

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

Факультет вычислительной математики и кибернетики

**СБОРНИК ТЕЗИСОВ
ЛУЧШИХ ВЫПУСКНЫХ
КВАЛИФИКАЦИОННЫХ РАБОТ
ФАКУЛЬТЕТА ВМК МГУ 2021 ГОДА**

МОСКВА

2021

Оглавление

Кафедра математической физики	9
<i>Д. С. Андрианов</i>	
Обратная задача для гиперболического уравнения с неизвестным источником	9
<i>О. П. Бучацкая</i>	
Метод анализа субъективного качества изображений при потере высокочастотной информации	11
<i>У. Ю. Вострикова</i>	
Исследование обратной задачи оптического секционирования	13
<i>О. А. Гаврилов</i>	
Метод решения задачи теплопроводности со специальным нелокальным условием	14
<i>Д. В. Гаврилова</i>	
Нейросетевой метод определения уровней артефактов на изображениях при потере высокочастотной информации	16
<i>А. М. Корнеева</i>	
Применение рекуррентных нейронных сетей для оценки динамики роста растений	18
<i>Д. В. Малов</i>	
Обратная задача частотного зондирования градиентно-слоистых сред	19
<i>В. А. Пятов</i>	
Нейросетевой метод совмещения биомедицинских изображений . . .	21
Кафедра вычислительных технологий и моделирования	23
<i>П. С. Чижов</i>	
Применение технологии глубокого обучения для прогнозирования заболеваемости	23
Кафедра вычислительных методов	25
<i>Д. В. Бочковенкова</i>	
Решение краевой задачи для уравнения Пуассона в прямоугольнике с помощью нейронных сетей	25

<i>А. В. Василенко</i>	
Математическое распределение тяжелых газов и жидкостей на орографически неоднородной поверхности	27
<i>О. К. Горбунова</i>	
Схема КАБАРЕ с улучшенными дисперсионными свойствами	28
Кафедра автоматизации научных исследований	29
<i>М. Б. Измайлов</i>	
Практическое применение вейвлет-преобразований для анализа и обработки данных	29
<i>В. Я. Кот</i>	
Применение методов машинного обучения для улучшения качества речевого сигнала	31
<i>И. И. Лысенко</i>	
Распознавание солирующего инструмента в полифонических музыкальных произведениях	33
Кафедра общей математики	34
<i>М. Л. Куркин</i>	
Оптимизация параметров численного метода для модели экологических сообществ	34
Кафедра исследования операций	38
<i>А. С. Селиванов</i>	
Исследование эффективности алгоритмов решения задачи о оптимальном разбиении графа	39
Кафедра оптимального управления	40
<i>А. С. Гуленко</i>	
Вариационный метод и асимптотические разложения в теории среднего поля Стонера	40
Кафедра нелинейных динамических систем и процессов управления	42
<i>Д. С. Зайцева</i>	
Разработка алгоритмов управления для нелинейной модели сердечно-сосудистой системы	42
Кафедра системного анализа	44
<i>Е. Д. Акимова</i>	
Сетевые модели экономического роста	44
<i>И. Б. Воробьева</i>	
Моделирование нефтяной отрасли на основе концепции игры среднего поля	46

<i>П. В. Карпикова</i>	
Оценка параметров модели Самуэльсона на основании дневной статистики торгов	47
<i>К. И. Салихова</i>	
Решение нелинейных уравнений в окрестности аномальной точки .	48
<i>Д. М. Сотников</i>	
Близость моделей Башелье и Самуэльсона для различных метрик . .	50
<i>И. Р. Удовиченко</i>	
Гамильтонов формализм для синтеза групповых наблюдений	51
<i>Н. А. Федяшин</i>	
Оптимальная политика торговли разрешениями на выбросы в динамической игре с трансграничным загрязнением	52
Кафедра математической статистики	54
<i>М. Г. Логванёва</i>	
Критерий диверсифицируемости в терминах ожидаемых средних потерь	54
<i>А. В. Максимова</i>	
Уточнение константы в неравенствах типа Эссеена-Розовского для случая независимых одинаково распределённых случайных слагаемых	56
<i>М. Б. Мухтаров</i>	
Методы прогнозирования временных рядов на примере прогнозирования выручки частной медицинской компании	57
Кафедра математических методов прогнозирования	59
<i>М. В. Мишустина</i>	
Учет глубины при стилизации изображений и автоматическая оценка качества стилизации	59
<i>А. С. Травникова</i>	
Методы сравнения и подгонки сканированных поверхностей	61
Кафедра математической кибернетики	62
<i>Аурангабадкар Удитангшу</i>	
О тестах относительно замен функциональных элементов	63
<i>М. А. Местецкий</i>	
Генетический алгоритм для поиска оптимальных и близких к ним схем от малого числа переменных	64
Кафедра интеллектуальных информационных технологий	66
<i>А. В. Березникер</i>	
Применение классических и нейросетевых методов построения модели пользователя в задаче динамической аутентификации пользователей на основе анализа работы с компьютерной мышью	66

<i>К. В. Малышев</i>	Нейросетевое вычисление карт доверия для методов оценки оптического потока в видео	67
<i>Т. З. Мамедов</i>	Усовершенствование алгоритма реидентификации для повышения качества видеоаналитики	69
<i>А. В. Москаленко</i>	Разработка нейросетевого метода заполнения областей изображений высокого разрешения	71
<i>А. В. Николаев</i>	Метод сжатия 3D моделей для параллельной обработки на GPU . . .	73
<i>А. В. Полевой</i>	Исследование алгоритмов инкрементального моделирования	75
Кафедра автоматизации систем вычислительных комплексов		76
<i>Д. И. Кукушкин</i>	Разработка прототипа планировщика для бессерверной платформы управления инфраструктурой	76
<i>А. А. Морквин</i>	Построение коммутируемых сетей для систем реального времени . . .	78
<i>Н. И. Никифоров</i>	Исследование применимости алгоритмов сжатия данных к таблицам классификации в сетевом процессорном устройстве	79
<i>Л. Е. Селезнев</i>	Разработка методов построения неоднородных нейронных сетей . . .	81
<i>В. М. Рябченков</i>	Аналитический метод оценки времени отклика задач в системах интегрированной модульной авионики в случае интервальной неопределенности длительности выполнения работ	82
<i>В. А. Шапошников</i>	Метод управления процессами реконфигурации ПКС при условии её полной непротиворечивости для выполнения конкурирующих запросов построения маршрутов	84
Кафедра суперкомпьютеров и квантовой информатики		85
<i>И. М. Камалов</i>	Исследование методов решения задачи обучения по нескольким примерам, основанных на применении ансамбля алгоритмов формирования синтетических обучающих данных	85
<i>Д. И. Личманов</i>	Суперкомпьютерный кодизайн задачи симуляции жидкокристаллических структур для современных вычислительных архитектур	88

<i>Е. С. Питанов</i>	
Исследование метрического метода в задаче обучения по нескольким примерам	90
<i>Н. В. Сорокина</i>	
Использование технологии InfiniBand для передачи данных системы мониторинга суперкомпьютера	91
<i>М. А. Цуканова</i>	
Разработка и исследование неоднородных эволюционных алгоритмов оптимизации в задачах машинного обучения	93
<i>Н. Д. Шалагин</i>	
Разработка и исследование архитектур рекуррентных нейронных сетей в задачах построения игровых стратегий	94
Кафедра алгоритмических языков	95
<i>Д. А. Авагян</i>	
Система автоматического обнаружения фейковых новостей на русском языке	95
<i>Н. М. Арсенин</i>	
Алгоритмы восходящего синтаксического анализа на основе детерминированных бесконтекстных L-графов	97
<i>В. Д. Дубинин</i>	
Программное средство рекомендации фильмов на основе их сюжетов и оценок пользователей	98
<i>Д. Д. Изюмова</i>	
Программная поддержка словообразовательного анализа	100
<i>Е. Б. Рязанова</i>	
Алгоритмы поиска маршрутов в графе с меняющейся структурой и их реализация в мобильном приложении	101
Кафедра системного программирования	102
<i>А. О. Богатенкова</i>	
Извлечение иерархической логической структуры из текстовых документов в формате docx	103
<i>В. М. Джиджоев</i>	
Динамическая компиляция пользовательских функций на языке PL/pgSQL	105
<i>В. О. Макеев</i>	
Определение устранимых зависимостей по данным для распараллеливания программ в системе автоматизированного распараллеливания	106

<i>Н. А. Оплачко</i>	
Оптимизация вычислений нейронных сетей на графических процессорах для мобильных устройств	108
<i>С. А. Фомин</i>	
Использование обучения с подкреплением в задаче автоматического тестирования мобильных приложений	110
Именной указатель	113

Обратная задача для гиперболического уравнения с неизвестным источником

Андреанов Даниил Сергеевич

*Кафедра математической физики
e-mail: justdaniel99@gmail.com*

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

В данной работе исследуется обратная задача для гиперболического уравнения, в которой неизвестной является функция, входящая в источник.

Рассмотрим задачу Коши для гиперболического уравнения:

$$u_{tt} = a^2 u_{xx} + \varepsilon q(x)u(x, t) + f(x)g(t), \quad (x, t) \in \Delta_T, \quad (1)$$

$$u(x, 0) = 0, \quad -aT \leq x \leq aT, \quad (2)$$

$$u_t(x, 0) = 0, \quad -aT \leq x \leq aT, \quad (3)$$

где $\Delta_T = \{(x, t) : -a(T - t) \leq x \leq a(T - t), 0 \leq t \leq T\}$, функции $q(x) \in C^1[-aT, aT]$, $f(x) \in C^1[-aT, aT]$, $g(t) \in C[0, T]$; a и ε - положительные числа.

Сформулируем обратную задачу. Пусть функции $q(x)$ и $f(x)$, числа a и ε заданы, а функция $g(t)$ неизвестна. Требуется определить функции $g(t)$ и $u(x, t)$, если задана дополнительная информация о решении задачи Коши:

$$u(0, t) = p(t), \quad 0 \leq t \leq T. \quad (4)$$

Определение. Функции $g(t)$ и $u(x, t)$ называются решением обратной задачи, если $u(x, t) \in C^2(\Delta_T)$, $g(t) \in C[0, T]$, и $u(x, t)$, $g(t)$ удовлетворяют (1)-(4).

Если функции $g(t)$ и $u(x, t)$ являются решением обратной задачи, то функции $u(x, t)$, $v(x, t) = u_x(x, t)$, $g(t)$ являются решениями системы интегральных уравнений:

$$u(x, t) = \frac{1}{2a} \int_0^t \int_{x-a(t-\tau)}^{x+a(t-\tau)} \varepsilon q(s)u(s, \tau) ds d\tau + \frac{1}{2a} \int_0^t \int_{x-a(t-\tau)}^{x+a(t-\tau)} f(s)g(\tau) ds d\tau, \quad (x, t) \in \Delta_T; \quad (5)$$

$$v(x, t) = \frac{\varepsilon}{2a} \int_0^t [q(x+a(t-\tau))u(x+a(t-\tau), \tau) - q(x-a(t-\tau))u(x-a(t-\tau), \tau)] d\tau + \frac{1}{2a} \int_0^t g(\tau)[f(x+a(t-\tau)) - f(x-a(t-\tau))] d\tau, \quad (x, t) \in \Delta_T; \quad (6)$$

$$g(t) = -\frac{a\varepsilon}{2f(0)} \int_0^t [q'(a(t-\tau))u(a(t-\tau), \tau) + q(a(t-\tau))v(a(t-\tau), \tau) - q'(-a(t-\tau))u(-a(t-\tau), \tau) - q(-a(t-\tau))v(-a(t-\tau), \tau)] d\tau - \frac{a}{2f(0)} \int_0^t g(\tau)[f'(a(t-\tau)) - f'(-a(t-\tau))] d\tau + \frac{p''(t)}{f(0)} - \frac{\varepsilon q(0)p(t)}{f(0)}, \quad 0 \leq t \leq T. \quad (7)$$

Теорема 1. Пусть $f(x) \in C^1[-aT, aT]$, $q(x) \in C^1[-aT, aT]$, $p(t) \in C^2[0, T]$ и $f(0) \neq 0$. Тогда существуют функции $u(x, t)$, $v(x, t)$ и $g(t)$, такие, что $u(x, t)$, $v(x, t) \in C(\Delta_T)$, $g(t) \in C[0, T]$, и $u(x, t)$, $v(x, t)$ и $g(t)$ являются решением системы интегральных уравнений (5)-(7).

Теорема 2. Пусть $f(x) \in C^1[-aT, aT]$, $q(x) \in C^1[-aT, aT]$, $p(t) \in C^2[0, T]$, $f(0) \neq 0$ и $p(0) = p'(0) = 0$. Тогда если функции $u_i(x, t)$, $g_i(t)$, $i = 1, 2$, являются решениями обратной задачи, то $u_1(x, t) = u_2(x, t)$ для $(x, t) \in \Delta_T$ и $g_1(t) = g_2(t)$ для $t \in [0, T]$.

Численный метод решения обратной задачи состоит в реализации итерационного метода, основанного на системе интегральных уравнений (5)-(7) при помощи метода последовательных приближений. Для вычисления двумерных и одномерных интегралов в итерационных формулах использовались квадратурные формулы прямоугольников. В итоге были получены дискретные аналоги итерационных формул, по которым и проводились вычисления. Численный метод был программно реализован на языке Python.

Схема вычислительных экспериментов выглядит следующим образом. Задавались функции $q(x)$, $f(x)$, $g(t)$ и числа ε , a и T . С ними решалась задача Коши (1)-(3) и находились функции $u(x, t)$ и $p(t)$. Затем с заданными $q(x)$, $f(x)$, $p(t)$, ε , a , T решалась обратная задача, и полученная приближенная функция $g_n(t)$, где n - номер итерации, сравнивалась с точной функцией $g(t)$.

Эффективность предложенного численного метода продемонстрирована рядом вычислительных экспериментов, показавших достаточно быструю сходимость и хорошую точность этого итерационного метода.

Литература

- [1] Тихонов А. Н., Арсенин В. Я. Методы решения некорректных задач. - Москва: Наука, 1986 - 288 с.
- [2] Денисов А. М. Введение в теорию обратных задач. // М.: Изд-во МГУ, 1994. - Т. 206.
- [3] Кабанихин С. И. Обратные и некорректные задачи. - Новосибирск: Сибирское научное издательство, 2009 - 457 с.
- [4] Баев А. В. Об одном методе решения обратной задачи рассеяния для волнового уравнения // Журнал вычислительной математики и математической физики. - 1988. - Т. 28. - № 1. - С. 25-33.
- [5] Денисов А. М. Итерационный метод решения обратной задачи для гиперболического уравнения с малым параметром при старшей производной. // Дифференциальные уравнения. - 2019. - Т. 55. - № 7. - С. 973-981.

Метод анализа субъективного качества изображений при потере высокочастотной информации

Бучацкая Ольга Павловна

Кафедра математической физики

e-mail: buchatskaya_olga_112@mail.ru

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

Визуальная оценка качества имеет множество применений на практике, а также играет центральную роль для тестирования, настройки и мониторинга алгоритмов обработки изображений.

Целью выпускной квалификационной работы является создание на основе моделирования процесса потери высокочастотной информации новой базы изображений, включающей в себя как изображения с артефактами шума и размытия, так и эффект ложного оконтуривания. Это позволит впервые провести анализ восприятия всех три указанных искажений одновременно. Особенностью разработанной базы должно являться сочетание эффектов, то есть они должны рассматриваться не отдельно, а в комбинации. Каждый из артефактов определяется отдельным параметром. Так же необходимо выполнять наложение эффектов в следующей очередности: сначала эффект ложного оконтуривания, потом выполняется свертка с круговым ядром для дефокуса и в конце добавляется шум. Далее необходимо провести опрос пользователей для получения субъективных оценок качества изображений, используя в качестве материала для опроса сформированную базу.

В качестве референсных взяты 24 изображения 512x384 пикселя из базы Kodak [1].

Эффект ложного оконтуривания представляет собой осцилляции от контуров на изображениях. Он возникает из-за искажения или потери высокочастотной информации, что делает его достаточно часто встречающимся на определенных классах изображений. Основной математической моделью, описывающей данный артефакт, является эффект Гиббса [3]. Моделировать эффект ложного оконтуривания можно с помощью различных алгоритмов обработки изображений, в данной работе артефакт получен через свертку с двумерным фильтром Ланцоша.

Ядро Ланцоша [2] представляет собой произведение функции $\text{sinc}(x)$ на оконную функцию Ланцоша, по определению равную нулю вне заданного параметром носителя функции a :

$$L(x) = \begin{cases} \text{sinc}(x) \cdot \text{sinc}\left(\frac{x}{a}\right), & -a < x < a \\ 0, & x \in (-\infty, -a] \cup [a, +\infty) \end{cases}$$

Двумерное ядро Ланцоша получается с помощью умножения одномерных:

$$L(x, y) = L(x)L(y)$$

Артефакт размытия получен с помощью свертки с круговым ядром. Для борьбы со ступенчатостью контуров на изображении, появляющейся при при-

менении ядра, фильтр создается с большим разрешением, а потом оно понижается [4]. Зашумление было сформировано с помощью суммы изображения и нормально распределенного шума [5].

В данной работе для опроса пользователей была выбрана double-stimulus система с показом референсных изображений, то есть человеку на экране показывались три изображения: оригинал и два искаженных. Задача пользователя - выбрать картинку, наименее отличающуюся по качеству от референсной. Для поставленной задачи была реализована программа с веб-интерфейсом: серверная часть написана на языке программирования Java версии 1.8 с использованием Spring Boot Framework, для клиентской части использованы Html-5 и CSS-3 для разметки документов и JavaScript. Разработанный алгоритм преобразования оценок пользователей и ранжирования картинок состоит из двух шагов:

- Вычисление модернизированного рейтинга Эло.
- Для всей статистики проводилось обновление рейтинга: если картинка A лучше картинки B , то из рейтинга B вычитался рейтинг A .

В выпускной квалификационной работе описано создание базы изображений с тремя типами искажений. В качестве искажений были выбраны эффект ложного оконтуривания, дефокусировка и зашумление. Модели данных артефактов были реализованы на языке программирования Python 3. Для каждого эффекта задавалась регулярная и случайная сетка параметров. Всего было сгенерировано 6480 изображений, по 270 для каждого исходного изображения. Выполнено ранжирование результатов опросов, проведенных с помощью разработанной на Java программы с веб-интерфейсом.

Литература

- [1] Kodak database,
<http://r0k.us/graphics/kodak/>
- [2] Duchon C. E. Lanczos filtering in one and two dimensions //Journal of applied meteorology. – 1979. – Т. 18. – №. 8. – С. 1016-1022.
- [3] Насонов А. В., Крылов А. С., Динг Й. Алгоритмы вычисления областей на изображениях для анализа явления ложного оконтуривания //REDS: Телекоммуникационные устройства и системы. – 2017. – Т. 7. – №. 3. – С. 374-377.
- [4] Krylov A., Nasonov A., Pchelintsev Y. Single parameter post-processing method for image deblurring //2017 Seventh International Conference on Image Processing Theory, Tools and Applications (IPTA). – IEEE, 2017. – С. 1-6.
- [5] Гонсалес Р., Вудс Р. Цифровая обработка изображений. – М.: Техносфера, 2005.

Исследование обратной задачи оптического секционирования

Вострикова Ульяна Юрьевна

Кафедра математической физики

e-mail: vostrikova.u.y@gmail.com

Научный руководитель — к.ф.-м.н. асс. Романенко Татьяна Евгеньевна

Рассматривается задача восстановления трехмерной структуры различных отделов глазного дна человека. Один из методов получения изображений структуры глаза [1] основан на быстрой перефокусировке изображающей системы для получения стека изображений глазного дна, находящихся на различной глубине. Получающиеся при этом изображения в каждой фокальной плоскости представляют собой сумму как истинного сечения трехмерного объекта в данной фокальной плоскости, так и внефокусные размытые изображения соседних по глубине сечений. Кроме того, на истинное изображение в фокальной плоскости накладываются различные искажения: как связанные с оптической системой глаза, так и вносимые снимающей системой. Изображение искомого объекта описывается уравнением трехмерной свертки

$$i(x, y, z) = o(x, y, z) * h(x, y, z) \quad (1).$$

Здесь $i(x, y, z)$ — наблюдаемое изображение, $o(x, y, z)$ — искомый объект, $h(x, y, z)$ — трехмерная функция рассеяния точки. Задача секционирования состоит в восстановлении стека слоев искомого трехмерного объекта $o(x, y, z)$. Решение затрудняется как тем, что обратная задача в общем случае задача некорректная, так и тем, что вид функции рассеяния точки, учитывающей аберрации оптической системы, для экспериментальных данных точно неизвестен.

Для решения обратной задачи (1) с известным оператором свою эффективность показали методы регуляризации в Фурье-пространстве [2]. В данной выпускной квалификационной работе используется неявный итерационный метод регуляризации [3].

Для решения же задач с неизвестными параметрами оператора перспективным подходом является использование сверточных нейронных сетей [4,5]. В данной работе предлагается подход, основанный на многократном решении прямой задачи по сгенерированным синтетическим данным для различных параметров функции рассеяния точки с обучением на них сверточной нейронной сети, который позволяет восстанавливать параметр функции рассеяния точки для его дальнейшего использования при решении обратной задачи спектральным итерационным методом. Применение предложенного подхода для восстановления на экспериментальных данных с неизвестной функцией рассеяния точки показали перспективность его дальнейшего развития.

Литература

- [1] Adaptive system for eye-fundus imaging / Larichev A. V., Ivanov P. V., Iroshnikov N. G., Shmalhauzen V. I., Otten L. J. // Quantum Electronics. 2002. Vol. 32. P. 902–908.

- [2] Fourier domain iterative approach to optical sectioning of 3D translucent objects for ophthalmology purposes / Razgulin A. V., Iroshnikov N. G., Larichev A. V., Romanenko T. E., Goncharov A. S. // The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences. 2017. Vol. 42. P. 173.
- [3] Бакушинский А. Б., Гончарский А. В. Итеративные методы решения некорректных задач. М. : Наука, 1989. 128 с.
- [4] Deep convolutional neural networks for estimating porous material parameters with ultrasound tomography / Lahivaara T., Karkkainen L., Huttunen J. M., Hesthaven J. S. // The Journal of the Acoustical Society of America. 2018. Vol. 143. P. 1148.
- [5] Fast Point Spread Function Modeling with Deep Learning / Herbel J., Kasprzak T., Amara A., Refregier A., Lucchib A. // Journal of Cosmology and Astroparticle Physics. 2018. Vol.2018, № 07. P. 54.

Метод решения задачи теплопроводности со специальным нелокальным условием

Гаврилов Олег Алексеевич

Кафедра математической физики

e-mail: oa.gavr@yandex.ru

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

В математической физике важную роль играет сильно некорректная параболическая задача с обратным направлением времени [1]. Для такой задачи в статье [2] предложен метод регуляризации при помощи специального нелокального условия, совмещающего начальное и финальное значения искомой функции. В статье [3] доказана теорема единственности для одномерной задачи теплопроводности с более общим нелокальным условием. Случай аналогичной нелокальной задачи на всей числовой прямой изучен в [4].

В настоящей работе проводится систематическое изучение модельной нелокальной задачи для одномерного уравнения теплопроводности. Уравнение рассматривается при $0 < x < l$ на конечном промежутке времени $0 < t \leq T$. При помощи нелокального по времени условия изучаются различные конструктивные методы восстановления решения. Проводится ряд численных расчетов с помощью программы на языке *MATLAB*, подтверждающих высокую надёжность предложенных алгоритмов.

Постановка задачи. Рассматривается задача следующего вида:

$$u_t = a^2 u_{xx} - c(x)u, \quad 0 < x < l, \quad 0 < t < T, \quad (1)$$

$$u(0, t) = 0, \quad u(l, t) = 0, \quad 0 \leq t \leq T, \quad (2)$$

$$\alpha u(x, 0) + (1 - \alpha)u(x, T) = \psi(x), \quad 0 \leq x \leq l. \quad (3)$$

Считаем заданными функции $c(x) \geq 0$, $\psi(x)$, коэффициент $a^2 > 0$, параметр $\alpha \in (0, 1)$. Условие (3) будем называть нелокальным условием по времени.

Требуется найти функцию $u = u(x, t)$. Для этого достаточно восстановить $\varphi(x) \equiv u(x, 0)$. Введённые обозначения $\psi(x)$, $\varphi(x)$ сохраняем всюду далее.

Алгоритм решения. Из нелокального условия (3) составляется операторное уравнение второго рода относительно неизвестной функции φ :

$$\varphi - \mathcal{B}_\alpha \varphi = \frac{\psi}{\alpha}. \quad (4)$$

Здесь оператор \mathcal{B}_α , зависящий от параметра α , вводится по правилу

$$(\mathcal{B}_\alpha f)(x) = - \left(\frac{1 - \alpha}{\alpha} \right) u(x, T; f), \quad f \in L_2[0, l],$$

где $u(x, t; f)$ обозначает решение следующей «прямой» задачи по отношению к «обратной» задаче (1)–(3):

$$u_t = a^2 u_{xx} - c(x)u, \quad 0 < x < l, \quad 0 < t < T, \quad (5)$$

$$u(0, t) = 0, \quad u(l, t) = 0, \quad 0 \leq t \leq T, \quad (6)$$

$$u(x, 0) = f(x), \quad 0 \leq x \leq l. \quad (7)$$

Можно убедиться, что при выполнении специальных условий решение уравнения (4) представимо в виде ряда Неймана

$$\varphi = \frac{1}{\alpha} \sum_{k=0}^{\infty} \mathcal{B}_\alpha^k \psi, \quad (8)$$

где степени оператора \mathcal{B}_α вычисляются следующим образом:

$$(\mathcal{B}_\alpha^k f)(x) = - \left(\frac{1 - \alpha}{\alpha} \right) u(x, kT; f), \quad k = 0, 1, 2, \dots, \quad f \in L_2[0, l].$$

Утверждение. Если $\alpha \in [\frac{1}{2}, 1)$, то ряд (8) сходится в пространстве $L_2[0, l]$ и его сумма является решением уравнения (4).

На основании описанного метода проведены вычислительные эксперименты, показывающие, что во многих случаях сходимость имеет место и при гораздо меньших значениях α , причём скорость сходимости существенно зависит от выбора функции $c(x)$.

Предлагаются две разностные схемы, аппроксимирующие задачу (5)–(7). Различие между ними состоит в выборе временного слоя для слагаемого, соответствующего $c(x)u$ в уравнении (5). Проведено аналитическое сравнение устойчивости обеих разностных схем.

Отдельно рассматривается частный случай задачи (1)–(3) при $c(x) \equiv 0$. Изучен способ аналитического решения (метод Фурье). Предложен комбинированный метод, сочетающий возможности и операторного подхода, и метода Фурье. Множество вычислительных экспериментов подтверждают высокую точность разработанных схем решения нелокальной задачи (1)–(3).

Литература

[1] Денисов А. М. Введение в теорию обратных задач. Изд-во МГУ. 1994.

- [2] Вабищевич П. Н. Нелокальные параболические задачи и обратная задача теплопроводности // Дифференциальные уравнения. 1981. Т. 17, № 7. С. 1193–1199.
- [3] Кангужин Б. Е. О единственности решения нелокальной по времени задачи для уравнения теплопроводности // Неклассические уравнения математической физики: Труды семинара, посвященного 60-летию профессора В. Н. Врагова. / Под. ред. А. И. Кожанова. — Новосибирск: Изд-во Ин-та математики. 2005. С. 130–132.
- [4] Королев Г. М. Специальная нелокальная задача для одномерного уравнения теплопроводности // Герценовские чтения, материалы научной конференции, 9–13 апреля 2018 г. — СПб.: Изд. РГПУ им. А. И. Герцена. 2018. С. 96–100.

Нейросетевой метод определения уровней артефактов на изображениях при потере высокочастотной информации

Гаврилова Дарья Владимировна

Кафедра математической физики

e-mail: darya_gavrilova1999@mail.ru

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

В современном мире в силу ограниченных технических возможностей изображения имеют артефакты, мешающие восприятию информации. Например, потеря высокочастотной информации и приобретение эффекта ложного оконтуривания (англ. - ringing), связанного с математической точки зрения с эффектом Гиббса [1]. Наиболее часто это встречается в результате получения ограниченной по спектру информации (например в МРТ снимках) или алгоритмах сжатия JPEG2000, которые могут сильно квантовать высокие частоты и, как следствие, огрублять мелкие детали. Артефакт ложного оконтуривания заключается в неприятном окаймлении возле чётких контуров. Также изображения могут иметь другие частные артефакты: размытия в результате потери фокуса при съёмке, шум. Подобные искажения изображений могут влиять на восприятие информации.

Целью дипломной работы является разработка нейросетевых алгоритмов определения параметров артефактов на изображениях, которые появляются при потере высокочастотной информации.

Изображения для обучения нейросетей были взяты из базы [2]. Изображения в базе содержат комбинацию артефактов: эффекта ложного оконтуривания, дефокуса и шума с различными параметрами.

Обучающие выборки были составлены из квадратных фрагментов изображений, взятых из базы, проявление артефактов в которых наиболее сильное. Для выбора областей наиболее заметного размытия были построены карты размытия для изображений из базы [2] и проведен анализ полученных карт. Для выделения областей, содержащих эффект ложного оконтуривания, строились

карты границ, используя алгоритм Канни [3], и выбирались области, содержащие границы между различными объектами или границы между объектом и фоном, так как артефакт ложного оконтуривания сильнее всего проявляется при резком изменении интенсивности. На полученных наборах данных обучались нейросетевые алгоритмы.

Для каждого артефакта строился свой нейросетевой алгоритм, определяющий параметр данного артефакта. Каждая нейросеть состояла из групп сверточных слоев и полносвязных, следующих друг за другом. Сверточные слои служат для выделения признаков на изображении, а полносвязные для соотнесения выделенных признаков с параметром артефакта. Ответом каждого нейросетевого алгоритма было вещественное число - оценка параметра артефакта. В качестве метода оптимизации использовался алгоритм Адам [4], а в качестве функции потерь выбрана MAE - mean absolute error:

$$MAE(y, y^p) = \frac{\sum_{i=1}^n |y_i - y_i^p|}{n}.$$

Подобранные архитектуры показали наилучшие результаты среди нескольких опробованных в ходе разработки нейросетевых алгоритмов, и дали достаточно хорошую оценку параметров артефактов на тестовом наборе данных.

Программа для обучения нейросетей написана на языке Python3, с использованием специальных библиотек Keras и Tensorflow. Обучение проходило на графическом процессоре с использованием сервиса Google Colab.

Литература

- [1] Umnov A. V., Krylov A. S. Sparse approach to image ringing detection and suppression // Pattern Recognition and Image Analysis: Advances in Mathematical Theory and Applications. – 2017. – Vol. 27
- [2] Бучацкая О. П. Разработка базы изображений с артефактами, возникающими при потере высокочастотной информации, Курсовая работа, ВМК, МГУ им. М. В. Ломоносова, 2020.
- [3] J. Canny, “A computational approach to edge detection,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. PAMI-8, no. 6, pp. 679–698, Nov 1986.
- [4] Adam: A method for stochastic optimization // arXiv preprint arXiv:1412.6980. – 2014.

Применение рекуррентных нейронных сетей для оценки динамики роста растений

Корнеева Александра Михайловна

Кафедра математической физики

e-mail: korneevasandra@gmail.com

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

В данной работе реализован метод, позволяющий обнаруживать отклонение роста растений по фотографиям от предсказанной динамики нормального развития растений.

Автоматизированный анализ изображений с помощью нейронных сетей может быть использован для классификации растений по их видам, физиологическому состоянию [1], а также для автоматической сегментации видимых органов растений, например листьев, в случаях, когда невозможно выделение этих элементов более простыми методами [2].

Такого рода задачи являются актуальными для автоматизированных производств, например, вертикальных ферм. На таких фермах сельскохозяйственные культуры выращиваются в изолированных, полностью контролируемых условиях с высоким уровнем автоматизации производственных процессов [3].

Площадь поверхности листьев можно рассматривать, как важнейшую характеристику, определяющую динамику и здоровый рост растений. В рамках данной работы рассматривается возможность автоматизированной оценки площади листьев и обнаружения их отклонений от предсказанной динамики роста.

Для сбора данных была разработана система с использованием IP-камеры, микрокомпьютера Raspberry, облачного хранилища и установлена на вертикальной ферме. Это позволило реализовать автоматизированный сбор данных для дальнейшего анализа.

В силу особенностей полученных фотографий: плохо отделимых листьев от фона из-за цвета вещества, в котором растут растения, наличия мха и отражений подсветки, оказываются неприменимы простые методы на основе работы с цветом изображений. Для выделения листьев реализована сверточная нейронная сеть, которая позволила решить задачу сегментации. Результатом работы данной сети являются маски сегментации, используя которые можно оценить текущую площадь листьев.

По оценённой площади листьев произведено прогнозирование динамики роста на промежуток времени от 1 до 4ч с помощью рекуррентных нейронных сетей, используя блоки LSTM и GRU. Полученные результаты метрики качества предсказания временного ряда позволяют говорить о том, что использование рекуррентных сетей с GRU блоками более предпочтительно. Таким образом, возможность прогнозирования динамики роста может быть использована для обнаружения отклонения развития растения.

Литература

- [1] Deep phenotyping: deep learning for temporal phenotype/genotype classification / S. T. Namin, M. Esmaeilzadeh, M. Najafi, T. B. Brown, J. O. Borevitz // Plant Methods, 2008.
- [2] Panicle-SEG: a robust image segmentation method for rice panicles in the field based on deep learning and superpixel optimization / X. Xiong, L. F. Duan, L. B. Liu, H. F. Tu, P. Yang, D. Wu // Plant Methods, 2017.
- [3] Иконописцева О. Г., Экоархитектура вертикальных ферм как новая типология агропромышленных зданий городского хозяйства будущего // Известия Самарского научного центра Российской академии наук. Социальные, гуманитарные, медико-биологические науки 2018. Т.20, № 3. С. 34–41.

Обратная задача частотного зондирования градиентно-слоистых сред

Малов Дмитрий Вячеславович

Кафедра математической физики

e-mail: malovdv@gmail.com

Научный руководитель — к.ф.-м.н н.с. Ингтем Женни Гастонова

Введение. В рамках данной работы разработано и реализовано решение обратной задачи частотного электромагнитного зондирования градиентно-слоистых сред и проведено исследование достаточных условий нахождения решения с допустимой погрешностью.

Определение. *Частотное электромагнитное зондирование (ЧЭМЗ) – метод геофизических исследований, при котором производится измерение напряженности гармонически изменяющегося электромагнитного поля в фиксированной точке на некотором удалении от источника поля для различных частот из заданного набора с целью получения информации о структуре геоэлектрического разреза земли.*

Постановка задачи. В рамках исследуемой модели структура разреза может быть формализована в виде вектора параметров $\mathbf{p} = [p_1, p_2, \dots, p_i]$, где каждый слой определяется его мощностью h и двумя значениями проводимости σ_s, σ_e на границах, т. к. проводимость линейно меняется по толщине (ранее подобная задача решалась для слоистых сред с постоянной проводимостью слоёв [1]). Для определения структуры разреза на основании экспериментальных данных необходимо решить обратную задачу, т. е. определить значения параметров среды \mathbf{p} , при которых формируется идентичное электромагнитное поле.

$$E_\phi(\mathbf{p}, \omega) = \frac{i\omega\mu_0 m_z}{4\pi} \int_0^\infty J_1(\lambda\rho)U(z, \lambda, \mathbf{p})d\lambda$$

Для численного интегрирования осциллирующей функции Бесселя использовался метод линейной фильтрации на основе алгоритма Андерсона [2].

На практике вместо напряженности поля используется значение кажущегося сопротивления $R(\mathbf{p}, \boldsymbol{\omega}) = \left| \frac{2\pi\rho^4}{3m_z} E_\phi(\mathbf{p}, \boldsymbol{\omega}) \right|$. Решение обратной задачи сводится к задаче минимизации функционала невязки $\Phi(\sigma(\mathbf{p}))$ (или целевого функционала): $\Phi(\sigma(\mathbf{p})) = \|R^T(\mathbf{p}) - R^E(\mathbf{p})\|$, где $R^T(\mathbf{p})$ – вектор значений кажущегося сопротивления, полученных при прямом решении задачи для текущих параметров \mathbf{p} , а $R^E(\mathbf{p})$ – вектор значений, полученных экспериментально при измерении поля для заданного набора частот $\boldsymbol{\omega}$.

Метод оптимизации. Целевой функционал является дифференцируемым, поэтому для минимизации могут быть использованы градиентные методы. В работе используется современная модификация метода градиентного спуска Adam [3], которая сочетает идеи сохранения момента движения и выбора индивидуального шага спуска для различных параметров.

Для нахождения градиента целевого функционала $\nabla\Phi(\mathbf{p})$ используется метод автоматического дифференцирования [4], который основан на том, что любую функцию можно представить в виде композиции элементарных функций и, используя правило дифференцирования сложной функции, можно по цепочке найти значения частных производных от целевого функционала в точке при обеспечении высокой вычислительной эффективности.

Регуляризация. Обратная задача ЧЭМЗ является некорректной, поэтому необходимо использование метода регуляризации путём введения дополнительного слагаемого $\Omega(\sigma(\mathbf{p}))$ с коэффициентом α для включения априорной информации в качестве дополнительного ограничения.

$$\Phi_\alpha(\sigma(\mathbf{p})) = \|R^T(\mathbf{p}) - R^E(\mathbf{p})\| + \alpha\Omega(\sigma(\mathbf{p}))$$

$$\Omega(\sigma(\mathbf{p})) = \sum_{i=1}^n \{a_{i1}(h_i - h_i^0)^2 + a_{i2}(\sigma_{is} - \sigma_{is}^0)^2 + a_{i3}(\sigma_{ie} - \sigma_{ie}^0)^2\}$$

В процессе вычислительных экспериментов наилучшие результаты были получены при значении параметра α в диапазоне $[0,1 - 0,01]$. При этом параметры разреза, близкие к исходному, могут быть успешно найдены для структуры геологического разреза, включающего 6 слоёв, когда известна приближенная информация о двух слоях с погрешностью до 40%.

В качестве дальнейшего направления развития осуществляется доработка алгоритма для определения разрезов с произвольным количеством слоёв.

Результаты. Разработан вычислительно эффективный параллелизуемый метод решения обратной задачи частотного зондирования для градиентно-слоистых сред, исследована его работоспособность в процессе серии вычислительных экспериментов, которые продемонстрировали возможность успешного решения обратной задачи для прикладного применения. Программное обеспечение разработано на языке Python 3.6 с использованием библиотеки NumPy и фреймворка автоматического дифференцирования PyTorch 1.8.1. Использование методов векторизации при разработке алгоритма и применение PyTorch позволяют осуществлять решение обратной задачи на многоядерных серверах как на обычных процессорах (CPU), так и на специализированных графических процессорах (GPU).

Литература

- [1] Дмитриев В. И. Морские электромагнитные зондирования, монография. Издательство «Аргамак-Медиа», 2014.
- [2] Anderson W. L. Computer program numerical integration of related Hankel transforms of orders 0 and 1 by adaptive digital filtering // Geophysic, 1979, Vol. 44. N 7. P. 1287-1305.
- [3] Diederik P. Kingma, Jimmy Ba. Adam: a method for stochastic optimization [Электронный ресурс], URL: <https://arxiv.org/abs/1412.6980> (дата обращения 15.01.2021).
- [4] Automatic differentiation in machine learning: a survey [Электронный ресурс], URL: <https://www.jmlr.org/papers/volume18/17-468/17-468.pdf> (дата обращения 01.03.2021).

Нейросетевой метод совмещения биомедицинских изображений

Пятов Владислав Андреевич

*Кафедра математической физики
e-mail: vladpyatov@yandex.ru*

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

Задача совмещения – одна из ключевых сфер анализа биомедицинских изображений. В клинической практике нередко приходится делать несколько снимков исследуемой области с целью анализа состояния пациента. Соответствующие данные могут быть получены, например, с помощью сонографии, рентгенографии или компьютерной томографии. Комбинация полученных изображений позволяет получить более полную информацию о пациенте, которая не видна на отдельных снимках.

В данной работе разработан и программно реализован комбинированный метод совмещения биомедицинских изображений – предварительное жесткое совмещение преобразованиями поворота и переноса, а также нежесткое совмещение с использованием методов глубокого обучения.

Объектом исследования является набор снимков, предоставленный факультетом фундаментальной медицины МГУ. Набор состоит из изображений тканей пациентов с гистологически подтвержденным диагнозом – аденокарцинома желудка. Каждому образцу соответствует от четырех до пяти снимков. Образцы имеют хрупкую структуру, поэтому на некоторых снимках ткань может быть деформирована.

Жесткое совмещение предлагается проводить с использованием классических алгоритмов. На первом шаге для каждого из изображений – фиксированного $I_F(x)$ и совмещаемого $I_M(x)$, методом RootSIFT [1] выделяются ключевые точки и соответствующие им дескрипторы. Далее выделенные наборы дескрипторов сопоставляются методом K ближайших соседей. Затем алгоритмом

RANSAC [2] по соответствующим парам ключевых точек вычисляется матрица трансформации.

В случае нежесткого совмещения изображений задача сводится к нахождению поля деформации $\varphi(x)$ совмещаемого изображения I_M . Для вычисления поля деформации используется сверточная нейронная сеть U-Net-подобной архитектуры. По входным изображениям – фиксированному и совмещаемому, нейронная сеть в процессе оптимизации учится генерировать поле деформации для совмещаемого изображения. Результатом работы алгоритма является совмещенное изображение $I_M \circ \varphi$.

Функция потерь нейронной сети имеет следующий вид:

$$\mathcal{L}(I_F, I_M, \varphi) = \frac{1}{2} \cdot \mathcal{L}_{\text{sim}}(I_F, I_M \circ \varphi) + \frac{1}{2} \cdot \mathcal{L}_{\text{sim}}(I_M, I_F \circ \varphi) + \lambda \cdot \mathcal{L}_{\text{smooth}}(\varphi) + \beta \cdot \mathcal{L}_{\text{cnt}}(\varphi) + \gamma \cdot \mathcal{L}_G(I_M \circ \varphi)$$

В качестве \mathcal{L}_{sim} используется нормализованная кросс-корреляция для определения меры сходства изображений. Регуляризация градиента $\mathcal{L}_{\text{smooth}}$ поля деформации отвечает за гладкость поля. Кроме того, предлагается использовать регуляризацию дивергенции \mathcal{L}_{cnt} поля деформации с целью поддержания несжимаемости результирующей трансформации. Для улучшения визуального качества выходных изображений используется отдельная нейронная сеть дискриминатор, классифицирующая изображения на обучающие и совмещенные. Ее функция потерь, а также компонента \mathcal{L}_G задаются перекрестной энтропией.

Ввиду большого размера изображений вычисление поля деформации предлагается выполнять по пересекающимся блокам. Таким образом удастся решить проблему ограниченности ресурсов графических ускорителей, используемых при обучении и непосредственном использовании нейронной сети.

С целью компенсации глобальных деформаций предлагается использовать многомасштабный анализ изображений. Согласно данному методу результирующее поле деформации является композицией полей деформации, вычисленных на различных масштабах.

Для оценки полученных результатов проведен количественный анализ качества совмещения изображений разработанным методом с использованием метрик, предложенных в [3]. Показано преимущество как всего метода, так и отдельного использования предложенного способа регуляризации дивергенции поля деформации над градиентной регуляризацией на примере метода VoxelMorph [4].

Литература

- [1] Arandjelović R., Zisserman A. Three things everyone should know to improve object retrieval // 2012 IEEE Conference on Computer Vision and Pattern Recognition. – IEEE, 2012. – P. 2911-2918.
- [2] Fischler M. A., Bolles R. C. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography // Communications of the ACM. – 1981. – Vol. 24. – N. 6. – P. 381-395.

- [3] Borovec J. et al. ANHIR: automatic non-rigid histological image registration challenge // IEEE transactions on medical imaging. – 2020. – Vol. 39. – N. 10. – P. 3042-3052.
- [4] VoxelMorph: a learning framework for deformable medical image registration / G. Balakrishnan, A. Zhao, M.R. Sabuncu, J. Guttag, A.V. Dalca // IEEE transactions on medical imaging. – 2019. – Vol. 38. – N. 8. – P. 1788-1800.

Применение технологии глубокого обучения для прогнозирования заболеваемости

Чижов Пётр Сергеевич

Кафедра вычислительных технологий и моделирования

e-mail: p.s.chizhov@gmail.com

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

В работе исследуется применимость методов глубокого обучения в прогнозировании динамики новой коронавирусной инфекции COVID-19. Эта болезнь характеризуется высокой летальностью и необходимостью госпитализировать значительное число больных. Поскольку динамика новой инфекции имеет волнообразный характер, для системы здравоохранения и экономики важно иметь двухнедельный прогноз, чтобы успеть развернуть дополнительные инфекционные стационары с увеличенным числом реанимационных блоков. Для построения таких прогнозов обычно используются компартментные модели [1]. Однако, их использование невозможно, так как заболевание активно передается на бессимптомной фазе, и оценить долю источников инфицирования не представляется возможным.

Применение технологии машинного обучения как раз уместно в малоизученных задачах, для которых пока не существует более точного и менее вычислительно затратного метода решения. Рекуррентные нейронные сети структурно учитывают последовательные во времени данные. В этой работе были построены рекуррентные сети, основанные на блоках LSTM [2] и GRU [3], которые позволяют частично решить проблему затухания и взрыва градиента, возникающую при обучении простых рекуррентных сетей [4].

Ключевыми компонентами LSTM являются вектор памяти C_t , который хранит информацию о прошлом, и вектор скрытого состояния h_t , формирующий выход на данной итерации по времени. Оба эти вектора в разном порядке проходят через три вычислительные структуры, называемыми фильтрами (1), (2), (5), которые по своей сути выполняют роль классификаторов, благодаря сигмоидальной функции, применяемой к выходу фильтров покомпонентно.

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \quad (1)$$

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \quad (2)$$

$$C'_t = \text{th}(W_C[h_{t-1}, x_t] + b_C) \quad (3)$$

$$C_t = f_t \otimes C_{t-1} + i_t \otimes C'_t \quad (4)$$

$$o_t = \sigma (W_o [h_{t-1}, x_t] + b_o) \quad (5)$$

$$h_t = o_t \otimes \text{th} (C_t) \quad (6)$$

Квадратные скобки означают конкатенацию двух векторов, функции активации – сигмоида (σ) и гиперболический тангенс (th) применяются к векторам и матрицам покомпонентно, \otimes - умножение Адамара.

Сеть GRU является некоторым упрощением LSTM. Вместо вектора памяти и вектора скрытого состояния используется только один вектор h_t . А также используются только два фильтра, устроенные также, как и фильтры LSTM. В этой задаче массив данных делится на 3 подвыборки: последний месяц данных выделяется в контрольную часть, чтобы затем можно было оценивать и сравнивать качество прогнозов уже обученных моделей, затем оставшаяся часть разделяется в соотношении 80% на 20% на тренировочную и проверочную части соответственно.

Были построены и обучены простые модели LSTM, GRU и их двунаправленные варианты в совокупности с тремя вариантами начальных преобразований ряда: нормализацией, переходом к ряду разниц с шагом в 1 день и с шагом в 7 дней. Тестирование и настройка гиперпараметров происходила в порядке увеличения количества весовых коэффициентов каждой модели с подсчетом невязки на тренировочной и проверочной частях данных.

В качестве оценки качества обучения модели используются среднее среднеквадратическое отклонение результата работы обученной модели от наблюдаемого временного ряда за одинаковый период по всем примерам тренировочной, проверочной и контрольной частей, до применения обратного преобразования. Для сравнения качества прогнозов различных моделей используется среднее абсолютное отклонение контрольной части ряда от прогноза за соответствующий период, над которым уже произведено обратное преобразование.

В работе при помощи методов глубокого обучения построены краткосрочные прогнозы динамики новой коронавирусной инфекции. Исследована эффективность обучения различных типов рекуррентных сетей и их модификаций на ежедневных данных по числу новых случаев заболевания за период в 1 год. Также изучено влияние различных преобразований исходных данных на качество прогноза. Показано, что наиболее точный прогноз реализует простая модель LSTM в сочетании с дифференцированием с шагом в один день и нормализацией.

Литература

- [1] Musulin J. et al. Application of Artificial Intelligence-Based Regression Methods in the Problem of COVID-19 Spread Prediction: A Systematic Review

//International Journal of Environmental Research and Public Health. – 2021. – Т. 18. – №. 8. – С. 4287.

- [2] Hochreiter S., Schmidhuber J. Long short-term memory //Neural computation. – 1997. – Т. 9. – №. 8. – С. 1735-1780.
- [3] Chung J. et al. Empirical evaluation of gated recurrent neural networks on sequence modeling //arXiv preprint arXiv:1412.3555. – 2014.
- [4] Greff K. et al. LSTM: A search space odyssey //IEEE transactions on neural networks and learning systems. – 2016. – Т. 28. – №. 10. – С. 2222-2232.

Решение краевой задачи для уравнения Пуассона в прямоугольнике с помощью нейронных сетей

Бочковенкова Диана Владиславовна

Кафедра вычислительных методов

e-mail: bochkdv@yandex.ru

Научный руководитель – к.ф.-м.н. доц. Хапаев Михаил Михайлович

В данной работе рассматривается задача Дирихле для уравнения Пуассона в прямоугольной области $\Omega \subset R^2$.

Рассмотрим задачу:

$$\begin{cases} \Delta u = -f(x), & x = (x_1, x_2) \in \Omega \\ u(x) = g(x), & x \in \partial\Omega \end{cases} \quad (1)$$

где Δu – оператор Лапласа, $f(x), g(x)$ – заданы на Ω и $\partial\Omega$ соответственно.

Сведем эту задачу к задаче вариационного исчисления. Обозначим через $C_g^1(\Omega)$ пространство всех гладких в Ω функций, которые на границе $\partial\Omega$ удовлетворяют граничным условиям задачи (1). Введем в $W_g^1(\Omega)$ (пространство Соболева для $C_g^1(\Omega)$) скалярное произведение и билинейный функционал:

$$(v, w) = \int_{\Omega} v(x)w(x)dx, \quad a(v, w) = \int_{\Omega} \frac{\partial v}{\partial x_1} \frac{\partial w}{\partial x_1} + \frac{\partial v}{\partial x_2} \frac{\partial w}{\partial x_2} dx$$

Определение. Функция $u \in W_g^1(\Omega)$ называется обобщенным решением задачи Дирихле (1), если

$$a(u, v) = (f, v), \quad \forall v \in W_g^1(\Omega)$$

Теорема 1. Обобщенное решение доставляет минимум в пространстве $W_g^1(\Omega)$ функционалу

$$J(v) = a(v, v) - 2(f, v) = \int_{\Omega} \left[\left(\frac{\partial v}{\partial x_1} \right)^2 + \left(\frac{\partial v}{\partial x_2} \right)^2 - 2fv \right] dx$$

Минимум этого функционала является обобщенным решением задачи (1) в пространстве $C_g^1(\Omega)$. Чтобы расширить область поиска минимума до $C^1(\Omega)$, будем искать решение, минимизируя функционал Ритца:

$$J(u) = \int_{\Omega} \left[\frac{1}{2} (\nabla_x u(x))^2 - f(x)u(x) \right] ds + \beta \int_{\partial\Omega} [u(x) - g(x)]^2 dl \quad (2)$$

где $\beta \gg 1$ (подбирается эмпирически).

Для приближенного вычисления (2) на $\bar{\Omega}$ вводится сетка, и функционал Ритца аппроксимируется на ней с помощью квадратурной формулы трапеций.

Нейронная сеть, используемая в данном исследовании, состоит из $n + 1$ входных узлов (включая узел смещения - bias), где $n = \dim(\Omega) = 2$, одного скрытого слоя (H узлов – гиперпараметр) с функциями активации *ReLU* и 1 выходного узла без активации. Точки введенной дискретизации являются обучающей выборкой. В качестве функции ошибки служит функционал Ритца, минимизация проводится с помощью алгоритма оптимизации Adam (Adaptive Moment Estimation), который сочетает в себе идею накопления движения и идею более слабого обновления весов для типичных признаков, поэтому является более эффективным, чем классический градиентный спуск. Используется стохастическая версия алгоритма.

При обучении нейронной сети точки сетки разбиваются на 3 набора данных: точки внутренней области, точки части границы, параллельной оси Ox , и точки части границы, параллельной оси Oy . Такое разделение удобно при численном вычислении интегралов в функции ошибки.

Описанный метод был программно реализован на языке Python. Задачи для верификации были подобраны исходя из сложности моделируемого аналитического решения. Гиперпараметры подбирались путем перебора по некоторой сетке значений перед началом обучения. В проведенных экспериментах область была разбита на 21 точку по обеим осям (обучение проводилось на 441 точке). При построении графиков размер исходного разбиения по обеим осям был увеличен в 10 раз, т. е. результат наблюдался на 40 401 точке.

Решения, получаемые данным методом, имеют максимальную погрешность 10^{-2} . В отличие от численных методов, этот способ позволяет получить непрерывное приближенное решение, но стоит отметить, что без ручной настройки архитектуры нейросети достижение желаемой точности не гарантируется. Т. к. данный метод обладает определенными преимуществами перед численными методами и может быть улучшен, он имеет перспективы применительно к решению задач математической физики.

Литература

- [1] Абакумов М. В. Лекции по численным методам математической физики : учеб. пособие / М. В. Абакумов, А. В. Гулин. — М. : ИНФРА-М, 2018.— 158с.

- [2] E W., Han J. Q., Jentzen A. Deep learning-based numerical methods for highdimensional parabolic partial differential equations and backward stochastic differential equations, submitted. arXiv:1706.04702.
- [3] Beck C., E W., Jentzen A. Machine learning approximation algorithms for highdimensional fully nonlinear partial differential equations and second-order backward stochastic differential equations, submitted. arXiv:1709.05963.
- [4] Kingma D. P., Ba J. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980, 2014

Математическое распределение тяжелых газов и жидкостей на орографически неоднородной поверхности

Василенко Анна Валерьевна

Кафедра вычислительных методов

e-mail: ann.v.vasilenko@mail.ru

Научный руководитель — д.ф.-м.н. в.н.с. Савенкова Надежда Петровна

В данной работе была затронута актуальная в современном мире проблема моделирования аварий на предприятиях, связанных с растеканием жидкостей и тяжелых газов таких, как нефть, хлор и других.

Существует много различных способов моделирования динамики промышленных аварий. Однако либо в них учитывается недостаточное количество внешних воздействий, либо многие формулы являются полуэмпирическими, и, следовательно, такие модели слабо адекватны. Целью настоящей работы является разработка математической модели, позволяющей достаточно быстро в реальных условиях распространения вещества по рельефу обширной области, с приемлемой точностью детализировать последствия произошедшей аварии.

В основе математической модели находятся усредненные по высоте трехмерные уравнения газовой динамики [1], что позволяет изучать движение вещества в среднем слое [2]. Для численной реализации модели предложен метод расщепления по физическим процессам [3]. При этом алгоритм решения разделяется на отдельные этапы. На каждом этапе учитывается один физический процесс. Причем решение предыдущего этапа служит начальным условием для расчета следующего этапа.

На первом этапе вычисляются значения скоростей, изменяющиеся из-за перепадов давления и рельефа. На втором этапе численно решается уравнение адвекции-диффузии. Далее проводятся вычисления, позволяющие учесть трение о подстилающую поверхность, наличие ветра, смешение облака газа с атмосферой, а также просачивание в почву в случае растекания жидкости.

Для численного решения задачи Коши на этапе учета трения, ветра, смешения с атмосферой и просачивания используется метод Эйлера—Коши [4]. Учет изменения давления и рельефа выполнен с помощью схемы в полуцелых точках. Уравнения адвекции численно решаются методом Мак—Кормака [5]. В работе были проведены одномерные эксперименты, направленные на выявления достоинств и недостатков данного метода. Установлено, что модель

является достаточно точной и достоверно описывает взаимодействия таких сред, как: газ-газ и жидкость-газ.

Разработанная математическая модель применялась к моделированию аварии на газопроводе с затухающим источником истечения газа и для моделирования промышленной аварии на нефтехранилище. Выполнена трехмерная визуализация результатов расчетов. Весь вычислительный комплекс для решения поставленной задачи и программы для одномерных исследований схемы Мак–Кормака реализованы на языке C++.

Литература

- [1] Тихонов А. Н., Самарский А. А. Уравнения математической физики, Издательство МГУ //Москва. – 1999.
- [2] Численный расчет растекания тяжелой жидкости в условиях орографически неоднородной поверхности / О. В. Доброчеев, А. А. Кулешов, Н. П. Савенкова, С. В. Филиппова // Математическое моделирование. – 1996. – Т. 8. – №. 5. – С. 92-105.
- [3] Марчук Г. И. Математическое моделирование в проблеме окружающей среды. – Наука. Гл. ред. физ.-мат. лит., 1982.
- [4] Амосов А. А., Дубинский Ю. А., Копченова Н. В. Численные методы для инженеров //М.: Высшая школа. – 1994. – 544 с.
- [5] Андерсон Д., Таннехилл Д., Плетчер Р. Вычислительная гидромеханика и теплообмен: В 2-х т.: Пер. с англ. – Мир, 1990.

Схема КАБАРЕ с улучшенными дисперсионными свойствами

Горбунова Ольга Константиновна

Кафедра вычислительных методов

e-mail: elyaishere@mail.ru

*Научные руководители — д.ф.-м.н. проф. Головизнин Василий Михайлович,
асп. м.н.с. Афанасьев Никита Александрович*

Для описания многих современных задач математической физики используют системы дифференциальных уравнений гиперболического типа. Одним из широко применяемых методов решения таких систем является схема КАБАРЕ, представляющая семейство балансно-характеристических разностных схем [1]. В качестве недостатка данной схемы можно выделить сильную аномальную дисперсию при стремлении чисел Куранта к нулю, что может привести к значительным искажением профиля решения. Данная работа посвящена улучшению дисперсионных свойств схемы КАБАРЕ.

Ранее была получена улучшенная схема для систем линейных дифференциальных уравнений гиперболического типа [2]. Особенность данного метода заключается в том, что для вычисления значений на новом временном слое необходимо знать полный набор собственных векторов матрицы системы,

постоянной в линейном случае. Основной задачей этой работы является обобщение данного метода на системы нелинейных дифференциальных уравнений гиперболического типа. Матрица в таких системах зависима от переменных задачи, а ее собственные значения в разностном случае становятся локальными для каждой ячейки сетки, что может приводить к малым числам Куранта в областях наиболее медленно переносящихся инвариантов Римана. Таким образом, задача построения схемы с улучшенными дисперсионными свойствами для систем нелинейных уравнений является крайне актуальной.

В данной работе была получена схема КАБАРЕ с улучшенными дисперсионными свойствами для системы нелинейных дифференциальных уравнений Сен-Венана, известной также как уравнения мелкой воды. Модель мелкой воды широко применяется для моделирования течений, глубина которых сильно меньше их линейных размеров. Новая схема была протестирована на задачах о распространении акустических возмущений и задаче распада разрыва. Результаты расчетов продемонстрировали, что дисперсионная ошибка улучшенной схемы значительно меньше, чем у стандартной схемы КАБАРЕ. В рамках исследования полученной схемы была проведена серия расчетов задачи о переносе промодулированной гауссианом акустической волны на последовательности сгущающихся сеток, в которой точность и порядок сходимости нового метода превысили аналогичные показатели для стандартной схемы.

Литература

- [1] Головизнин В. М., Карабасов С. А., Кобринский И. М. Балансно-характеристические схемы с разделенными консервативными и потоковыми переменными // Математическое моделирование. 2003. Т. 15, № 9. С. 29–48.
- [2] Афанасьев Н. А., Головизнин В. М., Соловьев А. В. Схема КАБАРЕ с улучшенными дисперсионными свойствами для систем линейных дифференциальных уравнений гиперболического типа // Выч. мет. программирование. 2021. Т. 22, № 1. С. 67–76.

Практическое применение вейвлет-преобразований для анализа и обработки данных

Измайлов Михаил Борисович

*Кафедра автоматизации научных исследований
e-mail: shishkin@cs.msu.ru, s02170619@gse.cs.msu.ru
Научные руководители — к.ф.-м.н. доц. Шагиров Эльдар Аминович,
д.ф.-м.н. профессор Шишкин Алексей Геннадиевич*

Быстрое развитие технологий привело к формированию многих процессов в информационной среде, хранению данных и их исследованию. Значительная часть информации такого рода представляет собой временные ряды. Анализ временных рядов в настоящее время посвящено большое число работ из-за его распространенности в различных областях: в науке, технике, медицине. К числу методов, применяемых для изучения временных рядов, относится кластеризация, которая является одной из наиболее важных концепций машинного

обучения и интеллектуального анализа данных, и которая позволяет определять структуру данных путем разделения немаркированных образов на однородные группы. Однако обычные алгоритмы кластеризации, такие как разделяющие и иерархические методы, неприменимы или неэффективны для временных рядов, поскольку они в основном предназначены для статистических данных. Кроме того, временные ряды имеют высокую размерность, высокую корреляцию признаков, а также часто бывают зашумленными. Это приводит к необходимости разработки алгоритмов, позволяющих решать указанные проблемы и при этом быть эффективными в вычислительном плане.

Разработанный алгоритм основан на модификации с помощью вейвлетов Хаара одного из наиболее популярных алгоритмов кластеризации — метода К-средних. Одним из главных недостатков алгоритма К-средних [1] является то, что качество кластеризации в значительной степени зависит от выбора исходных центроидов. Неудачный выбор центроидов может ухудшить качество обработки временных рядов, а также привести к более длительному времени выполнения. В разработанном алгоритме решены обе проблемы.

Идея, лежащая в основе нового метода, основана на том, что общая форма временных рядов часто может быть приблизительно описана с более низким разрешением с помощью вейвлет-преобразования [2]. Кратко разработанный алгоритм можно описать следующим образом:

- 1) Определить значение k
- 2) Определить центроиды k кластеров
- 3) Выполнить алгоритм К-средних на (i) -ом уровне
- 4) Использовать полученные центроиды на (i) -ом уровне для инициализации центроидов на уровне $(i+1)$
- 5) Если ни один из N объектов не переместился в другой кластер на последней итерации, закончить. В противном случае перейти к шагу 3

Постепенно переходя к более высоким разрешениям, начальное положение центроидов берется с предыдущего уровня. Во многих случаях оно довольно быстро устанавливается и затем не меняется при увеличении разрешения. Поэтому число итераций до момента сходимости обычно бывает намного меньше, чем в классическом методе К-средних.

Одна из основных серий экспериментов была проведена на общедоступном наборе данных «InsectWingbeatSound» [3]. Данные представляют собой спектр мощности звука взмаха крыльев различных насекомых. Свойства данных следующие: 11 классов, 220 временных рядов, длина ряда - 256 отсчетов.

Стоит отметить, что в качестве основных метрик оценки качества кластеризации были взяты Чистота, Коэффициент Рэнда (КР), Скорректированный коэффициент Рэнда (СКР) и F1-мера.

Для сравнения результатов, полученных с помощью разработанного метода, был применен традиционный алгоритм К-средних, а также метод DBSCAN и вычислены те же критерии качества кластеризации. Результаты показаны в Табл. 1.

	Чистота	КР	СКР	F1-мера
Разработанный алгоритм	0.55454	0.89273	0.34902	0.40788
Метод К-средних	0.51363	0.88389	0.30103	0.36475
Метод DBSCAN	0.31363	0.74719	0.10627	0.21983

Табл. 1: Результаты кластеризации

Видно, что коэффициенты качества кластеризации после обработки с помощью модифицированного метода имеют более высокие значения, чем после кластеризации с помощью традиционных алгоритмов.

В результате работы на основе традиционного метода К-средних и вейвлет преобразования Хаара был разработан новый метод кластеризации временных рядов. Для реализации этого метода центроиды, полученные на каждом разрешении вейвлет-преобразования, использовались в качестве начальных для следующего уровня разрешения. Проведенные эксперименты показали, что разработанный метод превосходит традиционные алгоритмы кластеризации временных рядов.

Литература

- [1] Chan, K. Fu, A. W. (1999). Efficient Time Series Matching by Wavelets.
- [2] Добеши, И. (2001). Десять лекций по вейвлетам.
- [3] Keogh, E. Folias, T. (2002). The UCR Time Series Data Mining Archive.

Применение методов машинного обучения для улучшения качества речевого сигнала

Кот Владислав Яковлевич

Кафедра автоматизации научных исследований

e-mail: vladislavkotx@gmail.com

Научный руководитель — д.ф.-м.н. проф. Шишкин Алексей Геннадиевич (shishkin@cs.msu.ru)

За последние несколько десятилетий были разработаны многочисленные методы улучшения речи. Большинство из этих методов после их настройки работают в конкретной среде с фиксированными параметрами и не в состоянии обрабатывать нестационарный шум для реальных сценариев в сложных акустических условиях. В связи с этим встает задача о разработке универсального метода устранения фонового шума для произвольных внешних условий, имеющего высокую эффективность и надежность. Данная работа посвящена применению нейросетевых подходов для решения задачи улучшения качества речевого сигнала.

На первом этапе из сигнала выделялись окна по 32 мс. с перекрытием в 16 мс. Далее применялось быстрое преобразование Фурье с оконной функцией Ханна для получения спектра сигнала. Выделялись его амплитудная и фазовая части. Так как человеческое ухо является нечувствительным к искажениям фазовой части спектра, то эта часть сохранялась для ее использования при восстановлении сигнала, и далее использовалась только амплитудная составляющая.

Логарифм квадрата амплитудной составляющей для нескольких последовательных окон сигнала подавался на вход нейросети. Для получения чистого сигнала выполнялось обратное преобразование для восстановления сигнала. А именно добавление фазовой части спектра с потенцированием и обратное преобразование Фурье с применением метода перекрытия с накоплением [1].

В качестве чистого сигнала использовались аудиозаписи из базы - TIMIT Acoustic-Phonetic Continuous Speech Corpus. Первая часть экспериментов проводилась с использованием гауссова шума в качестве аддитивного. Кроме искусственно созданного шума проводилось тестирование и оценка модели на реальном шуме. Для создания приближенных к реальным зашумленным аудиозаписей чистая речь была зашумлена с помощью аудиозаписей из базы данных - ESC: Dataset for Environmental Sound Classification

В данной работе использовалась нейронная сеть прямого распространения. Для оценки того, насколько качественно очищен наш сигнал нужна метрика, которая может оценивать сигнал согласно человеческому восприятию. Поэтому в качестве используемой метрики была выбрана перцепционная оценка качества речи (PESQ score) [2], которая позволяет оценить качество сигнала, основываясь на восприятии человека.

Проводились эксперименты с различными уровнями сигнал/шум для подбора оптимальных гиперпараметров модели. В качестве потенциальных гиперпараметров были исследованы следующие:

- Количество слоев: 4-7
- Количество нейронов на скрытых слоях: 100-1000
- Функция активации: сигмоида, ReLU (Rectified Linear Unit), LeakyReLU
- Оптимизаторы: RMSprop, Adam
- Использование нормализации весовых коэффициентов

На основе экспериментов с гауссовым шумом - модель, показавшая лучшие результаты, была выбрана для дальнейшего использования на реальных шумах.

Было выделено 9 разных классов аудиозаписей для использования их в качестве реального фонового шума: дождь, автомобиль, ветер, самолет, поезд, гром, морские волны, вертолет, пылесос

Модель показала отличные результаты при работе с гауссовым шумом. В рамках рассматриваемых отношений сигнал/шум: $\{-5, 0, 5, 10\}$ дБ при среднем качестве зашумленного сигнала 2.77 по шкале PESQ, качество обработанного сигнала повышалось до 3.45.

Результаты для 9 классов реальных шумов при различных отношениях сигнал/шум приведены в Табл. 1.

SNR (дБ)	-5	0	5	10
Зашумленный сигнал	1.525	2.254	2.63	3.2
Обработанный сигнал	1.965	2.586	2.8	3.24

Табл. 1: Результаты для различных уровней сигнал/шум по шкале PESQ.

Таким образом, разработана модель для устранения фонового шума из сигнала на основе многослойной нейронной сети прямого распространения, оптимизирована её архитектура для получения наилучшего результата и выполнена её программная реализация. Это позволило получить универсальный метод улучшения качества речевого сигнала путем устранения фонового шума с помощью методов машинного обучения, который показывает высокую эффективность независимо от внешних условий.

Литература

- [1] R. Crochiere «A weighted overlap-add method of short-time Fourier analysis/Synthesis» // IEEE Transactions on Acoustics, Speech, and Signal Processing (Volume: 28, Issue: 1, Feb 1980).
- [2] A. W. Rix, J. G. Beerends, M. P. Hollier, and A. P. Niekstra, «Perceptual evaluation of speech quality (PESQ) - a new method for speech quality assessment of telephone networks and codecs» // IEEE International Conference on Acoustics, Speech, and Signal Processing, 2001, vol. 2, pp. 749–752.

Распознавание солирующего инструмента в полифонических музыкальных произведениях

Лысенко Илья Игоревич

Кафедра автоматизации научных исследований

e-mail: s0217052@stud.cs.msu.ru

Научный руководитель — д.ф.-м.н. проф. Шишкин Алексей Геннадиевич

Музыка является широко распространенным явлением в современном мире. Люди слушают, исполняют её или занимаются её созданием. Поэтому особую актуальность приобретают задачи, связанные с поиском и аннотированием музыкальной информации.

Распознавание солирующего инструмента относится к классу таких задач. Задачи по распознаванию инструмента по изолированным нотам и по сольным выступлениям уже многократно решались различными способами, распознавание же музыкальных инструментов в мультиинструментальной музыке в данный момент является сложной проблемой, полного решения которой до сих пор не существует.

В данной работе рассматривается проблема распознавания отдельного музыкального инструмента, играющего ведущую партию на фоне других инструментов с помощью его спектральных особенностей с применением глубоких нейронных сетей.

В качестве характерных признаков для классификации были выбраны мел-спектрограмма и мел-частотные кепстральные коэффициенты (MFCC). Оба этих варианта показывают хорошие результаты в задачах, связанных с обработкой звуковых сигналов.

Аудиозаписи для обучения и тестирования брались из музыкальной базы iTunes. Все они имеют размеченный солирующий музыкальный инструмент. В качестве классификатора выбрана сверточная нейронная сеть.

Всего было проведено несколько экспериментов:

1. Нейросетевой классификатор для 5-ти инструментов с мел-спектрограммой в качестве характерного признака, обученный на 4-х разных наборах данных (правый, левый каналы, их сумма и разность)
2. Нейросетевой классификатор для 5-ти инструментов с MFCC в качестве характерного признака, обученный на 4х разных наборах данных (правый, левый каналы, их сумма и разность)
3. Нейросетевой классификатор для 6-ти инструментов с мел-спектрограммой в качестве характерного признака, обученный на полусумме каналов
4. Нейросетевой классификатор для 7-ти инструментов с мел-спектрограммой в качестве характерного признака, обученный на полусумме каналов
5. Классификатор случайный лес с MFCC в качестве входных данных

После многочисленных экспериментов с гиперпараметрами модели были получены следующие результаты:

	Виолончель	Гитара	Скрипка	Фортепиано	Саксофон
Мел-спектр	68.92%	79.87%	81.78%	69.31%	72.00%
MFCC	66.63%	80.21%	79.55%	59.92%	69.12%
Случайный лес	56.97%	65.21%	69.23%	53.20%	57.50%

Табл. 1: Результаты.

Как можно видеть из полученных результатов, нейросетевой классификатор на основе мел-спектрограммы показывает результаты лучше, чем нейросеть, входом которой служили мел-частотные кепстральные коэффициенты, а также случайный лес. Разработанная модель может быть использована для разметки и аннотации различных музыкальных баз данных. Информация о солирующих инструментах может оказаться ключевой для систем рекомендаций и автоподбора музыкальных произведений.

Литература

- [1] Рабинер, Голд Теория и применение цифровой обработки сигналов. Издательство Мир, 1978. 848 с.
- [2] Николенко С. И. Глубокое обучение. Спб.: Издательство Питер, 2018. 480 с.

Оптимизация параметров численного метода для модели экологических сообществ

Куркин Максим Леонидович

*Кафедра общей математики
e-mail: maxkurkin@inbox.ru*

Научный руководитель — к.ф.-м.н. доц. Никитин Алексей Антонович

В данной работе рассматривается модель самоструктурирующихся биологических сообществ, предложенная У. Дикманом и Р. Лоу в [1, 2], и является

продолжением развития тематики сосуществования экологических сообществ, описанной Б. Болкером и С. Пакала в [3] Ее преимуществом является использование информации о пространственной структуре сообщества. Модель учитывает поведение каждого индивида и его влияние на всю популяцию в целом, что позволяет более эффективно описывать динамику популяции и взаимодействие видов во многовидовых сообществах между собой.

Модель описывается следующими параметрами, тремя вещественными и двумя функциональными:

- $b \in \mathbb{R}, b > 0$ — интенсивность рождаемости
- $d \in \mathbb{R}, d \geq 0$ — агрессивность среды
- $d' \in \mathbb{R}, d' \geq 0$ — интенсивность конкуренции
- $m(x) : \mathbb{R}^n \rightarrow \mathbb{R}$ — ядро рождаемости
- $w(x) : \mathbb{R}^n \rightarrow \mathbb{R}$ — ядро смертности

В качестве ядер было предложено использовать гауссианы с нулевым смещением и дисперсиями σ_m, σ_w для ядер рождаемости и смертности соответственно.

Главным предметом изучения в текущей работе являются параметры замыкания третьего пространственного момента, приводящего к интегральному уравнению, которое описывает стационарное положение одновидового биологического сообщества. Данное уравнение решается далее численно. Оно естественным образом возникает при описании динамики популяции индивидов с использованием так называемых пространственных моментов: средней плотности популяции N на рассматриваемой области (первый момент); плотности пар, находящихся на расстоянии ξ друг от друга $C(\xi)$ (второй момент); плотности троек $T(\xi, \xi')$, образующих треугольник со сторонами ξ, ξ' (третий момент). Идея такого описания опирается на тот факт, что в рассматриваемой модели сообщество можно эффективно исследовать при помощи его усредненных пространственных характеристик вместо того, чтобы использовать стохастический процесс.

Состояние равновесия модели можно описать следующей системой интегральных уравнений, введенной в [2] и замыканием третьего момента для неё, предложенном в [4]:

$$\begin{cases} 0 = \frac{dN}{dt} = (b - d)N - d' \int_{\mathbb{R}} C(\xi)w(\xi)d\xi \\ 0 = \frac{dC(\xi)}{dt} = bm(\xi)N + \int_{\mathbb{R}} bm(\xi') C(\xi + \xi') d\xi' - (d + d'w(\xi)) C(\xi) \\ - \int_{\mathbb{R}} d'w(\xi') T(\xi, \xi') d\xi' \end{cases} \quad (1)$$

$$\begin{cases} T_{\alpha, \beta, \gamma}(\xi, \xi') = \frac{1}{\alpha + \gamma} \left(\alpha \frac{C(\xi)C(\xi')}{N} + \beta \frac{C(\xi)C(\xi - \xi')}{N} + \gamma \frac{C(\xi')C(\xi - \xi')}{N} - \beta N^3 \right), \\ s.t. \alpha + \beta + \gamma = 1 \end{cases}$$

Эта система решается численным методом рядов Неймана, сходимость которого обоснована в работах [2,8].

При симуляционном же подходе модель описывается точечным Пуассоновским процессом, рассмотренным подробно в статьях [3,7]. Значения моментов, полученные в ходе симуляций, считаются эталонным описанием модели (N_{sim}, C_{sim}) , однако, для их остановки и получения явного момента времени, при котором достигается состояние равновесия, не существует критериев, поэтому проведение симуляций для одной конфигурации модели может занимать несколько часов и даже суток. Поэтому в исследовании этой модели в [2] было предложено использовать аппроксимацию моментов, являющуюся решением системы (1) с трёхпараметрическим замыканием (2). Для минимизации погрешности аппроксимации решения численным методом был проведён априорный анализ значения ошибки в зависимости от параметров метода α, β и получены доверительные интервалы для лучших значений и области для выбора начального приближения для дальнейшего поиска оптимальных значений параметров, а так же разработан метод проекции градиента для адаптивного поиска оптимальных параметров замыкания в зависимости от заданного эталонного решения. Таким образом, решалась следующая оптимизационная задача:

$$\min_{\alpha, \beta, \gamma} \left\| \sum_{k=1}^m A_{\alpha, \beta, \gamma}(C) - C_{sim} \right\|_2^2 \text{ s.t. } \alpha + \beta + \gamma = 1$$

где m — фиксированное число итераций метода Неймана, оператор $A_{\alpha, \beta, \gamma}$ задаёт одну итерацию метода Неймана и может быть получен из системы (1) в виде

$$A_{\alpha, \beta, \gamma} Q = \frac{\frac{Y\bar{m}}{b-d} - \bar{w} + [\bar{m} * Q] - \frac{b-d}{(\alpha+\gamma)Y} \left(\beta(d' - 1) + (\beta(Q + 1) + \gamma)[\bar{w} * Q] + \gamma[\bar{w}Q * Q] \right)}{\bar{w} + d - \frac{b-d}{(\alpha+\gamma)Y} (\alpha Y - \beta d')},$$

$$Y = \frac{N}{b-d}$$

Метод проекции градиента был протестирован на 18 наборах биологических параметров модели, также был проведен анализ сходимости в зависимости от выбора длины шага. О методе и его сходимости в зависимости от длины шага можно узнать из [9]. Главной подзадачей для реализации метода проекции градиента являлась возможность вычисления градиента описанного целевого функционала. Аналитически вычислять его не представляется возможным ввиду сложности оператора $A_{\alpha, \beta, \gamma}$, а также рекурсивности самого метода Неймана: итоговая функция зависит от параметров замыкания неявно. Для этого было решено использовать библиотеку векторных вычислений PyTorch [5], позволяющую дифференцировать функции по их графовому представлению с помощью правила дифференцирования сложной функции.

Литература

- [1] Dieckmann U., Law R. Moment approximations of individual-based models // The Geometry of Ecological Interactions: Simplifying Spatial Complexity / Ed.

ID	σ_m	σ_w	b	d	d'	Выж.	α	β	γ
1	0.96	0.09	0.2	0.02	0.02	да	0.748	0.231	0.021
2	0.45	0.45	0.2	0.02	0.02	да	0.718	-0.103	0.385
3	0.09	0.96	0.2	0.02	0.02	да	0.412	-0.380	0.968
4	0.96	0.09	0.2	0.02	0.2	да	0.566	0.464	-0.030
5	0.45	0.45	0.2	0.02	0.2	нет	-0.762	2.170	-0.408
6	0.09	0.96	0.2	0.02	0.2	нет	0.343	1.059	-0.402
7	0.96	0.09	0.2	0.02	0.4	да	0.414	0.746	-0.160
8	0.45	0.45	0.2	0.02	0.4	нет	-0.239	1.570	-0.331
9	0.09	0.96	0.2	0.02	0.4	нет	0.365	1.038	-0.403
10	0.2	0.7	0.2	0.02	0.02	да	0.537	-0.314	0.776
11	0.7	0.2	0.2	0.02	0.02	да	0.799	0.242	-0.042
12	0.6	0.6	0.2	0.02	0.02	да	0.778	-0.098	0.319
13	0.2	0.7	0.2	0.02	0.2	нет	0.239	1.220	-0.460
14	0.7	0.2	0.2	0.02	0.2	нет	0.659	0.441	-0.101
15	0.6	0.6	0.2	0.02	0.2	нет	0.618	-0.346	0.727
16	0.2	0.7	0.2	0.02	0.4	нет	0.323	1.086	-0.408
17	0.7	0.2	0.2	0.02	0.4	да	0.488	0.746	-0.234
18	0.6	0.6	0.2	0.02	0.4	нет	-0.328	1.838	-0.510

Таблица 1: Значения биологических параметров, выживаемости вида и полученных оптимальных значений α, β, γ — параметров замыкания $T_{\alpha, \beta, \gamma}$

by U. Dieckmann, R. Law, J. Metz. Cambridge University Press. P. 252–270. **2000.**

- [2] Dieckmann U., Law R. Relaxation projections and the method of moments // The Geometry of Ecological Interactions: Simplifying Spatial Complexity / Ed. by U. Dieckmann, R. Law, J. Metz. Cambridge University Press. P. 412–455. **2000.**
- [3] Bolker B., Pacala S. Using moment equations to understand stochastically driven spatial pattern formation in ecological systems// Theor. Population Biol. 52. №3. P. 179–197. **1997.**
- [4] Murrell D. J., Dieckmann U. On moment closures for population dynamics in continuous space // J. Theor. Biology. **229.** PP. 421–432. **2004.**
- [5] Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library. // Advances in Neural Information Processing Systems 32 [Internet]. Curran Associates, P. 8024–35, **2019.**
- [6] Николаев М.В., Никитин А.А. Исследование интегрального уравнения равновесия с ядрами-куртозианами в пространствах различных размерностей // Вестник Московского университета. Сер 15: Вычисл. матем. и киберн. №3. С. 11–19. **2018.**

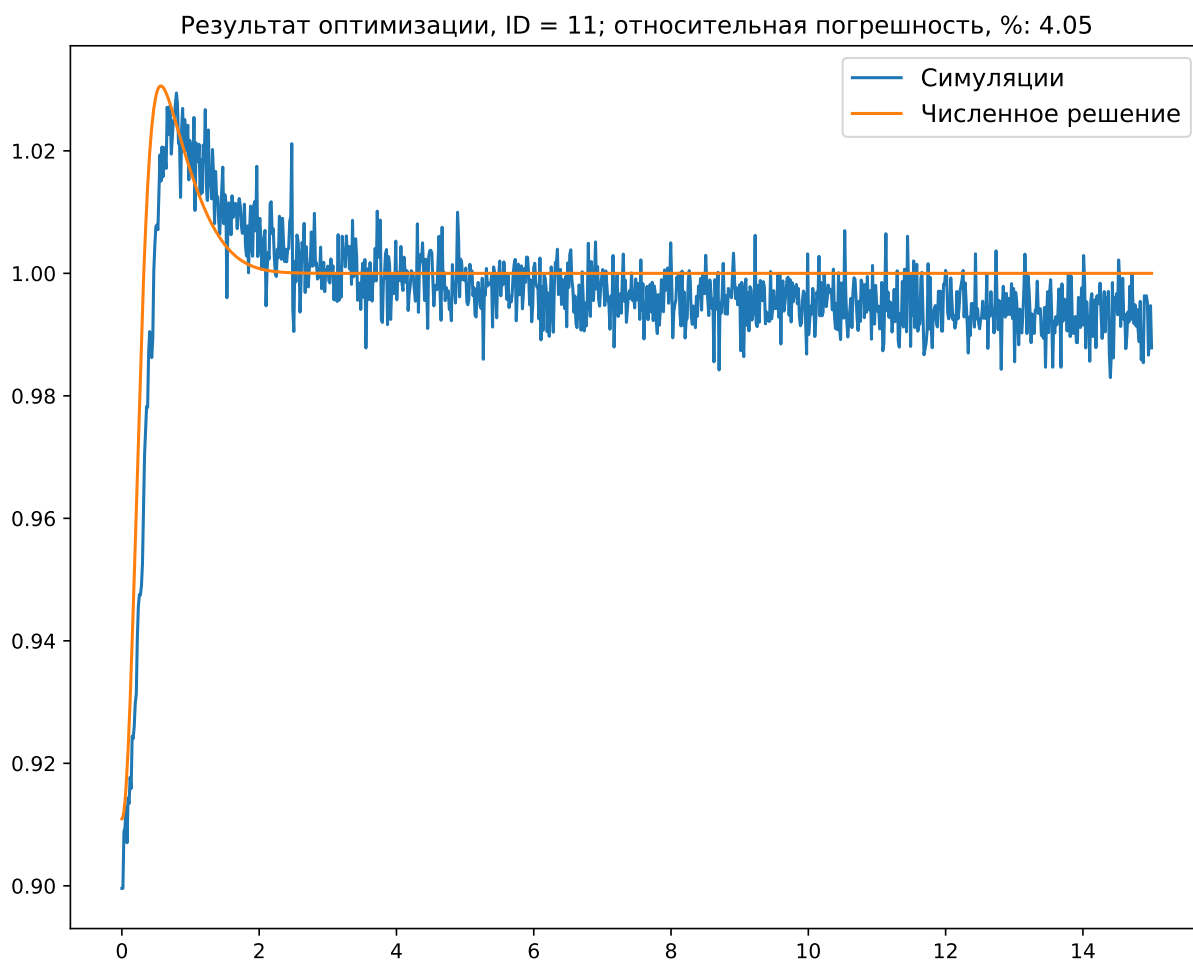


Рис. 1: Пример работы метода проекции градиента: сравнение эталонного значения второго момента $C_{sim}(\xi)$ и $C_{\alpha,\beta,\gamma}(\xi)$, полученного в результате работы метода рядов Неймана с подобранными методом проекции градиента параметрами α, β, γ .

- [7] Галкин Е.Г., Никитин А.А. Стохастическая геометрия для моделирования популяционной динамики: модель Дикмана с неподвижными особями Вестник Московского университета. Сер. 15: Вычисл. матем. и киберн., изд-во Моск. ун-та (М.), № 2, с. 11–18 **2020**.
- [8] Николаев М.В., Дикман У., Никитин А.А. Применение специальных функциональных пространств к исследованию нелинейных интегральных уравнений, возникающих в равновесной пространственной логистической динамике. // Доклады Академии Наук(в печати), **2021**
- [9] Nocedal J., Wright S. Numerical Optimization, PP. 30–65, **2006**.

Исследование эффективности алгоритмов решения задачи о оптимальном разбиении графа

Селиванов Антон Сергеевич

Кафедра исследования операций

e-mail: a.selivanov@yahoo.com

Научный руководитель — д.ф.-м.н. доц. Посыпкин Михаил Анатольевич

Графы естественным образом возникают в описании компьютерных сетей, социальных сетей, транспортных сетей, цифровых изображений. Редукция графа реального размера в меньший граф существенно упрощает его анализ [1], напр. позволяет его визуализировать, декомпозировать, рассчитывать на параллельном компьютере.

Сама по себе задача разбиения графа относится к классу NP-полных. И получение оптимальных решений при количестве вершин графа более 10^3 как правило невозможно за приемлемое время. Поэтому на практике ограничиваются субоптимальными решениями, полученными с использованием эвристик [2].

В исследовании рассмотрены 16 программ (алгоритмов-солверов), выполняющих разбиение. Из них отобраны 7 для исследования эффективности. Было проведено более 10 тыс. тестовых запусков солверов с различными параметрами разбиения на 100 графах конкурса DIMACS Implementation Challenge — Graph Partitioning and Graph Clustering [3]. Разрезы по характеристикам графов (количество вершин, рёбер, взвешенность) и разбиений (количество партий) обобщались в «точки», в которых солверы можно сравнивать между собой по эффективности (продолжительности счёта). Показано, что явного аутсайдера среди солверов нет — каждый был лучше остальных в отдельных случаях. Но и преобладающего алгоритма тоже нет — никто не выиграл хотя бы в половине разрезов.

Попытки вручную выделить наборы условий, благоприятствующих выбору того или иного алгоритма приводили к тому, что их число не снижалось менее нескольких десятков, границы оставались условными и имели исключения. Это неудобно в использовании. Поэтому было предложено построить метасолвер на основе поиска лучшего солвера в заданной локальной области. Так, к желаемым характеристикам разбиения (центр области) добавляется один параметр — допустимая доля вариации координат (радиус области). Показано, что радиус удобно выбирать достаточно большим, чтобы область включала точки с известной производительностью солверов, и одновременно достаточно малым, чтобы тенденции можно было приближать линейными функциями. В качестве таких функций в работе использовались простая средняя и решение модели множественной линейной регрессии.

Для верификации предложенного подхода был составлен эксперимент — из локальной области принудительно выкалывался центр, чтобы даже известный «правильный ответ» не мог повлиять на предсказание метасолвера. Затем обе

версии метасолвера запускались на всех ранее протестированных разбиениях и предсказанные ими солверы были сравнены с фактическими солверами—победителями. Оказалось, что метасолвером на основе средних удобно подбирать подходящий радиус, а метасолвер на основе регрессии правильно определил солвера—победителя в 2/3 всех точек. Для сравнения — максимальная доля точек, в которой «победил» один и тот же солвер — 1/3.

В результате исследования были рассмотрены особенности ряда практических алгоритмов решения задачи о оптимальном разбиении графа, определена их сравнительная эффективность, предложен подход по автоматическому поиску лучшего решения, выполнена его программная реализация. Тестирование показало существенный рост эффективности в сравнении с выбором отдельного солвера.

Литература

- [1] Graph Partition. URL: https://en.wikipedia.org/wiki/Graph_partition
- [2] Разбиение графа. URL: https://ru.wikipedia.org/wiki/Разбиение_графа
- [3] Graph Partitioning and Graph Clustering. 10th DIMACS Implementation Challenge. URL: <https://www.cc.gatech.edu/dimacs10/index.shtml>

Вариационный метод и асимптотические разложения в теории среднего поля Стонера

Гуленко Анна Сергеевна

Кафедра оптимального управления

e-mail: s02170011@gse.cs.msu.ru

Научный руководитель — д.ф.-м.н. доц. Мельников Николай Борисович

Теория среднего поля Стонера позволяет качественно исследовать магнитные свойства металлов при конечных температурах (см., напр., [1, 2]). Целью работы является исследование асимптотики намагниченности в модели Стонера при низких температурах.

Модель Стонера описывает металл как систему взаимодействующих электронов, каждый из которых обладает магнитным моментом (спином). С помощью вариационного принципа парное взаимодействие электронов заменяется взаимодействием со «средним полем», которое пропорционально среднему спину (намагниченности). Средний спин \bar{s} и среднее количество электронов n_e вычисляются по формулам

$$\bar{s} = \frac{1}{2}(n_{\uparrow} - n_{\downarrow}), \quad n_e = n_{\uparrow} + n_{\downarrow}, \quad (1)$$

где n_{\uparrow} и n_{\downarrow} — среднее количество электронов со спином 1/2 и -1/2, соответственно. Для системы невзаимодействующих электронов в среднем поле удаётся вычислить среднее число электронов с заданным спином:

$$n_{\sigma} = \int_{-\infty}^{+\infty} \nu_{\sigma}(\varepsilon, T) f(\varepsilon) d\varepsilon. \quad (2)$$

Здесь $\nu_\sigma(\varepsilon, T)$ — плотность электронных состояний, в которых могут находиться электроны со спином σ и энергией ε при температуре T , а

$$f(\varepsilon) = \left(1 + e^{(\varepsilon - \mu)/T}\right)^{-1},$$

— функция Ферми (μ — химический потенциал). Значение $\varepsilon_F = \mu(T = 0)$ называется уровнем Ферми. Плотности $\nu_\sigma(\varepsilon, T)$ вычисляются при помощи неполяризованной плотности состояний при нулевой температуре $\nu(\varepsilon)$:

$$\nu_\uparrow(\varepsilon, T) = \nu(\varepsilon + U\bar{s}), \quad \nu_\downarrow(\varepsilon, T) = \nu(\varepsilon - U\bar{s}),$$

где U — константа взаимодействия. Таким образом, величины n_\uparrow и n_\downarrow , заданные формулой (2), зависят от среднего спина \bar{s} и химического потенциала μ . Поэтому выражения (1) образуют систему уравнений с неизвестными \bar{s} и μ при каждом значении параметра T . Входными данными модели являются $\nu(\varepsilon)$, ε_F и U .

В работе получена асимптотика $\bar{s}(T)$ при $T \rightarrow 0$ для различных входных плотностей состояний $\nu(\varepsilon)$ и проведено сравнение этих асимптотических разложений с результатами численных расчётов.

Основной теоретический результат работы — асимптотика $\bar{s}(T)$ для случая, когда плотность $\nu(\varepsilon)$ имеет вид конечного прямоугольника. Этот случай обобщает результат для бесконечной прямоугольной плотности $\nu(\varepsilon)$, рассмотренной в книге [1].

Теорема. Пусть плотность состояний при нулевой температуре задаётся функцией

$$\nu(\varepsilon) = \begin{cases} \nu, & \varepsilon \in [0, 1], \\ 0, & \varepsilon \notin [0, 1]. \end{cases}$$

Тогда необходимое условие существования ненулевого решения уравнений Стонера — $\nu = 1/U$. В этом случае при $T \rightarrow 0$ имеет место разложение

$$\bar{s}(T) = \bar{s}_0 - \frac{T}{2U} e^{-\alpha_0/T} + \dots, \quad \alpha_0 = \text{const} > 0. \quad (3)$$

Для плотности $\nu(\varepsilon)$ в виде прямоугольника был выполнен численный расчёт зависимости $\bar{s}(T)$. Разница между рассчитанным средним спином и асимптотикой (3) составляет менее 1% для температур в пределах 20% от критической.

Прямоугольная плотность $\nu(\varepsilon)$ представляет собой вырожденный случай: она имеет нулевую производную в окрестности ε_F . Для типичных плотностей состояний $\nu'(\varepsilon_F) \neq 0$. В этом случае средний спин при $T \rightarrow 0$ описывается квадратичным законом (подробнее см. [2]):

$$\bar{s} = \bar{s}_0 - \alpha T^2 + \dots, \quad \alpha = \text{const} > 0. \quad (4)$$

Значение α зависит от производной $\nu'(\varepsilon)$ в окрестности ε_F . На примере плотности $\nu(\varepsilon)$ в виде эллипса показано, что рассчитанная численно зависимость $\bar{s}(T)$ действительно соответствует асимптотике (4).

Таким образом, вид асимптотического разложения $\bar{s}(T)$ при $T \rightarrow 0$ тесно связан с формой кривой $\nu(\varepsilon)$ и зависит от $\nu'(\varepsilon)$ в окрестности ε_F . Поскольку для типичной $\nu(\varepsilon)$ эта производная ненулевая, квадратичный закон (4) лучше согласуется с экспериментальными данными, чем асимптотика (3).

Численные расчёты проводились при помощи программного комплекса MAGPROP [3], написанного на языке Fortran 90. В ходе работы была модифицирована и дополнена часть комплекса, отвечающая за расчёт в модели Стонера.

Литература

- [1] Mohn P. Magnetism in the solid state. Leipzig : Springer, 2006. 229 p.
- [2] Melnikov N.B., Reser B.I. Dynamic spin fluctuation theory of metallic magnetism. Berlin : Springer, 2018. 287 p.
- [3] Резер Б. И., Парадеженко Г. В., Мельников Н. Б. Программный комплекс «MAGPROP 2.0». № 2018617208, РОСПАТЕНТ, 2018.

Разработка алгоритмов управления для нелинейной модели сердечно-сосудистой системы

Зайцева Дарья Сергеевна

*Кафедра нелинейных динамических систем и процессов управления
e-mail: s02170521@cs.msu.ru*

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

Введение. В данной работе реализован метод внутриаортальной контрпульсации (ВБК), используемый для лечения сердечной недостаточности. Принцип работы метода заключается в том, что введенный в аорту баллон сжимается и расширяется в соответствии с фазами сердечного цикла, обеспечивая временную поддержку насосной функции сердца.

Математическая модель. Моделируемая по модели Ю.В. Солодянникова [1] СК может находиться в двух состояниях (систола - сокращение желудочков и изгнание крови, диастола - расслабление желудочков и заполнение их кровью) и математически описывается переключаемой системой с вектором состояния $x = (x_1, \dots, x_{10})$, вектором параметров $A = (A_1, \dots, A_{35})$:

$$\begin{cases} \dot{x} = F^\sigma(x, u, A) \\ y = h^\sigma(x) \\ u = V_{BAL}(t) \end{cases} \quad x(0) = x_0, y \in R^1$$

Считаем, что значение $\sigma = 1$ – определяет систолу, $\sigma = 2$ – диастолу, и условия переключений задаются следующим образом (если ни одно из условий не выполнено, то σ не меняется):

$$\sigma : \begin{cases} 1, D_1(x, A) = 0, \\ 2, D_2(x, A) = 0, \end{cases} \quad \sigma(x_0, 0) = 1$$

Начальные условия в момент перехода задаются линейными зависимостями от x .

Управление контрпульсационным баллоном. Управление $V_{BAL}(t)$ задается законом изменения объема внутриаортального баллона от двух параметров V_{min}, V_{max} (минимальный и максимальный объем баллона) в пределах одного сердечного цикла:

$$V_{BAL}(t) = \begin{cases} V_{min}, & 0 \leq t \leq t_0 \\ V_{min} + (V_{max} - V_{min}) \frac{t-t_0}{t_1-t_0}, & t_0 < t \leq t_2 \\ V_{max}, & t_2 < t \leq t_3 \\ V_{max} - (V_{max} - V_{min}) \frac{t-t_2}{t_1-t_2}, & t_3 < t \leq t_1 \\ V_{min}, & t_1 < t \leq T \end{cases}$$

$t_i (i = \overline{0, 3})$ - переменные времени, обновляемые на каждой итерации сердечного цикла (T - его длительность).

В задачу управления входит выбор пары V_{min}, V_{max} из множества возможных значений так, чтобы модель верно имитировала переходные реакции СК при подключении баллона.

Для задания функционала качества J определим функции ([2]):

$$TTI = \int_0^{T_S} \phi_1(x, A, u) dt,$$

$$DPTI = \int_{T_S}^T \phi_1(x, A, u) dt,$$

$$E = \frac{DPTI}{TTI},$$

$$J = \max_{u \in U} (E - E_{bad})^2$$

$DPTI$ и TTI – индексы «напряжение – время», E – критерий энергообеспеченности сердца, T_S - длительность систолы текущего сердечного цикла, ϕ_1 - некоторая функция из основной системы, а E_{bad} обозначает критерий энергообеспеченности сердца в больном состоянии.

Формируется сетка, состоящая из пар (V_{min}, V_{max}) , ограниченных макс. и мин. возможным объемом баллона, и запускается алгоритм перебора по ней для нахождения оптимальной (в смысле описанного функционала) пары параметров.

Заключение. По итогу работы над поставленными задачами были достигнуты следующие цели:

- Реализован программный продукт, имитирующий кровеносную систему в средах разработки Matlab и Simulink.
- Смоделирован конкретный режим сердечного заболевания (инфаркт миокарда).
- Внедрен в модель контрпульсационный баллон.

- Реализован алгоритм подбора наилучшей пары параметров баллона для снижения нагрузки на сердце.

Литература

- [1] Солодянников Ю.В. Элементы математического моделирования и идентификация системы кровообращения // Самара: Самар. ун-т, 1994. 315 с.
- [2] А. П. Прошин, Ю. В. Солодянников. Математическое моделирование системы кровообращения и его практические применения // Автомат. и телемех., 2006, выпуск 2, 174–188.
- [3] Т. И. Остапенко, А. П. Прошин, Ю. В. Солодянников. Исследование идентифицируемости системы кровообращения // Автомат. и телемех., 2007, выпуск 7, 132–150.

Сетевые модели экономического роста

Акимова Евгения Дмитриевна

Кафедра системного анализа

e-mail: jane_muse@mail.ru

Научный руководитель — чл.-кор. РАН, д.ф.-м.н. профессор Шананин Александр Алексеевич

В данной работе проводился анализ межотраслевых связей и выделялись центральные отрасли российской экономики, определяющие эффективность этих связей. Похожий анализ проводился для экономики США, например, в статье [1]. В ней авторы использовали два подхода к нахождению центральных отраслей:

- С помощью связанных ориентированных графов, в которых вершины обозначают акторов, а дуги-связи между ними. В терминах экономической системы акторами являются отрасли, а дугами - поставки между отраслями. При этом, следуя работе [2], учитывается не только количество связей определенной отрасли с другими отраслями, но и рассматривается число вершин в окрестности второго порядка относительно нее.
- С помощью модели межотраслевого баланса, в которой авторы отошли от стандартной гипотезы Леонтьева о постоянстве норм затрат на выпуск продукции и стали учитывать взаимозаменяемость товаров и услуг.

Поскольку во многих крупных российских городах сейчас наблюдается тенденция к переходу от постоянного наращивания объемов производства к увеличению разнообразия товаров, то проведение схожего анализа для экономики России является актуальной задачей.

В своей работе я использовала симметричную таблицу межотраслевых связей России за 2016 год [3] и рассмотрела оба способа нахождения центральных отраслей. Сначала был построен связанный ориентированный граф, и были найдены центральные отрасли с точки зрения связей первого и второго порядка. Затем я рассмотрела более аналитичный подход с применением модели

межотраслевого баланса. С математической точки зрения построение модели означало решение обратной задачи, т.е. на основе нелинейного межотраслевого баланса строилась задача выпуклого программирования, решение которой воспроизводило симметричную таблицу межотраслевого баланса. При этом было рассмотрено влияние как фиксированного шока, так и случайного, имеющего логнормальное распределение, на изменение ВВП. Теоретические обоснования фактов, используемых в моей работе, приведены в статьях [4, 5].

Помимо нахождения центральных отраслей мною также был выделен наиболее влиятельный комплекс отраслей. Для этого целесообразно было разбить секторы на 4 группы:

- a) секторы с экспортным потенциалом(нефть, каменный уголь и пр.)
- b) секторы, конкурирующие с импортом(автомобилестроение, текстиль, изделия из кожи и пр.)
- c) инфраструктурные секторы(услуги по распространению электроэнергии, тепловая энергия, услуги по распределению воды и пр.)
- d) сфера услуг(торговля, ЖКХ и пр.)

Предполагалось, что в каждой группе отраслей одинаковая дисперсия.

Для проведения исследования была написана программа на языке python. С помощью обоих способов были получены непротиворечивые результаты, и стоит отметить положительные черты каждого из них:

Метод графов более наглядный и простой, а также позволяет рассмотреть связи второго порядка.

Тем не менее *метод межотраслевых балансов* предоставляет гораздо больше возможностей для исследования экономической системы. С его помощью можно:

- регулировать масштаб изменений каждой отрасли
- рассматривать агрегированные группы отраслей

Таким образом, оба метода помогают выделить ключевые отрасли экономики, однако, конкретизируя понятие центральности, можно выбирать тот или иной метод.

Литература

- [1] D. Acemoglu, A. Ozdaglar, A. Tahbaz-Salehi. The network origins of aggregate fluctuations. // *Econometrica*, 2012, v.80, №5, 1977-2016 p
- [2] Phillip Bonacich. Power and Centrality: A Family of Measures. *American Journal of Sociology*, Vol. 92, No. 5 (Mar., 1987), pp. 1170-1182
- [3] Url: <https://rosstat.gov.ru/accounts>
- [4] А. В. Рассоха, А. А. Шананин, “Обратные задачи анализа межотраслевых балансов”, *Матем. моделирование*, 33:3 (2021), 39–58
- [5] А. А. Шананин. Двойственность по Янгу и агрегирование балансов // Доклады РАН. Математика, информатика, процессы управления, 2020, т.493, с.81-85. DOI: 10.31857/S2686954320040177

Моделирование нефтяной отрасли на основе концепции игры среднего поля

Воробьева Ирина Борисовна

Кафедра системного анализа

e-mail: vib01@mail.ru

Научный руководитель — чл.-корр. РАН д.ф.-м.н. проф. Шананин Александр Алексеевич

В данной работе рассматриваются модели поведения агентов, производящих сланцевую нефть, так как их доля на рынке нефти с каждым годом растет, и их влияние необходимо учитывать. Для рассматриваемых агентов можно считать выполненными следующие предположения:

1. число агентов велико;
2. число агентов фиксировано;
3. выполнены условия идеальной конкуренции.

В представленной работе изучены три различных подхода к моделированию характеристик рынка исчерпаемых ресурсов. Первый подход использует модель, разработанную Харольдом Хотеллигом в 1931 году [2]. Его модель основана на предполагаемом характере изменения цен в условиях совершенной конкуренции и функции спроса специального вида.

Другие две модели основаны на концепции игр среднего поля: они позволяют моделировать поведение множества малых агентов, неспособных повлиять на поведение системы в целом. В рамках игр среднего поля основными считаются два уравнения в частных производных: уравнение Гамильтона—Якоби—Беллмана и уравнение Колмогорова—Фоккера—Планка. Первое позволяет моделировать поведение агентов, а второе учитывать изменение связи между ними.

Впервые концепция игр среднего поля для изучения агентов на рынке нефти была применена Пьером—Луи Лионсом, Жаном—Мишелем Ласре и Оливье Гейном в работе [1]. Была построена модель и получены соответствующие уравнения, связанные между собой через значение цены. В рамках данной модели предполагается, что агенты рациональны, то есть цена представляет собой функцию, заданную в начальный момент времени, значения которой известны агентам. Таким образом, зная будущие значения цен, агенты могут определять оптимальные объемы производства.

В рассматриваемой работе предлагается изучить случай ограниченной рациональности агентов: значения цен в будущем неизвестны, а цена является параметром, моделируемым вместе с остальными характеристиками, а не заданным ранее. Проводя соответствующие изменения была получена система, описывающая поведение таких агентов. Показано, что поведение характеристик рынка, таких как значение цен, объемы производства, момент полной остановки производства, при переходе от одной модели к другой существенно изменяется.

Также в работе было проведено исследование возможных способов внешнего регулирования, позволяющих отсрочить момент полного исчерпания ресурса. Такими мерами стали введение ограничения на значения цен и введение налога на прибыль. В работе приведены полученные результаты моделирования таких ограничений для трех различных подходов.

Литература

- [1] Guéant O., Lasry J.—M., Lions P.—L. Mean field games and applications in Paris-Princeton Lectures in Quantitative Finance, 2009.
- [2] Хотеллинг Х. Экономика исчерпаемых ресурсов // Вехи экономической мысли. Т. 3 / Под ред. В. М. Гальперина. — СПб.: Экономическая школа. 2000 (англ. The economics of exhaustible resources, 1931)
- [3] Kurzhanski A. B., Varaiya P. Dynamics and Control of Trajectory Tubes. Theory and Computation, Birkhauser, 2014.
- [4] Тихонов В. И., Миронов М. А. Марковские процессы. М., «Сов. радио», 1977.
- [5] Øksendal B., Stochastic differential equations, Springer, 1991.

Оценка параметров модели Самуэльсона на основании дневной статистики торгов

Карпикова Полина Владимировна

Кафедра системного анализа

e-mail: p.karpikova@mail.com

Научный руководитель — к.ф.-м.н. доц. Смирнов Сергей Николаевич

Для описания динамики цен на финансовом рынке используются различные модели. Одной из популярных является модель Самуэльсона, которая описывает цену актива X_t стохастическим процессом:

$$X_t = X_0 \exp[\sigma(\mu t + W_t)], \quad t \in [0, T], \quad (1)$$

где $(W_t)_{t \in [0, T]}$ — винеровский процесс, $\sigma\mu$ — коэффициент сноса, $\sigma > 0$ является постоянным параметром волатильности. В качестве статистики используются четыре величины: цены в начале дня, максимальной и минимальной цены за период, и цены в конце дня. Целью данной работы было проверить адекватность модели и оценить ее параметры.

Для получения оценки использовался метод максимального правдоподобия. Необходимая плотность не была задана явно, она была задана как функция распределения по одним переменным и как плотность по другим, и представлялась в виде ряда. Требуемая оценка была получена численно, с необходимым обоснованием, в частности, касающимся почленного дифференцирования и оценки погрешности при замене ряда на частичную сумму. Так как истинное значение параметра волатильности для реальных статистических данных неизвестно, то сравнение оценки с некоторыми известными оценками из работ [3],

[4], [5] было проведено на данных, полученных методом статистического моделирования (Монте–Карло). Оно показало, что данная оценка имеет наименьшую дисперсию среди сравниваемых оценок.

Для проверки сложной гипотезы о том, что цены подчиняются модели Самуэльсона, был использован критерий Пирсона для выборок, у которых параметры оценены по выборке. Множество значений наблюдаемых величин делится на области и вычисляются истинные вероятности областей и частоты попадания. Процедура разбиения на области и подсчет их вероятностей описаны в работе.

Литература

- [1] Бородин А. Н., Салминен П. Справочник по Броуновскому движению. Факты и формулы: Учебное пособие. Лань, 2016. 702 с.
- [2] Крамер Г. Математические методы статистики. Мир, 1975. 648 с.
- [3] Parkinson M. The extreme value method for estimating the variance of the rate of return // Journal of Business. 1980. N 53. P.61–68.
- [4] Garman M., Klass M. On the estimation of security price volatilities from historical data // Journal of Business. 1980. N 53. P.67–78.
- [5] Rogers L.C.G., Satchell S.E. Estimating variance from high, low and closing prices // Annals of Applied Probability. 1991. N 1. P.504–512.

Решение нелинейных уравнений в окрестности анормальной точки

Салихова Кария Илшатовна

Кафедра системного анализа

e-mail: s02170204@gse.cs.msu.ru

Научный руководитель – д.ф.-м.н. профессор Арутюнов Арам Владимирович

Пусть заданы отображение $G : \mathbb{R}^n \times \Sigma \rightarrow \mathbb{R}^m$, где Σ – это нормированное пространство с нормой $\|\cdot\|$, и точка (x_0, σ_0) такая, что $G(x_0, \sigma_0) = 0$. В работе исследуются условия на отображение G , при которых существует решение уравнения

$$G(x, \sigma) = 0, \tag{1}$$

относительно неизвестного x и параметра σ из некоторой окрестности $U(\sigma_0)$. Рассматривается случай, когда точка x_0 не является нормальной для отображения $G(\cdot, \sigma_0)$, то есть когда классическая теорема о неявной функции (см., например, [1, 2, 3]) неприменима.

Напомним конструкции и методы, предложенные в [4]. Обозначим через D множество неотрицательных n -мерных векторов d , а через $\widehat{D} \subset D$ – множество целочисленных векторов s . Пусть задан неотрицательный n -мерный вектор λ . Пусть также задано непустое конечное множество $S_i \subset \widehat{D}$, для которого выполняется

$$\exists \alpha_i > 0 : \langle \lambda, s \rangle = \alpha_i \quad \forall s \in S_i. \tag{2}$$

Обозначим через \mathcal{S} семейство множеств $S_i, i \in \{1, \dots, m\}$. Пусть также для каждого $i \in \{1, \dots, m\}$ заданы вещественные числа $p_{i,s}$, где $s \in S_i$. Множество этих чисел обозначим через \mathcal{P} .

Зададим отображение $P = (P_1, \dots, P_m) = P^{\mathcal{S}, \mathcal{P}} : \mathbb{R}^n \rightarrow \mathbb{R}^m$ следующим образом:

$$P_i(x) = \sum_{s \in S_i} p_{i,s} x^s, \text{ где } x^s = \prod_{k=1}^n x_k^{s_k}.$$

Определение 1. *Отображение $P = P^{\mathcal{S}, \mathcal{P}}$ называется λ -укорочением отображения $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$ в окрестности точки x_0 , если для всех $i \in \{1, \dots, m\}$ существует $D_i \subset D$ такое, что*

$$\langle \lambda, d \rangle > \langle \lambda, s \rangle \quad \forall d \in D_i, \quad \forall s \in S_i,$$

и для всех x из некоторой окрестности точки x_0 для F справедливо представление

$$F(x) = F(x_0) + P(x - x_0) + \Delta(x - x_0),$$

где вектор-функция Δ такова, что для всех x из некоторой окрестности x_0 выполняется оценка

$$|\Delta_i(x)| \leq \text{const} \sum_{d \in D_i} |x|^d, \text{ где } |x|^d = \prod_{k=1}^n |x_k|^{d_k}.$$

Определение 2. *Пусть задан вектор $h \in \mathbb{R}^n$. Будем говорить, что λ -укорочение $P = P^{\mathcal{S}, \mathcal{P}} : \mathbb{R}^n \rightarrow \mathbb{R}^m$ регулярно по направлению h , если $P(h) = 0, \text{ im} P'(h) = \mathbb{R}^m$.*

Сформулируем основные результаты исследования.

Теорема 1. *Пусть задан вектор $\lambda > 0$, и $P = P^{\mathcal{S}, \mathcal{P}}$ является λ -укорочением отображения $G(\cdot, \sigma_0)$ в окрестности точки x_0 , которое регулярно по некоторому направлению $h \in \mathbb{R}^n$. Пусть также существует $\kappa > 0$:*

$$|G(x, \sigma) - G(x, \sigma_0)| \leq \text{const} \|\sigma - \sigma_0\|^\kappa \quad \forall (x, \sigma) \in U(x_0, \sigma_0).$$

Тогда для любого $\alpha > \max \{\alpha_1/\kappa, \dots, \alpha_k/\kappa\}$, где α_i взяты из соотношений (2), существует такая окрестность $O(\sigma_0)$ точки σ_0 , что для любого $\sigma \in O(\sigma_0)$ существует решение уравнения (1) $x = x(\sigma) \in \mathbb{R}^n$ такое, что

$$|x_i(\sigma) - x_{0,i}| \leq \text{const} \|\sigma - \sigma_0\|^{\lambda_i/\alpha},$$

где $x_i(\sigma)$ и $x_{0,i}$ – координаты векторов $x(\sigma)$ и x_0 соответственно.

При усиленных предположениях гладкости отображения G справедлива более сильная априорная оценка неявной функции. А именно, имеет место следующее утверждение.

Теорема 2. Пусть выполнены предположения теоремы 1. Пусть также для всех (x, σ) из некоторой окрестности точки (x_0, σ_0) отображение G представимо в следующем виде:

$$G(x, \sigma) = G(x, \sigma_0) + G(x_0, \sigma) + \delta(x, \sigma),$$

причем для любого $\gamma > 0$ существует окрестность точки (x_0, σ_0) , зависящая от γ , такая, что:

$$|\delta(x, \sigma)| \leq \gamma |G(x_0, \sigma)| \quad \forall (x, \sigma) \in U_\gamma(x_0, \sigma_0).$$

Тогда существует такая окрестность $O(\sigma_0)$ точки σ_0 , что для любого $\sigma \in O(\sigma_0)$ у уравнения (1) существует решение $x = x(\sigma) \in \mathbb{R}^n$, для которого справедливы оценки

$$|x_i(\sigma) - x_{0,i}| \leq \text{const} |G(x_0, \sigma)|^{\lambda_i/\alpha}, \quad i = 1, \dots, n.$$

Литература

- [1] Dontchev A. L., Rockafellar R. T. *Implicit Functions and Solution Mappings*, 2nd edn. Springer, New York (2014).
- [2] Алексеев В. М., Тихомиров В. М., Фомин С. В. *Оптимальное управление*, М.: Наука, 1979.
- [3] Арутюнов А. В. Гладкие аномальные задачи теории экстремума и анализа, УМН, 67:3(405) (2012), 3–62.
- [4] Арутюнов А. В. Существование вещественных решений нелинейных уравнений без априорных предположений нормальности. Матем. заметки, 2021, 109:1, 3–18.

Близость моделей Башелье и Самуэльсона для различных метрик

Сотников Дмитрий Михайлович

Кафедра системного анализа

e-mail: dmitrysotni@gmail.com

Научный руководитель — к.ф.-м.н. доц. Смирнов Сергей Николаевич

Рассматриваются простейшие модели рынка с непрерывным временем. В рамках модели Башелье цена $(X_t)_{t \in [0, T]}$ актива на рынке описывается случайным процессом

$$X_t^B = X_0(1 + \alpha t + \sigma_B W_t), \quad t \in [0, T], \quad (1)$$

где $(W_t)_{t \in [0, T]}$ — винеровский процесс, $\alpha \in \mathbb{R}$, $\sigma_B > 0$.

Модель Самуэльсона использует геометрическое броуновское движение:

$$X_t^S = X_0 \exp(\gamma t + \sigma_S W_t), \quad t \in [0, T], \quad (2)$$

где $\gamma \in \mathbb{R}$, $\sigma_S > 0$. Эти модели описаны, например, в [1].

Решается задача вычисления расстояния между случайными величинами X_T^B и X_T^S в метриках Форте–Мурье, полной вариации и Колмогорова. Определения вышеупомянутых метрик можно найти в [2].

С помощью W -функции Ламберта (см. [3]) была получена формула для вычисления метрики Форте–Мурье. Показано, что вычисление метрик полной вариации и Колмогорова в риск-нейтральном случае может быть сведено к решению одного нелинейного уравнения.

Данные результаты были использованы для сравнения цен европейских опционов с липшицевыми или ограниченными функциями выплат. Получены неравенства, позволяющие оценить разность между ценами опциона в моделях (1), (2) с помощью соответствующей вероятностной метрики и нормы (или полунормы) функции выплат в подходящем функциональном пространстве. В отличие от некоторых других работ на данную тему, например [4], эти оценки применимы к достаточно широким классам европейских опционов.

Оценены параметры моделей для рынка нефти, для стандартных и бинарных опционов вычислена чувствительность цены к изменению этих параметров, позволяющая судить о влиянии ошибки оценивания на цену опциона. Показано, что в ряде случаев модели Башелье и Самуэльсона с практической точки зрения можно считать неразличимыми.

Литература

- [1] Ширяев А. Н. Основы стохастической финансовой математики. М.: ФАЗИС, 1998.
- [2] The methods of distances in the theory of probability and statistics / Rachev S. T., Klebanov L. B., Stoyanov S. V., Fabozzi F. G. // Springer, New York, 2013.
- [3] On the Lambert W function / Corless R. M., Gonnet G. H., Hare D. E. G., Jeffrey D. J., Knuth D. E. // Advances in Computational Mathematics. 1996. Vol. 5, N 1. P. 329–359.
- [4] Schachermayer W., Teichmann J. How close are the option pricing formulas of Bachelier and Black–Merton–Scholes // Mathematical Finance. 2005. Vol. 18, N 1. P. 55–76.

Гамильтонов формализм для синтеза групповых наблюдений

Удовиченко Игорь Романович

Кафедра системного анализа

e-mail: s02170232@stud.cs.msu.ru

Научный руководитель — акад. РАН, д.ф.-м.н. заф. каф. Куржанский Александр Борисович

В работе рассматривается задача наблюдения за движением группы (стаи) объектов (агентов). Динамика каждого члена стаи задается линейным нестационарным дифференциальным уравнением:

$$\ddot{x}^{(j)} = \tilde{A}(t)x^{(j)} + \tilde{C}(t)\dot{x}^{(j)} + \tilde{B}(t)u^{(j)}, \quad j = \overline{1, m}.$$

Группа представляет собой множество всех агентов, а состояние группы описывается матрицей, столбцами которой являются состояния агентов. Предполагается, что точные состояния агентов неизвестны, но даны зашумленные

измерения и квадратичная начально-интегральная мера неопределенности:

$$\sum_{j=1}^m \langle \mathbf{x}_0^{(j)}, \tilde{L} \mathbf{x}_0^{(j)} \rangle + \int_{t_0}^t \left(\langle u^{(j)}, \tilde{M}(\tau) u^{(j)} \rangle + \langle v^{(j)}, \tilde{N}(\tau) v^{(j)} \rangle \right) d\tau \leq \mu^2,$$

где $\mathbf{x}_0^{(j)}$ — вектор начального состояния j -го агента, $u^{(j)} = u^{(j)}(\tau)$ — управление, а $v^{(j)} = v^{(j)}(\tau)$ — помеха. μ может быть как известно, так и неизвестно. Требуется построить оценки состояний объектов в конечный момент времени t .

Поставленная задача решается следующим образом. Путем вытягивания матрицы состояний агентов в вектор и перехода к блочно-диагональным матрицам задача сводится к классической линейно-квадратичной задаче гарантированно-го оценивания. Для ее решения привлекается гамильтонов формализм: вводится функция цены, удовлетворяющая уравнению Гамильтона—Якоби—Беллмана. Решение данного уравнения выражается через решение матричного уравнения Риккати. Оценки состояний членов группы получаются проекцией оптимальной оценки на соответствующие подпространства. Также в случае, если мера неопределенности μ известна, становится возможным построение множества достижимости системы, являющегося эллипсоидом.

Литература

- [1] Kurzhanski A. B., Varaiya P. Dynamics and Control of Trajectory Tubes Birkhäuser // Boston, MA. – 2014.
- [2] Куржанский А. Б. Задача управления групповым движением. Общие соотношения // Доклады Академии наук. – Федеральное государственное бюджетное учреждение "Российская академия наук 2009. – Т. 426. – №. 1. – С. 20-25.
- [3] Куржанский А. Б. Гамильтонов формализм в задачах группового управления // Дифференциальные уравнения. – 2019. – Т. 55. – №. 4. – С. 546-554.
- [4] Bellman R. E., Dreyfus S. E. Applied dynamic programming. – Princeton university press, 2015. – Т. 2050.

Оптимальная политика торговли разрешениями на выбросы в динамической игре с трансграничным загрязнением

Федяшин Никита Александрович

Кафедра системного анализа

e-mail: Nikitafedyashin@gmail.com

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

В последние годы политики и ученые сосредоточили внимание на негативном влиянии быстрого технического прогресса и экономического роста, поскольку уровень загрязнения окружающей среды стал угрожающе высок. В частности особую опасность представляет глобальное потепление, вызванное парниковым эффектом от выбросов CO_2 в атмосферу. Так, например, в 2015 году было заключено Парижское соглашение, призванное регулировать меры

по снижению содержания углекислого газа в атмосфере с 2020 года. Существует много методов, которые в теории позволят снизить или хотя бы ограничить уровень парниковых газов в атмосфере. К ним относятся и те, согласно которым всем игрокам выделяются разрешения, которые и определяют сколько тонн CO_2 их владелец может выбросить в атмосферу за определённый период времени.

В работе проведено сравнение моделей, в которых разрешена и запрещена торговля квотами на выбросы. Анализ проводится с точки зрения их влияния на благосостояние игроков и экологию. Так же исследуется поведение цены на квоты в случае, если торговля разрешена.

Постановка задачи. Чтобы понять взаимосвязь между накоплением загрязнения и региональными стратегиями, отношения между регионами моделируются как дифференциальная игра. Цель игроков — максимизировать свою суммарную прибыль за некоторый период времени T в игре с симметричной информацией о трансграничном загрязнении. Они максимизируют прибыль выбирая оптимальный уровень энергозатрат $E_i(t)$ и объёмы борьбы с загрязнением $a_i(t)$. Согласно [1] полагаем, что уровень загрязнения меняется по следующему закону:

$$\dot{P}(t) = \sum_{i=1}^n (E_i(t) - a_i(t)) - \theta P(t),$$

где θ обозначает способность окружающей среды самоочищаться. Предполагаем, что игроки несут финансовые потери $D_i P(t)$, линейно зависящие от текущего уровня загрязнения. Так же игроки тратятся на очистку окружающей среды, покупку дополнительных разрешений, а доход получают от производства. Выручку игроков в зависимости от уровня энергопотребления выразим следующим образом:

$$R_i(E_i(t)) = A_i E_i(t) - B_i \frac{E_i^2(t)}{2}.$$

Затраты на покупку разрешений по цене $S(t)$:

$$S(t) \cdot (E_i(t) - E_{i0}(t) - a_i(t)),$$

где $E_{i0}(t)$ — квоты, выделенные регулятором в момент времени t . Как и в работе [2], будем считать что в конце периода игроки будут оштрафованы, если уровень загрязнения будет слишком велик.

Методы решения. Для решения обеих задач применяется метод динамического программирования. Удаётся найти одно из решений уравнений Гамильтона—Якоби—Беллмана. Это решение позволяет явно выписать выражения для оптимальных управлений всех игроков. Для игры с торговлей, из уравнения баланса удаётся построить выражение для цены на квоты в каждый момент

времени:

$$S(t) = \frac{\sum_{i:\psi_i(t)\geq 0} \left(\frac{A_i}{B_i} + \frac{l_i(t)}{B_i} \right) + \sum_{i=1}^n \frac{l_i(t)}{C_i} - E_{i0}(t)}{\sum_{i:\psi_i(t)\geq 0} \frac{1}{B_i} + \sum_{i=1}^n \frac{1}{C_i}},$$

где C_i отвечают за затраты на очистку окружающей среды i -го игрока, $\psi_i(t) = A_i - S(t) + l_i(t)$, а для $l_i(t)$ известна аналитическая формула.

Результаты. После вывода аналитических выражений было проведено численное моделирование на реальных данных. В качестве игроков рассмотрены 10 стран и регионов. Результаты экспериментов показали что модель с торговлей намного выгоднее для некоторых стран, например России и Японии. Для других же она во всяком случае не хуже своего аналога. Так же модель с торговлей соответствует более низкому индексу Джини, т. е. ведёт к меньшему расслоению между игроками.

Литература

- [1] Bertinelli L., Camacho C., Zou B. Carbon capture and storage and transboundary pollution: a differential game approach // Eur J Oper Res (2014) Vol.237, P. 721–728
- [2] Yeung D., Petrosyan L. A cooperative stochastic differential game of transboundary industrial pollution // Automatica. (2008) Vol. 44, P. 1532–1544.

Критерий диверсифицируемости в терминах ожидаемых средних потерь

Логванёва Мария Геннадьевна

Кафедра математической статистики

e-mail: masha.logvaneva@gmail.com

Научный руководитель — к.ф.-м.н. асс. Целищев Михаил Андреевич

В работе предлагаются несколько определений понятия диверсификации инвестиционного портфеля, где под последним понимается случайных доход (или убыток) от вложения капитала инвестора в один или несколько финансовых активов с фиксированным временным горизонтом.

Определение основано на предположении о том, что выпуклая линейная комбинация набора инвестиционных портфелей X_1, \dots, X_n не хуже, чем вероятностная смесь этих же активов с такими же весами, т. е.

$$\xi \stackrel{\text{div}}{\succ} \eta \iff \eta \stackrel{d}{=} \underset{\beta}{\text{mix}} X \text{ и } \xi \stackrel{d}{=} \sum_{i=1}^n \beta_i X_i, \text{ где } \beta_i \geq 0 \text{ и } \sum_{i=1}^n \beta_i = 1,$$

где под $\underset{\beta}{\text{mix}} X$ понимается случайная величина, которая имеет функцию распределения $\sum_{i=1}^n \beta_i F_{X_i}$.

Иными словами, считается, что распределение средств по разным активам предпочтительнее вложения всего капитала в случайно выбранный актив.

Лемма. Если $\mathbf{E}|\xi| < \infty$, $\mathbf{E}|\eta| < \infty$ и $\xi \succ^{\text{div}} \eta$, то $\mathbf{E}\xi = \mathbf{E}\eta$.

Из леммы следует, что данное определение позволяет сравнивать инвестиционные портфели только с равными математическими ожиданиями, поэтому, для того чтобы иметь возможность сравнивать большее число портфелей, было предложено другое определение диверсификации, которое связано со стохастической доминацией первого порядка:

$$\xi \succ^{\text{1sd}} \xi_0 \iff F_\xi(x) \leq F_{\xi_0}(x) \text{ для всех } x \in \mathbb{R}.$$

Будем говорить, что инвестиционный портфель ξ не хуже инвестиционного портфеля η в смысле диверсификации с положительной надбавкой, если существует такой портфель ξ_0 , что $\xi \succ^{\text{1sd}} \xi_0 \succ^{\text{div}} \eta$.

Далее строится замыкание предложенного отношения предпочтения на пространстве распределений на прямой с конечными первыми моментами по метрике Канторовича, которая задаётся следующим образом:

$$\kappa(\xi, \eta) = \kappa(F_\xi, F_\eta) = \int_{-\infty}^{+\infty} |F_\xi(x) - F_\eta(x)| dx = \int_0^1 |q_p(\xi) - q_p(\eta)| dp,$$

где $q_p(\xi) = \inf \{x \in \mathbb{R} : F_\xi(x) \geq p\}$ — нижняя квантиль распределения случайной величины ξ уровня p .

Оказывается, что построенные выше определения тесно связаны с широко используемой на практике мерой риска Expected Shortfall [1], определяемой как

$$\text{ES}_\alpha(\xi) = -\frac{1}{\alpha} \int_0^\alpha q_p(\xi) dp, \quad \alpha \in (0, 1].$$

Данная мера риска имеет смысл средних убытков по портфелю в худших $\alpha \cdot 100\%$ случаях.

С использованием результатов статьи [2], для итогового определения диверсификации был доказан следующий критерий, который значительно упрощает проверку определения:

Теорема. Пусть математические ожидания портфелей ξ и η конечны, тогда портфель ξ не хуже портфеля η в смысле замыкания отношения диверсификации с положительной надбавкой по метрике Канторовича, если и только если $\text{ES}_\alpha(\xi) \leq \text{ES}_\alpha(\eta)$ для всех $\alpha \in (0, 1]$.

Литература

- [1] Acerbi C., Tasche D. Expected Shortfall: a Natural Coherent Alternative to Value at Risk,
<https://arxiv.org/abs/cond-mat/0105191>.
- [2] Tselishchev M. On the Concavity of Expected Shortfall,
<https://arxiv.org/abs/1910.00640>.

Уточнение константы в неравенствах типа Эссеена-Розовского для случая независимых одинаково распределённых случайных слагаемых

Максимова Анастасия Викторовна

Кафедра математической статистики

e-mail: maksimova.anastasia@gmail.com

Научный руководитель — д.ф.-м.н. доц. Шевцова Ирина Геннадьевна

Работа посвящена уточнению оценок скорости сходимости в центральной предельной теореме. Рассматриваются такие оценки скорости сходимости как неравенства типа Эссеена-Розовского — обобщения неравенств Эссеена и Розовского, полученные в [1] с помощью введения параметра усечения ε и параметра балансировки γ . Особенность изучаемых оценок заключается в том, что они предполагают конечность только вторых моментов распределений случайных слагаемых, включая в себя сумму хвостов вторых моментов в виде дроби Линдеберга, а также абсолютное значение суммы усечённых третьих моментов распределений случайных слагаемых. В работе производится уточнение возникающих в этих оценках мультипликативных констант $C_E(\varepsilon, \gamma)$ и $C_R(\varepsilon, \gamma)$ при дополнительном предположении об одинаковой распределённости случайных слагаемых.

Для построения оценок этих констант используется метод характеристических функций, реализованный сглаживающим неравенством Правитца [2]. В его правой части фигурирует абсолютное значение разности характеристических функций нормированной суммы случайных слагаемых и стандартного нормального распределения. Полученные в работе оценки этой величины обладают свойством монотонного убывания по числу случайных слагаемых n . Это свойство наследуется и оценками констант в неравенствах типа Эссеена-Розовского.

Таким образом, при известных ограничениях снизу на число случайных слагаемых n и при их одинаковой распределённости можно получить более точные оценки скорости сходимости в центральной предельной теореме.

Например, для минимального значения константы $C_E(\varepsilon, \gamma)$ в неравенстве типа Эссеена в [1] было получено следующее значение:

$$\inf_{\varepsilon>0, \gamma>0} C_E(\varepsilon, \gamma) \leq 2.65,$$

однако если использовать предположение об одинаковой распределённости случайных слагаемых и ограничить их число n снизу, например, значением 10, то получим более точную оценку константы:

$$\inf_{\varepsilon>0, \gamma>0} C_E(\varepsilon, \gamma) \leq 2.21.$$

Аналогичные результаты получены для константы $C_R(\varepsilon, \gamma)$ в неравенствах типа Розовского. Для случая разнораспределённых случайных слагаемых, рас-

смаатриваемом в [1], верхняя оценка константы принимает следующее значение

$$\inf_{\varepsilon>0, \gamma>0} C_R(\varepsilon, \gamma) \leq 2.66.$$

Если же снова использовать знание об одинаковой распределённости случайных слагаемых, а также ограничение на их число, например, $n \geq 10$:

$$\inf_{\varepsilon>0, \gamma>0} C_R(\varepsilon, \gamma) \leq 2.27.$$

Также замечено, что при использовании тривиального ограничения $n \geq 1$ результаты работы совпадают с результатами [1].

Литература

- [1] Gabdullin R. A., Makarenko V. A., Shevtsova I. G. Esseen-Rozovskii type estimates for the rate of convergence in the Lindeberg theorem // Journal of Mathematical Sciences. 2018. Vol. 234, N 6. P. 847–885.
- [2] Prawitz H. Limits for a distribution, if the characteristic function is given in a finite domain // Skand. Actuarietidskr. 1972. Vol. 55, P. 138–154.

Методы прогнозирования временных рядов на примере прогнозирования выручки частной медицинской компании

Мухтаров Максим Булатович

Кафедра математической статистики

e-mail: mukhtarov.mb@gmail.com

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

Решение задачи прогнозирования временного ряда заключается в создании модели (практически всегда стохастической, т. к. прогнозист не обладает всей информацией о наблюдаемом явлении, чтобы построить детерминированную модель), приближающей значение целевой функции на выбранном прогнозном периоде в смысле некоторой оценки. Качество решения можно проверить только, когда прогнозный период наступит. Поэтому все оценки делаются на тех данных, что доступны исследователю.

Выделяют два вида оценок: 1) оценки, позволяющие сравнить аппроксимации целевой функции вне зависимости от класса моделей, реализацией которого эта аппроксимация является; 2) оценки, позволяющие сравнить аппроксимации целевой функции внутри класса моделей, реализацией которого эта аппроксимация является. Самые распространённые представители первого вида — MAE, RMSE, MAPE; второго вида — коэффициент детерминации R^2 и информационные критерии AIC, BIC [1]. Выбрав лучшую модель внутри своего класса с помощью оценок второго вида, можно переходить к сравнению моделей между классами с помощью оценок первого вида. Кроме того, оценки первого вида применимы и к не стохастическим моделям. Более того, оценки можно вычислять на разных периодах, в том числе с помощью кросс-валидации. Практическая часть показала, что для прогноза на разные шаги (конкретно на 1 шаг по времени вперёд и на 2 шага по времени вперёд) могут использоваться различные модели.

Саму целевую функцию, отражающую одно и то же явление, можно также выбирать по-разному. В целом это может происходить двумя способами: 1) за счёт трансформации данных, вытекающих из практического смысла задачи; 2) за счёт трансформации данных для обеспечения удобных для применения некоторых моделей свойств (например, временной ряд в результате становится стационарным). Самые распространённые трансформации первого вида — устранение неравномерностей временного ряда, переход к моделированию относительной величины, нормирование единиц измерения. Самые распространённые трансформации второго вида — логарифмо-степенные трансформации и переход к разностям.

К простейшим, и поэтому эталонным, моделям относятся среднее, наивный прогноз, наивный прогноз со смещением и наивный прогноз с учётом сезонности. Декомпозиция временного ряда заключается в задании аппроксимирующей функции как аддитивной или мультипликативной комбинации трёх компонент: тренда, сезонности и остатка. К основным методам декомпозиции временного ряда относятся: классический, X11, SEATS, STL, а также модели экспоненциального сглаживания [2]. Модели экспоненциального сглаживания классифицируются по трём параметрам: тренду (отсутствует, аддитивный, аддитивный с затуханием), сезонности (отсутствует, аддитивная, мультипликативная) и ошибкам (аддитивные и мультипликативные).

Авторегрессионные модели, в самом общем виде обозначаемые как SARIMA(p, d, q)(P, D, Q)m, основаны на идее линейной комбинации обычных и сезонных лагов наблюдаемой переменной и её ошибок, а также переходе к разностям [3]. Для SARIMA моделей возникает проблема оптимального выбора коэффициентов, которая чаще всего решается путём определения порядка вычитания с помощью специальных тестов, визуального анализа графиков автокорреляционной и частичной автокорреляционной функций и, в итоге, перебора различных комбинаций параметров для лагов. Авторегрессионные нейронные сети используют в качестве входных параметров лаги наблюдаемой переменной.

Кроме лагов наблюдаемой переменной и её ошибок в качестве предикторов можно выбирать и другие переменные, в результате чего привлекается теория хорошо изученной множественной линейной регрессии. В качестве объясняющих переменных для регрессий временных рядов могут выбираться тренд; булевы переменные для простой сезонности; экстремальных событий (разовых и приводящих к сдвигу); коэффициенты Фурье для сложной сезонности; переменные, отражающие календарные обстоятельства (количество рабочих дней, количество различных дней в неделе и т. п.); различные виды лагов предикторов (обычные, сезонные, накопительные). Кроме линейных моделей, можно рассматривать два самых распространённых случая нелинейных моделей. К первому относятся сплайны, которые возникают как альтернатива включению тренда более высокого порядка, чем первый, в качестве предиктора. Ко второму — логарифмические трансформации наблюдаемой и/или объясня-

ющей переменной. Для того, чтобы иметь возможность улучшить результаты SARIMA моделей вводится понятие динамической регрессии — множественная регрессия, ошибки которой подчиняются SARIMA модели.

Литература

- [1] Hyndman, R. J., Athanasopoulos, G. Forecasting: principles and practice. 2nd edition. Melbourne: OTexts. 2018. URL: <https://otexts.com/fpp2>.
- [2] Gardner, E. S. Exponential smoothing: The state of the art – Part II. Houston: University Of Houston. 2005. 81 p. URL: <https://www.bauer.uh.edu/gardner/Exponential-Smoothing.pdf>.
- [3] Time Series Analysis: Forecasting and Control / G. E. P. Box, G. M. Jenkins, G. C. Reinsel, G. M. Ljung. 5th edition. Hoboken, New Jersey: John Wiley and Sons Inc. 2015. 712 p. URL: http://www.ru.ac.bd/stat/wp-content/uploads/sites/25/2019/03/504_05_Box_Time-Series-Analysis-Forecasting-and-Control-2015.pdf BoxJenkins.

Учет глубины при стилизации изображений и автоматическая оценка качества стилизации

Мишустина Маргарита Владимировна

Кафедра математических методов прогнозирования

e-mail: mishustina.rita@yandex.ru

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

Задача стилизации изображений (англ. style transfer) заключается в отрисовке изображения в стиле, задаваемом другим изображением, например, картиной известного художника. Изображение, семантические черты которого нужно сохранить называется *контентом* или *стилизуемым изображением*. Изображение, которое задает стиль, называется *стилем* или *стилевым изображением*. Данная задача имеет приложения в индустрии развлечений, рекламе, дизайне, применении спец. эффектов в видеоиграх и средствах дополненной реальности, а также применяется для расширения обучающей выборки при настройке моделей машинного обучения. Например, приложение Prisma, осуществляющее стилизацию, имеет более 50 млн. скачиваний на Google Play, а компания Adobe включила нейросетевые фильтры в последнюю версию Photoshop 2021.

Стилизация связана с искажениями исходного изображения. Проблемой является то, что стандартные алгоритмы стилизации накладывают стиль равномерно по изображению, что может приводить к искажению существенных объектов, таких как люди и другие центральные объекты сцены.

Первая часть работы посвящена модификации метода [1] путем добавления учета глубины при стилизации. Анализируются результаты нового подхода, проводится сравнение результатов работы модифицированного и исходного решений. Преимущество измененного метода подтверждается на качественном уровне, а также опросом широкой группы пользователей. По результатам работы опубликована статья [2].

Другой важной проблемой стилизации является возможная несовместимость стилизуемого и стилевого изображений. Если эти изображения существенно отличаются по цветам, контурам, характерным паттернам, то результат стилизации может оказаться неудовлетворительным. Чтобы избежать подобной ситуации в работе предложен алгоритм автоматической оценки совместимости стилизуемого и стилевого изображений. Применение разработанного алгоритма позволит по входному изображению автоматически подбирать наиболее подходящие под него стили.

Поскольку данная задача является новой, для ее решения был собран набор данных, состоящий из 90000 изображений — 30000 троек (стиль, контент, результат). Для получения признаков были рассмотрены эвристические методы извлечения признаков (дескриптор HOG, метод главных компонент, среднее и дисперсия по цветовым каналам для изображения, фильтры Собеля и Лапласа), а также нейросетевые методы извлечения признаков (рассматривались промежуточные представления для изображений стиля и контента, полученные с помощью предобученной на наборе данных ImageNet нейронной сети VGG-16 [3]).

К полученным признакам применялись как традиционные методы машинного обучения — метод опорных векторов, градиентный бустинг; так и нейронные сети — были рассмотрены 2 архитектуры: полносвязная нейронная сеть, принимающая на вход эвристические признаки стилевого и контентного изображений; а также полносвязная нейронная сеть, принимающая на вход квадрат разности промежуточных представлений стиля и контента. Во второй архитектуре была также реализована идея skip-connection из одной из лучших нейронных сетей классификации изображений ResNet [4].

Данные методы были оценены, рассмотрены композиции алгоритмов. Результирующее качество по метрике accuracy (число правильных ответов к общему числу примеров) равняется 0.823.

Литература

- [1] Huang X., Belongie S. Arbitrary style transfer in real-time with adaptive instance normalization //Proceedings of the IEEE International Conference on Computer Vision. – 2017. – P. 1501-1510.
- [2] Kitov V., Kozlovtshev K., Mishustina M. Depth-aware arbitrary style transfer using instance normalization //arXiv preprint arXiv:1906.01123. – 2019.
- [3] Simonyan K., Zisserman A. Very deep convolutional networks for large-scale image recognition //arXiv preprint arXiv:1409.1556. – 2014.
- [4] He K. et al. Deep residual learning for image recognition //Proceedings of the IEEE conference on computer vision and pattern recognition. – 2016. – P. 770-778.

Методы сравнения и подгонки сканированных поверхностей

Травникова Арина Сергеевна

Кафедра математических методов прогнозирования

e-mail: travnikova_arina@mail.ru

Научный руководитель — д.т.н. проф. Местецкий Леонид Моисеевич

С развитием технологий 3D сканирования и повышением их доступности появляется все больше прикладных задач, связанных с обработкой дискретных моделей поверхностей, полученных в результате 3D сканирования. Одной из важных задач анализа поверхностей является их сравнение друг с другом — именно эта задача служит основным предметом исследования в данной работе; она заключается в том, чтобы по трехмерным облакам точек, задающим две однозначные поверхности, оценить меру их сходства и локализовать области наибольшего различия. Также в работе рассмотрена задача подгонки — поиска такого взаимоположения поверхностей, при котором они наиболее близки друг к другу. Ее необходимость объясняется тем, что съемки объектов исследования могут производиться с большим интервалом времени, из разного положения и даже разными приборами.

Математическая постановка задачи. Пусть даны две однозначные поверхности, заданные облаками точек $S_1 = \{(x_1^i, y_1^i, z_1^i)\}_{i=1}^{N_1}$, $S_2 = \{(x_2^i, y_2^i, z_2^i)\}_{i=1}^{N_2}$ в ПДСК $Oxyz$. Им соответствуют нерегулярные множества $g_1 = \{(x_1^i, y_1^i)\}_{i=1}^{N_1}$, $g_2 = \{(x_2^i, y_2^i)\}_{i=1}^{N_2}$, узлы которых есть проекции облаков точек на плоскость Oxy . В силу однозначности поверхностей S_1, S_2 , они представимы в виде функций двух переменных $z_1(x, y)$, $z_2(x, y)$, заданных на несовпадающих множествах g_1, g_2 . Задача заключается в том, чтобы по входным дискретным описаниям поверхностей определить количественную меру различия между ними и при необходимости выполнить подгонку поверхностей.

Предлагаемое решение основывается на следующих принципах:

- для обработки поверхности осуществляется переход от ее дискретного представления в виде облака точек к непрерывной триангуляционной модели, состоящей из множества пространственных треугольников.
- мера различия между поверхностями определяется как средняя разница истинных и интерполирующих значений функций высоты z_1, z_2 в узлах плоского множества $g_1 \cup g_2$: $\rho(S_1, S_2) = \sum_{(x,y) \in g_1 \cup g_2} \frac{|z_1(x,y) - z_2(x,y)|}{|g_1 \cup g_2|}$. Для определения неизвестных значений функций z_1, z_2 в узлах множеств g_2, g_1 в работе предлагается линейная интерполяция.
- задача подгонки сводится к задаче минимизации функции различия между поверхностями по параметрам преобразования одной из поверхностей: сдвигам вдоль осей координат и поворотам на наибольшие углы. Для ее решения используется аналог метода градиентного спуска.

Помимо решения теоретической задачи сравнения поверхностей, в работе рассмотрены два ее практических приложения:

Задача количественной оценки асимметрии человеческого лица. Под симметрией подразумевается зеркальная симметрия относительно плоскости, близкой к вертикальной. Задача заключается в том, чтобы найти плоскость симметрии лица и вычислить меру различия между поверхностями, задающими исходную модель лица и отраженную относительно найденной плоскости симметрии. В работе плоскость симметрии определяется как плоскость, для которой мера различия между исходной сканированной поверхностью и ее отраженной копией минимальна. Для проверки корректности работы разработанного алгоритма проведены вычислительные эксперименты на заведомо несимметричном мужском лице, симметричной модели лица актрисы Скарлетт Йоханссон, а так же лице статуи Давида (Микеланджело).

Задача локализации областей интенсивного изменения рельефа береговой зоны моря по типу эрозии по картам высот. Под эрозией понимается отступление береговой линии моря вглубь суши под воздействием волн и ветра. Идея решения задачи локализации заключается в определении меры различия между рельефами как расстоянии, на которое сдвинулась береговая линия за исследуемый промежуток времени. Для его вычисления достаточно совместить поверхности в пространстве, то есть решить задачу подгонки, и выделить береговую линию, например, поиском скачкообразного изменение угла между нормалью к поверхности и горизонтальной плоскостью в случае обрывистых берегов. На экспериментальном участке берега Азовского моря при помощи разработанного алгоритма удалось локализовать протяженные участки эрозии силой до 2м за период исследования 2016-2018гг., что по экспертной оценке является корректным результатом.

Результаты. В выпускной работе описана математическая модель, позволяющая эффективно работать с поверхностями, заданными дискретными облаками точек на нерегулярной сетке. В рамках исследований разработан и реализован метод сравнения и подгонки трехмерных поверхностей. Выбранная в работе метрика L_1 для сравнения поверхностей позволяет автоматически выполнять локализацию областей сравнения и дает широкие возможности для интерпретации результатов. Результаты работы использованы в проекте РФФИ 18-05-80082 для реализации метода оценки изменения рельефа береговой линии Азовского моря и локализации областей эрозии. По теме исследований в данной задаче опубликован доклад в сборнике конференции «Геоинформационные технологии и космический мониторинг» [1].

Литература

- [1] Травникова А. С., Мисиров С. А., Бердников С. В. Местецкий Л. М. Оценки изменения рельефа береговой зоны при абразии по данным дистанционного зондирования: методические подходы и предварительные результаты // Экология. Экономика. Информатика. Серия: Геоинформационные технологии и космический мониторинг. 2020. ВЫП. 5, С. 117–123.

О тестах относительно замен функциональных элементов

Аурангабадкар Удитангшу

Кафедра математической кибернетики

e-mail: aurudit@gmail.com

Научный руководитель — д.ф.-м.н. доц. Романов Дмитрий Сергеевич

В данной выпускной работе рассматривались специальные одиночные замены функциональных элементов в схемах из функциональных элементов с одним выходом, реализующих булевы функции. Рассматривались СФЭ над базисом $B = \{x \downarrow y, x \oplus y\}$, на которые действовал источник одиночных неисправностей замен элементов Z , заменяющий двухвходный элемент “стрелка Пирса” двухвходным элементом дизъюнкций, а двухвходный элемент “сумма по модулю 2” элементом конъюнкций.

Определение 1. *СФЭ S называется неизбыточной, тогда и только тогда, когда при любой одиночной неисправности СФЭ реализует функцию, неравную реализуемой этой схемой в отсутствие неисправностей.*

Определение 2. *Множество входных наборов называется проверяющим тестом для СФЭ S , реализующей функцию f , если оно позволяет отличить от исходной любую неравную ей функцию, реализуемой схемой S в результате неисправности. Мощность проверяющего теста называется его длиной.*

Определение 3. *Множество входных наборов называется диагностическим тестом для СФЭ S , реализующей функцию f , если оно позволяет отличить любые две неравные функции, реализуемые схемой S в результате неисправностей. Мощность диагностического теста называется его длиной.*

Определение 4. *Обозначим через $L^{detect}(Z, S)$ длину минимального проверяющего теста для схемы S относительно источника неисправностей Z . Обозначим через $L_B^{detect}(Z, f(\tilde{x}^n))$ минимум величины $L^{detect}(Z, S)$ по всем неизбыточным реализующим функцию $f(\tilde{x}^n)$ схемам S в базисе B . Тогда функция Шеннона длины единичного проверяющего теста называется:*

$$L_B^{detect}(Z, n) = \max_{f \in P_2(n)} (L_B^{detect}(Z, f(\tilde{x}^n)))$$

Определение 5. *Обозначим через $L^{diagnostic}(Z, S)$ длину минимального диагностического теста для схемы S относительно источника неисправностей Z . Обозначим через $L_B^{diagnostic}(Z, f(\tilde{x}^n))$ минимум величины $L^{diagnostic}(Z, S)$ по всем неизбыточным реализующим функцию $f(\tilde{x}^n)$ схемам S в базисе B . Тогда функция Шеннона длины единичного диагностического теста называется:*

$$L_B^{diagnostic}(Z, n) = \max_{f \in P_2(n)} (L_B^{diagnostic}(Z, f(\tilde{x}^n)))$$

К. А. Попковым в статье [1] был предложен метод построения легко диагностируемых схем относительно единичных неисправностей, где верхняя оценка

функции Шеннона длины единичного диагностического теста относительно инверсных или константных неисправностей на выходах или входах элементов превосходит не более чем на 2 верхнюю оценку функции Шеннона длины единичного проверяющего теста. В данной выпускной квалификационной работе был использован метод построения диагностических тестов, который был предложен в работе К. А. Попкова. Метод, предложенный К.А. Попковым, может быть использован для незадекларированного в статье случая, когда неисправности не являются константными или инверсными.

В данной работе были доказаны теоремы о значении функции Шеннона единичного проверяющего теста и об оценке функции Шеннона длины единичного диагностического теста.

Теорема 1. Для $\forall n \in \mathbb{N}$ справедливо равенство:

$$L_B^{detect}(Z, n) = 1.$$

Теорема 2. Для $\forall n \in \mathbb{N}$ справедливо неравенство:

$$L_B^{diagnostic}(Z, n) \leq 3.$$

Литература

- [1] Попков К. А. Метод построения легко диагностируемых схем из функциональных элементов относительно единичных неисправностей // Прикладная дискретная математика. 2019. № 46. С. 38–57.

Генетический алгоритм для поиска оптимальных и близких к ним схем от малого числа переменных

Местецкий Михаил Александрович

Кафедра математической кибернетики

e-mail: mixail1005001@mail.ru

Научный руководитель — к.ф.-м.н. доц. Шуплецов Михаил Сергеевич

Постановка задачи. Дискретные математические модели, описывающие работу и структуру сложных систем преобразования информации (интегральные схемы, программы и т. д.) часто используются в теории и на практике. Эти модели основываются на дискретных управляющих системах, которые задаются функцией алгебры логики (ФАЛ), структурно реализуемой в различных классах графов (т. е. схемах). При рассмотрении таких систем возникает *задача синтеза* — построение схемы, которая будет реализовывать некоторую заданную ФАЛ или систему ФАЛ. Кроме того, задача синтеза подразумевает оптимизацию данной схемы по какому-либо из параметров (сложность, глубина и т. д.). В работе в качестве параметров для оптимизации рассматривались статическая и динамическая активности.

Одним из подходов, используемых в задачах оптимизации, являются *генетические алгоритмы*. Это адаптивные методы поиска, которые используют

аналоги механизмов, встречающихся в природе, таких как наследственность, естественный отбор и др.

Таким образом, в работе исследуется возможность применения генетических алгоритмов к задаче синтеза.

Предлагаемое решение. Генетические алгоритмы симулируют эволюцию — процесс изменения генетического состава популяции в природе. Популяция состоит из особей, в данном случае особями являются схемы, являющие кандидатами на решение поставленной задачи синтеза. Таким образом, алгоритм сводится к созданию некоторой начальной популяции схем и получении из нее искомой схемы за счет эволюции. Для того, чтобы применять манипуляции, связанные с генетическими алгоритмами, используется специальное представление схем, описанное в [1].

Результаты. Алгоритм был реализован в виде программы на языке программирования Python. С программой были проведены эксперименты по синтезу функций: H — функция голосования, S — сумматор, U — умножитель. В ходе экспериментов были получены схемы, которые лучше других по одному из трех параметров или по их совокупности. Все лучшие схемы представлены в таблице 1. Существует ряд нижних оценок для статической и динамической активности [2]. Но данные оценки не репрезентативные на схемах небольшого размера. Тем не менее, результаты экспериментов показывают работоспособность данного подхода.

Функция	Стат. активность	Дин. активность	Размер схемы
$H(x_1, x_2, x_3)$	4	4	4
$H(x_1, x_2, x_3, x_4)$	11	11	11
	8	12	16
	9	12	14
$S(x_1, x_2, x_3)$	6	11	21
	8	13	18
	11	10	13
	10	8	19
$U(x_1, x_2, x_3, x_4)$	7	12	25
	8	14	22
	10	12	20
	15	11	21

Таблица 2: лучшие результаты экспериментов.

Литература

- [1] Тельпухов Д. В. Разработка методов генетического синтеза сбоеустойчивых комбинационных схем // Проблемы разработки перспективных микро- и наноэлектронных систем (МЭС). — 2018. — т. 1. — с. 45–49.

- [2] Dinesh K., Otiv S., Sarma J. New bounds for energy complexity of Boolean functions // Theoretical Computer Science. — 2020. — Vol. 845. — P. 59–75.

Применение классических и нейросетевых методов построения модели пользователя в задаче динамической аутентификации пользователей на основе анализа работы с компьютерной мышью

Березникер Алексей Витальевич

Кафедра интеллектуальных информационных технологий

e-mail: berezniker@mail.ru

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

Динамическая аутентификация пользователей на основе анализа работы с компьютерной мышью является одним из наиболее перспективных и актуальных направлений развития методов аутентификации в сфере компьютерной безопасности. Данный подход позволяет выявлять внутренние атаки в информационных системах в кратчайшие сроки и при минимальных затратах на дополнительное оборудование. В настоящее время существующие решения обладают рядом недостатков, в частности: контролируемые условия во время сбора данных, переобучением на исходных наборах данных и использованием бинарных классификаторов.

В работе рассматриваются существующие методы динамической аутентификации, основанные на использовании нейронных сетей «автокодировщик» и классических методов машинного обучения, таких как: одноклассовый метод опорных векторов, методы на основе решающих деревьев, локальной плотности точек и эллипсоидальной аппроксимации. Анализируются достоинства и недостатки существующих решений. Предлагаются подходы к предобработке данных и построению признакового пространства на основе градиентного бустинга, стандартной нормализации и динамического поиска выбросов. Комбинация предложенных методов поспособствовала улучшению качества распознавания. Предлагаются собственные архитектуры полносвязных, сверточных и рекуррентных нейронных сетей, сопоставимые по качеству работы с классическими методами. Для оценки производительности поведенческих биометрических алгоритмов используется комбинация пяти открытых наборов данных, содержащих информации о динамике работы с компьютерной мышью 105 пользователей.

На основе комбинации предложенных методов, демонстрирующих наилучшее качество работы ($ROC\ AUC = 0.82$), было спроектировано и реализовано кроссплатформенное приложение для динамической аутентификации пользователей на основе анализа работы с компьютерной мышью. Для принятия решения о легитимности пользователя используется предложенный метод, основанный на динамическом изменении уровня доверия текущего пользователя алгоритмом TrustModel [1] на базе предсказаний модели OneClassSVM. Данная

система и ее отдельные модули могут послужить основой для построения перспективных современных систем информационной безопасности.

Литература

- [1] Mondal S., Bours P. A computational approach to the continuous authentication biometric system // Information Sciences. – 2015. – Vol. 304. – P. 28-53.

Нейросетевое вычисление карт доверия для методов оценки оптического потока в видео

Малышев Кирилл Владимирович

Кафедра интеллектуальных информационных технологий

e-mail: kirill.malyshchev@graphics.cs.msu.ru

Научный руководитель — к.ф.-м.н. м.н.с. Ерофеев Михаил Викторович

Одной из основных задач обработки видео является оценка оптического потока, то есть векторов движения объектов между двумя последовательными кадрами. Карта оптического потока представляет собой матрицу векторов, в которой каждый элемент соответствует смещению пикселя между кадрами. Методы оценки оптического потока активно используются в различных областях обработки и анализа видео. В частности, информацию о смещении пикселей учитывают кодеки для достижения большей степени сжатия видеоданных. Также карты оптического потока используются в задачах компьютерного зрения для отслеживания и распознавания объектов.

К сожалению, не все области в кадрах можно успешно сопоставить. Области открытия/закрытия, области без текстуры, зеркальные и прозрачные поверхности могут привести к некорректным результатам работы методов в этих местах. Например, если какой-то объект перекрывает другой, который появляется только на одном из кадров, корректно определить его смещение не представляется возможным. Поэтому при использовании в какой-либо задаче метода построения оптического потока, нужно помнить о том, что в некоторых областях кадра эти данные могут быть неточны. Соответственно, возникает вопрос о надежности результатов работы методов вычисления карт оптического потока. То есть требуется для каждого пикселя кадра дать ответ, насколько надёжен результат работы метода сопоставления ракурсов. При этом ответ выражается в виде числа от 0 до 1, где 0 означает, что доверять результату нельзя, а 1 означает, что доверять результату можно. Матрицы, содержащие эту информацию, называются картами доверия.

В данной работе предлагается нейросетевой метод построения карт доверия для методов оценки оптического потока, в котором вычисление происходит по следующим данным:

- Согласованность первого и второго кадров [1]
- RGB-дисперсия кадра
- Ошибка сопоставления

— Разность между первым кадром и приведенным к нему с помощью оптического потока вторым кадром

Для обучения нейросетевой модели был усовершенствован способ генерации истинных карт доверия по сравнению с другим методом предсказания карт доверия LiteFlowNet3 [2]. Для этого подобран параметр a функции $f_{exp} = e^{-a*|Diff|}$ преобразования разности $Diff$ между истинным оптическим потоком и вычисленным в доверие на основе синтетического набора данных Sintel [3].

В качестве архитектуры нейронной сети выбрана архитектура типа кодировщик-декодировщик. Перекрестная валидация позволила подобрать гиперпараметры модели так, чтобы предсказываемые карты доверия были более точными.

Тестирование обученной модели производилось на валидационной выборке путём сравнения с другим нейросетевым методом вычисления карт доверия для метода оптического потока LiteFlowNet3. Точность вычисленных карт доверия проверялась на задаче оценки параллакса в стереоскопических видео [4]. Полученные результаты показывают, что предложенный метод может помочь оценить надёжность векторов движения. В случае сильно зашумлённого оценённого оптического потока вычисляемое доверие позволяет хорошо отфильтровать его так, чтобы некорректные векторы движения не учитывались алгоритмом обработки оптического потока. Прямое сравнение карт доверия двух разных методов для одного и того же оптического потока показало, что предлагаемая модель лучше справляется с задачей вычисления карт доверия.

	Среднеквадратичная ошибка
Без использования доверия	0.096
LiteFlowNet3	0.087
Предложенный метод	0.080

Табл. 1: Среднеквадратичная ошибка оценки параллакса для валидационной выборки.

Литература

- [1] Egnal G., Wildes R.P. Detecting binocular half-occlusions: empirical comparisons of five approaches // IEEE Transactions on Pattern Analysis and Machine Intelligence. IEEE Computer Society. 2002. Vol. 24, P. 1127–1133.
- [2] Hui T., Loy C. C. LiteFlowNet3: resolving correspondence ambiguity for more accurate optical flow estimation // Proceedings of XVI international conference «European Conference on Computer Vision» (Glasgow, 23–28 August 2020). Springer. 2020. Vol. 20, P. 169–184.
- [3] A naturalistic open source movie for optical flow evaluation / D.J. Butler, J. Wulff, G. B. Stanley, M. J. Black // Proceedings of XII international conference

«European Conference on Computer Vision» (Florence, 7–13 October 2012). Springer. 2012. Vol. 6, P. 611–625.

- [4] Methodology for stereoscopic motion-picture quality assessment / A. A. Voronov, D. S. Vatolin, D. A. Sumin, V. V. Napadovsky, A. P. Borisov // Proceedings of XXIV international conference «Stereoscopic Displays and Applications» (Burlingame, 3–7 February 2013). SPIE. 2013. P. 302–315.

Усовершенствование алгоритма реидентификации для повышения качества видеоаналитики

Мамедов Тимур Закирович

*Кафедра интеллектуальных информационных технологий
e-mail: timur.mamedov@graphics.cs.msu.ru*

Научный руководитель — к.ф.-м.н. доц. Конушин Антон Сергеевич

Данная работа посвящена задаче реидентификации людей, суть которой заключается в том, что по двум изображениям необходимо определить запечатлён ли на них один и тот же человек или нет. Рассматриваемая задача является практически важной, а алгоритмы её решения нашли широкое применение в видеоаналитике (например, в видеонаблюдении, задачах маркетинга, анализе спортивных мероприятий).

Задача реидентификации подразумевает работу с реальными данными, поэтому ввиду разнообразия жизненных ситуаций, при её решении приходится сталкиваться со множеством нюансов. Можно выделить 4 основные сложности, возникающие при решении данной задачи: обилие перекрытий реидентифицируемых объектов; сильное влияние заднего фона; проблема доменной переносимости (под переносимостью домена понимается изменение условий получения данных); малое количество данных, пригодных для обучения и тестирования алгоритмов реидентификации.

Особое развитие алгоритмы реидентификации получили с ростом популярности нейросетевых методов, которые условно можно разделить на три категории: с использованием масок сегментации объектов; с применением более совершенных функций потерь и архитектур нейронных сетей; методы, решающие проблему доменной переносимости. Первый и третий подходы затратны с точки зрения вычислительных ресурсов и времени, поэтому их весьма сложно применять на практике. По этой причине в данной работе рассматривается вторая категория методов.

В качестве базового алгоритма в данной работе используется решение из статьи [1]. Выбранный базовый алгоритм представляет собой нейросетевой классификатор, в основе которого лежит нейронная сеть ResNet50 [2]. Алгоритм имеет простую архитектуру, что позволяет ему работать достаточно быстро и без использования мощных вычислительных ресурсов. Как было отмечено выше, это одни из основополагающих факторов для применения решения в практических целях. Несмотря на свою простоту, выбранный алгоритм имеет достаточно высокие показатели качества реидентификации.

Все предложенные в данной работе модификации и изменения базового алгоритма можно разделить на 3 части: замена ResNet50 на Res2Net50; использование функции потерь FIDI Loss; внедрение модуля фильтрации на этапе тестирования.

Замена ResNet50 на Res2Net50. Первая предложенная в данной работе модификация заключается в замене упомянутой выше нейросети ResNet50 на более совершенную Res2Net50 [3], которая качественнее решает задачу классификации. При этом применение Res2Net50 не сильно сказывается на накладных расходах, связанных с вычислениями.

Использование функции потерь FIDI Loss. Следующая модификация базового алгоритма заключается в применении функции потерь FIDI Loss [4], предназначенной для задачи реидентификации. Её основная особенность заключается в сильном штрафе за мелкие различия между изображениями. Как показала практика, именно в них кроется основная сложность рассматриваемой задачи.

Внедрение модуля фильтрации на этапе тестирования. Проанализировав данные, используемые для обучения и тестирования алгоритмов реидентификации, был сделан вывод, что на вход алгоритму могут подаваться заведомо некорректные данные, поэтому их можно заранее отфильтровать, и тем самым повысить его устойчивость.

Некорректные данные не могут отражать настоящей картины, описывающей действительное качество рассматриваемого алгоритма, так как на них велика вероятность ошибки. Напротив, на таких входных изображениях стабильный алгоритм может выдать неверный ответ, а нестабильный — верный.

Для реализации идеи был разработан модуль фильтрации, представляющий собой легковесный бинарный нейросетевой классификатор, который несущественно влияет на скорость работы общего алгоритма.

В итоге весь этап тестирования предложенного алгоритма сводится к следующему: все изображения проходят через модуль фильтрации, а на вход нейронной сети для реидентификации поступают только те изображения, которые были признаны фильтром корректными.

Так как идея с фильтрацией входных данных нова, то необходимо было разработать критерии и инструкцию разметки данных для реидентификации на предмет корректности и некорректности, по которым двумя специально обученными людьми в течение 40 часов была произведена разметка 160479 изображений.

Экспериментальная оценка. Все модификации, предложенные в данной работе, благоприятно влияют на качество реидентификации, и имеется весомый с точки зрения задачи реидентификации прирост качества ($\approx 4\%$ по метрике $Rank_1$ и $\approx 5\%$ по метрике mAP).

Также в работе отдельно был изучен вопрос доменной переносимости. Экспериментальная оценка показала состоятельность предложенных изменений при сценарии, когда нейронная сеть обучена на одном наборе данных, а протестирована на другом.

стирована — на другом (прирост $\approx 7\%$ по метрике $Rank_1$ и $\approx 9\%$ по метрике mAP).

По теме работы была подана статья на международную конференцию «ГрафиКон-2021», труды которой индексируются в Scopus.

Литература

- [1] Bag of tricks and a strong baseline for deep person re-identification / H. Luo, Y. Gu, X. Liao, S. Lai, W. Jiang // 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). 2019. P. 1487-1495.
- [2] Deep residual learning for image recognition / K. He, X. Zhang, S. Ren, J. Sun // 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016. P. 770-778.
- [3] Res2Net: a new multi-scale backbone architecture / S. Gao, M. Cheng, K. Zhao, X. Zhang, M. Yang, P. Torr // IEEE Transactions on Pattern Analysis & Machine Intelligence. 2021. Vol. 43. N 2. P. 652-662.
- [4] Beyond Triplet Loss: person re-identification with fine-grained difference-aware pairwise loss / C. Yan, G. Pang, X. Bai, J. Zhou, L. Gu // arXiv preprint arXiv:2009.10295. 2020.

Разработка нейросетевого метода заполнения областей изображений высокого разрешения

Москаленко Андрей Викторович

*Кафедра интеллектуальных информационных технологий
e-mail: andrey.moskalenko@graphics.cs.msu.ru*

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

Задача заполнения областей изображения нацелена на восстановление поврежденной, удаляемой или неизвестной области. Входными данными алгоритма являются поврежденное изображение, а также маска области, где требуется восстановление. Выходными данными является восстановленное трехканальное изображение, где неизвестная область заполнена наиболее реалистичным способом.

В последние годы развитие нейросетевых подходов существенно способствовало появлению различных методов решения этой задачи. Однако нейросетевые подходы сильно привязаны к разрешению, на котором их обучали, из-за недостатка рецептивного поля. Большинство моделей имеет размер входа не превышающий 512 пикселей. В результате они не могут обрабатывать изображения произвольной формы, например, в интерактивных инструментах обработки изображений. Когда разрешение входного изображения повышается, у большинства подходов начинают проявляться артефакты.

В данной работе предложен новый нейросетевой метод заполнения областей изображений, менее подверженный появлению артефактов при увеличении разрешения из-за нехватки рецептивного поля. Предложенный метод осно-

ван на идее отдельного восстановления структуры изображения на низком разрешении, а текстуры на высоком.

На первой стадии алгоритм восстанавливает структуру изображения на низком разрешении. Сначала происходит уменьшение исходного изображения и маски до разрешения 512×512 , затем применяется предобученная нейросеть DFNet [1] для получения грубого результата заполнения на низком разрешении. Затем мы производим интерполяцию изображения до начального разрешения и выполняем замещение известной области из исходного изображения. Затем происходит генерация сдвигов исходного изображения в четырех направлениях: влево, вправо, вниз, вверх. В экспериментах использовались сдвиги на 20% от ширины изображения в пикселях. Отметим, что во время этого процесса мы также пересчитываем маски, помечая все пиксели в областях открытия как невалидные. Таким образом, после первой стадии получается 20-ти канальное изображение: пять RGB изображений (основное и четыре сдвига) и пять масок.

Полученный результат первой стадии в неизменном виде подается на вход сети восстановления из второй стадии. Вторая стадия нацелена на восстановление текстуры. Архитектура сети восстановления текстуры представляет собой U-Net [2] с расширенным числом входных каналов. Для предотвращения привязанности сети восстановления к структуре и ускорения процесса обучения, мы вырезаем случайный 512×512 квадрат из исходного тензора глубины 20, с условием, что маскированная область составляет не менее 10%, но не более 90% от размера этого квадрата. Отметим, что во время тестирования стадия выбора патча пропускается и просто передается изображение целиком в полном разрешении. Выход сети восстановления является финальным результатом заполнения.

Экспериментальная оценка показала превосходство предложенного подхода в сравнении с существующими нейросетевыми аналогами в экспертном сравнении с участием 150 зрителей, а также в объективном сравнении с использованием функционалов качества PSNR, SSIM, L1 при восстановлении деталей изображений высокого разрешения. Для сравнений использовался набор из 34 естественных изображений с полным разрешением в 2048×2048 . Результаты объективного сравнения представлены в Табл. 3.

Исследования выполнены при финансовой поддержке РФФИ в рамках научного проекта 19-01-00785 а.

Литература

- [1] Ronneberger O., Fischer P., Brox T. U-Net: convolutional networks for biomedical image segmentation // International Conference on Medical image computing and computer-assisted intervention. 2015.
- [2] Deep fusion network for image completion / X. Hong, P. Xiong, R. Ji, H. Fan // Proceedings of the 27th ACM International Conference on Multimedia. 2019.

Модель\Метрика	L1	PSNR	SSIM		L1	PSNR	SSIM
DeepFill v2	4.981	22.389	0.938		5.397	21.956	0.944
DFNet	3.592	24.910	0.946		4.132	23.836	0.946
HiFill	4.422	23.667	0.935		4.373	23.738	0.943
Photoshop 2020	4.115	23.789	0.944		13.320	23.756	0.949
ProFill	3.906	24.677	0.950		—	—	—
Предл. метод	3.524	25.175	0.944		3.474	25.311	0.950
Разрешение	1024x1024				1536x1536		

Модель\Метрика	L1	PSNR	SSIM		L1	PSNR	SSIM
DeepFill v2	5.546	21.834	0.946		5.669	21.697	0.948
DFNet	4.345	23.424	0.948		4.633	22.915	0.950
HiFill	4.403	23.696	0.944		4.391	23.710	0.947
Photoshop 2020	4.047	23.863	0.952		4.153	23.659	0.952
Предл. метод	3.447	25.374	0.952		3.434	25.412	0.954
Разрешение	1792x1792				2048x2048		

Таблица 3: Табл. 1: Результаты объективного сравнения.

Метод сжатия 3D моделей для параллельной обработки на GPU

Николаев Антон Владимирович

Кафедра интеллектуальных информационных технологий

e-mail: anton.nikolaev@graphics.cs.msu.ru

Научный руководитель — к.ф.-м.н. н.с. Фролов Владимир Александрович

Рост детализации 3D моделей, используемых в рендеринге, приводит к росту необходимых объёмов памяти GPU. Существующие методы сжатия позволяют производить декомпрессию модели только перед загрузкой в видеопамять, что не позволяет уменьшить потребление памяти GPU.

В данной работе предлагается метод сжатия 3D моделей с быстрой декомпрессией на GPU. Данный подход позволяет поместить от 3 до 8 раз больше геометрии в том же объёме памяти GPU и применяет декомпрессию на лету непосредственно в процессе растеризации или трассировки лучей. Для растеризации использовалась декомпрессия в двух вариантах, для трассировки лучей применялось аппаратное ускорение трассировки в Vulkan API и была предложена техника кэширования мешлетов.

Сжатие в предложенном методе. Предложенный метод сжатия предполагает использование трех шагов: разделения на мешлеты, сжатия данных о связях вершин по алгоритму Edgebreaker [1], сжатия атрибутов вершин при помощи квантизации. Параллельность декомпрессии при этом достигается за счет того, что восстановление каждого из мешлетов можно производить независимо от других.

Для первого шага на основе метода из работы [2] был предложен новый метод. Суть алгоритма заключается в повторяющемся выполнении двух шагов (выбора центров мешлетов и присоединения всех остальных треугольников к мешлетам) пока разбиение не перестанет меняться. В сравнении с [2] была изменена функция весов (добавлено слагаемое, сильно увеличивающее вес при превышении лимита в 64 треугольника на мешлет), а так же уменьшен средний размер мешлета.

Для сжатия данных о связях (шаг 2), был использован алгоритм Edgebreaker. Был проведен анализ частоты появления каждого из символов истории обхода и выбраны фиксированные коды символов, полученные с использованием кода Хаффмана.

Для сжатия атрибутов вершин (шаг 3) использовалась квантизация, для координат сжимались их изменения. С целью уменьшения количества не несущих информации ведущих нулей в двоичном представлении, возникающих из-за сжатия изменений с тем же количеством бит, что и для абсолютных значений, использовались две формы записи: полная и запись лишь младших бит. Выбор формы записи определялся дополнительным битом.

Параллельная декомпрессия в предложенном методе. Декомпрессия осуществлялась параллельно на GPU как при растеризации, так и при трассировке лучей. При растеризации использовались геометрические и меш шейдеры. В геометрических шейдерах каждый поток производил декомпрессию своего мешлета. При использовании меш шейдеров, параллельно производилась декомпрессия и в пределах одного мешлета, что позволило ускорить алгоритм в 1.5 - 2.

Для трассировки лучей был предложен кэш, в который сохраняются мешлеты, пересечение лучей с которыми проверяется наиболее часто, что позволяет значительно ускорить декомпрессию.

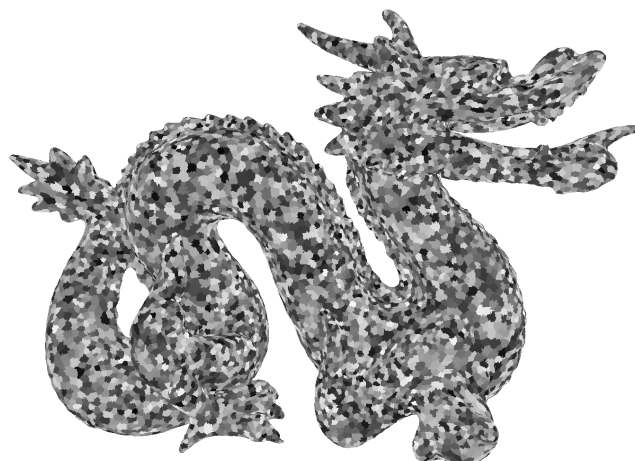


Рис. 1: Модель после декомпрессии, разными цветами показаны разные мешлеты.

Результаты. Был получен алгоритм, позволяющий разместить значительно больше (в 3-8 раз) данных 3D моделей в памяти GPU, а так же демонстриру-

Метод	Размер на диске	Размер на GPU	Время декомпрессии
Предложенный	398 Кб	398 Кб	1.61 мс
Draco	550 Кб	1694 Кб	4.8 мс
Corto	121 Кб	1694 Кб	3.5 мс
Без сжатия	1694 Кб	1694 Кб	0.25 мс

Табл. 1: Сравнение с существующими методами. Для версии без сжатия в качестве времени декомпрессии указано время отрисовки модели на GPU.

ющий более высокую скорость декомпрессии в сравнении с существующими (сравнение приведено в табл. 1). На рис. 1 показан результат работы алгоритма декомпрессии для одной из моделей.

Литература

- [1] Jarek Rossignac. Edgebreaker: Connectivity compression for triangle meshes. IEEE transactions on visualization and computer graphics, 5(1):47–61, 1999.
- [2] Sungyul Choe, Junho Kim, Haeyoung Lee, Seungyong Lee. Random accessible mesh compression using mesh chartification. IEEE Transactions on Visualization and Computer Graphics, 15(1):160–173, 2008

Исследование алгоритмов инкрементального моделирования

Полевой Антон Вячеславович

*Кафедра интеллектуальных информационных технологий
e-mail: polevoianton@bk.ru*

*Научные руководители — ст. спец. Герасимов Сергей Валерьевич,
к.ф.-м.н. Мецераков Александр Валерьевич*

Инкрементальное моделирование — это подход анализа данных, в котором решается задача классификации и отбора объектов, чувствительных к некоторому воздействию. В работе исследуются модели в области директ-маркетинга.

Новизна данной работы состоит, во-первых, в проведенном авторами систематическом сравнении точности широкого набора алгоритмов инкрементального моделирования на различных наборах данных академической (Criteo Lab [2]) и прикладной (соревнования X5 retail group [6]) направленности. Для данных X5 рассматривался полный набор агрегированных признаков и признаки, отобранные победителем соревнования. Для сравнения моделей в работе рассматривались базовые алгоритмы инкрементального моделирования [4]. Сравнение моделей проводилось на отложенном тестовом датасете по 4 различным метрикам: площадь под Qini (AUQC), площадь Uplift кривой (AUUC), коэффициент Кэндала и значение метрики Uplift30%. Во-вторых, в работе предложены и прошли сравнение новые алгоритмы инкрементального моделирования для области директ-маркетинга: предложен подход для построения *Uplift*-деревьев с использованием статистического теста [5] при разбиении вершины, для учёта дисбаланса классов в группе воздействия и контрольной

группе; предложены нейросетевые модели на базе сиамских сетей [1] с использованием архитектуры TabNet [3], и проведено сравнение их точности с другими подходами в задаче директ-маркетинга.

Таким образом, в рамках работе предложен подход для построения *Uplift*-деревьев с использованием статистического теста, а также усовершенствованы нейросетевые модели на базе сиамских сетей при помощи архитектуры TabNet. Использование статистического теста при построении *Uplift*-деревьев даёт прирост точности (на выборке X5 с отобранными признаками). Использование архитектуры TabNet вместо MLP позволяет улучшить точность моделей инкрементального моделирования на основе сиамских сетей. Проведено экспериментальное исследование и оценка качества предложенных и существующих подходов на различных наборах данных академической (Criteo Lab) и прикладной (соревнования X5 retail group) направленности. Результаты экспериментов показали, что универсальной модели, лучшей по всем метрикам нет — всё зависит от решаемой задачи (размера целевой группы).

Литература

- [1] Mouloud B., Olivier G., Ghaith K. Adapting Neural Networks for Uplift Models // arXiv preprint arXiv:2011.00041. 2020.
- [2] Diemert E. et al. A large scale benchmark for uplift modeling // KDD. 2018.
- [3] Arik S. O., Pfister T. Tabnet: Attentive interpretable tabular learning // arXiv preprint. arXiv:1908.07442. 2019.
- [4] Igor Schnakenburg, René Michel, Tobias von Martens Targeting Uplift: An Introduction to Net Scores // Springer International Publishing, 2019.
- [5] Uplift Regression: The R Package tools4uplift / Belbahri Mouloud, Alejandro Murua, Gandouet Olivier, Vahid Partovi Nia // arXiv preprint. arXiv:1901.10867v1. 2019.
- [6] X5. URL: <https://retailhero.ai>

Разработка прототипа планировщика для бессерверной платформы управления инфраструктурой

Кукушкин Денис Игоревич

Кафедра автоматизации систем вычислительных комплексов

e-mail: denis.cucu@yandex.ru

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

Бессерверные вычисления позволяют структурировать и гибко масштабировать программу, разбивая ее на отдельные функции. Возможность мгновенно перезапустить бессерверную функцию в случае ее отказа повышает надежность приложений построенных на бессерверной архитектуре. В то время как возможность исполнять эти функции на сторонней платформе с оплатой за время вычисления позволяет экономить на оборудовании и его поддержке. С ростом научного и коммерческого интереса и развитием модели бессерверных

вычислений была предложена концепция составных бессерверных функций. Эта концепция является достаточно органичным этапом развития парадигмы бессерверных вычислений, так как предлагает составлять архитектурно более сложные приложения из блоков *babel* — бессерверных функций. *Составная бессерверная функция babel* — это бессерверная функция вызывающая другие бессерверные функции, возможно также составные, в заданном порядке. Задача объединения бессерверных функций в программу несет в себе определенные трудности [1]. Платформы, позволяющие запускать составные бессерверные функции, тратят до 97% [2] вычислительного времени на накладные расходы связанные с запуском бессерверных функций входящих в составную бессерверную функцию и передачей информации между ними. Использование зависимостей по данным при разработке планировщика бессерверной платформы, управляющего выполнением составной бессерверной функции, может сократить накладные расходы. Главной идеей, лежащей в основе данной работы, является использование алгоритмов анализирующих зависимости по данным для того чтобы сократить накладные расходы на коммуникации, подготовку бессерверных функций и простой задач в ожидании выполнения.

В ходе изучения традиционных подходов к планированию с применением зависимостей по данным была предпринята попытка адаптировать уже известные модели и алгоритмы и применить их в бессерверных вычислениях. В результате исследования предложено разбиение проблемы планирования вычисления составной бессерверной функции на три этапа:

1. Анализ зависимостей по данным для своевременной подготовки бессерверных функций.
2. Анализ зависимостей по данным для составления близкого к оптимальному расписания вызовов частей составной бессерверной функции.
3. Выбор подходящих вычислительных узлов для исполнения вызванных бессерверных функций с целью сократить расходы на коммуникации.

В работе представлена математическая модель, формально описывающая данные этапы.

Был проведен обзор алгоритмов применяемых при планировании ресурсов компиляторами и при параллельных вычислениях в многопроцессорных системах, целью которого было определить наилучший алгоритм для реализации в прототипе планировщика. Для каждого алгоритма было указано, как именно он может быть применен при планировании ресурсов в бессерверных платформах. Результатом обзора является выбор эвристики критического пути для дальнейшей реализации прототипа.

В работе предложена архитектура прототипа на основе открытой бессерверной платформы *Fission*, реализующего эвристику критического пути. На основе архитектуры была написана реализация прототипа на языке *Go*, улучшающая работу планировщика системы *Fission* и предлагающая более продвинутый алгоритм планирования.

Было проведено экспериментальное исследование прототипа с целью оценить эффективность сокращения накладных расходов прототипом. Накладные расходы, полученные в результате работы прототипа, сравнивались с накладными расходами системы Fission без изменений, а также с результатами других известных бессерверных платформ [2]. В ходе экспериментального исследования выявлена тенденция к снижению доли накладных расходов относительно общего времени выполнения составной бессерверной функции при увеличении объема входных данных. Экспериментальное исследование подтверждает эффективность прототипа, в частности, зафиксирован прирост производительности до 100% на некоторых типах бессерверных функций. Экспериментальное сравнение прототипа с другими известными бессерверными платформами на схожих алгоритмах [2] показывает его конкурентоспособность.

Литература

- [1] Joseph Hellerstein, Jose Faleiro, Joseph Gonzalez et al. Serverless Computing: One Step Forward, Two Steps Back // ArXiv. 2018. Vol. abs/1812.03651.
- [2] Wen Jinfeng, Liu Yi. An Empirical Study on Serverless Workflow Service // ArXiv. 2021. Vol. abs/2101.03513.

Построение коммутируемых сетей для систем реального времени

Морквин Андрей Андреевич

Кафедра автоматизации систем вычислительных комплексов

e-mail: mr.andrej1102@yandex.ru

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

В работе рассматриваются коммутируемые сети, используемые для построения сетей обмена в информационно-управляющих системах реального времени (ИУС РВ) с архитектурой интегрированной модульной авионики (ИМА). Одной из проблем при переходе от ИУС РВ с федеративной архитектурой к ИУС РВ с интегрированной модульной архитектурой является значительное увеличение потока данных в сети обмена системы. Например, поток данных в сети обмена ИУС РВ от локационного комплекса увеличивается в 10^3 – 10^5 раз по сравнению с ИУС РВ с федеративной архитектурой [1]. Так же при построении сетей обмена ИУС РВ следует учитывать ограничения на возможные точки установки сетевого оборудования и ограничение на возможные варианты прокладки кабелей. С точки зрения надежности, чем меньше суммарная длина кабелей в сети, тем меньше перекрестные наводки, и, следовательно, меньше вероятность ошибок при передаче данных [2]. В связи со сказанным выше является актуальной проблема построения сетей обмена минимальной сложности, которые позволяют передавать сообщения в реальном времени.

В современных ИУС РВ с архитектурой ИМА для построения сетей обмена используются AFDX и TSN [3] сети.

В работе сформулированы задачи построения сетей обмена минимальной сложности, обеспечивающих передачу заданного набора периодических со-

общений в реальном времени с учетом ограничений на допустимые точки размещения сетевого оборудования, для AFDX и для TSN сетей.

Предложен алгоритм решения сформулированных задач. Он сочетает жадные стратегии и ограниченный перебор:

- на каждом шаге работы алгоритма делается локально-оптимальный выбор в соответствии с используемой жадной стратегией.
- на каждом шаге выполняется проверка того, что «жадный выбор не закрывает пути к оптимальному решению».
- вызов процедуры ограниченного перебора, если проверка условия «жадный выбор не закрывает пути к оптимальному решению» дала отрицательный результат.

В работе было проведено экспериментальное исследование точности алгоритма по критериям: количество размещенных сообщений из максимально возможного и отклонение сложности сети от минимально возможной. Эксперименты проводились на различных классах исходных данных характерных для бортовых авиационных систем. На всех классах данных алгоритм показал высокую точность.

Результаты выпускной работы для AFDX сетей были отражены в печатной работе [4].

Литература

- [1] Костенко В. А. Архитектура программно-аппаратных комплексов бортового оборудования // Изв. вузов. Приборостроение. 2017. Т. 60, № 3.
- [2] Al-Kuwaiti M., Kyriakopoulos N., Hussein S. A comparative analysis of network dependability, fault-tolerance, reliability, security, and survivability // IEEE Communications Surveys & Tutrial. 2009. Vol. 11, N 2.
- [3] Comparison of Time Sensitive Networking (TSN) and TTEthernet / Lin Zhao, Feng He, Ershuai Li, Jun Lu // IEEE/AIAA 37th Digital Avionics Systems Conference (DASC). 2018.
- [4] Костенко В. А., Морквин А. А. Построение бортовых коммутируемых сетей минимальной сложности // Программирование. 2021. № 4, С. 14–19.

Исследование применимости алгоритмов сжатия данных к таблицам классификации в сетевом процессорном устройстве

Никифоров Никита Игоревич

Кафедра автоматизации систем вычислительных комплексов

e-mail: nickiforov.nik@gmail.com

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

В настоящее время активно развиваются технологии программно-конфигурируемых сетей (ПКС) [1]. Для работы ПКС требуются высокопроизводитель-

ные коммутаторы, которые выполняют функцию передачи данных. Возникает задача разработки программируемого сетевого процессорного устройства (СПУ) [2], являющегося основным функциональным элементом коммутаторов. В работе рассматривается коммутатор функционирующий под управлением протокола OpenFlow [3]. Правила обработки пакетов в котором представляются в виде таблиц классификации, которые являются таблицами потоков в терминах протокола OpenFlow. В СПУ таблицы потоков представляются в виде программы обработки заголовков сетевых пакетов на языке ассемблера, получаемой с помощью транслятора таблиц потоков.

В данной работе рассматривался этап классификации пакетов. Под классификацией понимается процесс идентификации сетевого пакета по его признакам, определяемыми текущим протоколом. Таблицы потоков необходимы для процесса классификации пакетов. Таблица потоков — это набор правил, определенных протоколом OpenFlow. OpenFlow — один из наиболее распространенных протоколов для управления сетевым коммутатором ПКС. В данной работе рассматривается OpenFlow версии 1.3 [3]. Каждое правило содержит поле признака, битовую строку, по которой пакет может быть идентифицирован, и набор действий, которые СПУ выполняет с этим пакетом.

Данная работа посвящена разработке алгоритмов сжатия данных [4] для применения в трансляторе таблиц потоков в рамках рассматриваемой архитектуры сетевого процессорного устройства. Поскольку современные таблицы потоков занимают до нескольких десятков мегабайтов памяти [5]. Поэтому возникает задача сжатия таблиц потоков, для использования рассматриваемого СПУ в коммутаторах ПКС.

На основе проведённого обзора были выбраны алгоритм оптимального кеширования и алгоритм рекуррентного отсечения для дальнейшей реализации. Было проведено экспериментальное исследование, реализованных алгоритмов сжатия. В качестве исходных данных были использованы таблицы потоков по различным шаблонам, покрывающим большинство используемых протоколов.

Основной алгоритм сжатия таблицы потоков состоит из следующих шагов: получение таблицы потоков, оптимизация таблицы потоков, сжатие таблицы потоков, построение промежуточной структуры данных и трансляция таблицы потоков. Оптимизация таблиц потоков заключается в удаление незначачих правил и объединения похожих правил. Алгоритмы сжатия применяются на третьем шаге, к оптимизированной таблице потоков.

В рамках выпускной квалификационной работы разработаны алгоритмы сжатия таблиц потоков

OpenFlow. В ходе работы были достигнуты следующие результаты:

- Проведён обзор существующих алгоритмов сжатия данных, в результате обзора были выбраны для реализации алгоритм оптимального кеширования и алгоритм рекуррентного отсечения;
- Выбранные алгоритмы сжатия были адаптированы для внедрения в транслятор таблиц потоков;

- Разработана программная реализация алгоритмов сжатия данных в трансляторе таблиц потоков;
- Проведено экспериментальное исследование разработанных алгоритмов сжатия, использования которых позволило снизить объём памяти, затрачиваемой на хранение таблиц потоков;
- Статья по результатам работы опубликована на конференции SYRCoSE 2021 [6];

Экспериментальное исследование показало, что применения алгоритмов сжатия позволяет сократить объём используемой памяти в 1.5 – 2 раза.

Литература

- [1] Смелянский Р. Л. Программно-конфигурируемые сети // Открытые системы. 2012. № 9. С. 15–26
- [2] Беззубцев С. О., Васин В. В., Волканов Д. Ю. и др. Об одном подходе к построению сетевого процессорного устройства // Моделирование и анализ информационных систем. 2019. Т. 26, № 1. С. 39–62.
- [3] OpenFlow switch specification version 1.3.0 // Foundation Open Networking. 2012.
- [4] Gábor Rétvári, János Tapolcai, Attila Kőrösi et al. Compressing IP forwarding tables: Towards entropy bounds and beyond // ACM SIGCOMM Computer Communication Review. 2013. Vol. 43, no. 4. P. 111–122.
- [5] Rottenstreich, Ori and Tapolcai, János. Optimal rule caching and lossy compression for longest prefix matching. // IEEE/ACM Transactions on Networking. 2016. P. 864–878.
- [6] Nikita Nikiforov, Dmitry Volkanov. Data compression algorithms for flow tables in Network Processor RuNPU. // Труды института системного программирования РАН. 2021. № 3. (принято в печать).

Разработка методов построения неоднородных нейронных сетей

Селезнев Лев Евгеньевич

Кафедра автоматизации систем вычислительных комплексов

e-mail: lev4213@gmail.com

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

Современные нейросетевые алгоритмы нацелены на решение множества задач: от прогнозирования до классификации различных данных. Нейронные сети, решающие рассмотренные задачи зачастую имеют достаточно большое число параметров, что влияет на скорость работы сети и вычислительную нагрузку на систему в целом. Очевидно, что чем меньше сложность сети, тем быстрее сеть сможет выдать результат работы для определённых входных данных.

Задача построения неоднородных нейронных сетей в первую очередь решает проблему, связанную с большой вычислительной сложностью нейронных сетей. В последствии это позволяет применять неоднородные нейронные сети во встраиваемых системах, вычислительные ресурсы которых ограничены, и для которых важную роль играет время работы сети на входном наборе данных.

В рамках выпускной работы был разработан и протестирован алгоритм, решающий задачу построения неоднородных нейронных сетей минимальной сложности при соблюдении заданных ограничений на точность сети. В процессе разработки алгоритма были рассмотрены различные способы задания начального состояния сети, обучения сети (Отдельно был рассмотрен алгоритм случайного поиска с самообучением [1] в сравнении с классическими алгоритмами обучения), а также уменьшения сложности (или сжатия) сети.

Для проверки эффективности разработанного алгоритма произведено экспериментальное исследование, целью которого было сравнение полученной нейронной сети с полносвязными однородными нейронными сетями, а также сравнение качества работы предложенного алгоритма уменьшения сложности с существующей реализацией алгоритма удаления минимальных весов.

По результатам экспериментального исследования предложенный в работе алгоритм показал высокую эффективность в построении неоднородных нейронных сетей, в ряде случаев позволив добиться уменьшения размера сети более чем на 90%, допустив падение точности не более чем на 5% по сравнению с полносвязной однородной нейронной сетью. Таким образом описанный алгоритм решает задачу построения неоднородных сетей минимальной сложности с заданными ограничениями на величину точности.

Литература

- [1] Kostenko V. A., Seleznev L. E. Random search algorithm with self-learning for neural network training // Optical Memory and Neural Networks. 2021. Vol. 30, N. 3, P. 180–186.

Аналитический метод оценки времени отклика задач в системах интегрированной модульной авионики в случае интервальной неопределенности длительности выполнения работ

Рябченков Владимир Михайлович

Кафедра автоматизации систем вычислительных комплексов

e-mail: vultar1500@gmail.com

Научный руководитель — программист Глонина Алевтина Борисовна

В настоящее время основной концепцией построения бортовых вычислительных систем воздушных судов является архитектура интегрированной модульной авионики (ИМА). Системы ИМА представляют собой распределённые системы реального времени, состоящие из вычислительных модулей, соеди-

ненных коммутируемой сетью с поддержкой виртуальных каналов передачи данных.

Каждый модуль содержит один или несколько многоядерных процессоров. Вычислительные задачи для систем ИМА сгруппированы в разделы. Каждый раздел привязан к некоторому процессорному ядру. Все задачи являются периодическими; наименьшее общее кратное периодов всех задач называется интервалом планирования. За время одного периода, каждая задача выполняет работу или экземпляр задачи. Для каждого раздела на интервале планирования задан набор окон выполнения — набор интервалов времени, в течение которых могут выполняться только задачи из данного раздела. Выполнением работ раздела внутри окон управляет планировщик с фиксированными приоритетами и вытеснением. Стоит отметить, что одному процессорному ядру могут соответствовать сразу несколько разделов. Окна разделов привязанных к одному ядру не должны пересекаться.

Каждая вычислительная задача обладает следующими характеристиками: период выполнения, интервал длительности выполнения, фиксированный и уникальный в рамках раздела приоритет. Между задачами как из разных модулей, так и из одного модуля, могут существовать зависимости по данным, т. е. задача-получатель не может начать выполнение своей работы, пока не получит сообщения от соответствующих работ всех задач-отправителей.

Под конфигурацией системы ИМА понимается: набор вычислительных ядер; набор задач, сгруппированных в разделы; набор сообщений, передаваемых между задачами по сети; привязка разделов к ядрам; расписание окон разделов.

Время отклика задачи — это величина равная максимальному по всем работам этой задачи значению времен их завершения относительно начала периода. Наихудшее время отклика задачи (WCRT) — это максимально возможное время отклика задачи на данной конфигурации системы.

В данной ВКР рассматривается проблема оценки WCRT заданной задачи на заданной конфигурации системы. Известно, что уменьшение времени выполнения одной задачи может привести к увеличению времени завершения другой. В работе рассматриваются аналитические методы, учитывающие подобные аномалии.

Известные методы по нахождению точного значения WCRT имеют экспоненциальную сложность, а большинство аналитических методов, дающих верхние оценки, не учитывают наличие окон разделов или произвольных графов зависимостей по данным. Единственный аналитический метод, учитывающий все вышеперечисленные требования [1], дает оценки, на ряде примеров завышенные по сравнению с точными в несколько раз.

В работе предложен аналитический метод оценки времени отклика задач в системах ИМА, дающий оценку WCRT сверху. Данный метод основан на модификации алгоритма STBA, предложенного в работах [2, 3]. В отличие от базового алгоритма STBA, разработанный алгоритм дает более точные оцен-

ки, а также учитывает наличие окон разделов. Корректность предложенного алгоритма доказана.

Экспериментальное исследование показало, что предложенный алгоритм дает более точные оценки, чем алгоритм [1] как на искусственно сгенерированных, так и на реальных данных. Также исследование показало, что на большинстве задач из рассмотренных конфигураций предложенный алгоритм позволил получить точные значения WCRT.

Литература

- [1] Response-Time Analysis of Multipath Flows in Hierarchically-Scheduled Time-Partitioned Distributed Real-Time Systems / A. Amurrio, E. Azketa, J. Gutierrez, M. Aldea, M. G. Harbour // IEEE Access. 2020. Vol. 8. P. 196700–196711.
- [2] A novel analytical method for worst case response time estimation of distributed embedded systems / J. Kim, H. Oh, J. Choi, H. Ha, S. Ha // Proceedings of the 50th Annual Design Automation Conference. 2013. P. 1–10.
- [3] Choi J., Ha S. Worst-case response time analysis of a synchronous dataflow graph in a multiprocessor system with real-time tasks // ACM Transactions on Design Automation of Electronic Systems (TODAES). 2017. Vol. 22, N2. P. 1–26.

Метод управления процессами реконфигурации ПКС при условии её полной непротиворечивости для выполнения конкурирующих запросов построения маршрутов

Шапошников Владимир Александрович

Кафедра автоматизации систем вычислительных комплексов

e-mail: volodimirich@arccn.ru

Научный руководитель — с.н.с. Писковский Виктор Олегович

В данный момент технология Программно-Конфигурируемых сетей активно развивается, и проблема безопасности таких сетей стоит особенно остро [1]. При изменении правил маршрутизации на сетевых устройствах могут возникать конфликты, вызванные конкурирующими запросами маршрутизации потоков через объекты конфигурации. В отличие от «ручной» конфигурации традиционных сетей, в ПКС контроллер на основе пользовательских запросов в соответствии с принятыми политиками безопасности автоматически вносит изменения в конфигурацию сети. Ответственность за корректную работу сети возлагается на соответствующие приложения контроллера и реализованные в них методы и алгоритмы. Таким образом, наряду с культурой производства и внедрения приложений, управляющих сетью, большое значение приобретает теоретическое обоснование методов и алгоритмов, по которым работают названные приложения. В работе получены следующие основные результаты:

1. Выбран подход к управлению процессами реконфигурации ПКС при заданных условиях на основании выполненного обзора [2].

2. Сформулированы и доказаны условия непротиворечивой реконфигурации ПКС.
3. Сформулирован метод автоматического построения плана упорядочивания запросов для случая полной непротиворечивости, позволяющий балансировать нагрузку в сети для приведения потоков к независимому друг от друга виду
4. Условия непротиворечивой реконфигурации были проверены на экспериментальном стенде, разработанном на базе эмулятора компьютерных сетей ns3 [3] с использованием модуля ofswitch13 [4]. Эксперимент показал возможность применения подходов упорядоченности и блокировок, а также условия реконфигурации без потерь для интернета вещей.

Литература

- [1] Грушо А. А., Терёхина И. Ю. Анализ непротиворечивости реконфигурации программно-конфигурируемой сети // Системы и средства информ. 27:3 (2017). С. 12–22.
- [2] Шапошников В. А., Писковский В. О. Постановка задачи и краткий обзор решений для проблемы управления процессами согласованной реконфигурации ПКС // Программные системы и инструменты. Тематический сборник. № 20. Т. 20. С. 133–141.
- [3] NS-3: Network simulator 3. URL: <https://www.nsnam.org/>
- [4] NS-3 Ofswitch module. URL: <http://www.lrc.ic.unicamp.br/ofswitch13/>

Исследование методов решения задачи обучения по нескольким примерам, основанных на применении ансамбля алгоритмов формирования синтетических обучающих данных

Камалов Ирек Маратович

Кафедра суперкомпьютеров и квантовой информатики

e-mail: kamalovcmcmsu@yandex.ru

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

Многие задачи глубокого обучения требуют построения моделей по небольшому объёму данных [1, 2]. Например, такие задачи возникают в робототехнике [1] и в обработке естественных языков [2]. Существуют методы генерации вспомогательных примеров, которыми можно расширить небольшую выборку и свести задачу к классическим методам обучения. Целью работы был анализ таких решений и усовершенствование за счёт применения в ансамбле.

Рассматривается задача классификации изображений, но для каждого класса доступно только несколько размеченных изображений. Для решения используются методы генерации синтетических примеров. С их помощью формируется обучающая выборка для построения классификатора.

Задачей было отобрать несколько алгоритмов генерации, реализовать их совместное применение и исследовать его влияние на качество классификации. Было предложено строить синтетические выборки каждым генератором по отдельности и смешивать их для получения более репрезентативной. Целевая модель обучается на смешанной выборке. Такое применение генераторов понимается в работе под ансамблем. Предметом исследования также является оптимальная пропорция для создания смеси. В качестве методов генерации были отобраны нейросетевые модели, предложенные в работах [3, 4].

Было проведено тестирование ансамбля и классификатора на открытых базах изображений miniImageNet и CIFAR100. Выбирается 80 классов для обучения генераторов и 5 классов для тестирования. В задаче CIFAR100 были взяты 5 связанных категорий, принадлежащих классу транспортных средств и передвижной техники, - газонокосилка, ракета, трамвай, танк, трактор. В miniImageNet категории независимы: автобус, электронное табло, сцена, ваза, торт. Целевой классификатор - многоклассовая логистическая регрессия.

Для каждой базы изображений приведено 2 графика точности классификации при различных соотношениях смешиваемых выборок (рисунки 1 и 2). Левый и правый графики построены для случаев, когда для целевого класса дано, соответственно, по 1 и по 5 обучающих примеров. По оси абсцисс откладывается перевес числа примеров в пользу одного из генераторов при смешивании. Нуль означает смешивание в равном соотношении, то есть 512 к 512 при итоговой выборке из 1024 примеров. Точки, отдаляющиеся от нуля, соответствуют всё большему преобладанию примеров одного из генераторов. Для краткости центроидный генератор [4] обозначен как $G1$, Δ -кодировщик [3] как $G2$.

На основании экспериментов было сделано 2 основных вывода:

1. Смешивание выборок улучшает качество решения на 3-6% в случае 5 целевых классов
2. Оптимальное соотношение для смешивания зависит от задачи

В ходе работы был самостоятельно реализован ансамбль генераторов и входящие в него методы [3, 4], проведено исследование точности классификации на базах miniImageNet и CIFAR100 на примере многоклассовой логистической регрессии. Был получен прирост точности благодаря смешиванию выборок. Дополнительно было установлено, что оптимальная пропорция для создания смеси - гиперпараметр ансамбля, так как её выбор зависит от задачи.

Литература

- [1] Ayub A. Wagner A. R. Tell me what this is: Few-Shot Incremental Object Learning by a Robot. // In Proceedings of the 33rd International Conference on Intelligent Robots and Systems, Las Vegas, USA, 2020, P. 8344–8350
- [2] Brown T. Mann B. Ryder N. Subbiah M. Kaplan J. D. Dhariwal P. Neelakantan A. Shyam P. Sastry G. Askell A. Agarwal S. Herbert-Voss A. Krueger G. Henighan T. Child R. Ramesh A. Ziegler D. Wu J. Winter C. Hesse C. Chen M. Sigler E.

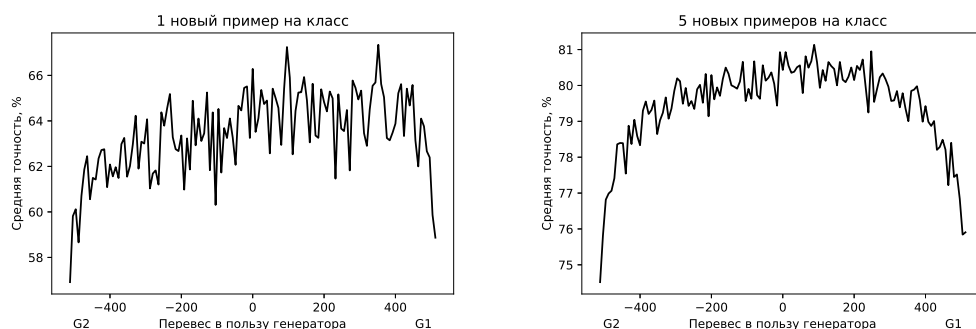


Рис. 1: *miniImageNet*. Точность для различных соотношений при смешивании.

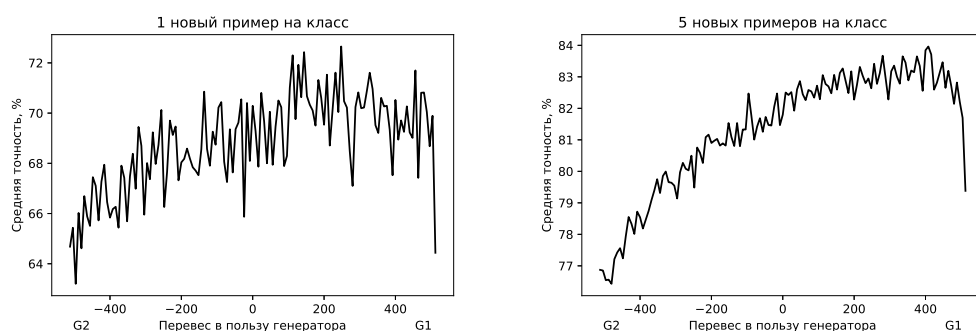


Рис. 2: *CIFAR100*. Точность для различных соотношений при смешивании.

Litwin M. Gray S. Chess B. Clark J. Berner C. McCandlish S. Radford A. Sutskever I. Amodei D. Language Models are Few-Shot Learners. // In Proceedings of the 33rd International Conference on Neural Information Processing Systems, Vancouver, Canada, 2020, P. 1877–1901

- [3] Wang Y. Yao Q. Kwok J. T. Ni L. M. Generalizing from a Few Examples: A Survey on Few-Shot Learning. // ACM Computing Surveys, 2020, Volume 53, Issue 3, P. 1–34
- [4] Schwartz E. Karlinsky L. Shtok J. Harary S. Marder M. Kumar A. Feris R. Giryes R. Bronstein A. M. Delta-encoder: an effective sample synthesis method for few-shot object recognition. // In Proceedings of the 32nd International Conference on Neural Information Processing Systems, Montreal, Canada, 2018, P. 2850–2860.
- [5] Hariharan B. Girshick R. Low-shot Visual Recognition by Shrinking and Hallucinating Features. // In Proceedings of the International Conference on Computer Vision, Venice, Italy, 2017, P. 3037–3046.

Суперкомпьютерный кодизайн задачи симуляции жидкокристаллических структур для современных вычислительных архитектур

Личманов Дмитрий Игоревич

Кафедра суперкомпьютеров и квантовой информатики

e-mail: dimlichmanov@gmail.com

Научный руководитель — чл.-корр. РАН, д.ф.-м.н. проф. Воеводин Владимир Валентинович

Суперкомпьютерный кодизайн включает в себя реализацию программного пакета на нескольких высокопроизводительных архитектурах и получение на каждой из архитектур наиболее оптимизированной версии пакета. Данная процедура исключительно полезна для эффективного решения ресурсоёмких задач и позволяет определять для подобных задач наиболее производительную архитектуру. Пакет прикладного программного обеспечения, используемый в выпускной работе для демонстрации недостатков существующих решений и преимуществ представленных оптимизаций, разработан для моделирования структуры и свойств капель жидких кристаллов [1]. В данном пакете, написанном на CUDA C++, стохастическая оптимизация определенного на кубической решетке функционала выполняется с помощью метода отжига - подкласса широко используемых методов Монте-Карло, действующих по схеме марковских цепей. Выпускная квалификационная работа посвящена как оптимизации частей данного программного обеспечения, так и формированию общего подхода к оптимизации описанного класса алгоритмов.

В работе исследуется производительность приложения на графических процессорах NVIDIA микроархитектур Volta, Pascal и Turing, векторных ускорителях NEC SX-Aurora TSUBASA, многоядерных процессорах Intel Xeon Skylake и ARM Kunpeng. Эти платформы принадлежат к значительному и представительному классу современных суперкомпьютерных архитектур, которые оснащены как множеством мощных вычислительных юнитов, позволяющих эффективно запускать многопоточный код, так и оперативной памятью с высокой пропускной способностью.

Для исследования узких мест исследуемого программного пакета и определения возможных оптимизаций были использованы как инструменты профилировки (NVIDIA Visual Profiler для версии на CUDA, Ftrace для NEC, Intel Vtune Profiler для Intel, perf для ARM и Intel), так и бенчмарки с различными шаблонами доступа к памяти, такие как STREAM [2].

Для CPU-версий программного пакета эффект от внедрения предложенных оптимизаций исследовался с помощью top-down анализа, позволяющего определять время нахождения различных процессорных микроопераций в слотах конвейера и причину появляющихся простоев микроопераций. Данный top-down анализ для платформ с процессорами Intel проводился с помощью Intel Advisor, для ARM аналогичные показатели были рассчитаны с помощью аппаратных метрик (PMU). Для версий типового ядра на Nvidia GPU и

NEC SX-Aurora TSUBASA эффект от применения оптимизаций был показан с помощью roofline модели, определяющей арифметическую интенсивность приложения. Roofline модель также использовалась на всех целевых платформах с целью определения предельных показателей пропускной способности различных уровней подсистемы памяти - DRAM, L1-кеш, LL-кеш.

Большое внимание при внедрении оптимизаций было уделено шаблону доступа к глобальной памяти. В результате, удалось увеличить эффективность транзакций к подсистеме памяти для GPU-версии программного пакета, использовать более эффективные операции загрузки данных на векторные регистры для NEC-версии, для CPU-версий удалось добиться крайне высокого соотношения активных микроопераций к общему их числу.

Для всех версий была применена оптимизация многозадачности - с помощью директив OpenMP, технологии CUDA Streams и динамического определения числа потоков, для приложения, оперирующего даже над небольшими решётками, удалось добиться предельных достижимых показателей пропускной способности памяти на каждой из архитектур, что является отличным результатом для memory-bound приложений.

В рамках выпускной работы были внедрены также оптимизации размера транзакций, лимитирование регистрового файла для большей резидентности CUDA-ядер, реформатирование структур данных, исследовано влияние точности вычислений на эффективность загрузки данных. В CPU-версии программного пакета была применена векторизация с использованием AVX-512 инструкций для процессоров Intel и ARM Neon инструкций для платформ с ARM. Исследована масштабируемость программного пакета на несколько NUMA-сокетов. Была проведена оптимизация редукции с использованием операций, выполняющихся внутри одного GPU-варпа.

После внедрения предложенных оптимизаций, был проведён сравнительный анализ скорости выполнения типового ядра на различных платформах. Наилучшую производительность показала платформа с NVIDIA V100 GPU. В результате, для приложений, запущенных на небольших физических решётках было получено ускорение в 4-5 раз, для больших - в 2-2.5 раза, что является отличным практическим результатом. Более того, была разработана и MultiGPU версия прикладного пакета, которая, в дополнение к уже достигнутому ускорению, позволила получить линейное, от числа используемых графических карт, ускорение.

Литература

- [1] V. Yu. Rudyak, A. V. Emelyanenko, and V. A. Loiko. Structure transitions in oblatenematic droplets // *Physical Review E*, 2013. Vol. 88, N 5.
- [2] Tom Deakin and Simon McIntosh-Smith. GPU-stream: Benchmarking the achievable memory bandwidth of graphics processing units // *IEEE ACM SuperComputing*, 2015. P.3202–3216.

Исследование метрического метода в задаче обучения по нескольким примерам

Питанов Елисей Сергеевич

Кафедра суперкомпьютеров и квантовой информатики

e-mail: piteles@mail.ru

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

Для многих задач, при решении которых применяются нейронные сети, характерен дефицит обучающих данных. [1] В такой ситуации стандартные алгоритмы не позволяют получить нейросетевые решения, обладающие достаточной точностью. Методы обучения по нескольким примерам разрабатываются для решения именно этой проблемы.

В данной работе был предложен новый алгоритм обучения по нескольким примерам для построения сетей классификации изображений. Известно, что для сверточной нейронной сети характерно формирование в своих слоях представлений, описывающих различные признаки исходных данных [2]. Этот факт лег в основу идеи предложенного в работе подхода. Класс входного изображения определяется близостью его внутренних представлений, полученных в некоторой базовой нейронной сети, и соответствующих представлений для изображений известных классов. Данный подход относится к метрическим методам обучения по нескольким примерам - для проведения классификации используется некоторое вложенное пространство. [3]

Основная идея заключается в использовании нескольких сетей. Первая сеть используется для создания промежуточных представлений, которые затем передаются в выбранную экспериментально функцию расстояния. Далее полученные числа попадают в сети коэффициентов, для каждого класса изображений настраивается своя такая сеть, в которой принимается решение о принадлежности изображений к одному и тому же классу. Выходы из сетей коэффициентов передаются в финальный классификатор, в котором происходит итоговое определение класса.

Для разработки данного подхода был проведен большой объем исследований, касающийся построения базовой сверточной нейронной сети, выбора метрик сравнения внутренних представлений и построения финального классификатора.

Для обеспечения экспериментальной проверки всех исследуемых вариантов была разработана программная реализация с поддержкой выполнения на графических ускорителях. Тестирование проводилось на двух базах изображений.

Промежуточные результаты, демонстрируемые разработанным алгоритмом при бинарной классификации, обладают достаточно высокой точностью, однако, финальное решение для пяти классов уступает современным подходам [4]. Был предложен способ для повышения конечной точности – построение специализированных базовых сетей, в которых получаемые внутренние представления

дополнительно преобразуются. Этому направлению предполагается посвятить будущие исследования.

Литература

- [1] Caruana et. al., Overfitting in Neural Nets: Backpropagation, Conjugate Gradient, and Early Stopping // <https://proceedings.neurips.cc/paper/2000/file/059fdcd96baeb75112f09fa1dcc740cc-Paper.pdf>
- [2] Li et. al., Revisiting Local Descriptor based Image-to-Class Measure for Few-Shot Learning // <https://arxiv.org/abs/1903.12290>
- [3] Wertheimer et. al., Few-Shot Learning with Localization in Realistic Settings // <https://arxiv.org/abs/1904.08502>
- [4] <https://paperswithcode.com/sota/image-classification-on-imagenet>

Использование технологии InfiniBand для передачи данных системы мониторинга суперкомпьютера

Сорокина Надежда Вячеславовна

Кафедра суперкомпьютеров и квантовой информатики

e-mail: sorenln@mail.ru

Научный руководитель — к.ф.-м.н. с.н.с. Стефанов Константин Сергеевич

На сегодняшний день наблюдается значительный рост обрабатываемых данных. С точки зрения сетевых технологий данная тенденция говорит о необходимости высокой пропускной способности канала передачи данных на всем пути. Применение технологии InfiniBand стало распространенным решением сложившейся проблемы транспортировки данных и снижения задержки на уровне программных приложений в суперкомпьютерных системах.

Введение. Функционирование суперкомпьютера напрямую зависит от сети передачи данных, и согласно списку Top500 более трети суперкомпьютеров используют технологию скоростной передачи данных InfiniBand. Кроме того, для правильной работы суперкомпьютера важна система мониторинга. Традиционно такие системы разрабатываются с использованием сети Ethernet, однако в мире существует и опыт использования InfiniBand. Например, система мониторинга LDMS [1], изначально реализованная вокруг этой технологии в университете Иллинойса. В рамках научной работы была проведена интеграция технологии InfiniBand в систему мониторинга DiMMon [2], уже функционирующую на суперкомпьютере «Ломоносов-2».

Преимущество использования технологии InfiniBand. Система мониторинга DiMMon — это распределенная модульная система, состоящая из агентов мониторинга. Каждый агент представляет собой набор модулей, которые имеют определенный функционал. Для того, чтобы связать работу модулей вместе, необходимо иметь модуль передачи данных, поскольку агенты мониторинга могут быть запущены на разных узлах суперкомпьютера.

В системе DiMMon модуль передачи данных использует сеть Ethernet. Особенность технологии Ethernet заключается в том, что для передачи данных приложению необходимо обратиться к ядру операционной системы за предоставлением сетевых ресурсов, после чего сами данные проходят через весь стек протоколов TCP/IP. Технология InfiniBand реализует совсем иной подход — приложение передает данные сразу сетевой карте, тем самым разгружая процессор и уменьшая время передачи сообщения.

Взаимодействие приложений по сети InfiniBand. При выборе транспортного режима в системе мониторинга, необходимо учитывать, что система работает на большом количестве узлов, хранение соединения с каждым из них потребует значительных накладных расходов. В связи с этим наиболее подходящим типом передачи данных является Unreliable Datagram, т. е. ненадежная передача сообщения без установления соединения. Программная реализация передачи данных по сети InfiniBand требует выделения набора рабочих структур и ассоциации их с противоположной стороной соединения. Таким образом создается канал обмена сообщениями InfiniBand, в который далее поступают запросы на передачу и прием данных.

Использование InfiniBand при разработке интерфейса передачи данных. Задача использования InfiniBand в системе DiMMon была разделена на две подзадачи — реализация интерфейса передачи данных по сети InfiniBand и интеграция разработанного интерфейса в систему. Для реализации интерфейса передачи данных использовался язык C++, поскольку он предоставляет механизм наследования. Это дает возможность для самостоятельно определять функционал сервера и выбирать режим работы серверной и клиентской части. Программная реализация передачи данных схематично представляется в виде обработки графа состояний канала сообщений InfiniBand. Однако используя разработанный интерфейс, пользователь может просто создать экземпляры классов Server и Client и отправлять запросы, при этом не зная, каким образом строится передача данных по сети InfiniBand и как этот процесс реализуется программно. Тестирование разработанного интерфейса производилось на узлах суперкомпьютера «Ломоносов-2». Пропускная способность канала сообщений приблизилась к 12,5 Гбит/секунду.

Интеграция интерфейса передачи данных в систему мониторинга. Разработанный программный интерфейс был внедрен в систему мониторинга DiMMon, а именно — в модуль передачи данных, который разбит на два подмодуля — Send и Receive, использующие экземпляры классов Client и Server соответственно. Система DiMMon контролирует работу подмодулей посредством управляющих сообщений. Например, при создании сервера системе мониторинга передаются файловые дескрипторы очереди, к которой подключаются клиенты, и каналов сообщений, в которые приходят данные. Система мониторинга отслеживает эти дескрипторы, и при получении новых данных отправляет события запущенному подмодулю. Тестирование системы мониторинга с интегрированной передачей данных по InfiniBand основывалось на

анализе загрузки процессора. Запуск системы DiMMon с модулями передачи данных и по InfiniBand, и по Ethernet показал одинаковую загрузку процессора, т.е. передача данных не оказывает существенного влияния на работу. При этом внедрение технологии InfiniBand позволяет повысить скорость передачи данных в системе мониторинга.

Литература

- [1] Agelastos A. и др. The Lightweight Distributed Metric Service: A Scalable Infrastructure for Continuous Monitoring of Large Scale Computing Systems and Applications // SC14: International Conference for High Performance Computing, Networking, Storage and Analysis. IEEE, 2014. С. 154–165.
- [2] Stefanov K. и др. Dynamically Reconfigurable Distributed Modular Monitoring System for Supercomputers (DiMMon) // 4th International Young Scientist Conference on Computational Science / Elsevier B. V., 2015. Т. 66. С. 625–634.

Разработка и исследование неоднородных эволюционных алгоритмов оптимизации в задачах машинного обучения

Цуканова Мария Александровна

Кафедра суперкомпьютеров и квантовой информатики

e-mail: cukanovaa93@gmail.com

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

В области искусственного интеллекта под эволюционными алгоритмами подразумевается эвристика поиска решений, основанная на естественном отборе. Как правило, эти алгоритмы применяются для задач с изменяющейся средой, где пространство поиска настолько огромно, что точное решение найти невозможно, такими являются задачи настройки весов искусственной нейронной сети и построения игровых стратегий.

Классической проблемой в практическом использовании эволюционных методов является сходимость к однородной (гомогенной) популяции, вследствие чего трудно найти глобальный оптимум. В данной работе на задаче обучения рекуррентных нейронных сетей и построения игровых стратегий сравнивается эффективность обучения с помощью классических эволюционных методов и неоднородных эволюционных алгоритмов.

Основной предмет исследования — неоднородный клеточный генетический алгоритм, суть которого заключается во введении зависимости величины мутации от положения особи в клеточном пространстве. Популяция представляется в виде связного графа, в котором каждая вершина является индивидуумом, который взаимодействует со своими ближайшими соседями. Таким образом достигается локальность взаимодействия между особями, что позволяет дольше поддерживать разнообразие в популяции.

Результаты проведенных вычислительных экспериментов показывают, что простые эволюционные алгоритмы уступают предложенным неоднородным клеточным алгоритмам. Разнообразие в популяции дольше поддерживается,

когда обучение происходит с помощью неоднородного клеточного алгоритма. Также в работе были предложены параллельные оптимизации процесса обучения на CPU и GPU, которые позволили значительно сократить время, отведенное на обучение.

Литература

- [1] Ершов Н. М., Попова Н. Н. Естественные модели параллельных вычислений // Компьютерные исследования и моделирование. 2015. Т. 7, № 3. С. 781–785.
- [2] Карпенко А. П. Современные алгоритмы поисковой оптимизации. Алгоритмы, вдохновленные природой. Изд-во: МГТУ, 2017.
- [3] Ершов Н. М. Неоднородные клеточные генетические алгоритмы // Компьютерные исследования и моделирование. 2015. Т. 7, № 3. С. 775–780. стендовые доклады.
- [4] Alba E., Dorronsoro B. Cellular Genetic Algorithms // Springer, 2008.

Разработка и исследование архитектур рекуррентных нейронных сетей в задачах построения игровых стратегий

Шалагин Никита Дмитриевич

Кафедра суперкомпьютеров и квантовой информатики

e-mail: shalagin.nikita99@gmail.com

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

Аннотация. В настоящий момент задачи построения игровых стратегий получили широкое распространение. Они появляются в обширном круге отраслей: от создания автопилотов для автомобилей до создания искусственных противников в компьютерных играх. Результаты современных исследований демонстрируют, что искусственные нейросети хорошо подходят на роль агентов, реализующих игровую стратегию. Среди достижений полученных на основе такого подхода стоит отметить: OpenAI Five от OpenAI — обыграл несколько профессиональных команд в компьютерную игру Dota 2; автопилот FSD от Tesla — способен свободно вести машину по дорогам общего пользования; AlphaZero от DeepMind — обыграл 18-кратного чемпиона мира в Го.

Одним из центральных свойств для хорошей нейросети является умение учитывать свои предыдущие действия в игре, чтобы придерживаться определенной стратегии в процессе принятия решений. В современных работах для решения данной задачи как правило встречается два подхода.

Первым из них является явная передача исторических данных на вход нейронной сети (например передача фиксированного количества предыдущих состояний игрового поля на вход нейросети). Данный подход имеет ряд недостатков: во-первых, для построения стратегии может требоваться больше ходов, чем было включено во входные данные; во-вторых, в задачах с большой размерностью игрового поля добавление каждого нового исторического поля будет провоцировать добавление большого количества новых весов, что в свою

очередь будет все сильнее и сильнее утяжелять процесс обучения/вычислений нейросети.

Вторым, и наиболее современным подходом в данной области являются нейронные сети, слои которых допускают не только прямое, но и обратное распространение сигнала. В общем случае такие сети называются рекуррентными нейронными сетями. Благодаря обратным связям, эти сети могут условно запоминать свои предыдущие шаги, и учитывать их при принятии текущих решений.

В данной работе исследуется эффективность различных архитектур рекуррентных нейронных сетей для задачи создания игровых ботов. Производится сравнительный анализ достигнутых результатов (насколько хорошо сети научились играть) и динамики обучения (при обучении с помощью эволюционных алгоритмов) трех различных архитектур нейросетей

Результаты. В данной работе были изучены и реализованы различные архитектуры рекуррентных нейронных сетей. На практике была исследована и подтверждена их применимость в задачах построения игровых стратегий при обучении с помощью неоднородных эволюционных методов оптимизации. Также были предложены параллельные оптимизации процесса обучения на CPU и GPU, которые позволили значительно сократить время, отведенное на обучение. Были проведены эксперименты, показывающие, что простые рекуррентные сети уступают LSTM/GRU сетям. Была разработана гибкая высокопроизводительная система, позволяющая быстро выстраивать эффективный процесс обучения нейросетевой модели эволюционными методами, с использованием нескольких gpu/cpu устройств. Были разработаны высоко оптимизированные игры, которые хорошо подходят в качестве среды для тренировки агента в задаче обучения игровой стратегии, и могут быть легко интегрированы в другие работы в ходе дальнейших исследований. Также для игр были разработаны визуализации, позволяющие наглядно показать, как играет нейронная сеть.

Система автоматического обнаружения фейковых новостей на русском языке

Авагян Давид Арменович

Кафедра алгоритмических языков

e-mail: david_avagyan@list.ru

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

Важнейшими свойствами информации являются актуальность и достоверность. Примером популярного источника информации может служить новостной ресурс. Роль задачи распознавания фейковых новостей возрастает с каждым днём, однако объёмы новостных потоков уже не позволяют производить модерацию новостей вручную; также не формализовано понятие фейка.

Целью данной работы является создание системы автоматического распознавания и классификации фейковых новостей на русском языке. В работе выдвигается гипотеза о том, что во многих случаях для определения достовер-

ности новости достаточно её лингвистического анализа. Рассмотрение медиа- и метаданных, относящихся к новости, выходит за рамки исследования.

Существующие определения и классификации фейков (например, изложенные в работах [1, 2]) не подходят для применения в автоматической обработке текстов (АОТ). Определения фейков встречаются и в законодательстве разных стран, включая Российскую Федерацию [3], но они также неприменимы в АОТ. В статье [4] изложены определение и классификация фейковых новостей, которые легли в основу данного исследования.

В работе *новость* — текст публицистического стиля, созданный с целью освещения событий и явлений действительности. Важными характеристиками достоверных новостей признаются наличие официального проверенного источника и объективность изложения. *Фейковая новость (фейк)* — веб-контент, распространяемый под видом новостей, но содержащий ложную информацию или преследующий иные цели, в отличие от достоверных новостей. В частности, фейки могут создаваться для дезинформации, выражения субъективной позиции автора, а также в рекламных или юмористических целях.

Выделяются следующие разновидности фейковых новостей (их определения и признаковые описания даны в статье [4]):

- *ложные новости* (англ. false news);
- *поляризованный контент* (англ. polarized content);
- *сатирические новости* (англ. satire);
- *непреднамеренно искажённые факты* (англ. misreporting);
- *комментарии и мнения* (англ. commentary);
- *нативная реклама* (англ. persuasive information, букв. «навязчивая информация»);
- *гражданская журналистика* (англ. citizen journalism).

В рамках данной работы автоматически собраны и размечены фейки четырёх типов: сатира, комментарии, нативная реклама и гражданская журналистика — для них не требуется решение сложной задачи проверки фактов. Более 57 000 новостей на русском языке, собранных из десяти источников*, объединены в датасет TopFakeSet, доступный в репозитории на GitLab[†]. Для сбора новостей реализованы краулеры для каждого источника. TopFakeSet является одной из малочисленных открытых коллекций новостей на русском языке и превосходит по объёму многие датасеты фейковых новостей.

Датасет TopFakeSet послужил выборкой для обучения и тестирования классификаторов новостей. Основной метрикой качества выбрана мягкая точность (англ. soft accuracy) — модификация метрики точности, частично поощряющая модель в случае отнесения фейков одного типа к другому, поскольку

* panorama.pub (сатира), wikinews.org (гражданская журналистика), vz.ru, newsru.com, ria.ru, iz.ru (комментарии), pikabu.ru, meduza.io (нативная реклама), interfax.ru, lenta.ru (достоверные новости).

[†]gitlab.com/Mathematician2000/topfakeset

фейк тем не менее распознан. Лучшим построенным классификатором стал ансамбль из трёх базовых моделей: градиентного бустинга, логистической регрессии и перцептрона — качество на тестовой выборке превысило 96.5%. Классификатор доступен в виде веб-приложения на PaaS-платформе Heroku[‡].

Существует несколько обозримых путей расширения выполненной работы. Например, возможно построение классификатора новостей, выделяющего в тексте n -граммы и конструкции, указывающие на его фейковость. Другое направление исследований состоит в уточнении классификации фейков. При этом большую роль сыграют лингвистические признаки и база знаний, позволяющая судить о справедливости фактов и логичности их изложения.

Литература

- [1] Defending against neural fake news / Zellers R., Holtzman A., Rashkin H., Bisk Y., Farhadi A., Roesner F., Choi Y. // Advances in Neural Information Processing Systems 32. 2019. P. 9054–9065.
- [2] Fake news detection on social media: a data mining perspective / Shu K., Silva A., Wang S., Tang J., Liu H. // ACM SIGKDD Explorations Newsletter. 2017. Vol. 19. N 1. P. 22–36.
- [3] О внесении изменений в статью 15-3 Федерального закона «Об информации, информационных технологиях и о защите информации»: федер. закон Рос. Федерации от 18 марта 2019 г. № 27-ФЗ: принят Гос. Думой Федер. Собр. Рос. Федерации 7 марта 2019 г.: одобр. Советом Федерации Федер. Собр. Рос. Федерации 13 марта 2019 г.
- [4] «Fake news» is not simply false information: a concept explication and taxonomy of online content / Molina M. D., Sundar S. S., Le T., Lee D. // American Behavioral Scientist. 2019. Vol. 65. N 2. P. 180–212.

Алгоритмы восходящего синтаксического анализа на основе детерминированных бесконтекстных L-графов

Арсенин Никита Михайлович

Кафедра алгоритмических языков

e-mail: veigh70@yandex.ru

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

Задача синтаксического анализа является одной из важнейших в теории формальных языков. Чаще всего данная задача рассматривается применительно к классу контекстно-свободных языков. Известны универсальные алгоритмы синтаксического анализа, основывающиеся на понятии контекстно-свободной (бесконтекстной) грамматики [1].

На практике часто используются подклассы бесконтекстных грамматик, для которых были разработаны более быстрые (по сравнению с универсальными) алгоритмы синтаксического анализа. Одним из таких классов является класс $LR(1)$ -грамматик. Доказано, что по любой грамматике из данного класса

[‡]top-fake-set.herokuapp.com

можно построить $LR(1)$ -анализатор, позволяющий проводить восходящий синтаксический анализ за линейное время от длины входной строки [2]. Однако внутреннее устройство такого анализатора является достаточно сложным и требует введения большого числа дополнительных понятий и конструкций. По этой причине актуальной является задача построения аналога восходящего анализатора с более простой внутренней структурой.

В данной работе рассматривается подход, основанный на бесконтекстных L -графах [3]. Разработан и обоснован алгоритм построения бесконтекстного L -графа по $LR(1)$ -грамматике с сохранением возможности восстановления синтаксических деревьев разбора. Также был разработан алгоритм детерминизации бесконтекстного L -графа, основанный на применении операций конкатенации, объединения и звезды Клини над регулярными выражениями над скобочными следами. Кроме того, приведен алгоритм восходящего синтаксического анализа, основанный на детерминированных бесконтекстных L -графах.

Литература

- [1] А. Ахо, Дж. Ульман. Теория синтаксического анализа, перевода и компиляции, том 1. М. : Мир, 1978. С. 352–369.
- [2] В. А. Серебряков. Теория и реализация языков программирования. М. : Издательский отдел факультета ВМК МГУ имени М.В.Ломоносова, 2011. С. 352–369.
- [3] А. А. Вылиток, П. Г. Сутырин. Характеризация формальных языков графами // Сборник тезисов научной конференции "Тихоновские чтения", Москва, МГУ имени М.В.Ломоносова, факультет ВМК, 25-29 октября 2010 г. - 2010.

Программное средство рекомендации фильмов на основе их сюжетов и оценок пользователей

Дубинин Владимир Дмитриевич

Кафедра алгоритмических языков

e-mail: vl.dubinin23@yandex.ru

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

В выпускной квалификационной работе рассмотрена задача создания рекомендательной системы. Система должна, используя минимальную информацию о пользователе, рекомендовать ему фильмы к просмотру. Для решения поставленной задачи необходимо было предложить методы предсказания предпочтения пользователя, где предпочтение — это числовая характеристика, которая показывает, насколько сильно тот или иной фильм должен понравиться пользователю.

Как правило, выделяют три основных подхода к предсказанию предпочтений в рекомендательных системах [1]:

1. **На основании признаков описаний (content-based)**: используется описательная информация о пользователях (демографическая информация) и фильмах (сюжет, жанр).

2. **Коллаборативная фильтрация (collaborative filtering):** рекомендации строятся исключительно на основании известных системе взаимодействий (проставление оценок) пользователей с фильмами.
3. **Гибридный подход:** используется композиция методов первых двух подходов.

Нами использовался гибридный подход, поскольку он позволяет устранять недостатки первых двух подходов, к которым относится вычислительно затратное предсказание предпочтения методами, опирающимися только на признаки описания, и проблема холодного старта для пользователя и фильма, которая характерна для методов коллаборативной фильтрации. Под холодным стартом понимается ситуация нехватки данных для корректной работы рекомендательной системы.

В рамках данной работы был проведен обзор существующих методов решения задачи предсказания предпочтения для пары пользователь-фильм и проблемы холодного старта пользователя и фильма. Для предсказания предпочтения был выбран метод SVD и его модификация SVD biased, потому что они показывают одни из лучших результатов на реальных данных [2]. Данные методы позволяют по оценкам пользователей строить вектора предпочтений для пользователей (какие фильмы понравятся пользователю) и фильмов (каким пользователям понравится фильм).

При решении проблемы холодного старта пользователя было предложено учитывать его жанровые предпочтения. На их основе из уже известных системе выбираются пользователи, наиболее похожие на нового. Для выбора используются алгоритмы KNN и K-Means; похожесть пользователей определяется как косинусная мера между соответствующими им векторами жанровых предпочтений.

При решении проблемы холодного старта фильма было принято решение использовать сюжеты фильмов, и с помощью языковой модели (RuBERT, MultilingualBERT) и нейросетевой модели (DNN, CNN, RNN) предсказывать его вектор предпочтения.

Для оценки качества работы выбранных методов был составлен набор данных, содержащий информацию о 4637 фильмах с русским и английским сюжетом, 170 тыс пользователей и их 17 млн предпочтений. Наилучшие результаты показали:

- для предсказания предпочтения — метод **SVD**;
- для решения проблемы холодного старта пользователя — метод **KNN**;
- для решения проблемы холодного старта фильма — методы **RuBERT + DNN** и **MultilingualBERT + DNN**.

На их основе была построена рекомендательная система на языке программирования Python 3.6, которая автоматически предсказывает предпочтение пользователей для известных системе фильмов. Для добавления новых пользователей достаточно получить от них их жанровые предпочтения, для добавле-

ния нового фильма достаточно загрузить в систему его сюжет. Исходный код программного средства и построенный набор данных выложены в открытый доступ: <https://github.com/MrKolbaskin/recommender-system>.

Литература

- [1] Adomavicius, G. Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions / G. Adomavicius, A. Tuzhilin // IEEETransactionson Knowledge and Data Engineering.—2005.—Vol. 17, no. 6.—с.734–749
- [2] Bennett, J. The netflix prize / James Bennett, Stan Lanning, Netflix Netflix // In KDD Cupand Workshop in conjunction with KDD.—2007.

Программная поддержка словообразовательного анализа

Изюмова Дарья Дмитриевна

Кафедра алгоритмических языков

e-mail: tannenwald@inbox.ru

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

В ряде задач автоматической обработки текстов (АОТ) на естественном языке может потребоваться поиск родственных слов для заданного слова: задачи перефразирования и поиска плагиата, исправления словообразовательных ошибок, обработка редких слов и новых слов (их смысл можно распознать путём сопоставления с родственными однокоренными словами). Для этих целей необходимо наличие словарных ресурсов, в которых представлены словообразовательные связи. В настоящее время подобные ресурсы для русского языка отсутствуют.

В данной работе рассматривалась задача автоматического построения словообразовательных деревьев и цепочек для слов русского языка. В качестве исходных данных были взяты данные из системы КроссЛексика [1], которые представляют собой наборы однокоренных слов, уже разобранных на морфы (корень, приставки, суффиксы, постфиксы): *дрож-ащ-ий, об-морож-ени-е*.

Для решения задачи был разработан метод, основанный на правилах и включающий несколько этапов: разбиение исходных данных на группы однокоренных слов, исключение из дальнейшей обработки слов с одинаковым лексическим значением, выделение в каждой группе слова-вершины и построение от него словообразовательного дерева. Построение отдельного дерева производилось в соответствии с особенностями словообразования русского языка [2].

Обработку слов затрудняло наличие корней-алломорфов, которые различаются произношением и написанием, но имеют одинаковое значение: *прилагать — переложить (лаг — лож), выращивать — растущий (ращ — раст), грех — грешник (грех — греш)*. Для их распознавания была создана коллекция наборов алломорфных корней.

На базе предложенного метода были реализованы несколько программных модулей на языке Python. По словарю данных КроссЛексики, содержащему

более 26 тыс. слов русского языка, было построено около 1780 словообразовательных деревьев. Полученные деревья можно использовать не только в задачах АОТ, но и при проведении статистических исследований для уточнения лингвистических моделей словообразования в русском языке.

Для удобства работы с созданным словообразовательным ресурсом реализованы прикладной и пользовательский интерфейсы. Пользователю предоставляются следующие возможности: поиск словообразовательного дерева по слову или корню, поиск поддерева по слову-вершине и поиск цепочки по слову-концу.

Литература

- [1] Большаков И. А. КроссЛексика — большой электронный словарь сочетаний и смысловых связей русских слов // Компьютерная лингвистика и интеллектуальные технологии: Труды Международной конференции «Диалог 2009». Вып. 8 (15) М.: РГГУ, 2009.
- [2] Лопатин В. В., Улуханов И. С. Словарь словообразовательных аффиксов современного русского языка. М. : Изд. Центр «Азбуковник», 2016.

Алгоритмы поиска маршрутов в графе с меняющейся структурой и их реализация в мобильном приложении

Рязанова Евгения Борисовна

Кафедра алгоритмических языков

e-mail: j.ryazanowa2010@yandex.ru

Научный руководитель — к.ф.-м.н. доц. Корухова Юлия Станиславовна

Описание задачи. Данная работа посвящена актуальной на текущий момент задаче реализации навигационной системы, выполняющей поиск кратчайшего маршрута внутри здания. В настоящее время в связи с развитием мобильных технологий подобные системы особенно востребованы в виде мобильных приложений.

В работе рассмотрены существующие подходы к решению данной задачи, а также изучены подходы к решению задачи ориентирования внутри здания, учитывающие ее особенности: структура здания может динамически меняться (например, при отключении лифта перемещение на лифте нельзя предлагать пользователю как часть маршрута), стены и перекрытия существенно ограничивают использование спутниковой навигации.

Архитектура системы. В системе реализован подход, учитывающий вышеупомянутые особенности задачи. Система реализована в виде клиент—серверного приложения, клиентский модуль которого запрограммирован для мобильных устройств на языке Java для операционной системы (далее — ОС) Android, серверный модуль построен на базе фреймворка FastAPI [1] на языке Python3. Представление структуры здания выполнено в виде взвешенного неориентированного графа, в котором вершины графа — различные помещения в здании, а его дуги — всевозможные коридоры, лифты, лестницы и т. д. Информация о графе хранится в виде базы данных на мобильном

устройстве с ОС Android [2], а также на сервере в формате JSON. Система поддерживает возможность динамического изменения структуры графа с помощью специального интерфейса посредством посылаемого запроса на сервер с целью редактирования хранящегося там файла формата JSON, с которым синхронизируется база данных на мобильном устройстве. Поиск кратчайшего пути осуществляется с помощью алгоритма A^* [3].

Описание работы системы. Прототипная реализация системы выполнена для помещений факультета ВМК, однако инструменты могут быть применены для любого здания, структура которого будет занесена в базу данных. В мобильном приложении реализованы два типа интерфейса: интерфейс пользователя и интерфейс администратора.

Интерфейс пользователя представляет из себя интерактивную карту здания факультета ВМК с размеченными помещениями. Текущий этаж можно сменить, нажав на нужный из списка. Выбрать начальную и конечную точку маршрута можно двумя способами: либо кликнув на них непосредственно на интерактивной карте, либо введя их названия в специально предназначенных для этого полях. Ввод осуществляется с помощью подсказок (автодополнений). Также имеется возможность выбрать способ перемещения между этажами: лифт, лестница или предоставить выбор системе. После ввода всей необходимой информации будет построен кратчайший маршрут через доступные промежуточные пути.

Интерфейс администратора предназначен для изменения структуры здания в приложении. Данный интерфейс представляет собой несколько полей: идентификатор помещения или прохода, этаж и статус, который необходимо присвоить объекту здания (открыт или закрыт). После заполнения полей на сервер будет отправлен запрос на изменение хранящегося там файла.

Результаты. Таким образом, в результате проведенной работы была реализована навигационная система на базе помещений факультета ВМК с клиентским модулем в виде мобильного приложения и возможностями обновления структуры здания.

Литература

- [1] Знакомство с FastAPI [Электронный ресурс]. — Электрон. дан. — URL: <https://itnan.ru/post.php?c=1&p=488468> (дата обращения: 10.03.2021)
- [2] Хранение данных в Android [Электронный ресурс]. — Электрон. дан. — URL: <https://code.tutsplus.com/ru/tutorials/android-from-scratch-how-to-store-application-data-locally--cms-26853> (дата обращения: 10.12.2020)
- [3] Introduction to the A^* Algorithm [Электронный ресурс]. — Электрон. дан. — URL: <https://www.redblobgames.com/pathfinding/a-star/introduction.html> (дата обращения: 04.02.2021)

Извлечение иерархической логической структуры из текстовых документов в формате docx

Богатенкова Анастасия Олеговна

Кафедра системного программирования

e-mail: bogatenkova.anastasiya@mail.ru

Научный руководитель — асс. Гомзин Андрей Геннадьевич

Для автоматической обработки документов (например, для суммаризации) необходимо представить документ в удобном для обработки виде, т. е. выделить его структуру. Для этих целей в ИСП РАН разрабатывается программный модуль docreader, основанный на проекте dedoc [1]. Одним из типов структуры документа является логическая структура [2], отражающая разбиение документа на компоненты, например, научные статьи состоят из секций, подсекций и т. д. В работе рассматривается логическая структура документа на примере технических заданий в формате docx.

Постановка задачи. Целью работы являлась формализация понятия иерархической логической структуры для технического задания, а также разработка и реализация метода извлечения логической структуры из технических заданий, представленных в формате docx. Кроме того, в рамках поставленной задачи требовалось реализовать извлечение текста и необходимых метаданных из документов в формате docx и встроить реализованный метод в открытый проект по обработке документов dedoc.

Извлекаемая структура. По итогам изучения существующих работ [3] и анализа примеров технических заданий[§], предложена структура ТЗ следующего вида: дерево произвольной глубины, в котором каждому узлу соответствует параграф документа с типом (заголовок, содержание, часть, элемент списка или текст). Параграфы — это составляющие документа в формате docx, разделённые переносом строки.

Общая схема метода. Предложен метод извлечения описанной структуры, состоящий из двух этапов:

1. Классификация параграфов документа на 5 классов, описанных выше;
2. Построение дерева документа, глубина вложенности каждого параграфа определяется на основе попарного сравнения параграфов.

Классификация параграфов. Для решения задачи классификации параграфов каждому параграфу сопоставляется вектор признаков. Признаки параграфа формируются на основе текста параграфа и некоторых метаданных, таких как жирность и размер шрифта, выравнивание и т. п. В качестве классификатора использовался алгоритм машинного обучения XGBoost.

Построение дерева. Построение дерева происходит следующим образом: рассматривается очередной параграф и сравнивается с предыдущим. Если текущий параграф меньше предыдущего (то есть структурно вложен в него)

[§]http://technicaldocs.ru/гост19/шаблоны/техническое_задание

или равен ему, в дерево добавляется новый узел, иначе происходит переход вверх по иерархии и сравнение с параграфом более высокого уровня. Сравнение параграфов рассматривалось как задача классификации пар параграфов на 3 класса: больше, меньше или равно. Для её решения формировались векторы признаков для пар параграфов и также использовался классификатор XGBoost.

Создание наборов данных. Так как готовых наборов технических заданий не было найдено, создано два размеченных набора из 22 технических заданий для классификации (1405 параграфов) и сравнения параграфов (4438 пар параграфов). Размеченные наборы данных использовались для обучения классификаторов и оценки качества метода.

Извлечение текста и метаданных из docx документов. На языке Python3 написан обработчик docx документов, позволяющий извлекать из документов текст и некоторые метаданные, такие как размер шрифта, отступ от края страницы, выравнивание, жирность, курсив и подчеркивание. В отличие от существующих библиотек языка, реализованный обработчик поддерживает корректное извлечение текста нумерации списков. На основе полученных метаданных формировались векторы признаков для параграфов.

Результаты. Разработанный метод определения типа параграфов показал на кросс-валидации достоверность (ассигасу) 0.96 по сравнению с достоверностью 0.91, которую показал адаптированный к нашей задаче метод [4]. Сравнение параграфов осуществлялось с достоверностью 0.98. Для проверки метода построения деревьев была размечена тестовая выборка, на которой усреднённое расстояние Робинсона-Фулдса [5] между построенными и размеченными деревьями оказалось равным 0.073.

Таким образом, был разработан и реализован метод по извлечению из документов структуры в виде дерева произвольной глубины с типизированными узлами. Для экспериментальной оценки метода составлен и размечен набор технических заданий. Реализовано извлечение текста и метаданных из docx документов, код обработчика встроен в открытый проект dedoc [1].

Литература

- [1] Dedoc. URL: <https://github.com/ispras/dedoc> (дата обращения: 23.04.2021).
- [2] Dori D. et al. The representation of document structure: A generic object-process analysis // Handbook of character recognition and document image analysis. 1997. С. 421–456.
- [3] Заболеева-Зотова А. В., Орлова Ю. А. Моделирование лексического анализа текста технического задания // Известия ВолгГТУ. 2007. С. 40.
- [4] Kim T. et al. A Machine-Learning Based Approach for Extracting Logical Structure of a Styled Document // TIIS. 2017. Т. 11. № 2. С. 1043–1056.
- [5] Robinson D. F., Foulds L. R. Comparison of phylogenetic trees // Mathematical biosciences. 1981. Т. 53, № 1-2. С. 131–147.

Динамическая компиляция пользовательских функций на языке PL/pgSQL

Джиджоев Владислав Муратович

Кафедра системного программирования

e-mail: dzhivl@ispras.ru

Научный руководитель — к.ф.-м.н. проф. Гайсарян Сергей Суренович

Научный консультант — м.н.с. Бучацкий Рубен Артурович

Реляционные системы управления базами данных (СУБД) стали основным инструментом хранения и анализа данных в различных сферах применения вычислительных систем. Неотъемлемой частью современных реляционных СУБД является SQL — язык описания запросов, обеспечивающий возможность манипуляции данными, хранящимися в СУБД, удобным для пользователей образом. В последнее время широкое распространение получили процедурные расширения языка SQL, позволяющие совместно использовать декларативные операторы SQL и концепции императивных языков программирования: переменные, ветвления, циклы.

PL/pgSQL — процедурное расширение языка SQL в СУБД PostgreSQL. Пользовательские функции на языке PL/pgSQL исполняются с использованием встроенного интерпретатора. Интерпретация сопряжена со значительными накладными расходами, связанными с анализом, планированием и выполнением кода пользовательской функции, снижающими производительность СУБД.

В выпускной квалификационной работе исследован и реализован метод динамической компиляции пользовательских функций на языке PL/pgSQL, призванный сократить накладные расходы на выполнения пользовательских функций в СУБД PostgreSQL.

В работе произведен подробный обзор существующих методов ускорения выполнения пользовательских функций (например, [1, 2]), подчеркиваются их преимущества и недостатки. Предлагается выполнять динамическую компиляцию пользовательской функции совместно с динамической компиляцией вызывающего ее SQL-запроса. Это позволяет генерировать более специализированный код и сократить накладные расходы на переключение контекстов между исполнителями пользовательской функции и SQL-запроса. Разработанный динамический компилятор пользовательских функций является компонентом динамического компилятора SQL-запросов, разрабатываемого в ИСП РАН [3].

Динамическая компиляция пользовательских функций реализована с использованием компиляторной инфраструктуры LLVM. LLVM предоставляет реализацию множества оптимизаций программного кода и динамической компиляции кода из промежуточного представления LLVM IR в машинный код. Использование LLVM позволяет свести задачу динамической компиляции пользовательской функции на языке PL/pgSQL к задаче преобразования абстрактного синтаксического дерева пользовательской функции в соответствующий ему по семантике код в промежуточном представлении LLVM IR.

Каждая вершина абстрактного синтаксического дерева пользовательской функции на языке PL/pgSQL соответствует некоторому оператору пользовательской функции. Компиляция пользовательской функции осуществляется путем полного обхода её абстрактного синтаксического дерева и вызова функции—генератора для каждой вершины, порождающей код соответствующего оператора в промежуточном представлении LLVM IR. Получившийся код в промежуточном представлении транслируется во время выполнения СУБД в эффективный машинный код, оптимизированный с учетом структуры конкретной функции в контексте вызывающего запроса.

В результате тестирования разработанного динамического компилятора было продемонстрировано ускорение выполнения синтетических вычислительно—интенсивных тестов в несколько раз. Реализована поддержка динамической компиляции большого числа операторов языка PL/pgSQL.

Исследование проводилось в Институте системного программирования им. В. П. Иванникова РАН. Результаты работы были опубликованы в [4].

Литература

- [1] Froid: Optimization of imperative programs in a relational database / K Ramachandra, P. Kwanghyun, E. K. Venkatesh, A. Halverson, C. Galindo—Legaria, C. Cunningham // Proceedings of the VLDB Endowment. 2017. Vol. 11, N 4, P. 432–444.
- [2] Duta C., Hirn D., Grust T. Compiling PL/SQL Away // 10th Conference on Innovative Data Systems Research (Amsterdam, The Netherlands, January 12-15, 2020). Online Proceedings.
- [3] Динамическая компиляция SQL-запросов для СУБД PostgreSQL / Р. А. Бучацкий, Е. Ю. Шарыгин, Л. В. Скворцов, Р. А. Жуйков, Д. М. Мельник, Р. В. Баев // Труды института системного программирования РАН. 2016. Т. 28, № 6. С. 37–48.
- [4] Динамическая компиляция пользовательских функций на языке PL/pgSQL / В. М. Джиджоев, Р. А. Бучацкий, М. В. Пантилимонов, А. Н. Томилин // Труды института системного программирования РАН. 2020. Т. 32, № 5. С. 67–80.

Определение устранимых зависимостей по данным для распараллеливания программ в системе автоматизированного распараллеливания

Макеев Виталий Олегович

Кафедра системного программирования

e-mail: v.mak33v@yandex.ru

Научный руководитель — к.ф.-м.н. доц. Бахтин Владимир Александрович

Современные вычислительные системы открывают возможность для решения большого количества прикладных и теоретических задач. Однако зачастую процесс построения этих решений неразрывно связан с разработкой эффективных

параллельных программ, предназначенных для выполнения на суперкомпьютерах. При распараллеливании программ основную проблему представляют инструкции, которые во время выполнения обращаются к одним и тем же участкам физической памяти. Выполнение таких инструкций на разных итерациях цикла порождает зависимость по данным между его итерациями и препятствует его параллельному выполнению.

SAPFOR [1] является инструментом для автоматизированного распараллеливания программ на языке Си и использует LLVM (вместе с Clang) для построения промежуточного представления программ и статического анализа кода. Некоторые виды зависимостей по данным (редукционные и приватизируемые переменные, переменные индукции) не препятствуют параллельному выполнению циклов программы. При этом для их обнаружения средствами анализа SAPFOR необходимо, чтобы обращения к соответствующим переменным были представлены в форме однократного присваивания (SSA). При построении SSA формы в LLVM используется проход SROA (scalar replacement of aggregates) [2], который пытается разделить агрегатную память такую как массивы константного размера и структуры на отдельные скаляры и переместить их на регистры. Кроме того, простые скалярные переменные также помещаются на регистры и для них удаляются инструкции чтения и записи в память. При использовании SROA теряются адреса преобразуемых переменных в памяти, поэтому данный проход никак не изменяет глобальные переменные, переменные-указатели и скаляры от которых берется адрес, что может затруднять поиск зависимостей по данным. В частности, указанная проблема возникает при использовании MPI, так как многие функции этой библиотеки оперируют указателями.

В данной работе были расширены возможности статического анализатора системы SAPFOR для улучшения анализа зависимостей по данным. Был добавлен проход преобразования, который для каждого гнезда циклов в функции пытается решить описанную выше проблему и поместить переменные, адреса которых используются в программе, на регистры. Для каждого цикла выполняются следующие шаги:

1. Поиск и сохранение переменных, которые до этого были помечены как имеющие зависимости по данным. На этом шаге рассматриваются только скалярные переменные, массивы и структуры не поддерживаются;
2. Для каждой переменной определение возможности проведения преобразования. Для этого необходимо гарантированное отсутствие записи через участок памяти, где находится адрес указателя, и отсутствие у цикла и переменной свойств LLVM, запрещающих преобразование;
3. Преобразование, похожее на проход SROA: загрузка значения переменной из памяти на регистры перед циклом, удаление LLVM инструкций load и store, добавление ϕ -функций в некоторые базовые блоки программы, загрузка обновленного значения переменной обратно в память в выходном блоке цикла. При этом использования инструкций load заменяются на использования регистров;

4. Добавление отладочной информации LLVM для корректного анализа с использованием дерева псевдонимов [3]. Все добавленные инструкции связываются с новой отладочной переменной, и сохраняется её связь с исходной переменной. На более позднем этапе анализа при обнаружении каких-либо свойств или зависимостей по данным этот результат распространяется на изначальную переменную.

Разработанный проход был протестирован на локальных тестах статического анализатора TSAR, который является частью SAPFOR. Кроме того, его использование помогло улучшить определение редукционных переменных в MPI версии тестов из набора NAS Parallel Benchmarks [4], в том числе EP и CG, что позволило выполнить дополнительное распараллеливание программ с помощью SAPFOR и задействовать графические процессоры в узле кластера.

Литература

- [1] Kataev N. LLVM Based Parallelization of C Programs for GPU //Russian Supercomputing Days. – Springer, Cham, 2020. – С. 436–448.
- [2] Jambor M. The new intraprocedural Scalar Replacement of Aggregates //GCC Developers' Summit. – 2010. – С. 47.
- [3] Kataev N. Application of the LLVM Compiler Infrastructure to the Program Analysis in SAPFOR //Russian Supercomputing Days. – Springer, Cham, 2018. – С. 487–499.
- [4] NAS Parallel Benchmarks
<https://www.nas.nasa.gov/publications/npb.html>

Оптимизация вычислений нейронных сетей на графических процессорах для мобильных устройств

Оплачко Николай Алексеевич

Кафедра системного программирования

e-mail: oplachko@ispras.ru

Научный руководитель — д.ф.-м.н. проф. ак. РАН Аветисян Арутюн Ишханович

В последние годы нейронные сети широко используются для решения задач в различных областях. Иногда возникает потребность запускать нейронные сети на мобильных устройствах в силу разных причин, в частности: защита данных пользователя, сокращение затрат на серверы, уменьшение временной задержки, обеспечение возможности использования приложения без доступа к сети Интернет.

Вследствие ограниченности ресурсов мобильных устройств критична оптимизация вычислений предсказаний нейронных сетей для сокращения расходов времени и используемой памяти.

Выпускная квалификационная работа включает в себя следующие задачи:

- исследование существующих подходов для оптимизации затраченного времени и используемой памяти при вычислении предсказаний нейронных сетей;

- разработка метода одновременной оптимизации времени и памяти;
- реализация и встраивание предложенного метода в библиотеку искусственного интеллекта с использованием прикладного программного интерфейса Vulkan.

В работе рассматриваются задача оптимизации памяти, решаемая эвристическими алгоритмами [1, 2], и задача параллельных вычислений на уровне ациклического графа вычислений, решаемая нахождением некоторого разбиения графа на группы независимых друг от друга вершин [3], далее называемых слоями.

Отмечается, что эти задачи конфликтуют между собой: параллелизация вычислений накладывает дополнительные ограничения на повторное использование памяти.

Предлагается метод, позволяющий находить компромисс в конфликте вышеупомянутых задач. Суть его состоит в разбиении найденных слоёв на несколько групп: с ростом числа групп ослабляются добавляемые ограничения на повторное использование памяти, при этом понижается степень параллелизма.

Возможность конфигурировать предложенный метод для используемой нейронной сети (варьировать количество групп при делении слоёв) позволяет разработчику выбрать компромисс между затрачиваемыми на вычисления временем и памятью.

В ходе тестирования предложенного метода для вычисления предсказания свёрточной нейронной сети ResNeXt [4] наилучший результат был получен при разбиении слоёв на две группы: ускорение составило 6.85% при сокращении потребления памяти на 72.4%. Оказалось, что при разбиении слоёв на одну группу вместо двух, кроме того, что на 2.8% увеличивается используемая память, значительно падает ускорение: 3.08% вместо 6.85%. Это означает, что особенности обращений в память графического процессора могут давать существенный вклад в итоговое время вычислений наравне с параллелизацией.

Тестирование предложенного метода производилось на одноплатном компьютере HiKey 970. Была подсчитана электрическая энергия во время вычисления предсказания нейронной сети ResNeXt при помощи цифрового измерителя мощности. Из результатов для этой нейронной сети следует, что с ростом степени параллелизма растут суммарные затраты электрической энергии, поэтому с целью баланса экономии затрат времени, памяти и электрической энергии может оказаться оптимальным разбиение слоёв более, чем на две группы. Также было произведено профилирование графического процессора мобильного устройства набором инструментов Arm Development Studio. Подсчёт числа тактов ожидания обращений в память подтвердил, что при разбиении слоёв на одну группу работа с памятью менее оптимальна, чем на две.

Основной результат данной работы состоит в разработке метода для одновременной оптимизации времени и памяти при вычислении предсказаний нейронных сетей, его реализации и встраивании в программно-аппаратную

часть библиотеки машинного обучения, а также получении положительных результатов при его тестировании.

Литература

- [1] Pisarchyk Y., Lee J. Efficient memory management for deep neural net inference // arXiv preprint arXiv:2001.03288. 2020.
- [2] Profile-guided memory optimization for deep neural networks / Sekiyama T., Imamichi T., Imai H., Raymond R. // arXiv preprint arXiv:1804.10001. 2018.
- [3] Node-level parallelization for deep neural networks with conditional independent graph / Zhou F., Wu F., Zhang Z., Dong M. // Neurocomputing, 267. 2017. P. 261–270.
- [4] Aggregated residual transformations for deep neural networks / Xie S., Girshick R., Dollár P., Tu Z., He K. // Proceedings of the IEEE conference on computer vision and pattern recognition. 2017. P. 1492–1500.

Использование обучения с подкреплением в задаче автоматического тестирования мобильных приложений

Фомин Сергей Александрович

Кафедра системного программирования

e-mail: totalchest@ya.ru

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

Мобильные приложения являются основным способом взаимодействия пользователя с мобильным устройством. Этот факт заставляет разработчиков уделять больше внимания качеству получаемого продукта. Любые сбои или зависания приложения во время его использования могут негативно влиять на рейтинг продукта. Различные виды тестирования помогают вовремя найти и исправить ошибки. В работе исследуются подходы автоматического тестирования мобильных приложений путем взаимодействия с графическим интерфейсом. Задача этого тестирования состоит в том, чтобы проверить отсутствие сбоев во время работы готового приложения.

Обучение с подкреплением. В работе рассматриваются подходы, которые генерируют взаимодействия с устройством на основе машинного обучения — обучения с подкреплением [1]. Техника обучения с подкреплением использует метод проб и ошибок для выработки правильной последовательности нажатий на экран, способной открывать новые состояния. Идея заключается в том, что каждое взаимодействие в каждом состоянии приложения оценивается некоторой функцией награды. Функция награды вознаграждает алгоритм за полезные действия и штрафует за бесполезные. В процессе исследования приложения алгоритм учится предсказывать какое действие приведет к большей суммарной награде и тем самым позволит исследовать как можно больше состояний.

Независимые от приложения стратегии. Для начала исследуются наиболее предпочтительные алгоритмы с независимыми от приложения подходами, то есть подходами, способными тестировать произвольное приложение после единоразового обучения модели. В качестве базового алгоритма была разработана модель, использующая некоторые идеи из [2]. Для совместимости со всеми приложениями в ней вводится ряд упрощений, а именно абстрактные состояния и абстрактные действия.

Для устранения недостатков базового алгоритма, связанных с использованием сильных упрощений, был разработан подход, который представляет состояние как изображение, а действие как бинарную маску в точке нажатия на этом изображении. Для преобразования состояния использовалась сверточная нейронная сеть, эмулирующая таблицу предсказания.

Зависимые от приложения стратегии. Для увеличения эффективности тестирования дальше разрабатывались зависимые от приложения стратегии, которые могут тестировать одно приложение, под которое они обучаются. При разработке базовой версии алгоритма учитывались работы других исследователей [3]. Здесь состояние описывается списком интерактивных элементов в текущий момент времени, действие — списком возможных взаимодействий в текущий момент времени. В базовом алгоритме была выбрана простая функция награды, описываемая формулой (1), то есть награда обратнопропорциональна количеству раз, которое действие a_t воспроизводилось в состоянии s_t .

$$R(s_t, a_t) = \frac{1}{\text{count}(s_t, a_t)}, \quad (1)$$

Также проводились различные модификации базового алгоритма: функция награды разрабатывалась таким образом, чтобы увеличивать эффективность тестирования. Был добавлен учет количества интерактивных элементов и учет расстояния между состояниями.

Чтобы улучшить последние стратегии использовались еще несколько идей из теории обучения с подкреплением и машинного обучения в целом: добавление случайности в выбор взаимодействия на первых этапах тестирования, разделение процесса тестирования приложения на эпизоды обучения и тестирования, объединение двух первых идей в одну стратегию.

Результаты. Для тестирования стратегий использовалось 10 современных приложения с платформы GooglePlay. Были выбраны приложения разных категорий и степеней сложности.

Как и ожидалось, зависимые от приложения стратегии превосходят независимые. Однако не все функции наград для зависимых приложений показали высокий результат. Улучшение зависимых стратегий позволило значимо повысить метрики качества. Анализ метрик качества позволил выделить лидирующий подход: стратегия с предобучением таблицы.

В результате сравнения с современным подходом к тестированию Humanoid [4], лучший из методов обучения с подкреплением оказался существенно эффективней в рамках поставленной задачи.

Литература

- [1] Szepesvári C. Algorithms for reinforcement learning //Synthesis lectures on artificial intelligence and machine learning. – 2010. – Т. 4. – №. 1. – С. 1-103.
- [2] Koroglu Y. et al. QBE: QLearning-based exploration of android applications //2018 IEEE 11th International Conference on Software Testing, Verification and Validation (ICST). – IEEE, 2018. – С. 105-115.
- [3] Adamo D. et al. Reinforcement learning for android gui testing //Proceedings of the 9th ACM SIGSOFT International Workshop on Automating TEST Case Design, Selection, and Evaluation. – 2018. – С. 2-8.
- [4] Li Y. et al. Humanoid: A deep learning-based approach to automated black-box android app testing //2019 34th IEEE/ACM International Conference on Automated Software Engineering (ASE). – IEEE, 2019. – С. 1070-1073.

Именной указатель

- Авагян Д. А., 95
Акимова Е. Д., 44
Андрианов Д. С., 9
Арсенин Н. М., 97
Аурангабадкар Удитангшу, 63
Березникер А. В., 66
Богатенкова А. О., 103
Бочковенкова Д. В., 25
Бучацкая О. П., 11
Василенко А. В., 27
Воробьева И. Б., 46
Вострикова У. Ю., 13
Гаврилов О. А., 14
Гаврилова Д. В., 16
Горбунова О. К., 28
Гуленко А. С., 40
Джиджоев В. М., 105
Дубинин В. Д., 98
Зайцева Д. С., 42
Измайлов М. Б., 29
Изюмова Д. Д., 100
Камалов И. М., 85
Карпикова П. В., 47
Корнеева А. М., 18
Кот В. Я., 31
Кукушкин Д. И., 76
Куркин М. Л., 34
Личманов Д. И., 88
Логванёва М. Г., 54
Лысенко И. И., 33
Макеев В. О., 106
Максимова А. В., 56
Малов Д. В., 19
Малышев К. В., 67
Мамедов Т. З., 69
Местецкий М. А., 64
Мишустина М. В., 59
Морквин А. А., 78
Москаленко А. В., 71
Мухтаров М. Б., 57
Никифоров Н. И., 79
Николаев А. В., 73
Оплачко Н. А., 108
Питанов Е. С., 90
Полевой А. В., 75
Пятов В. А., 21
Рябченков В. М., 82
Рязанова Е. Б., 101
Салихова К. И., 48
Селезнев Л. Е., 81
Селиванов А. С., 39
Сорокина Н. В., 91
Сотников Д. М., 50
Травникова А. С., 61
Удовиченко И. Р., 51
Федяшин Н. А., 52
Фомин С. А., 110
Цуканова М. А., 93
Чижов П. С., 23
Шалагин Н. Д., 94
Шапошников В. А., 84