобразование исходных данных и называется кодирующим покрытием указанной матрицы.

Число кодирующих покрытий быстро растет с ростом размеров обучающих данных, поэтому возникает сложная переборная задача поиска оптимального по качеству классификации кодирующего покрытия. На практике эта задача решается приближенно, в основном с использованием генетических алгоритмов. В частности, в работе [3] предложен и апробирован на реальных задачах генетический алгоритм КОДЗ, решающий задачу поиска оптимального корректного перекодирования.

Для реализации генетического алгоритма КОДЗ используется адаптированная схема генетического алгоритма из [4] с бинарным представлением особей и оригинальным оператором мутации с переменным числом мутирующих генов, что позволяет минимизировать влияние данного оператора на особь на

ранних этапах работы и усиливать его влияние с увеличением числа итераций.

Целью работы является разработка генетического алгоритма поиска оптимального корректного перекодирования обучающих данных, превосходящего по качеству классификации алгоритм КОДЗ. Фактически в работе предложены две модификации алгоритма КОДЗ, названные G-ENCODE и G-ENCODE+.

Одной из особенностей алгоритма G-ENCODE по сравнению с алгоритмом КОДЗ является использование иной функции приспособленности. В качестве значения функции приспособленности особи берется максимальная значность признака среди признаков в перекодированных данных, соответствующих этой особи. Другая особенность алгоритма G-ENCODE состоит в изменениях, внесенных в процедуру восстановления допустимости решения, которая заключается в выделении в кодирующем покрытии неприводимого покрытия. При выделении неприводимого покрытия столбцы матрицы просматриваются в особом порядке, что приводит к более эффективному понижению значности исходных данных.

В алгоритме G-ENCODE+ используется новый способ инициализации особей, основанный на применении алгоритма G-ENCODE к специальной популяции. После создания нулевой популяции каждой особи присваивается неотрицательный вес, характеризующий качество соответствующего решения. В дальнейшем эти веса также влияют на этапы скрещивания и селекции, что значительно повышает качество классификации.

В работе показано, что алгоритм G-ENCODE+ на большинстве тестируемых задач превосходит по качеству распознавания алгоритм КОДЗ. При этом вычислительные сложности указанных алгоритмов эквивалентны.

ЛИТЕРАТУРА

- [1] Обработка вещественнозначной информации логическими процедурами распознавания / Дюкова Е. В., Журавлев Ю. И., Песков Н. В., Сахаров А. А. // Искусственный интеллект. 2004. № 2. С. 80–85.
- [2] Combinatorial (Logical) Data Analysis in Pattern Recognition Problems / Djukova E. V., Inyakin A. S., Peskov N. V., Sakharov A. A. // Pattern Recognition and Image Analysis. 2005. Vol. 15 № 1. P. 46–48.
- [3] Дюкова Е. В., Сизов А. В., Сотнезов Р. М. Об оптимальном корректном перекодировании целочисленных данных в распознавании // Информатика и ее применения. 2011. Т. 6, № 4. С. 61–65.
- [4] Sotnezov R. M. Genetic Algorithms for Problems of Logical Data Analysis in Discrete Optimization and Image Recognition // Pattern Recognition and Image Analysis. 2009. Vol. 19, № 3. P. 469–477.

Проверка моделей для логики ветвящегося времени в исчислении индуктивных конструкций

Работа удостоена диплома 3 степени

Тищенко Анастасия Сергеевна

Кафедра математической кибернетики

e-mail: astisch.moscow@gmail.com

Научный руководитель — к.ф.-м.н. доц. Подымов Владислав Васильевич

Для решения задачи проверки правильности поведения системы с конечным числом состояний методом проверки моделей [1] часто используется представление этой системы особым конечным автоматом (моделью Крипке), а требований к системе — формулой подходящего логического языка, позволяющего описывать свойства правильности поведения системы во времени. В [2] был разработан алгоритм построения доказательства выполнимости формул некоторого фрагмента одного такого языка, СТЛ, на моделях Крипке, предназначенный для популярного средства автоматизации доказательств Соq и лежащего в его основе логического исчисления индуктивных конструкций. Цель данной работы — расширить полученные в [2] результаты на весь язык СТЛ.

Модель Крипке. Пусть задано множество атомарных высказываний AP . Модель Крипке — это четверка (S, S_0, T, L) , где: S — конечное множество состояний; S_0 — множество начальных состояний, $S_0 \subseteq S$; $T \subseteq S \times S$ — отношение переходов, обладающее свойством тотальности: для любого состояния s_1 существует состояние s_2 , такое что $(s_1, s_2) \in T$; L — функция разметки состояния, $L : S \rightarrow 2^{AP}$.

СТЛ. Формулы СТЛ строятся над множеством атомарных высказываний AP , булевыми операциями, темпоральными операторами линейной логики U («до тех пор, пока»), X («в следующий момент времени») и кванторами пути A («для любого пути из текущего момента времени»), E («существует путь...»). В позитивной нормальной форме (отрицание только над атомарными высказываниями) используется также вспомогательный темпоральный оператор W («всегда или до тех пор, пока»). Таким образом синтаксис языка можно задать следующей формой Бэкуса—Наура (БНФ):

$$\varphi ::= p \mid \neg p \mid \varphi \vee \varphi \mid \varphi \wedge \varphi \mid AX\varphi \mid EX\varphi \mid \varphi AU\varphi \mid \varphi EU\varphi \mid \varphi AW\varphi \mid \varphi EW\varphi,$$

где $p \in AP$.

Соq — это программное средство, предоставляющее логическую среду, позволяющую описывать математические теории и строить доказательства в интерактивном полуавтоматическом режиме. Согласно известному соответствию Карри—Говарда [3], теореме в логике предикатов можно понимать как тип, а её конструктивное доказательство — как обоснование того, что имеется хотя бы один терм этого типа. Этот терм тогда можно понимать как собственно доказательство. Под типом понимается особая запись, которой отвечает множество термов, имеющих этот тип. Правила вывода (способы

преобразования суждений о типах термов в ходе доказательства теорем, позволяющие строить термы-доказательства) в Coq называются тактиками. В Coq есть как встроенные тактики, заданным документированным способом преобразующие доказываемые утверждения, так и языки создания своих собственных тактик — например, Ltac [4].

Результаты. Суждения о типе имеют вид $\langle \Gamma \vdash A \rangle$, где Γ — контекст (множество объектов населенных типов, которое можно понимать как совокупность доказанных теорем) и A — тип, то есть доказываемая теорема. Доказательство теоремы — последовательность суждений о типе, связанных одним шагом доказательства. Под шагом доказательства понимается результат работы одной из базовых тактик или связки языка Ltac .

Центральное утверждение проверки моделей — утверждение о выполнимости формулы φ CTL в состоянии s модели Крипке M — записывается в виде суждения $\Gamma_{\text{CTL}} \vdash \text{satisfies } t_M t_\varphi t_s$, где Γ — набор определений, задающих отношение выполнимости *satisfies*, и t_M, t_φ, t_s — термы, представляющие M , φ и s соответственно.

В данной работе предлагается тактика *solver* на языке Ltac , индуктивно по структуре формулы доказывающая теорему *satisfies* $t_M t_\varphi t_s$, если формула φ действительно выполняется в состоянии s модели M .

Теорема. Пусть выполнены следующие условия:

1. t_M - терм, представляющий модель Крипке M .
2. t_φ - терм, соответствующий формуле CTL φ , представленной в позитивной нормальной форме.
3. t_s - терм, представляющий состояние s модели M .
4. Формула φ выполнима в состоянии s модели M .

Тогда *solver* успешно строит доказательство суждения *satisfies* $t_M t_\varphi t_s$.

ЛИТЕРАТУРА

- [1] Baier C., Katoen J. P. Principles of Model Checking. The MIT Press, 2008.
- [2] Ларочкин П. В. Решение задачи проверки моделей в исчислении индуктивных конструкций. // Магистерская диссертация. МГУ имени М.В. Ломоносова, 2023.
- [3] Sorensen M., Urzyczyn P. Lectures on the Curry–Howard Isomorphism // Studies in logic and the foundations of mathematics. Vol. 149. Elsevier B. V., 2006.
- [4] Jedynek W. Operational Semantics of Ltac. // Master Thesis. Uniwersytet Wroclawski, 2013.

О проверке биюнктивности полиномов Жегалкина

Работа удостоена диплома 3 степени

Ушаков Дмитрий Андреевич

Кафедра математической кибернетики

e-mail: s02190225@gse.cs.msu.ru

Научный руководитель — д.ф.-м.н. проф. Селезнева Светлана Николаевна

Рассматривается задача проверки свойства *биюнктивности* функции алгебры логики. В качестве представления функций алгебры логики используются *полиномы Жегалкина*.

В статье [1] доказано, что задача распознавания биюнктивности функций алгебры логики, записанных полиномами Жегалкина, имеет полиномиальную сложность относительно количества переменных полинома и его длины.

В работе произведена декомпозиция алгоритма из статьи [1]: для каждого из утверждений указана чёткая последовательность действий, строящая соответствующий объект или распознающая соответствующее свойство. Алгоритм запрограммирован на языке программирования Python, для каждого вспомогательного утверждения отведена отдельная процедура. (В таблице ниже данный алгоритм обозначен G .)

Также в работе автором предложен собственный алгоритм, проверяющий биюнктивность функции алгебры логики по её полиному Жегалкина. В сравнении с алгоритмом из [1] он проще по структуре, но хуже по порядку сложности. Алгоритм также запрограммирован на языке программирования Python. (В таблице ниже данный алгоритм обозначен A .)

Программы были запущены на полиномах всех функций трёх и четырёх переменных (тестирование выполнялось на 8-ядерном процессоре Apple M1 с частотой 3.20 ГГц). Краткая информация о длительности работы программ приведена в следующей таблице.

Местность	3	4
Наибольшее время работы G , мс	0.761	2.315
Наибольшее время работы A , мс	0.53	1.779
Среднее время работы G , мс	0.414	1.098
Среднее время работы A , мс	0.299	0.894

ЛИТЕРАТУРА

- [1] Горшков С. П. О сложности распознавания мультиаффинности, биюнктивности, слабой положительности и слабой отрицательности булевых функций // Обозрение прикладной и промышленной математики. Серия «Дискретная математика». 1997. Т. 4, № 2. С. 216–237.

Выявление устойчивых состояний и переходных процессов в эволюции сложных динамических систем на примере рядов генной экспрессии различных биологических состояний

Антонов Кирилл Николаевич

*Кафедра системного анализа
e-mail: prosto-kirich@mail.ru*

Научный руководитель — к.ф.-м.н. доц. Востриков Иван Васильевич

Научный консультант — д.ф.-м.н. доц. Смирнов Сергей Николаевич

Введение. В качестве динамической системы, подвергнутой анализу, был выбран процесс называемый эпителиально-мезенхимальным переходом. Он играет одну из ключевых ролей в клеточной биологии. В частности он определяет в значительной мере развитие рака, процесс фиброза в тканях и развитие организма. Любой биологический процесс можно представить в виде набора состояний и переходными процессами между ними. В клеточной биологии это утверждение иллюстрирует ландшафт Уоддингтона [1]. За любым динамическим процессом стоит та или иная управленческая структура, называемая регуляторной сетью. Эпителиально-мезенхимальный переход (ЭМП) представляет собой сегмент ландшафта Уоддингтона. В нем присутствуют два устойчивых терминальных состояния клеток: эпителий и мезенхима. Переход между этими состояниями осуществляется через ряд промежуточных состояний, в которых клетки характеризуются сложным профилем свойств, содержащим как свойства эпителия, так и свойства мезенхимы. Количество этих промежуточных состояний неизвестно и, как следствие, мы не можем эффективно управлять этим переходом. А учет того факта, что ЭМП играет огромную роль в развитии и течении рака, анализ и моделирование ЭМП с целью восстановления по экспериментальным данным его динамики по ландшафту Уоддингтона является крайне актуальной и важной задачей.

Цели и задачи. Цель моей работы: решить важную биологическую проблему — выяснить сколько состояний имеет клетка в процессе ЭМП. Для решения поставленной задачи необходимо было найти или придумать высокоточный метод, который позволяет обнаруживать состояния многомерной нелинейной динамической системы. На текущий момент все работы по данной проблеме, анализируя многомерную динамическую систему, оперируют в двумерных пространствах, выбирая из всех переменных системы наиболее важные на взгляд исследователей. Очевидно, что такой подход имеет существенные недостатки.

Дизайн исследования. Проведенное мной исследование делится на два последовательных этапа: теоретический и экспериментальный. Теоретический этап содержит в себе взятую из литературы регуляторную сеть и систему нелинейных дифференциальных уравнений, которая стоит за этой регуляторной сетью [2]. Для выбранной нелинейной динамической системы бы-

ли построены пробные траектории и проанализированы их временные срезы. Для выявления кластерной структуры временных срезов была выбрана искусственная нейронная сеть с обучением без учителя, которая позволяет без предварительных априорных гипотез обнаруживать в потоке векторов, поданных на вход, кластерную структуру, если она существует. Экспериментальный этап был связан с динамическими данными взятыми из базы Gene Expression Omnibus. На экспериментальных данных я попытался подтвердить методами машинного обучения, что количество выявленных кластеров в экспериментальных данных совпало с результатами теоретического этапа.

Теоретический этап. В основе теоретической модели ЭМП лежит регуляторная сеть. Это направленный взвешенный граф с обратными связями. В качестве уравнения для описания уровней экспрессии генов исследуемой сети выбрана функция Хилла.

$$F_i = \dot{x}_i = \frac{dx_i}{dt} = \sum_{j \in A_i} \frac{a_{ji} x_j^n}{s_{ji}^n + x_j^n} + \sum_{j \in B_i} \frac{b_{ji} s_{ji}^n}{s_{ji}^n + x_j^n} - k_i x_i. \quad (1)$$

На основе выписанной системы дифференциальных уравнений (1) были порождены траектории, сечения которых по времени были входными векторами для выбранной нейронной сети. Временные срезы построенных пробных траекторий были визуализированы с помощью самоорганизующихся карт Кохонена (СОК) [3]. СОК это нейронная сеть без учителя, предназначенная для визуализации многомерных данных и решения задач кластеризации. СОК — это двухстадийная процедура. На первой стадии выделяется кластерная структура, а на второй раскраска, позволяющая проинтерпретировать результаты. По итогам решения задачи визуализации многомерных данных был сделан вывод, что в рассматриваемой системе 6 состояний. Кроме визуализации многомерных данных СОК-карты позволяют увидеть какие генетические маркеры участвуют в формировании кластерной структуры. Было установлено, что маркеры miR-200, CDH1 и VIM наиболее точно описывают кластерную структуру. На основе значений этих переменных я получил биологическую интерпретацию для каждого из кластеров. А именно: для мезенхимы, для эпителия, и для 4 промежуточных состояний ЭМП.

Выявленные по кластерной структуре состояния мы интерпретируем как аттракторы. Так как нет теоретических методов классификации аттракторов в многомерном фазовом пространстве, проверка их устойчивости проводилась экспериментальным путем. Анализ показал, что каждый из аттракторов является устойчивым. Тогда встает вопрос о том, за счет чего осуществляются переходы между различными состояниями. Я высказываю гипотезу о том, что переходные процессы осуществляются за счет клеточного стресса и экспрессионного шума.

Экспериментальный этап. Для анализа экспериментальных данных из базы Gene Expression Omnibus были выбраны временные ряды геной экспрессии для того же набора маркеров, что и на теоретическом этапе. Поскольку говорилось, что экспрессионный шум является одним из механизмов

эпителиально-мезенхимального перехода, я ввел уровень шума 10 процентов, что вполне согласуется с экспериментальными данными. Для изучения нелинейной динамики экспериментальных временных рядов был выбран метод рекуррентных диаграмм. С помощью метода скользящего окна на основе рекуррентной диаграммы, построенной по экспериментальным рядам геной экспрессии, был подсчитан набор метрик [4]. Эти метрики характеризуют нелинейную динамику системы. К ним относятся, например, метрика рекуррентности, детерминизма и залипания входного сигнала. Они описывают разные стороны динамики процесса. В литературе принято считать, что ЭМП может иметь от 2 до 7 состояний. Чтобы ответить на вопрос о количестве состояний без априорной гипотезы, временные срезы новых рядов из метрик, описывающих динамические свойства системы, были кластеризованы с помощью метода k -ближайших соседей на разное количество кластеров. А именно от 2 до 7. Наибольшая точность метода была достигнута при кластеризации на 6 классов. Таким образом, мы получили совпадение результатов теоретического подхода и анализа экспериментальных данных.

Выводы. Было выделено 6 состояний в динамике регуляторной сети, в отличие от литературы, где количество состояний варьируется от 2 до 7. Выделенные кластеры интерпретируются как аттракторы динамической системы. Все они оказались устойчивыми аттракторами. Была высказана гипотеза о том, что регуляторную функцию в эпителиально-мезенхимальном переходе играют клеточный стресс и экспрессионный шум. Динамический анализ метрик, полученных при помощи рекуррентных диаграмм показал, что наибольшая точность классификации состояний регуляторной сети достигается для 6 кластеров. Таким образом был найден ответ на важный биологический вопрос.

ЛИТЕРАТУРА

- [1] Ladewig J., Koch P., Brüstle O. Leveling Waddington: the emergence of direct programming and the loss of cell fate hierarchies // *Molecular Cell Biology*. 2013. Vol. 14, P. 225–235.
- [2] Lang J., Nie Q., Li C. Landscape and kinetic path quantify critical transitions in epithelial-mesenchymal transition // *Biophysical Journal*. 2021. Vol. 120, P. 4484–4500.
- [3] Kohonen T. Self-organized formation of topologically correct feature maps // *Biological Cybernetics*. 1982. Vol. 43, P. 59–69.
- [4] Webber-jr C. L., Zbilut J. P. Recurrence quantification analysis of nonlinear dynamical systems // *Tutorials in Contemporary Nonlinear Methods for the Behavioral Sciences*. 2005. Vol. 94, P. 26–94.

Управление инерционным объектом в среде с препятствием

Евдокимов Евгений Дмитриевич

Кафедра оптимального управления

e-mail: evgeny.evd@yandex.ru

Научный руководитель — д.ф.-м.н. проф. Григоренко Николай Леонтьевич

Постановка задачи. Рассматривается модель гусеничного робота с управляемыми параметрами [1]:

$$\dot{x} = V \cos \theta, \quad \dot{y} = V \sin \theta, \quad m\dot{V} = u_1 - R_t, \quad \dot{\theta} = \omega, \quad I_z \dot{\omega} = \frac{Bu_2}{2} - M_r, \quad (1)$$

где x, y - координаты центра тяжести робота; θ - угловое перемещение робота, V - линейная скорость робота, ω - угловая скорость робота; B - расстояние между центральными линиями гусениц; I_z - момент инерции массы робота относительно вертикальной оси, проходящей через его центр тяжести; m - масса робота; R_t - результирующее сопротивление; M_r - момент сопротивления повороту; u_1, u_2 - параметры управления. Дано начальное положение системы (1) $(x_0, y_0, V_0, \theta_0, \omega_0)$ и целевое положение (x_T, y_T) . В плоскости (x, y) имеется препятствие $S(t)$. Его форма такова, что вычислима функция оценки снизу для расстояния от (x, y) до препятствия: $\text{dist}(x, y, t) \in C^3[\mathbb{R}^2 \times [0, T]]$ (трижды непрерывно дифференцируема при $\text{dist}(x, y, t) > 0$). Требуется найти позиционное управление [2] (u_1, u_2) , гарантирующее выполнение соотношения

$$\|(x_T, y_T) - (x(T), y(T))\| \leq c_{\text{tol}}; \quad (x(t), y(t)) \notin S(t), \quad t \in [0, T]. \quad (2)$$

Здесь $c_{\text{tol}} > 0$. Момент T прихода фазового вектора в целевую точку не фиксирован.

Схема решения. Система (1) для заданных граничных значений является управляемой. Назначим n опорных точек $\{(x_1, y_1), \dots, (x_n, y_n)\}$, $(x_1, y_1) = (x_0, y_0)$, $(x_n, y_n) = (x_T, y_T)$. Методом наименьших квадратов найдем полином $y(x) = A_0 + \sum_{i=1}^n A_i x^i$ проходящий через опорные точки. Искомые управления найдем с использованием метода аналитического конструирования агрегированных регуляторов (АКАР)[1]. Введем первую и третью макропеременные (Ψ_1, Ψ_3) и соответствующие функциональные уравнения ($\lambda_1, \lambda_4 > 0$):

$$\Psi_1 = \omega - \phi_1, \quad \Psi_3 = V - V_0, \quad \dot{\Psi}_1 + \lambda_1 \Psi_1 = 0, \quad \dot{\Psi}_3 + \lambda_4 \Psi_3 = 0.$$

Здесь ϕ_1 - внутреннее управление. Определим вторую макропеременную и соответствующее функциональное уравнение:

$$\Psi_2 = y - A_0 - \sum_{i=1}^n A_i x^i + \frac{\rho c_{\text{rep}}}{(\text{dist}(x, y, t) - c_{\text{dist}})^\xi}, \quad \ddot{\Psi}_2 + \lambda_2 \dot{\Psi}_2 + \lambda_3 \Psi_2 = 0$$

Управления ϕ_1, u_1, u_2 решающие задачу управления имеют вид :

$$\phi_1 = [V^2(\cos^2 \theta \gamma_{xx} + 2 \sin \theta \cos \theta \gamma_{xy} + \sin^2 \theta \gamma_{yy}) + (\lambda_4(V_0 - V) + \lambda_2 V)(\gamma_x \cos \theta + \gamma_y \sin \theta) + \lambda_3 \Psi_2 + 2V(\gamma_{xt} \cos \theta + \gamma_{yt} \sin \theta) + \lambda_2 \gamma_t + \gamma_{tt}] / [\gamma_x V \sin \theta - \gamma_y V \cos \theta],$$

$$u_2 = \frac{2}{B}(I_z(V \cos \theta \xi_x + V \sin \theta \xi_y + \lambda_4(V_0 - V)\xi_V + \omega \xi_\theta + \xi_t - \lambda_1(\omega - \phi_1)) + M_r),$$

$$u_1 = m\lambda_4(V_0 - V) + R_t, \text{ где } \gamma_x = \frac{\partial \Psi_2}{\partial x}, \gamma_y = \frac{\partial \Psi_2}{\partial y}, \gamma_{xx} = \frac{\partial^2 \Psi_2}{\partial x^2}, \gamma_{yy} = \frac{\partial^2 \Psi_2}{\partial y^2}, \gamma_{xy} = \gamma_{yx} = \frac{\partial^2 \Psi_2}{\partial x \partial y}, \gamma_t = \frac{\partial \Psi_2}{\partial t}, \gamma_{tt} = \frac{\partial^2 \Psi_2}{\partial t^2}, \gamma_{tx} = \gamma_{xt} = \frac{\partial^2 \Psi_2}{\partial x \partial t}, \gamma_{ty} = \gamma_{yt} = \frac{\partial^2 \Psi_2}{\partial y \partial t}, \xi_x = \frac{\partial \phi_1}{\partial x}, \xi_y = \frac{\partial \phi_1}{\partial y}, \xi_V = \frac{\partial \phi_1}{\partial V}, \xi_\theta = \frac{\partial \phi_1}{\partial \theta}, \xi_t = \frac{\partial \phi_1}{\partial t}.$$

Справедлива следующая лемма о точности достижения (x_T, y_T) :

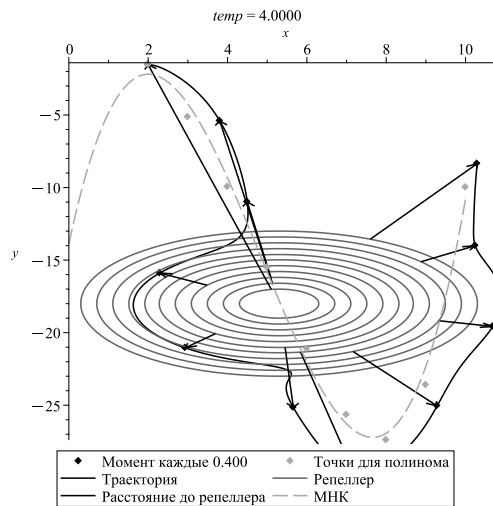
Лемма 1. Пусть $x(T) = x_T, |\Psi_2(x(T), y(T), T)| < \varepsilon, \text{dist}(x(T), y(T), T) > c_{dist}$ и ошибка аппроксимации МНК на x_T ($y_T - A_0 - \sum_{i=1}^n A_i x_T^i$) равна η . Тогда справедлива следующая оценка расстояния до целевой точки:

$$|y(T) - y_T| < \varepsilon + |\eta| + \frac{c_{rep}}{(\text{dist}(x(T), y(T), T) - c_{dist})^\xi}.$$

Численный эксперимент. Выберем опорные точки траектории следующим образом $(x_i, y_i) = (i + 1, e^{0.45(i+1)} - (i + 1)^2), i = \overline{1..9}$. В качестве препятствия рассмотрим расширяющуюся окружность радиуса $1 + t$:

$$\text{dist}(x, y, t) = \sqrt{(x - x_c)^2 + (y - y_c)^2} - (1 + t), (x_c, y_c) = (5.3, -18).$$

Порядок аппроксимирующего полинома положим $n = 3$. Параметры робота: $I_z = 2, B = 1.5, R_t = 0.1, M_r = 0.1, m = 4$. Параметры макропеременных: $\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = 10, V_0 = 14$. Начальные условия: $x(0) = x_1, y(0) = y_1, V(0) = 1, \theta(0) = 0, \omega(0) = 0$. Параметры окружности-репеллера: $\rho = 1, c_{rep} = 8, c_{dist} = 1, \xi = 1$. На рисунке опорная траектория обозначена пунктиром. Траектория движения - сплошной линией. Стрелки соответствуют расстоянию от расширяющегося препятствия до $(x(t), y(t))$ в текущий момент.



ЛИТЕРАТУРА

[1] Веселов Г. Е., Скляр А. А., Скляр С. А. Синергетический подход к управлению траекторным движением мобильных роботов в среде с препятствиями // Мехатроника, автоматизация, управление, № 7, 2013

[2] Красовский Н. Н., Управление динамической системой // М. Наука, 1985, 520 с.

Стабилизация квадрокоптера при наличии параметрических возмущений

Хмыров Алексей Константинович

Кафедра нелинейных динамических систем и процессов управления

e-mail: alexey.khmyrov@mail.ru

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

Постановка задачи. Стабилизация нелинейной динамической системы направлена на изучение свойств элементов системы и их взаимодействия, а также разработку алгоритма управления, обеспечивающего прогнозируемое или желаемое поведение объекта управления и его контролируемый переход на нужный динамический режим.

В рамках выпускной квалификационной работы разработаны математическая модель квадрокоптера, прошедшая тестирования в различных режимах полёта, и система управления, обеспечивающая автономный полет по подготовленной траектории и доставку груза в заданные координаты. Проанализированы методы стабилизации системы, и выбран линейный регулятор с квадратичным функционалом качества, как основной элемент управления ориентацией летательного аппарата. Проведены исследования устойчивости системы, выявившие определённые ограничения на управляющие воздействия.

Математическая модель объекта управления.

$$\begin{cases} \ddot{x} = \frac{P}{m}(\sin(\phi)\sin(\psi) - \cos(\phi)\sin(\theta)\cos(\psi)) \\ \ddot{y} = \frac{P}{m}\cos(\phi)\cos(\theta) - g \\ \ddot{z} = \frac{P}{m}(\sin(\phi)\cos(\psi) + \cos(\phi)\sin(\theta)\sin(\psi)) \\ \ddot{\phi} = \frac{M_x}{J_x} - \frac{J_z - J_y}{J_x}\dot{\psi}\dot{\theta} \\ \ddot{\psi} = \frac{M_y}{J_y} - \frac{J_x - J_z}{J_y}\dot{\phi}\dot{\theta} \\ \ddot{\theta} = \frac{M_z}{J_z} - \frac{J_y - J_x}{J_z}\dot{\phi}\dot{\psi} \end{cases}$$

Математическая модель силовой установки.

$$\begin{cases} \dot{\omega} = \frac{1}{J_r}K_t I - b_m \omega \\ \dot{I} = \frac{1}{L}(\omega_{cmd} - \omega) - R_m I - K_{emf} \omega \end{cases}$$

Линеаризация. Сложность нелинейных систем связана с зависимостью выходной величины от нелинейных функций, что обуславливает сложное поведение системы. Для анализа и синтеза таких систем используются разные методы, включая линеаризацию. Она важна при разработке систем управления, так как упрощает анализ и проектирование контура управления. Локальная линеаризация позволяет оценить устойчивость модели, что помогает определить границы применимости алгоритмов управления.

Метод стабилизации. В качестве метода стабилизации ориентации системы был выбран алгоритм синтеза линейного регулятора с квадратичным функционалом качества. Задачей синтеза является определение матрицы

коэффициентов обратной связи по состоянию $u = -Kx$, минимизирующей квадратичный функционал:

$$J(u) = \int_0^{\infty} ([u(t) - u_d(t)]^T R [u(t) - u_d(t)] + [x(t) - x_d(t)]^T Q [x(t) - x_d(t)]) dt$$

Матрица коэффициентов обратной связи (регулятора) находится из соотношения: $K = R^{-1} B^T S$, где S является решением нелинейного матричного уравнения Риккати:

$$A^T S + S A - S B R^{-1} B^T S + Q = 0$$

Для стабилизации позиционирования был выбран пропорциональный регулятор, параметры которого сформированы на основе исследования устойчивости системы.

Важные моменты и ограничения. Ограничения связаны с воздействием внешних факторов на летательный аппарат. Существенную роль играют исполнительные механизмы, представленные бесколлекторными двигателями постоянного тока. Для согласования математических моделей объекта управления и исполнительных механизмов проводились испытания системы в разных режимах полёта.

Линеаризация не всегда упрощает управление, но облегчает проектирование, так как для линейных систем существует больше методов и алгоритмов управления. Был выбран подход разделения контура управления на отдельные части для стабилизации системы в определённых режимах полёта.

Выводы. Исследование показало, что наличие параметрических возмущений значительно усложняет работу системы управления, когда модель становится более сложной и реалистичной.

Применение линейного регулятора с квадратичным функционалом качества обеспечивает достаточную стабилизацию системы при использовании квадрокоптера для доставки грузов в заданные координаты.

ЛИТЕРАТУРА

- [1] Ким Д. П. Теория автоматического управления. Т.1. Линейные системы. – М.: ФИЗМАТЛИТ, 2003.
- [2] Лебедев А.А., Чернобровкин Л.С. Динамика полета беспилотных летательных аппаратов. – М.: Машиностроение, 1973г.
- [3] Т.Ю. Лемешонок, А.А. Сизова, Н.Е. Баранов, В.А. Санников. Математические модели динамики движения летательных аппаратов. 2020г.
- [4] Горбатенко С. А. и др. Механика полета. Общие сведения. Уравнения движения. – М.: Машиностроение, 1969г.

Алгоритмы группового управления в условиях неопределенности

Ищенко Юлия Антоновна

Кафедра нелинейных динамических систем и процессов управления

e-mail: ishenko_y13@mail.ru

Научный руководитель — д.ф.-м.н. проф. Ильин Александр Владимирович

Постановка задачи. Задачей данной выпускной квалификационной работы является разработка простого алгоритма управления группой беспилотных летательных аппаратов (БПЛА) в пространстве с препятствиями, имеющими произвольную форму. В рамках работы была разобрана математическая модель БПЛА, проведена линеаризация данной модели и построен регулятор для управления движением летательного аппарата. Требуемый алгоритм был получен и промоделирован, с использованием пакета прикладных программ Matlab и среды динамического моделирования Simulink.

Математическая модель. Беспилотный летательный аппарат можно описать системой дифференциальных уравнений, представленной в статьях [1, 2]:

$$\begin{cases} \dot{q} = R(q)p \\ \dot{p} = M^{-1}(F_u + F_d) \end{cases} \quad (1)$$

где:

q — вектор, определяющий положение робота в земной системе координат;

p — вектор скоростей робота в связанных с мобильным роботом координатах;

$R(q)$ — матрица, описывающая кинематику робота;

M — матрица инерционных параметров;

F_u — вектор управления робота;

F_d — матрица прочих сил и моментов, действующих на мобильного робота.

Линеаризация и построение регулятора. Модель БПЛА (1) нелинейная, поэтому, на основе алгоритма, описанного в статье [3], она была приведена к линейному виду:

$$\begin{cases} \dot{X} = AX + BU \\ y = CX \end{cases} \quad (2)$$

где $X = [q \ p]^T$, $U = F_u$.

С использованием метода, описанного в статье [3], для полученной системы был построен линейно-квадратичный регулятор.

Описание алгоритма. Алгоритм был разделен на 4 основные части: представление препятствий в пространстве, огибание препятствий, организация строя и разделение пространства на секторы.

В пространстве препятствия представляются критической областью, имеющей форму сферы.

Огибание препятствий выполняется по кратчайшей траектории, лежащей на поверхности критической области данного препятствия.

Организация строя происходит следующим образом: множество агентов делится на подгруппы по 3 летательных аппарата, внутри каждой подгруппы агенты выстраиваются по окружности вокруг общего центра масс, данные шаги повторяются, но, вместо множества агентов группы, рассматривается множество подгрупп, полученных на предыдущем шаге.

Разделение пространства на секторы выполняется так, чтобы в одном секторе находились препятствия, расположенные на небольшом расстоянии друг от друга. Пролетая область, свободную от препятствий, группа поддерживает строй. При достижении сектора, содержащего препятствия, группа перестает поддерживать строй, и агенты поочередно пролетают данный сектор.

Выводы. В рамках данной выпускной работы был разработан простой алгоритм управления группой движения БПЛА. Полученный алгоритм гарантирует отсутствие коллизий и связность группы на промежутках пути, свободных от препятствий. Визуализация алгоритма помогает подтвердить его работоспособность.

ЛИТЕРАТУРА

- [1] Кульченко А. Е., Лазарев В. С., Медведев М. Ю. Метод управления движением гексакоптера в трехмерной среде с препятствиями на базе динамических отталкивающих сил // Инженерный вестник Дона №4. 2016.
- [2] Mathematical model of robot on base of airship / V. Pshikhopov, M. Medvedev, A. Gaiduk, R. Neydorf, V. Belyaev, R. Fedorenko, V. Krukhmalev // IEEE Conference on Decision and Control. 2013.
- [3] Sabatino F. Quadrotor control: modeling, nonlinear control design, and simulation // KTH Electrical Engineering. 2015.

Анализ алгоритмов управления группой с лидером

*Работа удостоена диплома 3 степени
Надольский Михаил Александрович*

*Кафедра нелинейных динамических систем и процессов управления
e-mail: nadolskyma@my.msu.ru*

Научный руководитель — д.ф.-м.н. зав. каф. Фомичёв Василий Владимирович

В данной выпускной квалификационной работе была поставлена задача моделирования движения группы роботов во главе с лидером, управляемым оператором. Движение осуществляется строем, когда роботы двигаются друг за другом на заданном равном расстоянии. Формирование строя определенной структуры и движение к целевой точке с сохранением заданной структуры — актуальная задача робототехники [1]. В качестве агентов рассматривались роботы с двумя ведущими колёсами. Роботы в группе пронумерованы, и каждый робот является локальным лидером для следующего за ним.

Для построения движения роботов с двумя ведущими колёсами была выбрана модель, известная как машина Дубинса [2], в которой движение осуществляется при постоянной скорости движения и с ограничением на угловую скорость поворота. Модели машины Дубинса находят применение при

построении траектории движения в горизонтальной плоскости [3]. Система уравнений, описывающая движение агента, имеет вид:

$$\begin{cases} \dot{x} = v \cos(\varphi) \\ \dot{y} = v \sin(\varphi) \\ \dot{\varphi} = u, u \in [-U, U] \end{cases}$$

Здесь x, y — координаты объекта, φ — угол наклона вектора скорости, u определяет мгновенную угловую скорость поворота, U — максимальное по модулю значение угловой скорости, величина линейной скорости v предполагается постоянной. В данной работе был реализован классический алгоритм, рассчитывающий оптимальную по быстродействию траекторию движения из заданного начального в заданное конечное положение.

Для решения задачи группового управления был предложен метод, когда процесс моделирования разбит на такты по времени, в которые каждый агент определяет свое целевое положение на основании положения своего локального лидера, рассчитывает оптимальную траекторию и начинает движение по ней. Таким образом, алгоритмы расчета для машины Дубинса рассматривались на определенных отрезках времени. На разных отрезках времени скорость может отличаться. Скорость постоянна в течение такта по времени.

Группа с лидером, управляемым оператором, движется друг за другом с расстоянием между машинами варьирующимся в диапазоне $[d_{min}, d_{max}]$. Скорость агента, начиная со второго, меняется по формуле:

$$v = \max(0, \min[k(d - d_{min}), v_{max}])$$

Здесь d — расстояние до локального лидера, d_{min} — минимальное разрешенное расстояние до локального лидера, k — коэффициент усиления, v_{max} — предельная скорость машины. Скорость лидера, управляемого оператором, принимает либо максимальное значение, либо ноль. Если в процессе движения агент отстал от своего лидера, то от догоняющей машины приходит сигнал к данному лидеру об отставании с требованием остановиться. Как только расстояние восстановится до разрешенного, движение возобновится.

Для визуализации описанных алгоритмов использовалась среда симуляции Gazebo. В ходе экспериментов в данной среде моделирования были выявлены ограничения на некоторые параметры, а также проблемы, связанные с точным движением по вычисленной траектории движения. Проведенные исследования показали, что разработанное управление группой роботов с использованием машины Дубинса обеспечивает стабильное удержание строя в группе из 3-4 агентов.

ЛИТЕРАТУРА

- [1] Морозова Н. С. Виртуальные формации и виртуальные лидеры в задаче о движении строем группы роботов // Вестник Санкт-Петербургского университета. Серия 10. Прикладная математика. Информатика. Процессы управления. 2015. № 1. С. 135–149

- [2] Dubins L. E. On Curves of Minimal Length with a Constraint on Average Curvature, and with Prescribed Initial and Terminal Positions and Tangents // American Journal of Mathematics. 1957. Vol. 79, N 3. P. 497–516
- [3] Хабаров С. П., Шилкина М. Л. Построение на базе задачи машины Дубинса опорных траекторий движения объектов с учетом постоянных внешних воздействий // Научнотехнический вестник информационных технологий, механики и оптики. 2022. Т. 22, N 1. С. 167–178

Метод потенциальных полей в групповой робототехнике для задачи построения маршрута на местности

Работа удостоена диплома 1 степени

Ямолдин Владимир Алексеевич

Кафедра нелинейных динамических систем и процессов управления

e-mail: yamoldinva@my.msu.ru

Научный руководитель — д.ф.-м.н. зав. каф. Фомичёв Василий Владимирович

В данной работе рассматривается задача построения маршрута группой мобильных неголономных роботов. Группе необходимо переместиться из заданной стартовой области плоского полигона в заданную целевую область этого полигона, поддерживая при этом свою связность. Роботы работают в неопределённой среде с препятствиями и имеют локальную информацию об окружающем мире, получаемую с датчиков в режиме реального времени. На основе этой информации роботы планируют свои движения. В группе имеется лидер, который обладает информацией о цели движения и притягивается к этой цели. Все остальные агенты при этом притягиваются к лидеру. Задача считается выполненной, если в какой-то момент вся группа локализуется в некоторой окрестности заданной целевой точки.

Динамика всех роботов основана на методе потенциальных полей, общее описание которого приведено в [1]. Идея метода состоит в движении роботов вдоль силовых линий векторного поля, потенциальная функция которого отражает форму и конфигурацию препятствий, цель движения и расположение роботов-соседей. Препятствия генерируют отталкивающие потенциалы, а целевая точка и соседние роботы — притягивающие потенциалы. Силы, действующие на роботов, определяются антиградиентами соответствующих потенциалов. Рассматривается группа из двухколёсных роботов, при этом управление каждым роботом осуществляется путём назначения скоростей вращения двум его колёсам. Эти скорости выбираются исходя из результирующей силы, действующей на робота. Применение метода потенциальных полей в групповой робототехнике можно найти в [2].

В [3] предлагается использовать притягивающий потенциал U_1 к целевой точке следующего вида: в точке q положить $U_1(q) = \xi \rho(q, q_0)$, где q_0 — целевая точка, $\rho(q, q_0) = \|q - q_0\|$, $\xi > 0$ — коэффициент усиления. При этом притягивающая сила $F_1 = -\nabla U_1$ постоянна по модулю всюду на полигоне:

$|F_1| = \xi$. Эту силу вычисляет только лидер группы. Отталкивающий потенциал U_2 задаётся формулой

$$U_2(q) = \begin{cases} \eta \left(\frac{1}{\rho(q, q_{obs})} - \frac{1}{\rho_0} \right)^2, & \rho(q, q_{obs}) \leq \rho_0 \\ 0, & \rho(q, q_{obs}) > \rho_0 \end{cases}$$

Здесь q_{obs} обозначает точку препятствия, ближайшую к точке q , а ρ_0 представляет собой безопасное расстояние, за пределами которого отсутствует влияние препятствия; $\eta > 0$ — коэффициент усиления. Этому потенциалу соответствует отталкивающая сила $F_2 = -\nabla U_2$:

$$F_2(q) = \begin{cases} -2\eta \left(\frac{1}{\rho(q, q_{obs})} - \frac{1}{\rho_0} \right) \frac{\nabla \rho(q, q_{obs})}{\rho^2(q, q_{obs})}, & \rho(q, q_{obs}) \leq \rho_0 \\ 0, & \rho(q, q_{obs}) > \rho_0 \end{cases}$$

Эта сила предохраняет роботов от столкновений с препятствиями и другими роботами и вычисляется для всех роботов группы.

Роботы поддерживают связность группы благодаря силам социального взаимодействия, применение которых в групповом управлении можно встретить в [4]. Эти силы являются притягивающими и в данной работе определяются для всех членов группы, кроме лидера. Все роботы могут обнаруживать в некоторой своей окрестности других роботов и измерять расстояния до них. Это позволяет ввести притягивающий потенциал U_3 как $U_3(q) = \frac{1}{2}\rho^2(q, q_{rob})$, где точка q_{rob} соответствует центру опознанного робота. При этом порождается сила притяжения $F_3(q) = q_{rob} - q$, направленная к этому роботу. Общая сила социального взаимодействия направлена в сторону локального центра масс, создаваемого обнаруженными агентами.

В данной работе поставленная задача решена путём моделирования сценариев группового поведения роботов в физическом симуляторе Autonomous Robots go Swarming (ARGoS). Было проведено 3300 независимых запусков разработанного алгоритма управления. Эксперименты проводились для групп, состоящих из 2–12 роботов. Результаты экспериментов показали, что эффективность алгоритма понижается с ростом размера группы. Также было установлено, что при фиксированном размере группы эффективность алгоритма понижается при увеличении размеров рабочего полигона, при увеличении насыщенности этого полигона препятствиями, а также при увеличении максимальных скоростей колёс самих роботов.

ЛИТЕРАТУРА

- [1] Khatib O. Real-time obstacle avoidance for manipulators and mobile robots // International Journal of Robotics Research. 1986. Vol. 5, N 1. P. 90–98.
- [2] Path planning of a group of robots with potential field approach: decentralized architecture / F. Matoui, B. Boussaid, B. Metoui, G. B. Frej, M. N. Abdelkrim // IFAC PapersOnLine 50–1. 2017. P. 11473–11478.

- [3] Ge S. S., Cui Y. J. New potential functions for mobile robot path planning // IEEE Transactions on Robotics and Automation. 2000. Vol. 16, N 5. P. 615–620.
- [4] Reif J. H., Wang H. Social potential fields: A distributed behavioral control for autonomous robots // Robotics and Autonomous Systems. 1999. N 27. P. 171–194.

Повышение эффективности алгоритма KRRT* за счет использования внешних оценок множеств достижимости

Работа удостоена диплома 2 степени

Заварзин Никита Юрьевич

Кафедра системного анализа

e-mail: s02200369@gse.cs.msu.ru

Научный руководитель — к.ф.-м.н. доц. Точилин Павел Александрович

Планирование траектории — фундаментальная задача робототехники, методы решения которой обеспечивают работоспособность роботизированных систем во множестве отраслей.

В данной работе рассматривается задача планирования траектории для линейной стационарной системы обыкновенных дифференциальных уравнений, с линейно-квадратичным критерием качества. Требуется с наименьшими затратами перевести систему из начального состояния в целевое множество, избегая при этом столкновений с препятствиями, представляющими собой неподвижные выпуклые многоугольники. Для этого предлагается модификация известного алгоритма KRRT* [1, 2], основанного на постепенном построении случайного графа в пространстве состояний системы.

Основная содержательная часть работы посвящена тому, как можно увеличить эффективность рассматриваемого метода за счет избегания обработки заведомо лишних конфигураций, случайно генерируемых при построении графа. Предлагается строить внешние эллипсоидальные оценки множества достижимости линейной системы. Такие оценки позволяют выделить ту часть фазового пространства, которая на каждом шаге работы алгоритма должна быть использована для формирования очередных случайных вершин графа.

В представленной статье поставлены и проанализированы следующие подзадачи:

1. Построение внешней эллипсоидальной оценки множества достижимости из произвольного состояния.

Для указанной подзадачи получена система дифференциальных уравнений, задающая аппроксимирующий эллипсоид с касанием множества достижимости по определенному заранее направлению.

2. Генерирование равномерного распределения в эллипсоиде.

Данная подзадача сведена к равномерному сэмплированию в единичной гиперсфере.

3. Построение траектории, связывающей два близких состояния (вершины случайного графа).

Через уравнение Беллмана выведено матричное уравнение Риккати, решением которого является искомая траектория.

4. Поиск возможных пересечений построенной части траектории и имеющих препятствий.

Эта подзадача эффективно разрешена при помощи полярной сортировки, бинарного поиска и ориентированной площади. Проверяемые отсчеты берутся по последовательности Ван Дер Корпута.

Программно были реализованы алгоритм KRRT* и его оптимизированная версия. Эмпирическое сравнение упомянутых алгоритмов показало следующие результаты:

1. При правильно подобранных параметрах предлагаемая модификация алгоритма исследует пространство также широко, как и базовая реализация KRRT*, но при этом делает это с большей степенью дискретизации.
2. По времени предложенный алгоритм работает дольше, чем KRRT*, однако, если сопоставить увеличение времени работы росту числа вершин в графе, то получится повышение эффективности алгоритма в среднем в 1.1 раз (при инициализации параметров случайным образом), а в отдельных случаях, для которых удалось подобрать параметры — в два и более раз.

ЛИТЕРАТУРА

- [1] Karaman S., Frazoli E. Optimal kinodynamic motion planning using incremental sampling-based methods // 49th IEEE Conference on Decision and Control. 2010. P. 7681–7687.
- [2] Webb D., Berg J. Kinodynamic RRT*: optimal motion planning for systems with linear differential constraints // IEEE Conference on Robotics and Automation. 2012.
- [3] Karaman S., Frazzoli E. Sampling-based algorithms for optimal motion planning // The International Journal of Robotics Research. 2011. Vol. 30. P. 846–894.

Идентификация параметров синхронного двигателя с постоянными магнитами методом наименьших квадратов

Закальская Ксения Вячеславовна

Кафедра нелинейных динамических систем и процессов управления

e-mail: zakalskaja.xen@yandex.ru

Научный руководитель — к.ф.-м.н. доц. Гончаров Олег Игоревич

Постановка задачи. Целью работы является разработка метода идентификации параметров синхронного двигателя с постоянными магнитами

(СДПМ) на основе метода наименьших квадратов (МНК). Неизвестными параметрами полагаются сопротивление R , индуктивность обмоток L для неявнополюсного двигателя или пара (L_d, L_q) для явнополюсного, потокосцепление ротора λ_{m1} и смещение электрического угла θ_0 . В литературе описано множество подходов к идентификации [1], особенностью разработанного метода является одновременное нахождение всех указанных параметров за один эксперимент, а также возможность идентификации смещения электрического угла.

Математическая модель. В ходе решения задачи идентификации использовались упрощенные модели неявнополюсного и явнополюсного СДПМ в неподвижной двухфазной системе координат, полученные из расширенной модели СДПМ и описанные в [2, разд. 2].

Упрощенная модель неявнополюсного двигателя может быть описана следующими уравнениями:

$$U_{\alpha\beta} = \dot{\lambda}_{\alpha\beta} + Ri_{\alpha\beta}, \quad (1)$$

$$\lambda_{\alpha\beta} = Li_{\alpha\beta} + \lambda_{m1} \begin{pmatrix} \cos(\phi + \theta_0) \\ \sin(\phi + \theta_0) \end{pmatrix}, \quad (2)$$

Модель явнополюсного двигателя включает уравнение (1) в неизменном виде, уравнение же (2) заменяется на следующее:

$$\lambda_{\alpha\beta} = L_q i_{\alpha\beta} + (\lambda_{m1} + (L_d - L_q)(\cos(\phi + \theta_0)i_\alpha - \sin(\phi + \theta_0)i_\beta) \begin{pmatrix} \cos(\phi + \theta_0) \\ \sin(\phi + \theta_0) \end{pmatrix},$$

где: $U_{\alpha\beta}$ — входные напряжения относительно осей α и β , $i_{\alpha\beta}$ — силы тока относительно осей α и β , $\lambda_{\alpha\beta}$ — суммарное потокосцепление, L_d, L_q — индуктивности относительно осей d и q в подвижной системе координат.

Методы идентификации. При решении задачи уравнения для обоих типов двигателей были приведены к виду регрессионной модели:

$$y[k] = \varphi[k]^T \Theta, \quad y[k] = \varphi[k, \Theta]^T \Theta,$$

где: $\varphi[k], \varphi[k, \Theta]$ — векторы регрессоров, Θ — оценка вектора неизвестных параметров.

В неявнополюсном случае регрессионная модель представляется в следующем виде:

$$\underbrace{\begin{pmatrix} U_\alpha[k] \\ U_\beta[k] \end{pmatrix}}_{y[k]} = \underbrace{\begin{pmatrix} I_\alpha[k] & \dot{I}_\alpha[k] & -\omega[k] \cos(\phi[k]) & -\omega[k] \sin(\phi[k]) \\ I_\beta[k] & \dot{I}_\beta[k] & -\omega[k] \sin(\phi[k]) & \omega[k] \cos(\phi[k]) \end{pmatrix}}_{\varphi^T[k]} \underbrace{\begin{pmatrix} R \\ L \\ \lambda_{m1} \sin(\theta_0) \\ \lambda_{m1} \cos(\theta_0) \end{pmatrix}}_{\Theta},$$

для явнополюсного СДПМ модель строится аналогично.

Для идентификации неявнополюсного СДПМ был выбран метод наименьших квадратов [3, с. 240], в случае же явнополюсного двигателя использовался рекуррентный алгоритм псевдолинейной регрессии [3, с. 274]. Также был предложен упрощенный метод идентификации, основанный на МНК.

Важным этапом в решении задачи идентификации являлся выбор способа оценки производной. В работе применялись два метода: оценка с помощью разностного отношения и получение сглаженных производных с помощью применения линейного фильтра.

Результаты. Описанные в работе алгоритмы тестировались на ряде моделей в программной среде Simulink, учитывающих наличие силы трения, шума измерения позиции и силы тока. Результаты экспериментов показали, что при моделировании без шума измерения идентификация неизвестных параметров происходит почти идеально: относительные ошибки меньше 0.1%. Наибольшие ошибки в обоих случаях наблюдались в экспериментах на моделях, учитывающих неидеальности измерительной системы силы тока, в которых для оценки производных сигналов применялось разностное отношение. В данном случае относительные ошибки по индуктивностям в среднем составляли 46% для неявнополюсного и 23–27% и 54–66% для явнополюсного СДПМ. С данной проблемой удалось справиться с помощью использования сглаживающего предварительного фильтра, в результате применения которого ошибки по индуктивности получилось снизить до 5% и 2–4% и 5–7% для неявнополюсного и явнополюсного случая соответственно.

Оценивая результаты экспериментов, можно заключить, что представленные в работе методы показали приемлемое качество идентификации моделей в описанных условиях.

ЛИТЕРАТУРА

- [1] A Review of State-of-the-art Techniques for PMSM Parameter Identification / Н. Ahn, Н. Park, С. Kim, Н. Lee // Journal of Electrical Engineering & Technology. 2020. Vol. 15, N 3. P. 1177–1187.
- [2] Perera P.C. Sensorless Control of Permanent-Magnet Synchronous Motor Drives : Ph.D. Thesis, IET Aalborg University. 2002.
- [3] Льюнг Л. Идентификация систем. Теория для пользователя : Пер. с англ. / Под ред. Я.З.Цыпкина. Наука, 1991. 432 с.

Разработка метода оценки устойчивости к состоятельным атакам нейросетевых алгоритмов предсказания качества изображений и видео

Работа удостоена диплома 3 степени

Абуд Халед Набиль

Кафедра интеллектуальных информационных технологий

e-mail: abud.khaled.n@gmail.com

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

Нейросетевые модели в последние годы показывают значительный прогресс во многих областях компьютерного зрения, включая задачу оценки субъективного визуального качества изображений и видео. Однако многие

работы [1,2] показывают существенную уязвимость таких подходов к злонамеренной манипуляции входными данными — состязательным атакам. Устойчивость методов оценки качества изображений к ним остается недостаточно изученной.

В рамках данной работы было проведено исследование устойчивости нейросетевых алгоритмов, решающих данную задачу. Для этого предложена методология сравнения устойчивости безэталонных моделей для оценки визуального качества, а также новый метод состязательной атаки, ориентированный на данную задачу и минимизирующий заметность атакующих искажений.

Разработанная методология включает в себя несколько ключевых шагов: создание набора атакованных данных с помощью алгоритма состязательной атаки, вычисление визуальных потерь полученных данных относительно исходных, вычисление значений атакуемых моделей до и после атаки, применение к значениям моделей преобразования домена, позволяющего сравнивать их между собой, и вычисление на основе этих данных нескольких итоговых оценок устойчивости моделей. Важным элементом методологии является разработанный метод состязательной атаки, позволяющий генерировать для исследуемых моделей атакованные данные, почти не отличимые от исходных. Для снижения визуальной заметности атаки были разработаны два подхода: шаг коррекции градиента и высокочастотное маскирование атаки. Первый подход заключается в проекции градиента атакуемой модели на подпространство, ортогональное градиентам нескольких различных функций расстояния между исходным изображением и атакованным. Вторым подходом предотвращается появление высокочастотных шумов (свойственных состязательным атакам) в областях, в которых этот шум будет наиболее заметен человеческому глазу.

С помощью разработанной методологии и метода генерации атакованных данных было проведено сравнение устойчивости шести популярных нейросетевых моделей. В результате сравнения была выявлена их уязвимость к разработанной атаке, а также проанализированы причины различий в устойчивости разных моделей. Для более подробного анализа воздействия состязательной атаки на поведение исследуемых моделей, из них были извлечены два набора активаций с внутренних слоев сети: один был получен при прохождении через сеть «чистых» данных, а другой — атакованных. Путем анализа различий между этими наборами были выявлены корреляции между некоторыми свойствами внутренних активаций сети и эффективностью атаки.

Помимо этого, была оценена переносимость состязательных примеров между различными моделями. Также был проведен анализ применимости к задаче оценки качества некоторых методов «очистки» изображений от состязательных добавок [3] в адаптивном (в процессе атаки известно о применяемом методе защиты) и неадаптивном сценариях.

Предложенная методология сравнения устойчивости моделей оценки качества изображений и видео с небольшими изменениями вошла в статью

[4], опубликованную в сборнике трудов конференции AAAI Conference on Artificial Intelligence 2024 уровня A*.

ЛИТЕРАТУРА

- [1] Goodfellow I., Shlens J., Szegedy C. Explaining and harnessing adversarial examples // Proceedings of 3rd International Conference on Learning Representations. 2015. С. 1–11.
- [2] Perceptual Attacks of No-Reference Image Quality Models with Human-in-the-Loop / Z. Weixia, L. Dingquan, M. Xiongkuo, Z. Guangtao, G. Guodong, Y. Xiaokang, M. Kede // Advances in Neural Information Processing Systems. 2022. С. 1–14.
- [3] Diffusion Models for Adversarial Purification / W. Nie, B. Guo, Y. Huang, C. Xiao, A. Vahdat, A. Anandkumar // Proceedings of the 39th International Conference on Machine Learning. 2022., С. 1–8.
- [4] Comparing the Robustness of Modern No-Reference Image- and Video-Quality Metrics to Adversarial Attacks / A. Antsiferova, K. Abud, A. Gushchin, E. Shumitskaya, S. Lavrushkin, D. Vatolin // Proceedings of the AAAI Conference on Artificial Intelligence. 2024. Т. 38, № 1. С. 700–708.

Разработка мультязычной модели для извлечения данных из новостных веб-страниц

Бедрин Павел Александрович

Кафедра системного программирования

e-mail: pbedrin@ispras.ru

Научный руководитель — Яцков Александр Константинович

Научный консультант — Варламов Максим Игоревич

С каждым днём растёт объём информации, генерируемой в сети Интернет. В частности, на множестве сайтов разных областей публикуются новостные статьи. Для решения различных прикладных задач, таких как анализ новостной повестки и агрегация новостей, возникла потребность в их массовом сборе и анализе. Однако веб-страницы ориентированы на пользователя браузера и не предоставляют структурированных интерфейсов доступа к информации. В связи с этим возникает потребность автоматически извлекать атрибуты (заголовок, дату, текст и др.) из произвольной новостной страницы.

Используемые для этого в актуальных исследованиях нейросетевые модели-трансформеры обучены на англоязычных данных и для страниц на других языках требуют ресурсозатратного перевода. Работа посвящена разработке мультязычной модели, способной извлекать заданный набор атрибутов из HTML-страницы с новостью на любом языке.

Для обучения и оценки моделей был создан мультязычный набор данных новостных страниц с размеченными атрибутами. В его основу легли карты сайтов СМИ разных стран, созданные в рамках платформы ИСП РАН Talisman. Карты задают сопоставление между атрибутами и CSS-селекторами

HTML-узлов на сайте. После валидации карт, сбора и предобработки данных было получено 5 языков (en, de, ar, ko, zh), в каждом — 9–10 сайтов по 50 страниц. Кроме этого, мы располагали русскоязычным набором данных — 722 страницы со 112 сайтов. О его создании была опубликована статья [1].

Была рассмотрена модель MarkupLM [2]. Среди актуальных решений она предобучалась на наиболее объёмном наборе данных — более 24 млн англоязычных страниц набора данных CommonCrawl. Авторами была опубликована предобученная модель для дообучения под конечные задачи, в нашем случае — классификацию HTML-узлов. Несмотря на англоязычность модели, мы предположили, что благодаря признакам HTML-разметки и возможности ВРЕ-токенизатора кодировать любые символы, MarkupLM мог бы поддерживать страницы произвольных языков. Однако эксперименты с дообучением и оценкой внутри отдельных языков показали, что предварительный перевод улучшает качество, а одной лишь информации разметки недостаточно для его стабильно высокого уровня. Необходимо предобучение на мультязычных данных.

Для мультязычного сценария была выбрана модель DOM-LM [3]. На классическом в задаче извлечения наборе данных SWDE она показывает аналогичные MarkupLM результаты, при этом требует в сотни раз меньший объём предобучающих данных. Мы инициализировали DOM-LM из мультязычной языковой модели XLM-RoBERTa. Были собраны и предобработаны данные предобучения, в их составе — наш мультязычный набор и более 37 000 новостных страниц разных языков из открытого набора CommonCrawl-News. Далее модель дообучалась аналогично MarkupLM.

Методика оценки качества методов — сравнение извлечённого и эталонного текста. Многословные атрибуты (заголовки, текст) сравниваются как мешки n -грамм. Многозначные атрибуты (авторы, даты, теги) нормализуются и сравниваются как множества значений. У даты, заголовка и автора выбирается узел с наибольшей вероятностью.

При оценке внутри отдельных языков мультязычный DOM-LM показал результаты, близкие или превосходящие MarkupLM, при этом ему не требуется предварительный перевод страниц на английский.

В следующей оценке был задействован весь мультязычный набор данных, то есть в обучающей и тестовой выборке содержалась смесь страниц разных языков. Это практический вариант использования модели, когда она обучается на всех имеющихся данных. В этом эксперименте (таблица 2) DOM-LM превзошёл по большинству атрибутов MarkupLM и эвристические инструменты — Trafilatura, Newspaper, Newsplease.

ЛИТЕРАТУРА

- [1] A Dataset for Information Extraction from News Web Pages / M. Varlamov, D. Galanin, P. Bedrin, S. Duda, V. Lazarev, A. Yatskov // 2022 Ivannikov Ispras Open Conference (ISPRAS). 2022. P. 100–106.

Метод / Атрибут	Заголовок	Дата	Текст	Авторы	Теги
Trafilatura	0.88	0.36	0.85	0.29	0.18
Newspaper	0.89	0.21	0.72	0.17	0.55
Newsplease	0.91	0.35	0.70	0.17	0.00
MarkupLM	0.90	0.79	0.90	0.30	0.47
MarkupLM-EN	0.95	0.88	0.91	0.40	0.67
DOM-LM	0.93	0.87	0.93	0.41	0.73

Таблица 2: Оценка качества на смеси языков в данных (F1-мера).

- [2] MarkupLM: Pre-training of Text and Markup Language for Visually Rich Document Understanding / J. Li, Y. Xu, L. Cui, F. Wei // ACL (1). 2022. P. 6078–6087.
- [3] DOM-LM: Learning Generalizable Representations for HTML Documents / X. Deng, P. Shiralkar, C. Lockard, B. Huang, H. Sun // CoRR. 2022. Vol. abs/2201.10608.

Разработка метода получения эталонных карт внимания видеопоследовательностей с помощью краудсорсинга

Брынцев Алексей Юрьевич

Кафедра интеллектуальных информационных технологий

e-mail: alexey.bryncev@graphics.cs.msu.ru

Научный руководитель — к.ф.-м.н. инж. Куликов Дмитрий Леонидович

Введение. Карта внимания представляет собой тепловую карту, более яркие участки которой отвечают за области повышенного внимания. Информация о наиболее интересных областях видео может использоваться при сжатии видео, при оценке качества видео, в исследованиях особенностей визуального восприятия и в рекламной отрасли.

Цель работы. Главная цель данной работы заключается в разработке метода постобработки данных, собранных для задачи моделирования карт внимания видеопоследовательностей при помощи технологий краудсорсинга.

Обзор литературы и существующих решений. Существуют методы получения эталонных данных при помощи таких устройств как, профессиональный ай-трекер, компьютерная мышь и веб камера:

- **Ай-трекер.** Устройство позволяет точно вычислять направление взгляда человека. Метод сбора, использующий ай-трекер, является наиболее качественным, но дороговизна этого устройства серьезно ограничивает объем доступный для сбора.
- **Веб камера.** Алгоритм *EFE* [1] принимает на вход лицо наблюдателя, а возвращает точку просмотра на экране. Использование такого алгоритма может улучшить базовый метод, однако, это требует использования дополнительных наборов данных.

— **Компьютерная мышь.** Компьютерная мышь использовалась в соревновании *SALICON* [2] и методе из работы [3]. Метод представляет собой сбор информации о местоположении курсора. Участнику эксперимента видна только небольшая область видео вокруг курсора, что мотивирует перемещать курсор в поисках областей интереса. Такие методы предоставляют возможность использовать стандартную периферию компьютера и открывают возможности к созданию более объемных наборов данных сравнительного качества. В моей работе описывается разработка и применение алгоритма улучшения последнего метода.

Гипотезы. Среди рассматриваемых гипотез есть предположение о том, что вначале просмотра видео человеку необходимо некоторое время для того, чтобы найти область внимания и начать ее отслеживать. Еще предполагается, что существует некоторая временная задержка в получаемых фиксациях. Более того предполагается, что эту задержку можно устранить при помощи инвертированных видео. Еще есть предположение о том, что в базовом методе использовалось субоптимальное количество наблюдателей на видео. Так же делается предположение о том, что используемых наблюдателей можно отсеять несколькими способами для исключения выбросов.

Результаты. В табл. [1] приведены результаты всех упомянутых экспериментов. Использование каждой модификации дало прирост качества, а использование всех модификаций улучшило базовую модель на 0.16 по *CC*, на 0.11 по *SIM*, на 0.09 по *AUC Judd*, на 0.85 по *NSS* и на 1.04 по *KLDiv*.

Модель	<i>CC</i> ↑	<i>SIM</i> ↑	<i>AUC Judd</i> ↑	<i>NSS</i> ↑	<i>KLDiv</i> ↓
Базовый метод (24Н)	0,70	0,63	0,833	1,91	1,51
24Н+БМ	0,77	0,67	0,895	2,44	1,07
74Н+БМ	0,79	0,69	0,901	2,48	0,80
74Н+БМ+ЗВ	0,81	0,70	0,902	2,55	0,76
74Н+БМ+ЗВ+ОО	0,80	0,69	0,906	2,48	0,58
74Н+БМ+ЗВ+ВС	0,84	0,73	0,911	2,72	0,61
74Н+БМ+ЗВ+ОО+ВС	0,86	0,74	0,915	2,76	0,47

Табл. 1: Результаты проведенных экспериментов

Сравнение с автоматическими методами. В ходе работы было проведено масштабное исследование качества различных автоматических методов моделирования карт внимания видеопоследовательностей. В сравнении принимали участие 27 моделей из 14 работ, а также результаты из предыдущего [3] и текущего сбора и модель *Center Prior*. По итогам сравнения новый метод сбора эталонных карт внимания оказался лучше всех рассмотренных автоматических методов моделирования карт внимания.

Новый набор данных. Также был создан новый набор данных по рассматриваемой задаче с использованием предложенной методологии. Для но-

вого набора карт внимания использовалось 1500 20 секундных 1080P видео из мультимодальных наборов *YouTube UGC* и *Vimeo*. В ходе эксперимента проверялась корректная работа звука у участников, так как использовались видео со звуковой дорожкой. Данный набор является одним из самых больших наборов карт внимания и обладает хорошим качеством данных.

ЛИТЕРАТУРА

- [1] Haldun B., Seonwook P., Xi W., Xucong Z., Otmar H. Efe: End-to-end frame-to-gaze estimation // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023. С. 2687–2696.
- [2] Ming J., Shengsheng H., Juanyong D., Qi Z. Salicon: Saliency in context // Proceedings of the IEEE conference on computer vision and pattern recognition. 2015. С. 1072–1080.
- [3] Lyudvichenko V. A., Vatolin D. S. Predicting Video Saliency Using Crowdsourced Mouse-Tracking Data // CEUR Workshop Proceedings. 2019.

Применение методов глубокого обучения в задачах сегментации текстовых изображений

Работа удостоена диплома 2 степени

Бурикова Анна Георгиевна

Кафедра Суперкомпьютеров и Квантовой Информатики

e-mail: burikova.ag@gmail.com

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

Сегментация текстовых изображений, то есть выделение на них текстовых блоков, является одним из этапов предобработки изображений текстовых документов с целью распознавания текста на этих изображениях и дальнейшего форматирования распознанного текста в соответствии с оригинальным документом. Эта задача все еще остается актуальной, особенно для работы с историческими документами, например, для оцифровки, которая очень важна в наши дни, так как это путь к сохранению исторического и культурного наследия, а так же быстрого и удобного доступа к обработанным материалам.

Целью работы является разработка, реализация и исследование нейросетевых методов решения задачи поиска и выделения текстовых блоков на изображениях печатных документов. Текстовым блоком будем называть минимальную область на изображении, содержащую внутри себя только связный текст, и отделенную от других блоков пустым пространством. Примерами текстовых блоков являются: колонки в газетных страницах, заголовки, колонтитулы и т.п. На вход нейронной сети должно подаваться RGB-изображение страницы печатного документа, а её выходом должна быть чёрно-белая маска сегментации, где чёрный цвет означает фон, белый — искомые текстовые блоки.

Был выполнен обзор различных методов общей задачи сегментации изображений, рассмотрены как классические, так и современные нейросетевые

подходы решения данной задачи. Отдельно были рассмотрены методы сегментации текстовых изображений, в ходе чего было выявлено, что многие из этих методов никак не учитывают структуру текстовых изображений, которая важна для решения рассматриваемой задачи. Были рассмотрены существующие программные продукты для выделения текстовых блоков на изображениях документов и газет. В большинстве случаев разработчики продуктов не описывают деталей реализации, кроме того, практически все эти программы являются платными в использовании. Таким образом, создание собственного решения с открытым кодом является актуальной исследовательской и практически важной задачей.

Для применения подхода с использованием нейронных сетей к решению задачи сегментации текстовых блоков был создан набор обучающих данных на основе размеченных вручную изображений реальных газет. Также была разработана и реализована программно автоматическая генерация обучающих данных и с её помощью сгенерирован набор из 250 размеченных искусственных текстовых изображений.

По результатам проведенного обзора были выбраны, настроены и обучены на сгенерированных данных три нейросетевых модели — сети SegNet [1] и UNet [2], а также предобученная на медицинских изображениях сеть UNet. Обучение происходило с помощью функции потерь бинарной кросс-энтропии, качество решения измерялось по метрике Intersection over Union. Проведенное тестирование моделей показало, что лучшей для решения поставленной задачи является предобученная модель UNet. Для улучшения выделения границ блоков и распознавания тонких границ между блоками была предложена модифицированная взвешенная функция потерь на основе стандартной бинарной кросс-энтропии. Дообученная с новой функцией потерь модель была протестирована на изображениях реальных газет, рукописных документов, а также на изображениях с текстами на различных языках и с текстовыми блоками различной формы.

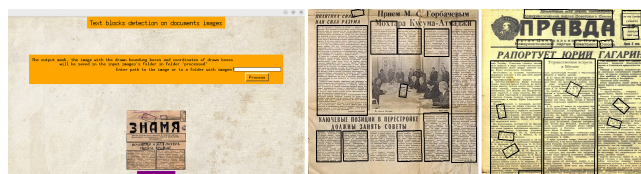


Рис. 1: Интерфейс программы и примеры ее работы

Для удобства использования разработанного решения задачи выделения текстовых блоков на текстовых изображениях, было разработано приложение с графическим интерфейсом с помощью библиотеки Tkinter на языке программирования Python. Для получения координат прямоугольных текстовых блоков на основе выходной карты сегментации использовалась библиотека OpenCV. Само приложение и необходимые для его работы данные

выложены в открытом доступе на веб-сервисе GitHub [3]. Пример результата работы на изображениях газетных страниц представлен на рис. 1.

ЛИТЕРАТУРА

- [1] Badrinarayanan V., Kendall A., Cipolla R., SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, Vol. 39, No. 12, PP. 2481-2495, DOI: 10.1109/TPAMI.2016.2644615.
- [2] Ronneberger O., Fischer P., Brox T., U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab N., Hornegger J., Wells W., Frangi A. (eds), Medical Image Computing and Computer-Assisted Intervention, MICCAI 2015. Lecture Notes in Computer Science, vol 935. DOI: 10.1007//978-3-319-24574-4_28.
- [3] TextBlocksDetection // URL: <https://github.com/AnnaBurikova/TextBlocksDetection> (дата обращения: 10.06.2024).

Автоматическое преобразование текста к формальному стилю при помощи большой языковой модели

Работа удостоена диплома 2 степени

Васильева Надежда Андреевна

Кафедра алгоритмических языков

e-mail: kvna0908@gmail.com

Научный руководитель — к.ф.-м.н. доц. Волкова Ирина Анатольевна

В современном мире автоматизация текстовой обработки становится все более востребованной. Использование больших языковых моделей показывает высокую эффективность в разнообразных NLP-задачах.

В данной работе рассматривается построение модели, способной качественно преобразовывать тексты в неформальном стиле в формальный стиль, что особенно полезно для профессиональной коммуникации и публикации контента.

Для решения поставленной задачи выбрана модель T5 (Text-to-Text Transfer Transformer), а точнее ее версия – T5-small [1]. Для обучения и оценки качества использован корпус данных GYAFC [2], включающий параллельные предложения в неформальном и формальном стилях. Обучение модели проведено с настройкой гиперпараметров. Качество модели оценено по метрикам BLEU [3] и Ассигасу [4]. Решение задачи преобразования стиля текста сведено к решению подзадач, описанных ниже.

Подготовка данных. Реализована токенизация текстов и их разбиение на тренировочные, валидационные и тестовые выборки.

Обучение модели. Используются функция потерь кросс-энтропии и оптимизатор Adam для обновления весов модели.

Настройка гиперпараметров. Подбор оптимальных значений количества эпох, скорости обучения, размера батча для улучшения сходимости и производительности модели.

Оценка качества. Расчет метрик BLEU и Accuracy (при помощи двух классификаторов) для проверки точности преобразования стиля и сохранения смысла текста.

Проведение экспериментов. Эксперименты показали, что, модель достигла конкурентоспособных результатов, демонстрируя высокие значения метрик BLEU и Accuracy. Реализованная модель показала следующие результаты: по метрике BLEU с корпусом эталонных текстов – 72.78%, по метрике Accuracy при помощи классификатора TextCNN [4] – 85.87%, по метрике Accuracy при помощи предобученного классификатора на основе BERT (Bidirectional Encoder Representations from Transformers) [5] – 88.67%.

Произведено сравнение результатов, полученных с помощью T5-small с результатами полученными с помощью систем, основанных на моделях T5-large и BART-large, показывающее, что модель T5-small может конкурировать с более крупными моделями, оставаясь при этом более экономичной.

ЛИТЕРАТУРА

- [1] Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer / C. Raffel, N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, Y. Zhou, W. Li, P. J. Liu // Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing (EMNLP 2019).
- [2] Rao S., Tetreault J. Dear Sir or Madam, May I Introduce the GYAFC Dataset: Corpus, Benchmarks and Metrics for Formality Style Transfer // Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT 2018).
- [3] BLEU: a Method for Automatic Evaluation of Machine Translation / K. Papineni, S. Roukos, T. Ward, W.-J. Zhu // Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL 2002).
- [4] Liu A., Wang A., Okazaki N. Semi-Supervised Formality Style Transfer with Consistency Training // Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing (EMNLP 2022).
- [5] Dementieva D., N. Babakov N., A. Panchenko A. Detecting Text Formality: A Study of Text Classification Approaches // Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing (EMNLP 2022).

Разработка алгоритмов автоматизации сбора задержек на основе анализа внутренней структуры серии измерений задержки в сети

Работа удостоена диплома 1 степени

Волчанинов Алексей Павлович

Кафедра автоматизации систем вычислительных комплексов

e-mail: mr.volchaninov@yandex.ru

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

Одной из наиболее важных составляющих современных распределённых вычислительных комплексов является сетевая инфраструктура. Для того, чтобы изучать особенности поведения сети вычислительного кластера и контролировать её состояние существует широкий набор программных средств. Среди подобных средств одними из наиболее часто используемых являются *MPI-бэнчмарки* [1]. Средства данного типа позволяют собирать данные о задержках при передаче данных с помощью *MPI-функций* между процессами, расположенными на различных узлах вычислительного кластера.

Выпускная Квалификационная Работа (*ВКР*) реализуется на основе проекта *clustbench* ([2],[3],[4]), который является *MPI-бэнчмарком*. Обычно при запуске теста пользователь указывает некоторый набор параметров, среди которых есть параметр, определяющий *статическое* количество измерений задержки. В связи с этим возникает дилемма: слишком малое количество измерений может привести к неинформативности собранных данных, а слишком большое может привести к nepозволительному росту времени тестирования. Для разрешения данной дилеммы в *ВКР* предложены алгоритмы, автоматически определяющие количество измерений задержки.

В *ВКР* представлены:

- результат обзора некоторых *MPI-бэнчмарков*;
- результаты исследования данных о задержках [5], собранных с помощью проекта *clustbench* с различных сетевых инфраструктур с помощью различных режимов тестирования;
- описание математических методов оценки собираемых данных, спроектированных с учётом характерных особенностей данных, выявленных в рамках исследования;
- описание алгоритмов, автоматически определяющих количество измерений задержки, необходимое для достижения заданного порогового значения предложенных оценок;
- описание программной реализации предложенных алгоритмов в рамках проекта *clustbench*;
- результаты экспериментального исследования реализованных алгоритмов на трёх различных сетевых инфраструктурах.

В результате исследования данных были выявлены некоторые повторяющиеся структуры и обнаружен следующий факт: в векторе мощностей гармоник, полученном с помощью Дискретного Преобразования Фурье (ДПФ), константная гармоника оказывается наиболее мощной. Методы оценки собираемых данных основаны на выделении некоторых участков данных (окон) и их сравнении. Предложено два концептуально разных способа сравнения окон: сравнение значений скалярных характеристик окон и сравнении векторов мощностей гармоник, получаемых с помощью некоторого спектрального представления. Измерения задержки прекращаются, если при очередной проверке достигается заданное пороговое значение используемой оценки, либо при достижении максимального допустимого количества измерений задержки. По результатам экспериментального исследования в качестве наиболее удачных для использования на практике выделены следующие алгоритмы: сравнивающий значения дисперсии в различных окнах и сравнивающий векторы мощностей гармоник, полученные с помощью применения ДПФ без учёта константной гармоники.

Благодаря реализации алгоритмов, автоматически определяющих количество измерений задержки, пользователь проекта *clustbench* получил возможность оценки собираемых данных и ускорения процесса сбора данных.

ЛИТЕРАТУРА

- [1] Hunold, Sascha and Carpen-Amarie, Alexandra. MPI Benchmarking Revisited: Experimental Design and Reproducibility// 2015.
- [2] Clustbench. Benchmarks and analysis of interconnection in HPC cluster (дата обр. 24.04.2023) – URL: https://github.com/clustbench/network-tests2/tree/new-tests_as_plugins.
- [3] Salnikov A.N., Andreev D.Yu., Lebedev R.D. Toolkit for analyzing the communication environment characteristics of a computational cluster based on MPI standard function// Moscow University Computational Mathematics and Cybernetics 2012. P. 41–49.
- [4] Сальников А.Н., Andreev D.Yu., Бегаев А.А.. Инструментальная система покомпонентного тестирования сети вычислительного кластера//Труды Международной конференции «Параллельные вычислительные технологии (ПаВТ) 2021. С. 162–175.
- [5] Волчанинов, Алексей (2023), “Clustbench launches”, Mendeley Data, V2, doi: 10.17632/t4hn5d8wmх.2.

Методы статической аутентификации пользователей по данным датчиков движения мобильного устройства

Работа удостоена диплома 2 степени

Чикин Олег Павлович

Кафедра интеллектуальных информационных технологий

e-mail: chikin.oleg@gmail.com

Научный руководитель — к.ф.-м.н. доц. Казачук Мария Андреевна

В современном мире с учетом ощутимого роста популярности мобильных устройств критически важными становятся вопросы, связанные с обеспечением безопасности хранимой в них информации. При этом одним из основных инструментов обеспечения безопасности является процедура аутентификации. Данная работа посвящена исследованию и разработке алгоритмов статической аутентификации пользователей по жесту, проделанному с мобильным устройством в руке. При этом анализируются показания датчиков движения – акселерометра и гироскопа. Особенностью задачи является высокая зашумленность анализируемых данных, представляющих собой временные ряды – проекции кажущегося ускорения и угловой скорости по трем осям. Для решения проблемы зашумленности, в работе предлагается схема многошаговой фильтрации данных (применение фильтра Баттерворта, медианного фильтра, вейвлет-преобразования, кубического сглаживающего сплайна). Для построения модели пользователя в работе предлагаются два алгоритма: алгоритм на основе комбинации собственного подхода к построению признакового пространства и нечеткого метода поиска исключений на основе потенциальных функций (Fuzzy), а также алгоритм на основе комбинации беспризнакового распознавания с использованием алгоритма DTW и метода Fuzzy. Ввиду отсутствия открытого набора данных для проведения исследований, реализован программный прототип кроссплатформенной системы аутентификации пользователей на основе жеста, проделываемого с мобильным устройством в руке, позволивший собрать необходимый набор данных и протестировать на нем разработанные алгоритмы. По результатам проведенной серии экспериментов, предложенные алгоритмы превзошли по качеству работы существующие [1, 2, 3, 4], достигнув среднего значения метрики ROC AUC, равного 0.99, и могут активно применяться на практике.

ЛИТЕРАТУРА

- [1] Shen C. et al. Waving gesture analysis for user authentication in the mobile environment //IEEE Network. – 2020. – Т. 34. – №. 2. – С. 57-63.
- [2] Yanna W. Research on Time-frequency Feature Fusion Method based on Dynamic Gesture Identity Authentication //2021 IEEE 4th Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC). – IEEE, 2021. – Т. 4. – С. 523-527.
- [3] Huang E. et al. A New Dataset for Smartphone Gesture-based Authentication //ICISSP. – 2021. – С. 771-780.

- [4] Liu J. et al. uWave: Accelerometer-based personalized gesture recognition and its applications // Pervasive and Mobile Computing. – 2009. – Т. 5. – №. 6. – С. 657-675.

Построение виртуальных анализаторов для производственных процессов

Гриненко Анна Андреевна

Кафедра интеллектуальных информационных технологий

e-mail: grinenko.aa@mail.ru

Научный руководитель — к.ф.-м.н. доц. Петровский Михаил Игоревич

Научный консультант —

Современные технологические процессы требуют автоматизации контроля и управления. В работе рассматриваются данные с производственного комплекса нефтепереработки, с которого с фиксированным периодом снимаются показания физических датчиков. Виртуальные анализаторы (Soft Sensors) - математические модели лабораторных показателей, которые используют значения физических датчиков и позволяют улучшить процесс оценки качества продукта, снизить зависимость от затратных по времени или ресурсам лабораторных исследований.

Основным объектом исследования является набор данных с реальной технологической установки, включающий ежеминутные показания полусотни датчиков за четыре года (более 2 млн записей) и 4 лабораторные переменные, которые поступают ежедневно или через день (количество варьируется от 800 до 1200 записей для каждой переменной).

В условиях разреженности данных и малого их объема поставлена задача аппроксимации лабораторных показателей на основе физических исторических изменений в виде виртуальных анализаторов.

В качестве основных моделей прогнозирования были выбраны рекуррентные нейронные сети (в частности, LSTM) и полносвязные нейронные сети.

Для решения проблемы маленького объема данных используются генеративно-сопоставительные сети (Generative Adversarial Network) [1], позволяющие аугментировать набор данных. Также исследуются модификации архитектуры rCWGAN (Conditional Wasserstein Generative Adversarial Networks with regressor) [2]. В отличие от классической генеративно-сопоставительной модели, генератору и дискриминатору подается на вход условия – истинные значения лабораторных переменных или результаты регрессора. Также отличием является совместное обучение регрессора и GAN.

Кроме того, в работе исследуется разбиение данных на стабильные периоды и аномальные. Последние могут ухудшать точность моделей и их устойчивость, а в рамках при построении виртуальных анализаторов рассматривается задача оценки качества при стабильном процессе.

В работе рассматривается модификация подхода rCGAN и предлагается метод настройки таких моделей на примере моделей, построенных на основе

полносвязных и рекуррентных нейронных сетей. Также исследуется применимость разработанных подходов в условиях малых объемов данных, оценивается зависимость качества модели от объема данных. Результаты экспериментов на реальном наборе данных показывают, что предложенный подход rCGAN обладает хорошей аппроксимирующей способностью и демонстрирует более высокую точность прогнозирования и на кросс-валидации, и при небольшом объеме выборки, чем соответствующие модели регрессоров.

Результаты, представленные в работе, представлялись автором на научной конференции "Ломоносовские чтения — 2024".

ЛИТЕРАТУРА

- [1] Jansen S. Machine Learning for Algorithmic Trading: Predictive models to extract signals from market and alternative data for systematic trading strategies with Python. – Packt Publishing Ltd, 2020, с. 649-678
- [2] Y. L. et al. A novel virtual sample generation method based on a modified conditional Wasserstein GAN to address the small sample size problem in soft sensing // Journal of Process Control. – 2022. – Т. 113. – С. 18-28.

Параллельный поиск наиболее близкого элемента в коллекции трехмерных объектов, устойчивый к повороту

Дьяченко Данил Сергеевич

Кафедра суперкомпьютеров и квантовой информатики

e-mail: danil.dyach@gmail.com

Научный руководитель — к.ф.-м.н. доц. Никольский Илья Михайлович

Научный консультант — к.ф.-м.н. доц. Никольский Илья Михайлович

В современном мире, где объемы данных растут с каждым днем, возрастает значимость эффективных методов их обработки. Одной из центральных проблем в обработке трехмерных объектов является их сравнение и поиск наиболее близких элементов в условиях, когда объекты могут быть повернуты или иначе трансформированы. В данной работе акцентируется внимание на исследовании и разработке параллельного алгоритма поиска наиболее близкого элемента в коллекции трехмерных объектов, устойчивого к повороту.

Традиционные алгоритмы, работающие в последовательном режиме, часто оказываются неэффективными из-за высокой вычислительной сложности задачи. В этом контексте, использование параллельных вычислений и, в частности, технологий, таких как CUDA [1] от NVIDIA, предоставляет значительные преимущества в скорости обработки данных, открывая новые возможности для решения поставленных задач. Помимо этого, существующие алгоритмы показывают малую точность при поиске объектов со сложными формами и угле поворота более 30° .

В рамках данной работы был разработан и исследован параллельный алгоритм поиска наиболее близкого элемента в коллекции трехмерных объектов, устойчивый к поворотам и другим трансформациям. Был реализован параллельный алгоритм на основе ICP (Iterative Closest Point) [2], который показал высокую устойчивость к поворотам и другим трансформациям. Проведено сравнительное исследование двух популярных алгоритмов выравнивания облаков точек: ICP и PCA (Principal Component Analysis) [3]. Было выявлено, что алгоритм ICP обладает рядом преимуществ при использовании параллельных вычислений на GPU. Были исследованы различные алгоритмы поиска ближайших соседей, предложены оптимальные реализации для использования на GPU. Экспериментально подтверждена высокая производительность параллельной версии алгоритма ICP при использовании технологии CUDA. Разработанный алгоритм показал высокую точность и производительность на популярной коллекции трехмерных объектов ModelNet40 [4].

ЛИТЕРАТУРА

- [1] Scalable parallel programming with CUDA / J. Nickolls, I. Buck, M. Garland, and K. Skadron in ACM Queue, vol. 6, no. 2, pp. 40-53, 2008.
- [2] A Survey of Rigid 3D Pointcloud Registration Algorithms / B. Bellekens, V. Spruyt, R. Berkvens, M. Weyn // AMBIENT 2014 : The Fourth International Conference on Ambient Computing, Applications, Services and Technologies, pp 8-13
- [3] P. J. Besl and N. D. McKay, A method for registration of 3-d shapes in IEEE transactions on pattern analysis and machine intelligence, vol. 14, no. 2, pp. 239-256, 1992.
- [4] 3D ShapeNets: A Deep Representation for Volumetric Shape Modeling / Z. Wu, S. Song, A. Khosla, F. Yu, L. Zhang, X. Tang and J. Xiao // Proceedings of 28th IEEE Conference on Computer Vision and Pattern Recognition (CVPR2015), 2015.

Влияние аугментаций на качество классификации электрокардиограмм с помощью сравнительного обучения

Дятлинко Екатерина Сергеевна

Кафедра системного программирования

e-mail: ms.dyatlinko@mail.ru

Научный руководитель — д.ф.-м.н. зав. каф. Аветисян Арутюн Ишханович

Научный консультант — Аветисян Арам Арутюнович

Электрокардиограмма (ЭКГ) – это графическое представление разности потенциалов, возникающих в результате сокращения клеток сердца и передающихся через тело человека к поверхностным покровам тела. ЭКГ сигналы используют для диагностики сосудисто-сердечных заболеваний. Нейросетевая модель, решающая задачу классификации ЭКГ сигналов на сердечные

патологии, может быть использована в качестве второго мнения при выставлении диагноза.

Разметка медицинских данных это трудоемкая задача, поэтому хорошо размеченных данных такого типа, в частности ЭКГ сигналов, мало. Чтобы увеличить количество данных для обучения и потенциально повысить качество классификации, часто применяют сравнительное обучение. Суть такого вида машинного обучения заключается в том, чтобы по паре объектов определить насколько эти объекты похожи. Исходный объект и его вид, после применения аугментаций должны как можно ближе располагаться в эмбединговом пространстве (такую пару объектов называют положительной), а измененные виды разных сигналов – как можно дальше. Так, аугментации играют большую роль в сравнительном обучении. В рамках выпускной квалификационной работы было выяснено как влияют часто используемые для временных рядов аугментации на качество мультиклассовой классификации ЭКГ сигналов при использовании сравнительного обучения.

Сравнительное обучение можно поделить на 2 этапа: предобучение на неразмеченных данных и дообучение на хорошо размеченных данных. Измерение качества было проведено на дообученных моделях, которые решают задачу мультиклассовой классификации. Из-за специфики медицинских данных были выбраны метрики G-means, F1-score и Precision.

В данной работе были выбраны и реализованы следующие распространённые для временных рядов аугментации данных: добавление шума (также называют фазовым дрожанием цифрового сигнала данных), переворот исходного сигнала относительно горизонтальной оси, зануление случайных промежутков временного сигнала, сужение и растягивание частей сигнала, перемешивание частей сигнала между собой.

Рассматривалось 3 современных метода сравнительного обучения: SimCLR [1], TS-TCC [2] и PCLR [3]. SimCLR часто берут в качестве некоего начального решения для сравнительного обучения, хотя и этот подход показывает неплохие результаты. Архитектура TS-TCC была выбрана так как на момент написания работы именно она показывала наилучшие результаты классификации временных рядов с использованием сравнительного обучения. Подход PCLR был выбран, так как он учитывал медицинскую особенность задачи: так, в этом методе в качестве положительной пары брались не аугментированные виды одного и того же сигнала, а аугментированные виды сигналов, принадлежащие одному и тому же пациенту. Стоит заметить, что так как изначально эти методы создавались для решения других задач, был написан код для реализации данных методов и встраивания их в общую последовательность действий обучения для задачи классификации ЭКГ сигналов.

Был поставлен ряд экспериментов, в которых к каждому из выбранных методов сравнительного обучения применялась одна из аугментаций для временных рядов. Исходя из полученных результатов поставленных эксперимен-

тов был сделан вывод о том, что при тестировании на двух наборах данных, наилучший прирост в качестве, смотря на метрики Gmean, F1-score и Precision с взвешенным и макро усреднением, для каждого сравнительного метода был получен при использовании фазового дрожания сигнала. Второй по качеству классификации результат достигается при перемешивании частей ЭКГ сигнала.

Сравнительные методы самообучения SimCLR, TS-TCC и PCLR демонстрируют более высокое качество классификации по сравнению с традиционным обучением на размеченных данных. Среди этих методов TS-TCC показывает наилучшие результаты, в то время как PCLR уступает другим подходам. На используемых наборах данных эти методы значительно улучшают метрики качества относительно модели, обученной с учителем. Так, на используемых данных для тестирования Gmean увеличивается на 3.6–10.7%, F1-score — на 8.3–30.4%, а Precision — на 10.3–41.3%. Таким образом, применение методов самообучения позволяет существенно повысить эффективность мультиклассовой классификации по сравнению со стандартным подходом обучения на размеченных данных.

ЛИТЕРАТУРА

- [1] Chen T. et al. A simple framework for contrastive learning of visual representations // International conference on machine learning. – PMLR, 2020. – P. 1597-1607.
- [2] Eldele E. et al. Time-series representation learning via temporal and contextual contrasting // arXiv preprint arXiv:2106.14112. – 2021.
- [3] Diamant N. et al. Patient contrastive learning: A performant, expressive, and practical approach to electrocardiogram modeling // PLoS computational biology. – 2022. – Vol. 18. – № 2. – P. e1009862.

Лексическая нормализация текстов на русском языке

Работа удостоена диплома 3 степени

Коляскина Ирина Сергеевна

Кафедра алгоритмических языков

e-mail: ira.koliaskina@gmail.com

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

В выпускной квалификационной работе рассматривается задача лексической нормализации, заключающаяся в преобразовании текста к лексической норме. Под лексической нормой обычно подразумевают отсутствие в тексте орфографических ошибок и опечаток, аббревиатур и других сокращений, сленга, просторечий, архаизмов, жаргонизмов и прочих аномальных элементов. Например, сейчас лексической нормой предложения «*Студента спалили со шпорой*» является предложение «*Студента заметили со шпаргалкой*». Лексическая нормализация необходима для уменьшения лексической и орфографической вариации текста, что облегчает его последующую обработку,

анализ и понимание. Для русского языка задача лексической нормализации исследована недостаточно, в открытом доступе отсутствуют программы, осуществляющие ее.

Целью выпускной квалификационной работы стало создание программного модуля лексической нормализации текстов на русском языке. Для этого были проанализированы современные методы нормализации и по результатам анализа было решено провести дообучение нейросетевой модели ВуТ5-small [1]. Модель имеет архитектуру Transformer [2], адаптированную для обработки байтовых последовательностей, и изначально была обучена на большом корпусе неразмеченных многоязычных данных mC4 [3]; на вход модели подается последовательность байтов UTF-8 без предварительной обработки.

Для настройки модели ВуТ5-small на решение задачи лексической нормализации текстов на русском языке был написан программный модуль на языке Python 3, реализующий ее дообучение и вычисление значения метрик accuracy, precision, recall, f1-мера и ERR:

$$ERR = \frac{accuracy_{system} - accuracy_{leave-as-is}}{1.0 - accuracy_{leave-as-is}},$$

где system – это оцениваемая система, а leave-as-is – система, которая не нормализует ни одно слово. Дообучение проводилось на сформированном наборе русскоязычных данных. Рассматривались данные социальной сети Twitter. Для автоматизации сбора и разметки набора была написана программа на языке Python 3. Процесс разметки состоял из двух последовательных этапов: первый этап осуществлялся автоматически, второй этап включал ручную корректировку полученных результатов.

На первом этапе для ускорения процесса и улучшения качества нормализации твитов было решено воспользоваться возможностями ChatGPT [4]. Для этого была написана программа, которая посылает запрос с промптом, содержащим твит и инструкцию, как нужно обработать этот твит, и получает ответ. Однако, поскольку ChatGPT не всегда правильно нормализовал текст, на втором этапе набор данных прошёл ручную корректировку. Таким образом был получен набор из 1115 предложений, содержащих более 10800 слов.

После дообучения модели было проведено экспериментальное исследование ее работы. Полученные результаты сравнивались с результатами работы Yandex Speller [5] (инструмент Yandex, позволяющий находить и исправлять орфографические ошибки) и модели ChatGPT (большая генеративная языковая модель, разработанная OpenAI). Экспериментальное исследование показало, что дообученная модель ВуТ5-small существенно превосходит Yandex Speller и ChatGPT, что говорит о перспективности ее использования для решения задачи лексической нормализации на русском языке.

ЛИТЕРАТУРА

- [1] Xue L. et al. Byt5: Towards a token-free future with pre-trained byte-to-byte models //Transactions of the Association for Computational Linguistics. – 2022. – Т. 10. – С. 291-306.
- [2] Vaswani A. et al. Attention is all you need //Advances in neural information processing systems. – 2017. – Т. 30.
- [3] Roberts A. et al. Exploring the limits of transfer learning with a unified text-to-text transformer //Google, Tech. Rep. – 2019.
- [4] Teubner T. et al. Welcome to the era of chatgpt et al. the prospects of large language models //Business & Information Systems Engineering. – 2023. – Т. 65. – №. 2. – С. 95-101.
- [5] Yandex Speller. [Электронный ресурс]. – Электрон.дан. – URL: <https://yandex.ru/dev/speller/> (дата обращения 10.05.2024).

Исследование свойств признакового пространства для повышения эффективности обучения по нескольким примерам

Работа удостоена диплома 1 степени

Кучеров Василий Дмитриевич

Кафедра Суперкомпьютеров и Квантовой Информатики

e-mail: kuchierov_2002@mail.ru

Научный руководитель — к.ф.-м.н. доц. Буряк Дмитрий Юрьевич

Направление обучения по нескольким примерам (FSL или Few Shot Learning) становится все более актуальным из-за необходимости решения задач с ограниченным набором тренировочных данных, таких как добавление новых пользовательских голосовых команд в системах умного дома или голосовых помощниках или аутентификация по биометрии. Помимо ограничений на объем доступных данных, следует учитывать их конфиденциальный характер, что побуждает решать данные задачи на конечных пользовательских устройствах, зачастую маломощных.

Классическим подходом к решению задач обучения по нескольким примерам является построение промежуточного пространства признаков[1] с высокими качествами кластеризации, в котором объекты одного класса расположены близко друг к другу, а объекты разных классов - далеко друг от друга. Для построения таких пространств обучается энкодер на похожей задаче с большим количеством тренировочных данных. В полученном признаковом пространстве решается исходная задача обучения по нескольким примерам с помощью классификатора.

В данной работе исследуются методы оценки признаковых пространств по отношению к точности последующего решения задачи обучения по нескольким примерам. В качестве методов оценки качества пространств выбраны метрики кластеризации Feature Clustering, Hyperplane Variance, Silhouette

Score и Davies-Bouldin Index. Роль энкодера выполняет сверточная нейронная сеть DS-CNN[2]. В качестве классификатора используется прототипическая сеть[3], которая в процессе обучения создает хранилище прототипов. Для этого для каждого тренировочного примера с помощью энкодера формируется признаковый вектор, а затем вычисляется среднее значение признаковых векторов для каждого класса - это и есть прототипы классов. В процессе классификации для тестового примера также формируется признаковый вектор, после чего вычисляется расстояние от него до каждого прототипа класса, и выбирается наиболее близкий прототип.

Экспериментальное сравнение проводится на примере задачи распознавания ключевых слов на наборах данных Google Speech Commands[4] и MLCommons MSWC. Энкодеры обучаются на различных подвыборках из этих наборов с использованием функций потерь семейства contrastive loss, направленных на улучшение качества кластеризации итогового пространства. После обучения энкодеров отбираются наилучшие эпохи по каждой из выбранных метрик. На этих эпохах проводятся эксперименты по задаче классификации с малым числом образцов (FSL) для 10 классов в сценариях 1-shot и 5-shot. В сценарии 1-shot для обучения используется только один пример каждого тренировочного класса, а в сценарии 5-shot используются пять примеров для каждого класса.

Наибольшее значение модуля корреляции Пирсона между значением метрики кластеризации и точностью решения задачи FSL показали метрики Silhouette Score и Davies Bouldin Index со значениями 0.655 и 0.608. При этом корреляция сохранялась и в кросс-языковых экспериментах, когда энкодер обучался на одном языке, а классификатор на другом. На основе метрики Silhouette Score была построена функция потерь Silhouette Margin Loss, которая превзошла рассматриваемые функции потерь contrastive-loss в среднем более чем на 3% как в 1-shot, так и в 5-shot сценариях.

ЛИТЕРАТУРА

- [1] Wang, Yaqing. Generalizing from a Few Examples: A Survey on Few-Shot Learning // <https://arxiv.org/abs/1904.05046v3>. — 2020.
- [2] Chollet, Francois. Xception: Deep Learning with Depthwise Separable Convolutions // <https://arxiv.org/abs/1610.02357> — 2017.
- [3] Jake Snell, Kevin Swersky, Richard S. Zemel. Prototypical Networks for Few-shot Learning // <https://arxiv.org/abs/1703.05175v2> — 2017.
- [4] Warden, Pete. Speech Commands: A Dataset for Limited-Vocabulary Speech Recognition // <https://arxiv.org/abs/1804.03209v1> — 2018.

Исследование методов кластеризации в задаче

извлечения смысла слова

Работа удостоена диплома 1 степени

Куклин Михаил Игоревич

Кафедра алгоритмических языков

e-mail: kuklinmike02n@gmail.com

Научный руководитель — к.ф.-м.н. доц. Полякова Ирина Николаевна

Научный консультант — д.ф.-м.н. проф. Полякова Ирина Николаевна

Задача извлечения смысла слова является одной из ключевых проблем в области компьютерной лингвистики и имеет широкое прикладное применение. Дано множество примеров употреблений многозначных слов. Требуется для каждого слова кластеризовать примеры употреблений по его значению. Подавляющее большинство исследований, посвященных данной проблеме, сосредоточено на построении моделей векторизации примеров. Данная работа направлена на исследование и реализацию методов кластеризации в задаче извлечения смысла слова, их анализ и сравнение между собой.

В работе были детально исследованы плотностные методы кластеризации: DBSCAN [1] и HDBSCAN [2]. Для этих методов проведен тщательный подбор гиперпараметров и проведено сравнение методов на различных языках. В дополнение было исследовано, как число примеров влияет на качество работы плотностных методов кластеризации.

Весомой частью работы является разработка методов кластеризации с использованием словаря. В данной постановке дополнительно имеется словарь, в котором описаны некоторые, но не обязательно все, значения слов. Значения в словаре могут быть описаны как определениями, так и примерами словоупотреблений с общим значением.

В работе разработаны и реализованы два метода кластеризации с использованием словаря. Первый способ, названный «Agglomerative+M», является модификацией агломеративной кластеризации [3] и использует примеры из словаря. Второй способ «WSD+WSI+OD» использует определения из словаря. Одним из его компонент является распознаватель выбросов – модель, предсказывающая присутствует ли в словаре подходящее определение для данного словоупотребления или нет. В качестве распознавателя выбросов была обучена логистическая регрессия.

В результате проделанной работы было получено первое место в соревновательной дорожке «AXOLOTL-24 Shared Task on Explainable Semantic Change Modeling» [4].

ЛИТЕРАТУРА

- [1] A density-based algorithm for discovering clusters in large spatial databases with noise / Ester Martin, Kriegel Hans-Peter, Sander Jorg, Xiaowei Xu // Proceedings of the Second International Conference on Knowledge Discovery and Data Mining. — KDD'96. — AAAI Press, 1996. — P. 226–231.

- [2] Malzer Claudia, Baum Marcus. A Hybrid Approach To Hierarchical Density-based Cluster Selection // 2020 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI). — IEEE, 2020
- [3] Daniel Mullner. Modern hierarchical, agglomerative clustering algorithms. arXiv:1109.2378v1
- [4] 5th International Workshop on Computational Approaches to Historical Language Change 2024 (LChange'24) [Электронный ресурс]. — Электрон. дан. — URL: <https://www.changeiskey.org/event/2024-acl-lchange/> (дата обращения 15.04.2024)

Методы защиты нейронных сетей от атак, основанные на интерпретируемости

Работа удостоена диплома 3 степени

Кузьмина Елизавета Васильевна

Кафедра системного программирования

e-mail: kuzminaev.289@gmail.com

Научный руководитель — Архипенко Константин Владимирович

Глубокие нейронные сети (англ. Deep Neural Networks — DNN) продемонстрировали впечатляющую производительность в широком спектре задач. Однако стоит заметить, что нейронные сети ведут себя как «чёрные ящики», по результатам работы которых пользователь не может понять ни как принималось решение, ни причины принятия того или иного решения. Именно эта непрозрачность моделей машинного обучения повлияла на доверие человека к ним и распространение их в приложения, касающиеся важных сторон общественной жизни, так как для полного доверия модели к машинного обучения пользователю крайне важно понимать, как она приходит к принятию того или иного решения.

Интерпретировать — значит объяснить или показать в понятных терминах. В контексте модели машинного обучения интерпретируемость — это способность объяснить предсказание модели и представить это объяснение в естественном для человека виде.

Все исследования по интерпретируемости моделей глубоко машинного обучения можно разделить на *апостериорные* методы и *интерпретируемые по построению* DNN. Цель методов апостериорной интерпретации — дать представление об уже обученных *произвольных* моделях. Они стремятся раскрыть либо значение изученных характеристик, либо обоснование модельных решений. Эти методы можно применять без изменения модели и её повторного обучения. Использование методов объяснения позволяет повысить доверие пользователей к моделям, так как они помогают искать в них уязвимости, а благодаря хорошим результатам объяснения можно снизить риски переобучения. Однако использование «чёрных ящиков» в совокупности с моделями их объяснений также имеет ряд проблем.

Таким образом, для повышения доверия пользователей к моделям машинного обучения использования вместе с «чёрными ящиками» апостериорных методов их объяснения недостаточно. Для создания действительно «прозрачных» моделей в них с самого начала необходимо закладывать механизмы интерпретируемости: их объяснения будут всегда соответствовать решению модели, а процесс рассуждения будет понятен пользователю.

В работе был исследован вопрос устойчивости интерпретируемых нейросетевых архитектур, представленных моделями Deformable-ProtoPNet [1], EMD-Corr [2] и SHM-Corr [2], к атакам.

Обучение данных архитектур на чистых данных дало понять, что их главное преимущество перед обычными сверточными моделями, например, ResNet, заключается в предоставлении объяснений своих предсказаний практически без потери точности. Более того, модель Deformable-ProtoPNet работает с точностью гораздо выше своей оригинальной версии ProtoPNet. Такие результаты достигнуты за счёт гибкости активированных прототипов, которые из-за этого наименее чувствительны к различным пространственным искажениям.

После проведения экспериментов на чистых данных каждая из исследуемых моделей была атакована методом с ограниченными по норме возмущениями (PGD) [3] и методом с сильным, но локальным возмущением — состязательной патч-атакой [4]. Результаты экспериментов говорят о том, что данные модели не являются более устойчивыми к самим атакам, чем обычные сверточные сети, однако за счёт встроенных механизмов интерпретируемости, они предоставляют объяснения своим предсказаниям, по которым можно сделать вывод о факте совершения атаки.

Для автоматизации обнаружения атакованных данных было решено разработать детектор. Были реализованы три разные версии детектора, в основе каждой из которых лежит свёрточная сеть со встроенным обучаемым слоем сопоставления Хафа [5]. Результаты проведенных экспериментов говорят о том, что максимальные значения $F1$ -меры, равные 73% на совершенных атаках проективного градиентного спуска и 75% на атаках с состязательными патчами, достигаются при реализации, в которой с пороговым значением сравнивается среднее расстояние от активированных участков тестового изображения до точек, которым соотнесла свёрточная модель сопоставления Хафа наиболее значимые участки активированных прототипов.

Таким образом, можно сделать вывод о том, что за счёт механизмов объяснения интерпретируемые модели действительно более устойчивы к атакам, так как факт их совершения можно обнаружить. Благодаря этому есть возможности для разработки различных методов защиты и детекторов этих атак, один из которых был также представлен в работе.

ЛИТЕРАТУРА

- [1] Donnelly J. et al. Deformable ProtoPNet: An Interpretable Image Classifier Using Deformable Prototypes // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2022). P. 10265–10275.
- [2] Giang N. et al. Visual correspondence-based explanations improve AI robustness and human-AI team accuracy // (2023).
- [3] Kurakin A. et al. Adversarial examples in the physical world // (2017).
- [4] Tom B. et al. Adversarial Patch // (2017).
- [5] Juhong M. et al. Convolutional Hough Matching Networks // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (June 2021). P. 2940-2950.

Полифония полостей в квантовой электродинамике

Работа удостоена диплома 3 степени

Кузьминский Андрей Владиславович

Кафедра суперкомпьютеров и квантовой информатики

e-mail: s02200417@gse.cs.msu.ru

Научный руководитель — д.ф.-м.н. профессор Ожигов Юрий Игоревич

Введение. Исследование квантовых эффектов в вопросах химии и биологии являются приоритетными, так как они помогают лучше понять природу микромира живых существ. Люк Монтенье предсказывал передачу биологических динамических процессов между разными системами посредством квантовой механики [1], что, в свою очередь, позволит лучше понять, как их контролировать и модифицировать.

Также, данное исследование даёт теоретические предсказания экспериментам для системы полостей из микрорезонаторов, которые используются для квантовых гейтов в квантовом компьютере (аналог транзисторов в обычном). Экспериментально и аналитически хорошо исследован случай двух полостей [6]. Больше же число не имеет аналитического решения, а эксперименты без него в области квантовой механики бесполезны из-за природы измерений.

Постановка задачи. В данной работе требовалось решить следующие задачи

- Определить теоретическую возможность передачи сценария из полости в полость (в квантовых гейтах полость - это сам микрорезонатор, в квантовой химии - некоторая область пространства) [2]
- Определить условия передачи сценария
- Определить зависимости этой передачи от разного рода параметров в графе полостей, а именно: амплитуды волновода, его длины и формы графа

Реализация. Для решения не только этой, но и многих других задач в области квантовой механики, была реализована библиотека QComputations [3], призванная упростить и ускорить исследования в данной области. Библиотека полностью распараллелена в силу экспоненциального роста сложности

вычислений и требуемой памяти. С её помощью был полностью исследован вопрос данной ВКР на базе многих моделей, но в результатах [7] для демонстрации приведены лишь 2: Рабиевские осцилляций и неустойчивая связь.

Результаты. По результатам исследований определены следующие закономерности:

— Амплитуда волноводов напрямую влияет на скорость передачи динамики и на саму динамику в целом, но даже в видоизменённом виде динамика передаётся в полной мере при разных амплитудах в графе в абсолютно таком же виде. На факт передачи динамики **не** влияет. (Передаётся всегда)

— Форма графа влияет на амплитуду в целевой полости, так как частицы могут улетать дальше по графу. Также была замечена закономерность. При одинаковых длинах (в смысле числа ветвей) кратчайших путей, но при их разном числе, чтобы динамика в целевых полостях были в 2 графах **полностью** одинаковы, должна выполняться следующая формула: $pa^r = qb^r$, где

р - число путей в 1 графе

а - амплитуда волноводов в 1 графе

q - число путей в 2 графе

b - амплитуда волноводов в 2 графе

г - длина кратчайших путей (число ветвей).

На факт передачи динамики **не** влияет. (Передаётся всегда)

— Длина волновода влияет на факт передачи динамики по той же самой причине, что и в тёмных состояниях [4]. Также влияет на скорость передачи динамики, но это зависит полностью от самого сценария

— Реализована библиотека для квантовых вычислений на C++ - QComputations [3], с помощью которой можно работать в квантовых нотациях. Библиотека полностью распараллелена. Документация и инструкция написаны.

Вся работа лежит в открытом доступе по ссылке [7].

ЛИТЕРАТУРА

[1] Luc Montagnier, Electromagnetic signals are produced by aqueous nanostructures derived from bacterial DNA sequences, 2009, doi:10.1007/s12539-009-0036-7

[2] Ю.И.Ожигов., Квантовый компьютер, 2020, - ISBN 978-5-317-06403-7

[3] <https://github.com/KuzmaLine/QComputations>

[4] Ozhigov Y.I., Space of dark states in Tavis-Cummings model, 2019

[5] Кулагин Алексей Владимирович, Афанасьев Виталий Юрьевич, Ваньшунь Ли, Чжен Кэли, Мяо Хуэйхуэй, Плужников Иван, Ожигов Юрий Игоревич, Викторина Надежда Борисовна. Химический квантовый компьютер, Ломоносовские чтения 2021

- [6] Pulsed Rabi oscillations in quantum two-level systems: beyond the area theorem
- [7] <https://github.com/KuzmaLine/PCQED>

Прогноз выполнения задач видео по запросу и систем управления данными с использованием периферийных вычислений

Лычева Екатерина Олеговна

Кафедра автоматизации систем вычислительных комплексов

e-mail: s02200433@gse.cs.msu.ru

Научный руководитель — к.ф.-м.н. в.н.с. Писковский Виктор Олегович

Развитие технологии распределенных облачных вычислений, реализация отказоустойчивых сетевых сервисов в виде распределенных сетевых услуг приводят к необходимости перехода от атомарных систем, локализованных на выделенных для этого аппаратных единицах, к так называемым «логическим» или «сетевым» распределенным системам, которые «живы» пока есть обеспечивающая их существование сеть с достаточной степенью избыточности. Такие системы используют преимущества технологии виртуализации и распределены по различным серверным мощностям.

Обеспечение функционирования распределенных вычислительных систем в соответствии с предъявляемыми к ним техническими требованиями определяется эффективностью размещения сетевого сервиса на оборудовании в сети.

В работе исследуются методы прогнозирования временных характеристик прикладных сетевых сервисов [1] в зависимости от данных об эксплуатации сервиса на используемых физических ресурсах, текущего и прогнозируемого состояния аппаратного и программного обеспечения. Рассматривается влияние способов понижения размерности пространства признаков на результат предсказания характеристик уровня предоставляемого сервиса. Для прогнозирования применяется машинное обучение моделей случайного леса [2]. Для понижения размерности пространства признаков используются анализ главных компонент (PCA) [3] и усечённое сингулярное разложение (TruncatedSVD) [4].

Представлены результаты сравнительного анализа эффективности методов понижения размерности пространства признаков в контексте задачи прогнозирования параметров сетевых приложений с использованием машинного обучения.

ЛИТЕРАТУРА

- [1] Data traces from a data center testbed – Kaggle. [Электронный ресурс] - URL: <https://www.kaggle.com/datasets/jaliltaghia/data-traces-from-a-data-center-testbed>

- [2] A service-agnostic method for predicting service metrics in real time / R. Yanggratoke, J. Ahmed, J. Ardelius, C. Flinta, A. Johnsson, D. Gillblad, R. Stadler // International Journal of Network Management September 2017
- [3] Bishop C.M. Pattern Recognition and Machine Learning // Springer Science+Business Media, LLC 2006
- [4] Сайфутдинов Р. А. Исследование алгоритмов уменьшения размерности данных для задачи классификации // СПбГУ, 2014

Разработка и реализация переносимого решения для проверки корректности данных от системы мониторинга суперкомпьютеров

Матвеев Владимир Александрович

Кафедра Суперкомпьютеров и Квантовой Информатики

e-mail: maculaali@yandex.ru

Научный руководитель — к.ф.-м.н. зав.лаб. НИВЦ МГУ Воеводин Вадим Владимирович

Введение. Система мониторинга является важным инструментом анализа поведения суперкомпьютерных систем в целом и эффективности выполнения запускаемых на суперкомпьютере программ в частности. Неотъемлемым свойством систем мониторинга должна быть корректность выдаваемых ими данных, на проверку которой и направлено данное исследование.

Актуальность данной работы обуславливается тем фактом, что в настоящее время в открытом доступе отсутствуют полноценные, переносимые и автоматизированные решения по проверке корректности данных, выдаваемых системами мониторинга суперкомпьютеров.

Проведённая работа. Выпускная квалификационная работа описывает проведённые доработки, модификации и расширение функционала разработанной в 2023 году базовой системы проверки корректности данных от системы мониторинга производительности [1]. Базовая версия включала в себя 10 тестов, каждый из которых призван проверять корректность определенной характеристики производительности приложений. Проведённая в данном исследовании работа включала в себя разработку недостающих тестов, исправление ошибок в уже готовых тестах, написание подробной документации. Также была проведена автоматизация процесса тестирования и существенно повышена переносимость разработанного решения.

Результаты. В рамках данного исследования были разработаны, реализованы, апробированы и верифицированы методы тестирования для следующих характеристик работы приложений:

- количество открытых/закрытых файлов;
- число IPC (выполненных инструкций за такт), количество выполненных инструкций и тактов процессора;
- nice уровень загрузки процессора;
- загрузка памяти графического ускорителя;

– любые доступные PAPI [2] датчики, в т.ч. аппаратные.

Каждый тест выдает ожидаемое значение проверяемой характеристики производительности, которое затем можно сравнить с данными мониторинга, полученными во время работы теста. Если значения совпадают — данные по рассматриваемой характеристике выдаются корректно; если различаются — в работе системы мониторинга присутствуют ошибки.

В процессе разработки тестов было необходимо решить различные подзадачи. Так, получение тестом данных от аппаратных процессорных датчиков не позволяет одновременно собирать эти данные системой мониторинга. Поэтому в тесте IPC было невозможно получать данные о числе выполненных инструкций. Было принято решение определять значение этой характеристики аналитически. Для обеспечения точности такого подхода, в код теста были добавлены ассемблерные вставки, а сам тест разделён на 2 версии — для процессоров с архитектурой x86 и ARM. Также был разработан и реализован механизм тестирования, обеспечивающий корректную работу системы мониторинга и тестов, которым необходим одновременный доступ к аппаратным датчикам PAPI.

Для всех тестов была выполнена интеграция в рамках общего решения. Проведено тестирование корректности работы системы на суперкомпьютере Ломоносов-2, ПК Polus и нескольких локальных ЭВМ с различной компонентной базой. Помимо этого, был разработан механизм простого добавления новых тестов в общую систему. Также была выполнена автоматизация процесса сравнения данных мониторинга и данных от системы тестирования, что позволяет заметно упростить проверку корректности данных мониторинга с помощью разработанного решения.

При помощи полученного решения была проверена корректность данных, выдаваемых системой мониторинга производительности DiMMon [3] на суперкомпьютере Ломоносов-2.

Помимо этого, была обеспечена и проверена переносимость созданного решения, в том числе на разных аппаратных архитектурах и разных ОС (тестировалась на процессорах Intel, AMD и ARM; на дистрибутивах Linux на базе Debian, Arch и решений от RedHat).

Код проекта доступен в GitLab НИВЦ МГУ по ссылке:

<https://gitlab.srcc.msu.ru/maculaali/supmv>.

ЛИТЕРАТУРА

- [1] Kashin D., Voevodin V. Verifying the Correctness of HPC Performance Monitoring Data // Parallel Computing Technologies, 17th International Conference, PaCT 2023. Vol. 14098. P. 10–13.
- [2] A Portable Programming Interface for Performance Evaluation on Modern Processors // International Journal of High Performance Computing Applications archive. Aug. 2000. Vol. 14, N 3. P. 189–204.
- [3] Dynamically reconfigurable distributed modular monitoring system for supercomputers (dimmon) / K. Stefanov, V. Voevodin, S. Zhumatiy,

V. Voevodin // 4th International Young Scientist Conference on Computational Science. Procedia Computer Science. Elsevier B.V Netherlands. 2015. Vol. 66. P. 625–634.

Детектирование нарушений ограничений контрольных групп ядра операционной системы Linux в системах контейнеризации

Пчелкин Фёдор Алексеевич

Кафедра системного программирования

e-mail: pchelkin@ispras.ru

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

Системы контейнеризации приобретают всё большую популярность в облачных средах, поддерживая приемлемый уровень изоляции между процессами и эффективно используя имеющиеся аппаратные ресурсы вычислительной инфраструктуры [1]. Контрольные группы лежат в основе систем контейнеризации, обеспечивая разграничение и доступ к различным типам разделяемых системных ресурсов как между несколькими соседствующими контейнерами, так и между контейнерами и другими процессами хост-системы. Это необходимо для того, чтобы предотвратить потенциальную возможность исчерпания всех ресурсов хост-системы одним или несколькими зловредными контейнерами.

Тем не менее, в статье [2] был представлен ряд архитектурных недостатков механизма контрольных групп, при эксплуатации которых зловредный контейнер способен выйти за рамки наложенных на него ресурсных ограничений. Основная идея заключается в делегировании дополнительной нагрузки потокам выполнения в системе, принадлежащим другим контрольным группам и на которые действуют другие, зачастую более слабые, ресурсные ограничения. В таких случаях механизм контрольных групп не способен ассоциировать с зловредным контейнером сверхпотребление ресурсов остальными компонентами системы.

В данной работе предложен метод детектирования нарушений ограничений контрольных групп ядра операционной системы Linux в системах контейнеризации, состоящий из следующих шагов:

- определение типов ресурсов, контроль ограничений которых будет производиться;
- развёртывание на хост-машине разработанной тестовой системы;
- изначальная подготовка ОС-пакетов и модификация конфигураций ограничений тестовых контейнеров;
- реализация и интеграция в тестовую систему агентов по сбору метрик использования заявленного ресурса и выявлению аномальных показателей его потребления;

— запуск системы и анализ сценариев, спровоцировавших сверхпотребление ресурса.

Метод базируется на использовании спроектированной тестовой системы, за основу которой взят инструмент Syzkaller фаззинг-тестирования ядра Linux. В ходе работы в устройство Syzkaller была включена поддержка сред выполнения контейнеров для проведения фаззинг-тестирования в необходимых условиях, а также добавлена инфраструктура для внедрения агентов-измерителей потребления ресурсов гостевыми окружениями и агентов-детекторов аномального использования соответствующих ресурсов.

Метод был апробирован для реализации детектирования нарушений ограничений `cru` и `cgroup` ресурс-контроллеров. Получение данных потребления ресурсов ЦПУ в различных контекстах исполнения компонентами гостевой системы было реализовано посредством технологий интроспекции виртуальных машин.

Были проведены запуски системы для валидации обнаружения аномальных нагрузок на известных сценариях нарушений ограничений указанных `cru` и `cgroup` контроллеров. Также производились запуски системы в рабочем режиме, в результате которых было выявлено несколько ранее неизвестных сценариев нарушений ограничений контрольных групп, в частности, ведущих к делегированию дополнительной нагрузки `init`-процессу системы и сервисному `udev`-демону.

ЛИТЕРАТУРА

- [1] A. Randal. The ideal versus the real: revisiting the history of virtual machines and containers // ACM Computing Surveys. 2020. Vol. 53, N5. Pp. 1–31. DOI: 10.1145/3365199.
- [2] Houdini’s escape: breaking the resource rein of Linux control groups / X. Gao, Z. Gu, Z. Li, H. Jamjoom, C. Wang // Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security. 2019. Pp. 1073–1086. DOI: 10.1145/3319535.3354227.

Моделирование оптических свойств многослойных тонких плёнок в задачах синтеза изображений

Родионов Роман Олегович

Кафедра интеллектуальных информационных технологий

e-mail: roman.rodionov@graphics.cs.msu.ru

Научный руководитель — к.ф.-м.н. м.н.с. Санжаров Вадим Владимирович

При разработке систем фотореалистичного синтеза изображений (рендеринга) одной из важнейших задач является реализация моделей материалов, отражающих свойства объектов реального мира. Особыми оптическими свойствами обладают объекты, поверхность которых покрыта тонкими плёнками — слоями материалов толщиной от нанометра до нескольких микрометров, что сопоставимо с длинами волн видимого спектра.

Объекты, покрытые тонкими плёнками, встречаются в природе и быту: мыльные пузыри, крылья насекомых, кожа змей. Кроме того, их наносят на оптические приборы с целью получения отражающих и просветляющих покрытий, добавляют в состав автомобильных красок. Моделирование подобных материалов актуально как в задачах фотореалистичного рендеринга, так и в проектировании оптических систем. В данной работе решается задача разработки и реализации математической модели материала многослойных тонких плёнок, применимой для визуализации широкого класса объектов реального мира, имеющей открытую программную реализацию и поддерживающей вычисления на графическом процессоре.

В ходе исследования был проведён обзор существующих работ. Были рассмотрены модели для однослойных и многослойных плёнок, в том числе из проводящих материалов. За основу взяты два аналитических подхода для моделирования многослойных тонких плёнок: метод матриц переноса [1] и рекуррентный метод [2]. Для визуализации шероховатых плёнок используется физически корректная модель, основанная на теории микрограней [3].

Реализованная модель материала, поддерживающая вычисления с помощью выбранных базовых методов, была интегрирована в систему фотореалистичного рендеринга Hydra Renderer лаборатории компьютерной графики и мультимедиа кафедры ИИТ факультета ВМК МГУ. Модель объединяет в себе возможности, доступные в отдельности в разных существующих работах. Учитываются такие параметры материала, как шероховатость поверхности, толщина, показатель преломления и показатель поглощения для каждого слоя. Допускается комбинация нескольких слоёв с различными параметрами. Для ускорения синтеза изображений добавлена возможность предварительного расчёта значений отражательной и пропускательной способности материала в зависимости от угла падения луча и длины волны. Для рендеринга не только в спектральном режиме, но и в режиме RGB, используется перевод предрасчитанных значений в цветовой формат. Кроме того, поддерживается локальное задание параметра толщины плёнки с помощью специальных текстур. Реализация модели была адаптирована для работы на графическом процессоре с использованием инструмента Kernel Slicer [4].

Для сравнения производительности выбранных базовых методов проведено исследование зависимости времени рендеринга от числа слоёв, включённых в материал многослойной тонкой плёнки. Эксперимент показал, что время рендеринга возрастает линейно с ростом числа слоёв для обоих методов. При этом реализованный рекуррентный метод оказался более производительным. Перенос вычислений на графический процессор уменьшил время рендеринга тестовой сцены до 66 раз.

Реализация материала в составе рендер-системы Hydra Renderer выложена в открытом доступе [5].

ЛИТЕРАТУРА

- [1] Benamira A., Pattanaik S. Application of the transfer matrix method to anti-reflective coating rendering // Proceedings of the Computer Graphics International Conference. 2020. P. 83–95.
- [2] Spectral rendering of interference phenomena caused by multilayer films under global illumination environment / S. Ikeda, S. Watanabe, B. Raytchev, T. Tamaki, K. Kaned // ITE Transactions on Media Technology and Applications. 2015. Vol. 3. P. 76–84.
- [3] Wu F., Zhen C. Microfacet-based interference simulation for multilayer films // Graphical Models. 2015. Vol. 78. P. 26–35.
- [4] Инструмент для генерации кода для API Vulkan [Электронный ресурс] // URL: https://github.com/Ray-Tracing-Systems/kernel_slicer
- [5] Программная реализация материала в составе рендер-системы [Электронный ресурс] // URL: <https://github.com/RomanRodionov/HydraCore3>

Исследование и разработка методов классификации изображений мобильного пользовательского интерфейса

Работа удостоена диплома 2 степени

Семеняк Георгий Алексеевич

Кафедра системного программирования

e-mail: georgesemenyak@gmail.com

Научный руководитель — к.ф.-м.н. доц. Турдаков Денис Юрьевич

Научный консультант — Сорокин Константин Сергеевич

Современный рынок мобильного программного обеспечения характеризуется большими масштабами и высокой конкуренцией, соответственно, кто будет поставлять наиболее быстрые и корректные приложения, тот и будет выигрывать у конкурентов и получать клиентов. Современные мобильные приложения — это сложные программные продукты, которые содержат ошибки, требующие быстрого обнаружения. Один из видов автоматизированного тестирования приложений, популярный в настоящее время [1] — *исследовательское тестирование* (exploratory testing, GUI testing). Этот вид тестирования осуществляется путём имитации тестирующей системой взаимодействия пользователя с графическим интерфейсом приложения: тестирующая система получает данные о текущем состоянии устройства и отправляет на него обратно команды для выполнения необходимых действий.

Процесс исследовательского тестирования имеет ряд недостатков. Информация об экране часто доступна только в виде самого изображения, а не *дерева пользовательского интерфейса* — дерева, узлами которого являются элементы пользовательского интерфейса, а рёбра указывают на отношение непосредственной вложенности. Узлы содержат в себе такие важные атрибуты как класс элемента из API Android и информацию об интерактивности элемента и значении его текстового поля. Отсутствие дерева пользова-

тельского интерфейса означает и отсутствие семантической информации об экране и его элементах.

Целью данной работы является реализация семантического классификатора изображений пользовательского интерфейса для его использования в инструменте исследовательского тестирования. Это позволит повысить эффективность и качество тестирования мобильных приложений путем автоматической классификации экранов и приоритизации действий инструмента тестирования. Также это поможет решить проблему отсутствия деревьев элементов пользовательского интерфейса.

Актуальность данной работы подтверждается и тем, что такие мировые IT-гиганты, как Apple и Google, в настоящее время занимаются задачами, смежными с представленной, о чём можно судить из публикаций [2] и [3].

В качестве исходного набора данных был использован Engiso [4], состоящий из 1460 изображений мобильного пользовательского интерфейса, размеченных на 20 классов. Он был изучен, и в нём был выявлен ряд проблем, таких как дисбаланс классов, устаревший графический пользовательский интерфейс (2017 года) и присутствующие ошибки в разметке. Исходя из данных проблем, был проделан ряд работ: произведена полная переразметка набора данных, убраны 5 существующих классов, которые не являются полезными для исследовательского тестирования, и добавлены 4 новых класса, наоборот, полезных. В дополнение, были собраны и размечены 1767 новых изображений из более 150 Android-приложений с современным графическим пользовательским интерфейсом.

Для решения поставленной задачи был реализован ряд свёрточных нейросетевых моделей на основе Resnet34. Разработаны базовая модель-классификатор, модель контрастного обучения с учителем и серия моделей для решения задач предобучения: восстановитель закрашенных регионов на изображении, автокодировщик изображений и модель контрастного обучения без учителя. Предобученные модели были доучены на целевую задачу классификации при помощи тонкой донастройки методом обучения с учителем.

Наконец, была реализована модель многозадачного обучения, которая одновременно выполняет несколько задач: классификация изображений, автокодирование, восстановление закрашенных регионов на изображении и контрастное обучение с учителем. Особенности архитектуры этой модели являются общие слои для извлечения признаков, в качестве которых был взят тот же Resnet34; 4 разные «головы» для каждой из задач, обучаемые одновременно, и взвешенная функция потерь с весом 0.7 для задачи классификации и весами по 0.1 для остальных трёх задач.

Была проведена оценка качества работы возможных комбинаций разработанных моделей, и лучшей из них, а именно, моделью многозадачного обучения, получен результат классификации 0.71 macro f1-score, что на 7 и более

пунктов превосходит результаты существующих реализаций, решающих схожую задачу.

ЛИТЕРАТУРА

- [1] Improving Automated GUI Exploration of Android Apps via Static Dependency Analysis / Guo W., Shen L., Su T., Peng X., Xie W. // 2020 IEEE International Conference on Software Maintenance and Evolution (ICSME). 2020. P.557–568.
- [2] Ferret: Refer and ground anything anywhere at any granularity / You H., Zhang H., Gan Z., Du X., Zhang B., Wang Z., Cao L., Chang CF., Yang Y. // arXiv preprint arXiv:2310.07704. 2023.
- [3] Li G., Li Y. Spotlight: Mobile UI Understanding using Vision-Language Models with a Focus // arXiv preprint arXiv:2209.14927. 2022
- [4] Leiva L. A., Hota A., Qulasvirta A. Enrico: A dataset for topic modeling of mobile UI designs // 22nd International Conference on Human-Computer Interaction with Mobile Devices and Services. 2020. P.1–4.

Исследование оптимальных параметров для повышения производительности блокчейн-хранилища для контроллера RunOS

Шибает Павел Павлович

Кафедра автоматизации систем вычислительных комплексов

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

Научный руководитель — к.ф.-м.н. Писковский Виктор Олегович

Имеется самообучающийся коммутатор второго уровня (l2-learning-switch, далее – L2 LS) в виде приложения для контроллера программно-конфигурируемых сетей RunOS[1] и блокчейн для хранения данных приложения. Задача состоит в том, чтобы выявить зависимости, которые позволили бы настраивать блокчейн-сеть (в частности, параметры консенсуса в этой сети) оптимальным образом.

Для достижения этой цели определим следующие переменные:

- пропускная способность блокчейн-сети (T) — количество транзакций, попадающих в распределенный реестр, в единицу времени;
- интенсивность потока транзакций (f) — количество поступающих на узлы-валидаторы в течение эксперимента транзакций в единицу времени;
- доля отклоненных транзакций ($Q = 1 - \frac{T}{f}$) — доля транзакций в единицу времени, попавших в сеть, но не включенных в блоки;
- средний размер транзакции (s) — средний размер транзакции в потоке транзакций;
- конфигурация алгоритма консенсуса ($C_i, i \in 1, 2, 3$) есть кортеж (b, p, τ) , b — скорость байтового потока передачи транзакций, p — количество постоянных соединений, τ — интервал голосования.

Требуется выявить зависимости $T(C_i, f)$ и $Q(C_i, f)$, $i \in 1, 2, 3$. Для лучшей конфигурации исследовать зависимости $T(s, f)$, $Q(s, f)$. Табличные данные об этих зависимостях позволяют аппроксимировать функции $f(s, Q)$ и $s(f, Q)$ для того, чтобы найти решение двух оптимизационных задач (1) и (2).

$$\begin{cases} \max f(s_0, Q) \\ Q \leq Q_0 \end{cases} \quad (1)$$

$$\begin{cases} \max s(f_0, Q) \\ Q \leq Q_0 \end{cases} \quad (2)$$

На языке Go реализовано блокчейн-приложение, использующее алгоритм консенсуса Tendermint. Блокчейн-приложение развернуто в 16 экземплярах на отдельных виртуальных машинах (кластер работает на процессорах Intel Xeon E5645, Intel Xeon E5-2630 v3; каждая машина имеет доступ к 2 ядрам ЦПУ и 4 Гб ОЗУ) с помощью скриптов Ansible в соответствии с подходом «инфраструктура как код». Блокчейн-приложение скомпилировано с использованием Cosmos SDK версии v0.50. Также присутствует экземпляр сетевого контроллера RunOS, на котором запущено модифицированное приложение L2 Learning Switch, написанное на языке C++. Экземпляр RunOS взаимодействует с OpenvSwitch Mininet v2.2.2. Mininet — ПО для разворачивания виртуальных сетей. В блокчейне хранятся кортежи вида $(dp_id, eth_addr, port)$.

Проведено три эксперимента: 1) сравнение производительности операций чтения-записи централизованного хранилища Redis и блокчейн-приложения 2) сбор значений T и Q для различных конфигураций 3) сбор значений для выявления зависимостей $f(s, Q)$, $s(f, Q)$. Во всех экспериментах для создания транзакций на запись использовалась утилита на основе tm-load-test. Для операции чтения использовалась CLI утилита, вызывающая модуль запросов блокчейн-приложения.

Алгоритм консенсуса Tendermint [2] оказался устойчив к экспоненциальному росту потока транзакций при среднем размере этих транзакций до 10000 байт. Наиболее устойчивой по доле отклоненных транзакций конфигурацией является конфигурация C_3 , которую следует использовать для разработки конфигураций в дальнейших исследованиях. Конфигурация C_3 может применяться в приложениях, которые генерируют до 3200 транзакций в секунду при размере транзакций до 1 Мб.

На основе анализа полученных экспериментальных результатов разработан метод приблизительного решения поставленных математических задач ((1) и (2)), который основан на сплайн-аппроксимации.

Результаты исследования обсуждались на ряде конференций, таких как: Ломоносовские Чтения – 2023 ; XXX Международная научная конференция студентов, аспирантов и молодых учёных «Ломоносов-2023», XVI-я Всероссийская мультikonференция по проблемам управления «Управление в распределенных и сетевых системах (УРСС-2023)», «Тихоновские чтения 2023», «Ломоносовские чтения - 2024». Результаты работы были опубликованы в тематическом сборнике «Программные системы и инструменты» и легли в осно-

ву доклада, поданного для участия в международной конференции «MoNeTec-2024».

ЛИТЕРАТУРА

- [1] The Runos OpenFlow Controller / Shalimov A., Nizovtsev S., Morkovnik D., Smeliansky R. // Proceedings — European Workshop on Software Defined Networks EWSDN : 4 Bilbao 30 сентября 2015 года. — Bilbao, 2015. — P. 103–104. — DOI 10.1109/EWSDN.2015.69. — EDN WWBTAZ.
- [2] Correctness of tendermint-core blockchains / Y. Amoussou-Guenou, A. D. Pozzo, S. Tucci-Piergiovanni, M. Potop-Butucaru // Leibniz International Proceedings in Informatics, LIPIcs : 22, Hong Kong, 17–19 декабря 2018 года. — Hong Kong, 2019. — P. 16. — DOI 10.4230/LIPIcs.OPODIS.2018.16. — EDN DCVONC.

Исследование и разработка методов кластеризации топологии сети передачи данных и отображения сети кластеров на среду распределенного моделирования NPS (network prototype simulator)

Работа удостоена диплома 2 степени

Тюшев Максим Владимирович

Кафедра автоматизации систем вычислительных комплексов

e-mail: s02200498@gse.cs.msu.ru

Научный руководитель — д.ф.-м.н. чл.-корр. РАН Смелянский Руслан Леонидович

Научный консультант — д.ф.-м.н. проф. Ефимов Иван Влерьевич

Одна из основных проблем распределенного имитационного моделирования – отображение моделируемой сети на распределенную среду моделирования. Для эффективной работы системы распределенного имитационного моделирования NPS, разработанной на кафедре АСВК в 2015 году, а именно обеспечения согласованности модельного времени между кластерами, минимизации отклонения модельного времени от астрономического времени, необходимо разработать метод отображения моделируемой сети передачи данных (далее граф-модель) на среду распределенного имитационного моделирования (далее РСИМ). Эта задача рассматривается как суперпозиция решений двух задач:

- задача кластеризации графа, представляющего топологию граф-модели, где каждый узел – это точка обработки данных;
- задача отображения графа, где вершины – кластеры графа, представляющую модельную сеть (далее кластер-граф), на граф, представляющий топологию РСИМ, где вершины – вычислительные сервера.

При следующих предположениях:

- характеристики всех узлов граф-модели одинаковы, и при моделировании каждый узел граф-модели создает одинаковую вычислительную нагрузку на сервер РСИМ;
- все сервера РСИМ одинаковы;
- Сеть РСИМ неоднородна (пропускные способности каналов могут быть разными).

В работе представлены:

- математическая постановка задачи;
- результаты сравнительного анализа разнообразных алгоритмов для разбиения граф-модели на кластеры и алгоритмов для отображения кластер-графов на граф РСИМ, в результате которого были отобраны алгоритмы кластеризации METIS, СТСЕНС и алгоритмы отображения ISORANK и MAGNA++;
- результаты экспериментального исследования суперпозиции алгоритмов кластеризации и отображения графов ISORANK и MAGNA++ и METIS и MAGNA++;
- показано, что суперпозиция METIS и MAGNA++ является наилучшей по критериям: равномерность распределения вычислительной нагрузки по серверам распределенной среды имитационного моделирования, минимизация отклонения модельного времени от астрономического.

ЛИТЕРАТУРА

- [1] Antonenko, Vitaly A. Large scale network simulation based on hi-fi approach / Vitaly A Antonenko, Ruslan L Smelyanskiy, Andrey Nikolaev // SummerSim. — 2014. — P. 4.
- [2] Ma, Yan, Lin, Hongren, Wang, Yan et al. A multi-stage hierarchical clustering algorithm based on centroid of tree and cut edge constraint // Information Sciences. — 2021. — Vol. 557. — Pp. 194–219.
- [3] Karypis, George. METIS: A software package for partitioning unstructured graphs, partitioning meshes, and computing fill-reducing orderings of sparse matrices / George Karypis, Vipin Kumar. — 1997.
- [4] Liao, ChungShou, Lu, Kanghao, Baym, Michael et al. IsoRankN: spectral methods for global alignment of multiple protein networks // Bioinformatics. — 2009. — Vol. 25, no. 12. — Pp. i253–i258.
- [5] Vijayan, Vipin. MAGNA++: maximizing accuracy in global network alignment via both node and edge conservation / Vipin Vijayan, Vikram Saraph, Tijana Milenković // Bioinformatics. — 2015. — Vol. 31, no. 14. — Pp. 2409–2411.

Исследование и разработка методов анализа выживаемости для прогнозирования сбоя оборудования

Работа удостоена диплома 1 степени

Филимонова Ирина Олеговна

Кафедра интеллектуальных информационных технологий

e-mail: s02200574@gse.cs.msu.ru

Научный руководитель — к.ф.-м.н. доц. Петровский Михаил Игоревич

В современном мире, где технологии становятся все более сложными и интегрированными, обеспечение надежной работы систем является критически важным условием их эксплуатации. Анализ надежности позволяет предсказать вероятность сбоев и определить факторы, влияющие на безотказную работу оборудования. Одним из главных инструментов анализа надежности является метод анализа выживаемости, заимствованный из биомедицинской статистики. Первоначально предназначенный для изучения выживаемости пациентов и факторов, влияющих на нее, этот метод нашел широкое применение в анализе надежности инженерных систем.

В рамках работы была поставлена задача исследования и разработки расширений существующих методов анализа выживаемости для решения задачи анализа надежности. На основе рассмотренных моделей была построена общая стратификационная модель с использованием дерева выживаемости [1] в качестве стратификатора и исследована его обобщающая способность. Также в рамках исследования нейросетевых моделей анализа выживаемости была реализована генеративно-состязательная модель, которая позволяет прогнозировать ожидаемое время события на основе ковариат. В качестве расширения данного подхода был предложен собственный подход прогнозирования непрерывных функций выживания и риска на основе множества точечных прогнозов. Для обогащения признакового пространства и заполнения пропусков в данных предложен метод ансамблирования дерева выживаемости и нейросетевых моделей.

Чтобы оценить качество предложенных методов на реальных данных был собран и предобработан один из крупнейших доступных наборов открытых квартальных отчетов параметров S.M.A.R.T. (Self-Monitoring, Analysis and Reporting Technology) компании Backblaze [2].

Было проведено сравнение предложенных методов с существующими подходами, включающими как статистические методы [3], так и нейросетевые, такие как DeepSurv [4], CoxCC и CoxTime [5]. Экспериментальное исследование показало, что предложенный метод стекинга превосходит существующие подходы по метрикам качества CI, IBS_REMAIN, AUPRC на собранных наборах данных, что подтверждает его эффективность и практическую применимость в прогнозировании сбоев оборудования.

ЛИТЕРАТУРА

- [1] Vasilev I., Petrovskiy M., Mashechkin I. V. Survival Analysis Algorithms based on Decision Trees with Weighted Log-rank Criteria //ICPRAM. 2022. С. 132–140.
- [2] Backblaze N. Hard drive data and stats. 2023.
- [3] Wang P., Li Y., Reddy C. K. Machine learning for survival analysis : A survey //ACM Computing Surveys (CSUR). 2019. Т. 51. №. 6. С. 1–36.
- [4] Katzman J. L. et al. DeepSurv : personalized treatment recommender system using a Cox proportional hazards deep neural network //BMC medical research methodology. 2018. Т. 18. С. 1–12.
- [5] Kvamme H., Borgan Ø., Scheel I. Time-to-event prediction with neural networks and Cox regression //Journal of machine learning research. 2019. Т. 20. №. 129. С. 1–30.

Прогнозирование характеристик очередей задач пользователей на вычислительных кластерах с использованием глубоких нейронных сетей

Ященко Игорь Сергеевич

Кафедра автоматизации систем вычислительных комплексов

e-mail: yashch.igor@gmail.com

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

Прогнозирование времени ожидания задач в очереди в вычислительных кластерах важно для эффективного управления ресурсами. Развитие методов в данной области необходимо, так как на данный момент существует малое количество исследований по данной теме, встроенный прогнозатор Slurm — системы управления ресурсами, используемой в 60% суперкомпьютеров из списка TOP500, — несовершенен, увеличивается количество неопытных пользователей вычислительных кластеров, и существует необходимость выбора наилучшего кластера под задачу. Улучшение методов прогнозирования позволит оптимизировать распределение задач, уменьшить время ожидания и повысить общую производительность системы, что особенно важно в условиях растущей сложности и разнообразия вычислительных задач.

В ходе настоящей работы была разработана система прогнозирования времени ожидания задачи в очереди до начала ее выполнения. Она включает клиентский плагин Slurm и сервер с доступом к базе данных Slurm, производящий прогноз времени ожидания и возвращающий его плагину. Для вычисления прогнозов были обучены разнообразные модели машинного обучения, включая нейронные сети, с различными характеристиками точности и скорости вычисления прогнозов для разных типов вычислительных кластеров. Наилучшее качество прогнозов в метрике MAE было достигнуто с помощью двухмодульной нейронной сети, использующей механизм внимания [1]. Для обучения и проверки качества моделей использовались два набора реальных

данных: данные о результатах мониторинга системы MIT Supercloud [2] и данные о работе кластера Mustang [3] лаборатории Лос-Аламос.

Также в данной работе был реализован тестовый стенд, на котором было проведено экспериментальное тестирование системы прогнозирования с помощью сгенерированного набора синтетических данных. Таким образом, была показана применимость разработанной системы в реальных условиях, а также показана точность предложенного метода, превосходящая точность встроенной оценки времени ожидания Slurm на 80%.

Часть результатов данной работы представлена на международной конференции «Distributed Computing and Grid Technologies in Science and Education (GRID-2023)», по материалам которой в журнале «Physics of Particles and Nuclei» опубликована статья «Software for prediction of task start moment in computer cluster by statistical analysis of jobs queue history».

ЛИТЕРАТУРА

- [1] Vaswani A., Shazeer N., Parmar N. Attention is all you need // Advances in neural information processing systems. 2017. Vol. 30, N 2. P. 5998–6008.
- [2] Li B., Arora R., Samsi S. AI-Enabling Workloads on Large-Scale GPU-Accelerated System: Characterization, Opportunities, and Implications // 2022 IEEE International Symposium on High-Performance Computer Architecture (HPCA). 2022. P. 1224–1237.
- [3] Amvrosiadis G., Park J.W., Ganger G. On the diversity of cluster workloads and its impact on research results // Proceedings of the 2018 USENIX Annual Technical Conference. 2018.

Повышение точности моделирования библиотечных функций в статическом анализаторе исходного кода

*Работа удостоена диплома 2 степени
Биктимиров Михаил Геннадьевич*

*Кафедра системного программирования
e-mail: bikmish@mail.ru*

*Научный руководитель — к.ф.-м.н. Игнатьев Валерий Николаевич
Научный консультант — Беляев Михаил Владимирович*

В работе представлен один из методов, используемых статическим анализатором исходного кода SharpChecker [1], для моделирования библиотечных функций с использованием спецификаций C# в механизме символьного исполнения. Такой подход позволяет с достаточной точностью учитывать при анализе важные побочные эффекты и другую полезную для детекторов ошибок информацию. Также описан способ анализа пользовательских лямбда-функций, передаваемых в библиотечные методы в качестве параметров, что особенно важно для часто используемых в C# операторов запроса LINQ.

Описание предметной области. Статический анализ кода – это метод анализа исходного кода, при котором код анализируется без его фактического выполнения. Он основывается на анализе синтаксических и семантических

свойств кода и позволяет обнаружить ошибки, потенциальные проблемы и уязвимости до того, как код будет запущен.

Одной из проблем статического анализа исходного кода является обработка вызовов библиотечных функций. Суть проблемы заключается в том, что статический анализ использует только исходный код пользовательской программы и не имеет доступа к исходному коду вызываемых библиотечных функций. Таким образом, нет возможности анализировать эти функции как пользовательский код, и невозможно собрать необходимую информацию, например, возможные исключения, которые они могут выдать, возвращают ли они значение `null` и т. д. Это делает невозможным предсказать результат и побочные эффекты их выполнения для дальнейшего анализа пользовательского кода. Такая нехватка семантики библиотечных методов отрицательно сказывается на полноте результатов анализа, поскольку, например, при разыменовании возвращаемого значения детектор может не знать, что соответствующая библиотечная функция возвращает `null` в текущих условиях, из-за чего не будет выявлено ошибки разыменования `null`.

Таким образом, качество моделирования библиотечных функций существенно влияет на точность и полноту результатов статического анализа, что подтверждает необходимость в разработке подходов к моделированию библиотечных функций.

Подсистема спецификации библиотечных функций. Для обоснования выбора метода моделирования с использованием сокращённых спецификаций в работе был проведён всесторонний обзор основных существующих подходов к библиотечному моделированию, демонстрирующий достоинства и недостатки каждого из подходов. Было выделено четыре основных способа моделирования:

1. *Семантика, явно заданная в коде движка.* Гибкость, эффективное моделирование отдельных свойств, крайняя немасштабируемость;
2. *База данных аннотаций.* Эффективный подход, легковесные модели, однако ограниченная выразительность (невозможно учитывать нетривиальные контекстно-зависимые побочные эффекты);
3. *Предварительный анализ библиотек.* Увеличенная сложность анализа, низкая производительность;
4. *Сокращённые C# спецификации.*
 - спецификации являются масштабируемыми и легкими, предоставляют информацию, дополненную предикатами, перекрестными зависимостями и побочными эффектами;
 - нет лишней семантической сложности;
 - нет потерь производительности.

Обзор существующих решений показал, что ни одно из них не решает достаточно эффективно и полно поставленную задачу моделирования.

В работе предложен новый подход для моделирования библиотечных функций на основе спецификаций на языке С#. В статическом анализаторе SharpChecker была создана подсистема моделирования с использованием данного подхода с использованием атрибута, который позволил пометить только «интересующие» для моделирования функции, а остальные рассматривать как неизвестные. Это позволило избежать ошибок компиляции и гарантировать масштабируемость подхода, поскольку атрибут позволяет писать спецификации не ко всем функциям библиотеки, а только к необходимым. Для данной системы моделирования было написано 89 спецификаций библиотечных функций: 65 спецификаций для моделирования функций работы с коллекциями, 15 – для функций операторов запросов LINQ, 9 – для функций, которые работают с пользовательским вводом/выводом.

Анализ вызовов делегатов и подстановок лямбда-выражений. Также в контексте задачи моделирования библиотечных функций в рамках межпроцедурного анализа появляется необходимость в поддержке анализа функций, переданных в качестве аргументов другим функциям с помощью делегатов С#. Поэтому был разработан алгоритм анализа вызовов делегатов для поддержки лямбда-выражений, переданных в качестве аргументов функций, а также алгоритм учета лямбда-выражений, которые передаются в предопределенные LINQ операторы в качестве предикатов.

Тестирование. SharpChecker с описанными в работе улучшениями тестировался на наборе различных проектов с открытым исходным кодом, который состоит из 6 миллионов строк кода. Для оценки эффективности были включены спецификации для нескольких основных методов класса `TextReader`, классов коллекций `List<T>`, `Stack<T>` (конструкторы, добавление элементов, подсчет количества элементов), а также для некоторых операторов запроса LINQ.

Тестирование не показало никаких ухудшений в производительности, что подтверждает легковесность спецификаций и эффективность алгоритмов анализа делегатов и подстановок лямбда-выражений. Всего было проанализировано около 20000 различных вызовов делегатов. Соответствующие им резюме были применены к контекстам вызвавших их функций в зависимости от количества вызовов каждого делегата. Среди них 2000 делегатов были вызваны один раз, и в точности их резюме были применены к контексту. Примерно 18000 делегатов вызывались ноль и более раз, поэтому их резюме применялись с упрощениями (анализатор присвоил новые значения из резюме переданной функции только тем захваченным ею переменным, которые представляют из себя константы, а остальным переменным были сопоставлены некоторые неизвестные значения).

Моделирование всего нескольких библиотечных функций в совокупности с использованием алгоритмов анализа вызовов делегатов и подстановки лямбда-выражений позволило при анализе проектов из набора ПО, который состоит из 6 миллионов строк кода, избавиться от 111 ложных предупрежде-

ний и найти 74 новых истинных срабатываний. При этом было потеряно 21 истинное срабатывание и приобретено 20 новых ложных. Из общего числа ушедших ложных срабатываний 19 ушло благодаря работе алгоритма анализа вызовов делегатов, также этот алгоритм способствовал нахождению 4 (из 74 всего пришедших) новых истинных срабатываний, однако пришло 3 (из 20 всего) новых ложных срабатывания и ушло 3 (из 21 всего ушедших) истинных.

Описанные в работе улучшения анализа дали положительный результат: изменения позволяют обнаруживать новые истинные срабатывания и избавляться от ложных предупреждений без потери производительности. При этом демонстрируется эффективность использования спецификаций с разработанными алгоритмами, поскольку они не оказывают влияния на время анализа. Поэтому, добавляя спецификации для остальных часто используемых библиотечных функций, можно еще больше повысить точность и эффективность анализа.

Результаты работы представлены на международной конференции ISPRAS OPEN – 2023 [2].

ЛИТЕРАТУРА

- [1] V.K.Koshelev, V.N. Ignatiev, A.I. Borzilov, A. A. Belevantsev. Static analysis tool for C# programs // Programming and Computer Software, vol. 43, issue 4. 2017. Pp. 268–276.
- [2] Biktimirov M. G., Ignatyev V. N., Belyaev M. V. Improving the Accuracy of Library Function Modeling in the Static Analyzer // 2023 Ivannikov Ispras Open Conference (ISPRAS). 2023. Pp. 26–32.

Балансно-характеристический метод расчета гемодинамики в сосуде с подвижными стенками

Конопляников Василий Викторович

Кафедра вычислительных методов

e-mail: vaskonopl@mail.ru

Научный руководитель — Майоров Петр Александрович

Введение. Точные расчеты многомерных потоков крови в сосудах необходимы для решения медицинских и физиологических задач. Особенно важно моделирование кровотока в отдельном сосуде, учитывая взаимодействие общей и локальной гемодинамики. Основные сложности включают разработку эффективного численного алгоритма для многомерных уравнений, создание моделей для замкнутой системы кровообращения и объединение многомерной модели с квазиодномерной. В данной выпускной квалификационной работе решена первая проблема: выведена система уравнений для жидкости в сосуде с подвижными стенками и разработан эффективный численный алгоритм расчета кровотока с использованием схемы КАБАРЕ.

Задачи. В рамках работы требовалось решить следующие задачи:

- Вывести систему дифференциальных уравнений, описывающих динамику жидкости в отдельном сосуде в смешанных эйлерово-лагранжевых (СЭЛ) координатах;
- Разработать численный алгоритм, предназначенный для расчета течения крови в объемном сосуде;
- Протестировать полученный алгоритм;
- Проанализировать результаты тестов.

Система дифференциальных уравнений. В работе из законов сохранения (массы и импульса) была выведена система дифференциальных уравнений, описывающая трёхмерное осесимметричное течение слабосжимаемой вязкой баротропной жидкости в цилиндрических смешанных эйлерово-лагранжевых (СЭЛ) координатах, безвихревое вдоль направления течения. Полученная система уравнений учитывает подвижность стенок рассматриваемой области.

Численная схема. Численная схема построена по методике КАБАРЕ [1] и состоит из трех фаз:

Фаза 1 и фаза 3 - балансные, в них используется метод конечного объёма. В этих фазах вычисляются консервативные переменные на промежуточном слое по времени $t_{n+1/2}$ и на новом слое по времени t_{n+1} , соответственно, а также происходит учёт вязкости;

Фаза 2 - характеристическая, в ней происходит экстраполяция локальных инвариантов Римана на новый слой по времени t_{n+1} . К новым инвариантам применяется процедура монотонизации [2], и по полученным инвариантам вычисляются потоковые переменные на новом слое по времени.

Тесты. Полученный алгоритм был протестирован на двух тестах:

Классический тест о течении Пуазейля. Результат был сопоставлен с аналитическим решением, максимальная погрешность составила $1.6 \cdot 10^{-4}$, практически точное совпадение. Также на этом тесте была получена оценка порядка точности алгоритма: согласно методу, описанному в [3] метод имеет второй порядок точности.

Тест канала с подвижными стенками. В рамках этого теста была продемонстрирована работоспособность вычислительного алгоритма при подвижных границах. Результаты подтвердили сохранение полной массы и импульса жидкости, а также свойство малой диссипативности разработанной схемы.

Результаты. В рамках выпускной квалификационной работы реализовано:

- Была выведена система дифференциальных уравнений, описывающих динамику жидкости в отдельном сосуде с подвижными стенками в цилиндрических координатах с учетом осевой симметрии в смешанных эйлерово-лагранжевых (СЭЛ) координатах;
- Для данной системы дифференциальных уравнений была разработана балансно-характеристическая схема, основанная на методике КАБАРЕ;

- Был разработан численный алгоритм, предназначенный для расчета течения крови в объемном сосуде;
- В работе также приведены результаты расчетов тестовых задач, демонстрирующие работоспособность разработанного алгоритма, эффективность и точность предложенного численного метода, а также малая диссипативность и консервативность построенной схемы;
- В ходе расчётов был подтверждён второй порядок точности численного метода.

ЛИТЕРАТУРА

- [1] Головизнин В. М. и др. Новые алгоритмы вычислительной гидродинамики для многопроцессорных вычислительных комплексов. – 2013.
- [2] Головизнин В. М., Карабасов С. А. Нелинейная коррекция схемы Кабаре // Математическое моделирование. – 1998. – Т. 10. – №. 12. – С. 107-123.
- [3] Brock J., Kamm J. Evaluating Convergence Rates for Calculation-Verification Analyses // 46th AIAA Aerospace Sciences Meeting and Exhibit. – 2008. – P. 728.

Разработка системы управления сбором данных из социальных сетей

Русанов Антон Евгеньевич

Кафедра системного программирования

e-mail: antonrusan@ispras.ru

Научный руководитель — к.ф.-м.н. доц. Корухова Людмила Сергеевна

Научный консультант — Лазарев Владимир Александрович, Яцков Александр Константинович

Современные методы машинного обучения требуют больших наборов данных из различных источников. Специфика задачи сбора данных из социальной сети заключается в том, что ее решение требует постоянного мониторинга заданного объекта и связанных с ним сущностей. Такие задачи на сбор данных неравномерно распределены во времени и требуют разное число отправляемых запросов. Для сбора данных из сети Интернет используются специальные программы - сборщики.

Организацией работы сборщиков занимаются системы управления сбором данных. Одной из таких систем является система Talisman [1]. В этой системе посредством обращения в сервисы Scrapy и Crawlk8s создаётся запуск сборщика для каждой задачи сбора данных. Данный процесс в обоих случаях включает в себя несколько этапов, что в совокупности занимает значительное количество времени. Работа посвящена разработке системы управления сбором данных, реализующей легкие задачи сбора, где неделимой единицей сбора является один запрос к веб-ресурсу. Такая система реализуется на основе передачи небольших сообщений постоянно работающему сборщику.

Были рассмотрены существующие системы управления сбором данных: Scrapy Cluster, Frontera, Scrapy Redis, Talisman [1, 2]. Исходя из анализа их

преимуществ и недостатков сделан вывод, что данные системы не подходят для решения поставленной задачи, так как не могут обрабатывать нестандартные заявки на сбор, а также поддерживают ограниченный набор технологий написания сборщика. Для передачи сообщений между сборщиками и системой управления сбором данных используется такой компонент, как брокер сообщений. В работе были рассмотрены следующие брокеры сообщений: RabbitMQ, Kafka и Redis [3]. При построении системы был выбран брокер RabbitMQ, так как он обладает гибкой маршрутизацией, встроенным приоритетом сообщений и функционалом подтверждения обработки сообщений.

Основными компонентами системы являются: программный интерфейс Frontier-api, принимающий пользовательские GraphQL запросы, и кластер системы контейнерной оркестрации Kubernetes, в котором развёрнуты: брокер сообщений RabbitMQ, база данных, реестр образов и реплики загруженных сборщиков.

Когда пользователь добавляет новый сборщик в систему с помощью отправки запроса в программном интерфейсе Frontier-api, то архив со сборщиком загружается в систему. Происходит сохранение информации в базу данных, а также сборка и отправка образа сборщика в реестр образов. Затем внутри кластера Kubernetes производится развертывание необходимого числа экземпляров нового сборщика.

В случае добавления новой заявки на сбор в системе происходит создание соответствующего запроса на сбор и отправка сообщения с ним в брокер, откуда оно будет прочитано из очереди нужным сборщиком. Затем сообщение с результатом сбора будет передано сборщиком через брокер обратно в систему.

Для того чтобы сборщик мог быть добавлен в систему, была написана спецификация для загружаемых сборщиков. Она определяет, какие файлы поставляются вместе со сборщиком и аргументы, которые принимает программа для работы с брокером сообщений.

Реализованная система управления сбором данных предоставляет следующую функциональность:

- возможность загрузить сборщик, написанный на любой технологии согласно спецификации;
- горизонтальная масштабируемость реплик сборщика в зависимости от числа заявок на сбор;
- изолированность библиотек сборщиков друг от друга;
- функциональность создания заявок на сбор данных с приоритетом;
- возможность обработки вложенных веб-запросов;
- мониторинг выполнения заявок и запросов на сбор.

Для тестирования полученной системы были реализованы тестовые сборщики средствами языков Python и Java согласно написанной спецификации. Было проведено сравнение производительности разработанной системы

управления сбором данных Frontier-api и системы Talisman. Оно показало, что много небольших запросов на сбор эффективно распределяются между экземплярами сборщиков в Frontier-api, тем самым уменьшая время обработки задачи на сбор.

ЛИТЕРАТУРА

- [1] Лазарев В. А., Варламов М. И., Яцков А. К. Система управления заданиями автоматизированного сбора данных из сети Интернет // Труды Института системного программирования РАН. 2022. Т. 34, № 2. С. 111–122.
- [2] Эйзенах Д.С., Рамейков А.С., Никифоров И.В. Эйзенах Д.С., Рамейков А.С., Никифоров И.В. Высокопроизводительный распределенный веб-скрапер. // Труды Института системного программирования РАН. 2021. Т. 33, № 3. С. 87–100.
- [3] Sharvari T., SowmyaNag K. A study on Modern Messaging Systems - Kafka, RabbitMQ and NATS Streaming // ArXiv. 08.12.2019.

Исследование и разработка генераторов псевдослучайных чисел на основе моделей динамического хаоса

Работа удостоена диплома 1 степени

Шеханов Алексей Сергеевич

Кафедра алгоритмических языков

e-mail: alexeysh2002@yandex.ru

Научный руководитель — к.ф.-м.н. доц. Полякова Ирина Николаевна

Генераторы псевдослучайных чисел (ГПСЧ) – это алгоритмы, которые производят последовательность чисел, которые выглядят случайными, но на самом деле являются детерминированными и могут быть воспроизведены.

Разработка генераторов псевдослучайных чисел является актуальной задачей на сегодняшний день, поскольку ГПСЧ используются повсюду, от научных исследований, до игровой индустрии.

Существует множество алгоритмов для генерации псевдослучайных чисел, каждый из которых имеет свои достоинства и недостатки. В одних случаях важна высокая криптографическая стойкость, в других – максимальная скорость генерации чисел, поэтому важно тщательно выбирать ГПСЧ для каждой конкретной задачи. Цель данной работы – разработка генератора, обеспечивающего наибольшую степень случайности.

В рамках работы рассматриваются два метода построения генераторов псевдослучайных чисел.

Первый - с помощью логистического отображения, математическая формула которого выглядит следующим образом:

$$x_{n+1} = 4 * r * x_n * (x_n - 1)$$

И второй - на основе теоремы Вейля:

Обозначим через $\{x\}$ дробную часть числа x . Если ξ — иррациональное число, то последовательность $\{\xi\}, \{2\xi\}, \{3\xi\}, \dots, \{n\xi\}, \dots$ равномерно распределена в $(0; 1)$. [1]

В качестве ξ рассматриваются иррациональные числа, такие как $\sqrt{2}$, π , e , $\ln(2)$, для получения наибольшей степени случайности.

Также предлагается метод тестирования и сравнения ГПСЧ на основе энтропии. [2]

Предложенные генераторы псевдослучайных чисел сравниваются с линейным конгруэнтным генератором и вихрем Мерсенна. Все они тестируются с помощью теста на основе энтропии и гистограммы распределения. Результаты визуализируются на графиках и диаграммах.

Авторы выражают благодарность Ульянову М. В. за постановку задачи и участие в получении результатов.

ЛИТЕРАТУРА

- [1] И. Циглер, Г. Хельмберг, Новейшее развитие теории равномерного распределения, Математика – 1963. № 3. - Т. 7. - С. 3–46.
- [2] Сметанин Ю.Г., Ульянов М.В., Пестова А.С. Энтропийный подход к построению меры символьного разнообразия слов и его применение к кластеризации геномов растений // Математическая биология и биоинформатика — 2016, № 1 — Т. 11 — С. 114-126

Моделирование поверхностных квантовых эффектов в плазмонных наноструктурах методом дискретных

ИСТОЧНИКОВ

Работа удостоена диплома 2 степени

Арьянова Анна

Кафедра математической физики

e-mail: annaaryanova@gmail.com

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

Поверхностные плазмоны (ПП) представляют собой коллективные колебания свободных электронов на границе раздела внешней среды и металла, как правило, возбуждаемые светом видимого и инфракрасного спектра. Благодаря явлению плазмонного резонанса, позволяющему получать значительное усиление рассеянного поля по сравнению с первичным, поверхностные плазмоны нашли широкое практическое применение и изучение их свойств является актуальной задачей. В случае локализованных ПП, возникающих в металлических наночастицах, наблюдаются поверхностные квантовые эффекты, не учитываемые в классической теории Максвелла [1] и оказывающие влияние на характеристики резонанса.

В работе рассмотрена задача дифракции плоской р-поляризованной электромагнитной волны на бесконечном плазмонном цилиндре кругового сечения. Для описания квантовых эффектов использована модель мезоскопических граничных условий с параметрами Фейбельмана, предложенная в [2].

Идея этого подхода состоит в том, чтобы дополнить классические граничные условия с помощью функций поверхностного отклика, называемых параметрами Фейбельмана и определяемых экспериментально для различных материалов. Классические граничные условия для рассматриваемой задачи — условия непрерывности касательных компонент полей на поверхности цилиндра — имеют вид:

$$\begin{aligned}\vec{n} \times [\vec{H}_0 + \vec{H}_e] &= \vec{n} \times \vec{H}_i, \\ \vec{n} \times [\vec{E}_0 + \vec{E}_e] &= \vec{n} \times \vec{E}_i,\end{aligned}\tag{1}$$

в то время как мезоскопические граничные условия ставятся в следующей форме [3]:

$$\begin{aligned}\vec{n} \times [\vec{H}_0 + \vec{H}_e] &= \vec{n} \times \vec{H}_i, \\ \vec{n} \times [\vec{E}_i - \vec{E}_0 - \vec{E}_e] &= -d_{\perp} \vec{n} \times \nabla \{ \vec{n} \cdot [\vec{E}_i - \vec{E}_0 - \vec{E}_e] \},\end{aligned}\tag{2}$$

где d_{\perp} — параметр Фейбельмана, \vec{n} — внешняя нормаль к поверхности цилиндра, \vec{E}_0, \vec{H}_0 — поле падающей волны, \vec{E}_i, \vec{H}_i — поле внутри цилиндра, \vec{E}_e, \vec{H}_e — поле рассеянной волны.

Значение параметра d_{\perp} является комплексной величиной, вещественная часть которой соответствует центру масс электронного облака индуцированных зарядов, а мнимая описывает затухание Ландау.

Для численного исследования влияния квантовых эффектов был выбран метод дискретных источников (МДИ) — строгий численно-аналитический метод, основанный на представлении неизвестных полей в виде конечной линейной комбинации полей элементарных источников [4]. Для верификации результатов МДИ использовался метод собственных функций, позволяющий получить аналитическое решение задачи дифракции на цилиндре и заключающийся в представлении неизвестных полей в виде разложений в бесконечный ряд по собственным функциям.

Оба метода дают разложения, удовлетворяющие системе уравнений Максвелла и условиям излучения на бесконечности. Выполнение граничных условий (1) или (2) достигается за счет вычисления неизвестных коэффициентов данных разложений. Сравнительный анализ показал полное соответствие результатов, полученных с помощью МДИ и с помощью метода разложения по собственным функциям.

Реализован расчет сечения рассеяния для различных длин волн падающего излучения и выполнен анализ влияния таких параметров задачи как внешняя среда, радиус частицы и материал частицы на амплитуду и положение плазмонного резонанса.

Сравнение результатов моделирования, полученных с использованием классических и мезоскопических граничных условий, показало, что учет квантовых эффектов приводит к смещению положения резонанса в сторону коротких длин волн, при этом резонансный пик становится более плавным и уменьшается его амплитуда.

ЛИТЕРАТУРА

- [1] Еремин Ю. А., Свешников А. Г. Влияние асимметрии геометрии слоистой частицы на подложке на оптические характеристики с учетом пространственной дисперсии // Вестн. Моск. Ун-та. Физ. Астрон. 2020. № 5. С. 90–96.
- [2] A general theoretical and experimental framework for nanoscale electromagnetism / Y. Yang, D. Zhu, W. Yan, A. Agarwal, M. Zheng, J.D. Joannopoulos, P. Lalanne, T. Christensen, K.K. Berggren, M. Soljačić//Nature. 2019. N 576 (7786). P. 248–252.
- [3] Еремин Ю. А., Лопушенко В. В. Влияние поверхностных квантовых эффектов на оптические характеристики пары плазмонных наночастиц // Оптика и спектроскопия. 2023. Т. 131. № 8.
- [4] Еремин Ю. А., Свешников А. Г. Математические модели задач нанооптики и биофотоники на основе метода дискретных источников // Журнал вычисл. математики и мат. физики. 2007. Т. 47. № 2. С. 269–287.

Полностью неявный метод решения уравнения переноса на основе схем WENO

*Работа удостоена диплома 1 степени
Мартиросян Арутюн Артушович*

Кафедра ВВО

e-mail: marutyun@bk.ru

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

Одним из лучших методов разрешения по пространству для систем дифференциальных уравнений гиперболического типа является дискретизация по пространству с использованием схемы WENO (weighted essentially non-oscillatory) [1] для уравнений газовой динамики. В настоящей работе рассматривается важный вопрос применимости данных схем к простейшим модельным гиперболическим уравнениям в частных производных при использовании неявной пространственной дискретизации. Данный вопрос слабо освещен в литературе и представлен лишь работами [1, 2, 3], в которых коэффициенты схем WENO предварительно замораживались, что не позволяло поднять длину шага по времени выше такой, что число Куранта не превосходило 1.

В работе представлены решения уравнения переноса и уравнения Хопфа полностью неявными методами Рунге—Кутты (без замораживания коэффициентов) с использованием аппроксимации по пространству схемами WENO 3-го и 5-го порядков. То есть в данном случае на каждой итерации метода Ньютона рассматривается неявность и по коэффициентам схемы WENO, и по оператору потока.

Для нахождения потоков на границах ячеек решалась задача Римана о распаде произвольно разрыва (Риман-солвер). В работе представлены два Риман-солвера: UPWIND и Lax—Friedrich.

Среди неявных методов Рунге—Кутты выбраны 16 методов разного порядка аппроксимации по времени семейств Lobatto, Radau, Gauss, SDIRK. Данные методы представлены в виде таблиц Бутчера. Рассматриваемые методы можно разделить на две основные группы: IRK (Implicit Runge—Kutta) и SDIRK (Single Diagonally Implicit Runge—Kutta). Каждый метод был исследован на А и В-устойчивости.

Для решения СЛАУ (систем линейных алгебраических уравнений) разработан алгоритм метода Ньютона для IRK и SDIRK методов Рунге—Кутты. Основным отличием в составлении СЛАУ между IRK и SDIRK методами является то, что для IRK методов решаются блочные СЛАУ на каждой итерации метода Ньютона. Для повышения области сходимости метода Ньютона введен фактор глобализации.

При численном исследовании проводилось сравнение с аналитическим решением каждого уравнения для заданного начального условия. Аналитическое решение уравнений получено методом характеристик. Всего исследовано по два начальных условия для каждого уравнения. В ходе численного исследования варьировались числа CFL (Courant—Friedrich—Lewy) и время моделирования. Для оценки эффективности проводилась оценка точности каждого из 16 методов по относительной ошибке по норме L_2 и длительности расчета.

Разработан универсальный алгоритм тестирования произвольных методов Рунге—Кутты, задаваемых на основе таблиц Бутчера. В дальнейшем данный алгоритм может быть применен к другим видам пространственной дискретизации. В ходе исследования получено быстрое действие SDIRK методов относительно других рассматриваемых методов, возможность получить высокий порядок аппроксимации и приемлемую точность на монотонных и немонотонных участках. Не отмечено сильного влияния свойств А и В-устойчивости методов Рунге—Кутты в рассматриваемых задачах. Таким образом наиболее логичным является применение методов типа SDIRK совместно с WENO-пространственной аппроксимацией.

При рассмотрении образования ударной волны для уравнения Хопфа замечено расхождение всех схем, начиная с некоторого значения CFL. Данный факт требует дополнительного исследования.

В целом стоит рекомендовать рассмотренный подход для гиперболических законов сохранения (схемы WENO + неявные методы Рунге—Кутты) только в областях простой геометрии при умеренных числах CFL или при поиске стационарного решения методом установления. Численные расчеты задач, предполагающих нестационарные решения, при больших значениях CFL могут приводить к образованию нефизических осцилляций и расхождению всего процесса расчета. С учетом вычислительных затрат и применения неявных схем в данных ситуациях рекомендуется использовать динамическое измельчение сетки с применением менее сложных схем пространственной аппроксимации.

ЛИТЕРАТУРА

- [1] Liu X.-D., Osher S., Chan T. Weighted Essentially Non-Oscillatory Schemes // Journal of Computational Physics, 1994. Vol. 115, P. 200-212.
- [2] Евстигнеев Н. М. О построении и свойствах WENO-схем пятого, седьмого, девятого, одиннадцатого и тринадцатого порядков. Часть 1. Построение и устойчивость // Компьютерные исследования и моделирование 2016. Т. 8, № 5 С. 721–753
- [3] Евстигнеев Н. М. О построении и свойствах WENO-схем пятого, седьмого, девятого, одиннадцатого и тринадцатого порядков. Часть 2. Численные примеры // Компьютерные исследования и моделирование 2016. Т. 8, № 5 С. 885–910

Алгоритмы обучения федеративных классификаторов

Работа удостоена диплома 3 степени

Селезнева Алиса Евгеньевна

Кафедра автоматизации систем вычислительных комплексов

e-mail: alice.in.moskow@gmail.com

Научный руководитель — к.ф.-м.н. доц. Костенко Валерий Алексеевич

Федеративное обучение — это модель обучения с сохранением конфиденциальности в распределенных сетях. Такое обучение позволяет создать общую модель прогнозирования, сохраняя все обученные данные независимо на обучаемых устройствах [1]. Локальным устройством будем называть любой компьютер, на котором хранится достаточное для обучения количество данных и на котором возможен запуск алгоритма обучения и передача локальной модели по какому-либо виду связи.

Центральный сервер — компьютер, который оркестрирует обучение на локальных устройствах и содержит главный алгоритм обучения. Центральный (глобальный) классификатор - классификатор, работающий на сервере. Локальный классификатор - это алгоритм обучения одного из локальных устройств.

Транзакцией или банковской операцией будем называть любой совершенный клиентом платеж.

Федеративное обучение необходимо на практике, например, в банковском секторе, когда необходимо передавать данные обучения, полученные в разных отделениях банка, городах или странах, чтобы иметь одну большую общую базу. Раздельное обучение имеет свои уязвимости в безопасности. Злоумышленник может захватить один или несколько локальных классификаторов с целью нарушить функционирование глобального классификатора. Поэтому при использовании раздельного обучения стоит учитывать возможность атак на глобальный классификатор [3].

В работе рассматривается решение задачи бинарной классификации - определение типа операций: является ли банковская транзакция обычной или мошеннической, оценивается устойчивость федеративного классификатора к

захвату локальных классификаторов. Алгоритмы логистической регрессии и нейронной сети прямого распространения взяты для обучения на локальных устройствах. Алгоритм голосования по большинству взят как алгоритм для центрального сервера.

Датасет для обучения был взят с сайта kagle - Credit Fraud Detection [2] - это набор данных, содержащий информацию о банковских транзакциях. Набор данных содержит информацию о транзакциях - 8 количественных и булевских параметров. Датасет предварительно обработан, классы были сбалансированы за счет понижения значений доминирующего класса, случайным отбрасыванием значений выборки из доминирующего класса.

Метрики качества обучения: точность классификации (**recall**), основанная на количественном соотношении правильных и неправильных предсказаний и полнота (**precision**) - соотношение позитивных и негативных. Функции нахождения ошибок были взяты как среднеквадратичная ошибка и квадратичная ошибка.

В таблице 1 представлены лучшие результаты работы алгоритмов, где alg - алгоритм (lg - логистическая регрессия, n - нейросеть прямого распространения), LR - шаг градиента, для нейронной сети указано количество нейронов в 1 и 2 скрытом слое (L1, L2). Нейронная сеть прямого распространения получила большую точность, чем логистическая регрессия.

L1	L2	alg	LR	Accurancy	mse	precision	recall
-	-	lg	-	0.940910	0.059090	0.935200	0.948080
10	-	n	0.01	0.99469	0.005310	0.99098	0.998510
10	-	n	0.001	0.98561	0.014387	0.97749	0.994220
10	-	n	0.0001	0.94573	0.054275	0.92388	0.971840
10	-	n	0.00001	0.92355	0.076449	0.91301	0.936850
9	8	n	0.01	0.997500	0.002495	0.995760	0.999270
7	2	n	0.001	0.991370	0.008632	0.986430	0.996400
4	6	n	0.0001	0.947000	0.052996	0.923870	0.974380
10	4	n	0.00001	0.506370	0.493628	0.832610	0.210740

Табл. 1: Лучшие результаты обучения алгоритмов.

Федеративное обучение происходило с разным количеством локальных устройств и захваченных локальных устройств. Сначала федеративным классификатором классифицировались полностью исправные данные от локальных классификаторов. По одному происходил захват локальных устройств. Каждый раз, после захвата нового устройства происходила новая классификация классификатором, пока все устройства не были захвачены. Для каждого случая проходил полный алгоритм обучения. В результате для различного количества локальных устройств (10, 15 и 20) - злоумышленнику возможно захватить центральный классификатор и снизить точность, только если он контролирует более 50% локальных устройств.

По итогам выполненной работы можно заключить, что задача бинарной классификации решается как алгоритмом логистической регрессии, так и с помощью нейросетей прямого распространения с почти одинаковой точностью и что центральный классификатор устойчив к атакам при захвате до 50% локальных классификаторов.

ЛИТЕРАТУРА

- [1] Federated Learning: Collaborative Machine Learning without Centralized Training Data. [Электронный ресурс] // URL: <https://ai.googleblog.com/2017/04/federated-learning-collaborative.html>.
- [2] Credit Card Fraud Detection. [Электронный ресурс] // URL: <https://www.kaggle.com/datasets/dhanushnarayananr/credit-card-fraud>.
- [3] Костенко В.А., Селезнева А.Е. Виды атак на федеративные нейросети и способы защиты // Труды Института системного программирования РАН, том 36, вып. 1, 2024, стр. 35-44. DOI: 10.15514/ISPRAS-2024-36(1)-3.

Анализ возможностей инструментов статического анализа исходного кода для поиска недостатков класса инъекция с помощью генерации примеров уязвимого кода

Работа удостоена диплома 2 степени

Зубрилин Антон Николаевич

Кафедра информационной безопасности

e-mail: s02200373@gse.cs.msu.ru

Научный руководитель — Самосадный Кирилл Алексеевич

Согласно рейтингу OWASP Top 10 наличие недостатков класса «инъекция» является одним из наиболее критичных рисков в современных веб-приложениях. Для поиска и предотвращения возникновения таких недостатков могут использоваться автоматизированные инструменты статического анализа кода. Для правильного применения этих инструментов важно уметь оценивать их возможности по поиску инъекций.

В рамках данной работы исследуется проблема оценки возможностей инструментов статического анализа по поиску инъекций в контексте модели недостатка, основанной на анализе помеченных данных. Пусть есть программа, и в ней определены следующие множества ее операторов: истоки, соответствующие источникам недоверенных данных, и стоки, соответствующие критичным узлам программы, например, обращения во внешние подсистемы. Кроме того, для каждого стока определено множество санитизаций — преобразований данных, делающих их безопасными для использования в данном стоке. Тогда модель инъекции — такая зависимость по данным стока от истока, которая не проходит ни через одну соответствующую санитизацию.

Для этой модели были определены желаемые свойства методов оценки статических анализаторов, обнаруживающих недостатки класса «инъекция»:

1. Возможность оценки множества истоков и стоков
2. Возможность оценки множеств санитизаций
3. Возможность объяснения ошибок инструмента при построении зависимостей по данным

В рамках исследования проведен обзор существующих методов оценки инструментов статического анализа для поиска инъекций [1, 2, 3]. В результате анализа этих методов установлено, что среди них есть метод, обладающий свойствами 1 и 2, но ни один из них не обладает свойством 3. Все рассмотренные методы основаны на запуске оцениваемого инструмента на некоторой тестовой выборке, которая может состоять как из синтетических примеров, так и из примеров реальных приложений, содержащих известные уязвимости.

В данной работе предлагается новый метод оценки возможностей статических анализаторов для поиска инъекций, обладающий свойством 3. Метод основан на генерации тестовых примеров особой структуры, отличающихся друг от друга отдельными составными частями, и сравнении поведения анализатора на этих примерах. Множества возможных составных частей задаются заранее, исходя из знаний о синтаксисе и семантике языка и о стандартной библиотеке. Сначала с помощью перебора строится минимальный, в котором инструмент находит недостаток. Затем в этот пример вводится зависимость по данным через исследуемую конструкцию языка. Изменения поведения анализатора на модифицированных таким образом примерах позволяет делать выводы о его возможностях по моделированию зависимостей по данным через различные конструкции языка.

Предложенный метод реализован для языка программирования PHP и протестирован на 7 современных инструментах статического анализа, предназначенных для поиска инъекций в коде программ на языке PHP: Semgrep, Progpilot, Psalm, Checkmarx, AppScreeener, Fortify и DeepTective. Результаты эксперимента показывают, что метод позволяет оценивать корректность моделирования инструментом зависимостей по данным на уровне отдельных синтаксических конструкций языка, учитывая полноту и точность этого моделирования. Кроме того, было выявлено, что в текущей реализации метод не позволяет работать с анализаторами, использующими машинное обучение.

Было выделено 3 направления дальнейших исследований: построение полного набора синтаксических конструкций языка программирования, адаптация предложенного метода под инструменты, использующие машинное обучение, и исследование зависимости между предсказываемыми возможностями инструмента и вероятностью ложных срабатываний инструмента при его промышленном применении.

ЛИТЕРАТУРА

- [1] Goseva-Popstojanova K., Perhinschi A. On the capability of static code analysis to detect security vulnerabilities // Information and Software Technology. 2015. Vol. 68. P. 18–33.

-
- [2] Benchmarking Static Analysis Tools for Web Security / P. Nunes, I. Medeiros, J. C. Fonseca, N. Neves, M. Correia, M. Vieira // IEEE Transactions on Reliability. 2018. Vol. 67. N. 3. P. 1159–1175.
 - [3] Schuckert F., Langweg H., Katt B. Systematic Generation of XSS and SQLi Vulnerabilities in PHP as Test Cases for Static Code Analysis // 2022 IEEE International Conference on Software Testing, Verification and Validation Workshops (ICSTW). 2022. P. 261-268.