Министерство науки и высшего образования РФ

Национальный исследовательский
Нижегородский государственный университет им. Н.И. Лобачевского



МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ И СУПЕРКОМПЬЮТЕРНЫЕ ТЕХНОЛОГИИ

ТРУДЫ XXIV МЕЖДУНАРОДНОЙ КОНФЕРЕНЦИИ

Нижний Новгород, 18–21 ноября 2024 г.

Нижний Новгород
Издательство Нижегородского госуниверситета
2024

УДК 004.942+519.876.5+519.6
ББК 22.18я43+22.19я43
М34

М34 **Математическое моделирование и суперкомпьютерные технологии.** Труды XXIV Международной конференции (Н. Новгород, 18–21 ноября 2024 г.) / Под ред. проф. Д.В. Баландина. – Нижний Новгород: Изд-во Нижегородского госуниверситета, 2024. – 186 с.

Отв. за выпуск К.А. Баркалов

ISBN 978-5-91326-925-6

Сборник материалов Двадцать четвертой Международной конференции «Математическое моделирование и суперкомпьютерные технологии», состоявшейся 18−21 ноября 2024 г. на базе Нижегородского государственного университета им. Н.И. Лобачевского, содержит тезисы докладов и короткие статьи, посвященные математическому моделированию сложных процессов и явлений, численным методам их исследования, а также проблемам разработки методов суперкомпьютерных вычислений для решения актуальных задач в различных областях науки, промышленности и образования.

Подробную информацию о конференции можно найти в сети Интернет по адресу http://agora.guru.ru/hpc2024.

Поддержка конференции

Программа академического лидерства Приоритет 2030

Научно-образовательный математический центр
«Математика технологий будущего»

ISBN 978-5-91326-925-6

УДК 004.942+519.876.5+519.6
ББК 22.18я43+22.19я43

© Авторы статей, 2024
© ННГУ им. Н.И. Лобачевского, 2024

ПРЕДИСЛОВИЕ

18–21 ноября 2024 г. в рамках международного конгресса «Суперкомпьютерные дни в России» на базе Нижегородского государственного университета им. Н.И. Лобачевского при поддержке программы академического лидерства «Приоритет 2030» и научно-образовательного математического центра «Математика технологий будущего» проведена XXIV Международная конференция и молодежная школа «Математическое моделирование и суперкомпьютерные технологии».

Тематика конференции охватывала основные направления в математического моделирования и суперкомпьютерных технологий:

- математическое моделирование динамики систем и процессов управления;
- теория динамических систем и бифуркаций;
- колебательные процессы в динамических сетях;
- модели и методы поддержки принятия решений;
- модели и методы искусственного интеллекта;
- математическое модели в биологии и медицине;
- математическое моделирование природных процессов;
- алгебра, геометрия и дискретная математика;
- технологии параллельных и распределенных вычислений;
- применение суперкомпьютерных технологий для решения вычислительно-сложных задач.

В рамках молодежной школы была организована двухдневная интенсивная онлайн-программа обучения современным технологиям высокопроизводительных вычислений, гетерогенного и параллельного программирования.

Основными задачами проведения мероприятия являлись представление актуальных результатов в области математического моделирования сложных процессов и явлений, обсуждение различных аспектов организации суперкомпьютерных вычислений, расширение контактов между специалистами для решения ресурсоемких прикладных задач, обмен опытом научно-образовательной деятельности при подготовке специалистов в области математического моделирования и параллельных вычислений.

Все перечисленные темы нашли отражение в настоящем сборнике тезисов докладов и коротких статей, представленных на конференцию (публикуются частично в авторской редакции).

ЗАДАЧА О БАЗИСАХ В ФУНКЦИОНАЛЬНОЙ СИСТЕМЕ РАЦИОНАЛЬНЫХ ФУНКЦИЙ С РАЦИОНАЛЬНЫМИ КОЭФФИЦИЕНТАМИ (АЛГОРИТМИЧЕСКИЙ ПОДХОД)

***Н.Ф. Алексиадис*[1,2]**

[1]*Московский государвстенный университет им. М.В. Ломоносова,*
[2]*Национальный исследовательский университет МЭИ*

Функциональная система (ф.с.) представляет собой множество функций с некоторым набором операций, применяемых к этим функциям и приводящих к получению других функций из этого же множества. Проблематика ф.с. обширна. Одной из основных проблем является проблема полноты, состоящая в описании таких подсистем функций, которые являются полными, т.е. из этих функций с помощью заданных операций над ними можно получить все функции. С проблемой полноты связана задача о базисах, т.е. изучение вопроса о минимальных полных системах. Целью настоящего доклада является исследование задачи о нахождении базиса конечной полной системы для ф.с. рациональных функций с рациональными коэффициентами с алгоритмической точки зрения: существует ли алгоритм, который в ф.с. рациональных функций с рациональными коэффициентами находит базис любой конечной полной системы?

Ключевые слова: функциональная система, полная система, проблема полноты, базис, алгоритмическая неразрешимость, рациональная функция.

1. Введение

Несмотря на то, что мы используем стандартные обозначения и общеизвестные понятия дискретной математики (в частности, теории функциональных систем), с целью корректного понимания изложенного, все-таки следует уточнить некоторые моменты.

Приведем предварительные сведения из теории функциональных систем, необходимые для дальнейшего изложения. При изложении материала в основном используется терминология книг [9,10].

Мы будем использовать следующие стандартные обозначения:

– N – множество всех натуральных чисел (включая число 0),
– Q – множество всех рациональных чисел.

Для удобства полагаем, что $0^0 = 1$.

Говорят, что функция $f(x_1, \cdots, x_{i-1}, x_i, x_{i+1}, \cdots, x_n)$ *существенно зависит от переменной* x_i, если существуют такие два набора (из области определения этой функции)

$$c_1, \cdots, c_{i-1}, a, c_{i+1}, \cdots, c_n \text{ и } c_1, \cdots, c_{i-1}, b, c_{i+1}, \cdots, c_n$$

значений переменных $x_1, \cdots, x_{i-1}, x_i, x_{i+1}, \cdots, x_n$, что

$$f(c_1, \cdots, c_{i-1}, a, c_{i+1}, \cdots, c_n) \neq f(c_1, \cdots, c_{i-1}, b, c_{i+1}, \cdots, c_n).$$

В этом случае мы говорим, что x_i является *существенной переменной функции* $f(x_1, \cdots, x_{i-1}, x_i, x_{i+1}, \cdots, x_n)$.

Если x_i не является существенной переменной $f(x_1, \cdots, x_{i-1}, x_i, x_{i+1}, \cdots, x_n)$ то она называется *фиктивной* (*несущественной*) *переменной* функции $f(x_1, \cdots, x_{i-1}, x_i, x_{i+1}, \cdots, x_n)$.

Пусть $f(x_1, \cdots, x_{i-1}, x_i, x_{i+1}, \cdots, x_n)$ и $g(x_1, \cdots, x_{i-1}, x_{i+1}, \cdots, x_n)$ две произвольные заданные функции и пусть x_i фиктивная переменная функции $f(x_1, \cdots, x_{i-1}, x_i, x_{i+1}, \cdots, x_n)$.

Если для любых $c_1, \cdots, c_{i-1}, c_i, c_{i+1}, \cdots, c_n$ значений переменных $x_1, \cdots, x_{i-1}, x_i, x_{i+1}, \cdots, x_n$ имеем

$$f(c_1, \cdots, c_{i-1}, c_i, c_{i+1}, \cdots, c_n) = g(c_1, \cdots, c_{i-1}, c_{i+1}, \cdots, c_n),$$

то говорят, что функция $g(x_1, \cdots, x_{i-1}, x_{i+1}, \cdots, x_n)$ получается из функции $f(x_1, \cdots, x_{i-1}, x_i, x_{i+1}, \cdots, x_n)$ *удалением* (*изъятием*) *фиктивной переменной* x_i и, наоборот, функция $f(x_1, \cdots, x_{i-1}, x_i, x_{i+1}, \cdots, x_n)$ получена из функции $g(x_1, \cdots, x_{i-1}, x_{i+1}, \cdots, x_n)$ *добавлением фиктивной переменной* x_i.

Если одна функция может быть получена из другой функции путем добавления или изъятия некоторых фиктивных переменных, то эти две функции являются *равными*.

Замечание 1. В дальнейшем будем считать, что вместе с функцией *f* заданы и все равные ей функции, т.е. функции рассматриваем с точностью до фиктивных переменных.

Замечание 2. Если дана конечная система функций $f_1, \cdots, f_m$ $(m \geq 1)$, то можно считать, что все они зависят от одних и тех же переменных $x_1, \cdots, x_n$, т.е. имеют вид

$$f_1(x_1, \cdots, x_n), \cdots, f_m(x_1, \cdots, x_n).$$

Замечание 3. Если дана функция, отличная от константы, то путем отождествления переменных из нее можно получить равную ей функцию, все переменные которой являются существенными.

Функциональная система (*ф.с.*) представляет собой множество функций с некоторым набором операций, применяемых к этим функциям и приводящих к получению других функций из этого же множества, т.е. *функциональная система* – это пара вида **F** = (*F*, *O*), где *F* – множество функций, а *O* – множество операций над этими функциями, при этом каждая операция из *O* замкнута относительно множества *F*.

Для произвольного подмножества $A \subseteq F$ обозначим через $[A]$ множество всех функций из *F*, которые получаются из функций множества *A* с помощью конечного числа применения операций из *O*. Множество $[A]$ называется *замыканием множества A*.

Множество *A* (A ⊆ F) называется *замкнутым* в функциональной системе **F**, если $[A] = A$.

Замкнутое множество принято называть *замкнутым классом*.

Множество *A* (A ⊆ F) называется *полным* в функциональной системе **F**, если $[A] = F$.

Полное множество принято называть *полной системой*.

Система функций *A* ($A \subseteq F$) называется *базисом* в ф.с. **F**, если *A* полная система и никакая ее собственная подсистема не является полной в **F**, т.е. можно сказать, что базис – это минимальная полная подсистема.

Подмножество *A* множества *F* называется *предполным классом* (*максимальной подалгеброй*) в **F**, если $[A] \neq F$, но для любой функции *f* из F\A выполнено $[A \cup \{f\}] = F$.

Система *K* замкнутых подмножеств множества *F* называется *критериальной системой* в ф.с. **F** = (*F*, *O*), если любое множество *M* ($M \subseteq F$) является полным в **F** тогда и только тогда, когда оно целиком не содержится ни в одном классе системы *K*.

В ф.с. **F** = (*F*, *O*) критериальной системой является, например, множество всех замкнутых классов, отличных от всего *F*. В общем случае последняя критериальная система является "избыточной". Это позволяет перейти к рассмотрению более "экономных" критериальных систем и с этой точки зрения может быть уточнено строение критериальной системы.

Критериальная система называется *приведенной*, если она не содержит собственных подсистем, являющихся критериальными.

Замечание 4. Следует отметить, что в функциональной системе **F** приведенная система, если она существует, определяется однозначно и состоит из всех предполных классов в **F** (см., например, [9]).

2. Определение функциональной системы рациональных функций с рациональными коэффициентами

Выражение вида $cx_1^{k_1} \cdots x_n^{k_n}$, где $n, k_1, \cdots, k_n \in N$, а $c \in Q$, называется *мономом с рациональным коэффициентом*, зависящим от *n* переменных $x_1, \cdots, x_n$; при этом, когда *n* = 0, тогда заданный моном является просто константой *c*, т.е. мономом с рациональным коэффициентом, зависящим от 0-го числа переменных.

Конечная сумма мономов с рациональными коэффициентами называется *полиномом с рациональными коэффициентами*.

Функция вида

$$f(x_1, \cdots, x_n): \underbrace{Q \times \cdots \times Q}_{n} \to Q,$$

которая задается с помощью полинома с рациональными коэффициентами, называется *полиномиальной функцией с рациональными коэффициентами*.

Следует обратить внимание на то, что аргументы полиномиальных функций с рациональными коэффициентами и сами эти функции принимают только рациональные значения.

Функция вида

$$f(x_1, \cdots, x_n) = \frac{g(x_1, \cdots, x_n)}{h(x_1, \cdots, x_n)},$$

где $g(x_1, \cdots, x_n)$ и $h(x_1, \cdots, x_n)$ – полиномиальные функции с рациональными коэффициентами, называется *рациональной функцией с рациональными коэффициентами*.

Рациональные функции с рациональными коэффициентами будем также называть *rq-функциями*.

Следует обратить внимание на то, что аргументы рациональных функций с рациональными коэффициентами и сами эти функции принимают только рациональные значения.

Обозначим через F_{RQ} множество всех рациональных функций с рациональными коэффициентами.

Функциональная система рациональных функций с рациональными коэффициентами $\mathbf{F}_{RQ}$ – это пара $\mathbf{F}_{RQ} = (F_{RQ}, O)$, где F_{RQ} – множество всех рациональных функций с рациональными коэффициентами, а O – множество операций суперпозиции над рациональными функциями с рациональными коэффициентами, которое включает в себя:

1. перестановку переменных,
2. переименование переменных без отождествления,
3. отождествление переменных,
4. введение фиктивной переменной,
5. удаление фиктивной переменной,
6. подстановку одной функции в другую.

Заметим, что определение функциональной системы $\mathbf{F}_{RQ} = (F_{RQ}, O)$ корректное, так как любая суперпозиция функций из F_{RQ} является опять функцией из F_{RQ}.

3. Постановка задачи

Одной из основных проблем в теории функциональных систем является *проблема полноты, состоящая в описании всех подмножеств данного множества F, которые являются полными в* **F** = (*F*, *O*).

Выделяют два подхода к решению проблемы полноты: *алгебраический* и *алгоритмический*. В первом случае ставится вопрос о получении критериев (т.е. необходимых и достаточных условий) полноты. Во втором случае задача заключается в существовании алгоритма, устанавливающего полноту или неполноту заданных систем функций.

С проблемой полноты связана и *задача о базисе, т.е. изучение вопроса о минимальных полных системах*.

При изучении базиса также выделяют два подхода: *алгебраический* и *алгоритмический*. В первом случае ставится вопрос о получении критериев (т.е. необходимых и достаточных условий) для того, чтобы данная подсистема была базисом. Во втором случае задача заключается в существовании алгоритма, который выделяет базис из полной системы.

В настоящей работе будет рассмотрен алгоритмический вариант задачи о базисе для функциональной системы рациональных функций с рациональными коэффициентами $\mathbf{F}_{RQ} = (F_{RQ}, O)$, который звучит так:

Существует ли алгоритм, который из любой конечной полной системы выделит базис; другими словами: задача о базисе алгоритмически разрешима или нет в ф.с. $\mathbf{F}_{RQ}$?

Алгебраический подход задачи о базисах было рассмотрено в докладе [2] и далее расширенная версия этого доклада была опубликована в [3].

4. Решение задачи о базисах (алгоритмический подход) для рациональных функций с рациональными коэффициентами

При изучении проблем полноты и базисов с алгебраической точки зрения для функциональной системы рациональных функций с рациональными коэффициентами $\mathbf{F}_{RQ} = (F_{RQ}, O)$ автором ранее было установлено справедливость следующих теорем [1, 3, 4].

Теорема 1. *В функциональной системе рациональных функций с рациональными коэффициентами* $\mathbf{F}_{RQ} = (F_{RQ}, O)$ *множество всех предполных классов является (приведенной) критериальной системой, т.е. произвольная система rq-функций является полной тогда и только тогда, когда она целиком не содержится ни в одном предполном классе*.

Теорема 2. *В функциональной системе рациональных функций с рациональными коэффициентами* $\mathbf{F}_{RQ} = (F_{RQ}, O)$ *мощность критериальной системы равна континууму, т.е. существует континуальное число предполных классов.*

Теорема 3. *В функциональной системе рациональных функций с рациональными коэффициентами* $\mathbf{F}_{RQ} = (F_{RQ}, O)$ *каждая полная система имеет базис, причем конечный.*

Теорема 4. *В функциональной системе рациональных функций с рациональными коэффициентами* $\mathbf{F}_{RQ} = (F_{RQ}, O)$ *для любого натурального* $n \geq 1$ *существует базис, состоящий ровно из* n *функций.*

Теорема 5. *В функциональной системе рациональных функций с рациональными коэффициентами* $\mathbf{F}_{RQ} = (F_{RQ}, O)$ *проблема полноты алгоритмически неразрешима.*

При дальнейшем исследовании задачи о нахождении базиса любой конечной полной системы с алгоритмической точки зрения для функциональной системы рациональных функций с рациональными коэффициентами было получено следующее утверждение:

Теорема 6. *В функциональной системе рациональных функций с рациональными коэффициентами* $\mathbf{F}_{RQ} = (F_{RQ}, O)$ *задача о нахождении базиса конечных полных систем алгоритмически неразрешима.*

5. Заключение

Функциональные системы являются одним из основных объектов дискретной математики и математической кибернетики и отражают следующие основные особенности реальных и абстрактных управляющих систем: функционирование (в функциональных системах - это функции), правила построения более сложных управляющих систем из заданных и описание функционирования сложных систем по функционированию их компонент (последние два момента отражены в операциях функциональных систем).

В настоящей работе рассматривается функциональная система рациональных функций с рациональными коэффициентами и для этой функциональной системы решается задача о базисе (алгоритмический подход), которая играет ключевую роль не только в самой дискретной математике и математической кибернетике, но и во многих других областях математики, например, в теории функций (аппроксимационные теоремы Чебышева и Вейерштрасса), в вычислительной математике и технике (построение и анализ вычислительных чипов и нейронных сетей). Актуальность полученных результатов также состоит в развитии самой теории функциональных систем как в плане охвата новых модельных объектов типа рациональных функций, так и в вычленении позитивных результатов, и в отсечении негативных ситуаций типа алгоритмической неразрешимости задачи о базисе.

Целью настоящего доклада, как уже было сказано выше, было изучение задачи о базисе с алгоритмической точки зрения в ф.с. $\mathbf{F}_{RQ} = (F_{RQ}, O)$. Полученная на этом пути теорема (6) об алгоритмической неразрешимости задачи о базисе в ф.с. $\mathbf{F}_{RQ} = (F_{RQ}, O)$ является основным результатом данной работы.

В заключении отметим, что представленный в настоящем докладе результат (Теорема 6) является логическим продолжением ранее полученных автором результатов (Теорем 1-4), т.е. продолжением цикла работ [1-8].

Литература

1. Алексиадис Н.Ф. О проблеме полноты рациональных функций с рациональными коэф-фициентами // Тезисы Международной конференции "Мальцевские чтения". Новоси-бирск, 2021. С. 143.
2. Алексиадис Н.Ф. О базисах рациональных функций с рациональными коэффициентами // Материалы XIX Международной конференции “Алгебра, теория чисел, дискретная геометрия и многомасштабное моделирование: современные проблемы, приложения и проблемы истории”, посвященной 130-летию со дня рождения академика И.М. Вино-градова - Тула, 2021. С. 77-80.
3. Aleksiadis N.F. Bases of complete systems of rational functions with rational coefficients. Chebyshevskii Sbornik. 2023; 24(2):5-14. (In Russ.) DOI 10.22405/2226-8383-2023-24-2-5-14.
4. Алексиадис Н.Ф. О проблеме полноты для функциональной системы рациональных функций с рациональными коэффициентами (алгоритмический подход) // Труды XXIII Международной конференции «Математическое моделирование и суперкомпьютерные технологии» – ННГУ им. Н.И. Лобачевского, 2023. С. 8-12.
5. Алексиадис Н.Ф. О рациональных А-функциях с рациональными коэффициентами //XIX Международная конференция «Алгебра, теория чисел, дискретная геометрия и много-масштабное моделирование: современные проблемы, приложения и проблемы исто-рии», посвященная двухсотлетию со дня рождения академика П. Л. Чебышева - Тула, 2021. С. 97-101.
6. Алексиадис Н.Ф. Рациональные А-функции с рациональными коэффициентами. 2022. // Чебышевский сборник. - Том 23. Выпуск 3С. 2-9. DOI 10.22405/2226-8383-2022-23-3-2-10.
7. Алексиадис Н.Ф. Об аналоге теоремы Колмогорова о суперпозициях непрерывных функций для функциональных систем полиномиальных и рациональных функций // Ма-териалы XXII Международной конференции «Алгебра, теория чисел, дискретная гео-метрия и многомасштабное моделирование: современные проблемы, приложения и проблемы истории», посвященной 120-летию со дня рождения академика А. Н. Колмо-горова – ТГПУ им. Л. Н. Толстого, 2023. С. 140-144.
8. Aleksiadis N.Ph. An analogue of Kolmogorov’s theorem on superpositions of continu-ousfunctions for functional systems of polynomial and rational functions // Chebyshevskii sbornik. 2024. V. 25, No. 2. P. 29–42. DOI 10.22405/2226-8383-2024-25-2-29-42.
9. Кудрявцев В.Б. Функциональные системы. – М.: Изд--во МГУ, 1982. 157 с.
10. Яблонский С.В. Введение в дискретную математику. – М.: Изд--во Наука, 1986. 384 с.

МНОГОКРИТЕРИАЛЬНЫЕ МИНИМАКСНЫЕ ЗАДАЧИ ОПТИМАЛЬНОГО УПРАВЛЕНИЯ[1*]

Д.В. Баландин, Р.С. Бирюков, М.М. Коган

Нижегородский государственный университет им. Н.И. Лобачевского,
Научно-технологический университет «Сириус»

Рассматриваются задачи многокритериального минимаксного управления динамическими системами с критериями, представленными как максимальные уклонения выходов динамической системы по отношению к ее входу и начальному состоянию. Показано, что синтез оптимальных по Парето управлений может быть осуществлен в терминах линейных матричных неравенств на основе оптимизации свертки Гермейера, которая оказывается максимальным уклонением выхода, составленного из целевых выходов, умноженных на скаляры. Приводятся примеры решения задач оптимального управления механическими системами: оптимальная защита объектов от вибрационных и ударных воздействий; оптимальное гашение колебаний вращающего в электромагнитных подшипниках ротора; отслеживание заданной траектории мобильным сферическим роботом.

Ключевые слова: многокритериальная оптимизация, множество Парето, обобщенная H_2 норма, максимальное уклонение выхода, виброзащита, виброизоляция, жесткий ротор, сферический робот.

1. Введение

Реальные задачи управления всегда многокритериальны. Нахождение множества Парето, а, следовательно, и оптимальных по Парето решений, т.е. неулучшаемых одновременно для всех критериев, представляет собой сложную задачу. Известно, что замена многокритериальной задачи на однокритериальную с критерием в виде свертки Гермейера исходных критериев обеспечивает в общем случае нахождение некоторого подмножества множества Парето. В случае, когда отдельные исходные критерии представляют собой обобщенные H_2 нормы, то есть являются максимальными уклонениями целевых выходов рассматриваемой динамической системы по отношению к ее входу и начальному состоянию, решение многокритериальной задачи может быть получено через решение вспомогательной задачи полуопределенного программирования с помощью аппарата линейных матричных неравенств.

В настоящей работе приводится краткий обзор решения многокритериальных задач и рассматриваются примеры применения данного подхода. В качестве таких примеров выбраны задача виброзащиты, задача стабилизации вращения жесткого вертикального ротора в электромагнитных подшипниках и задача отслеживания траектории движения сферического мобильного робота.

2. Синтез оптимальных по Парето управлений

Рассмотрим линейную непрерывную нестационарную систему на конечном горизонте

$$\begin{aligned} &\dot{x} = A(t)x + B_u(t)u + B_v(t)v,\ x(t_0) = x_0,\ t \in [t_0, t_f], \\ &z_1 = C_1(t)x + D_{u,1}(t)u, \\ &z_2 = C_2(t)x + D_{u,1}(t)u, \end{aligned} \tag{1}$$

где $x \in \mathbb{R}^{n_x}$ - состояние, $v \in \mathbb{R}^{n_v}$ - возмущение и $z_1, z_2 \in \mathbb{R}^{n_z}$ - целевые выходы. Также введем в рассмотрение два функционала

[1*] Работа выполнена при финансовой поддержке РНФ (проект № 24-11-20023).

$$J_1\big(\Theta(t)\big) = \sup_{x_0,v} \frac{\sup\limits_{t\in[t_0,t_f]} |z_1(t)|^2}{x_0^T R^{-1} x_0 + \|v\|^2}, \qquad J_2\big(\Theta(t)\big) = \sup_{x_0,v} \frac{\sup\limits_{t\in[t_0,t_f]} |z_2(t)|^2}{x_0^T R^{-1} x_0 + \|v\|^2},$$

здесь весовая матрица $R = R^T > 0$ отражает компромисс при учёте неопределённости в начальных условиях и внешнем возмущении. Отметим, что данные функционалы представляют собой квадраты обобщённых H_2 норм системы от входа v к целевому выходу z_1 и z_2 соответственно. Требуется найти оптимальное по Парето управление в виде линейной нестационарной обратной связи $u = \Theta(t)x$, минимизирующее функционалы J_1 и J_2.

Самым распространенным методом нахождения оптимальных по Парето решений многокритериальной задачи является метод скаляризации. Для этого определим на траекториях системы (1) вспомогательный функционал

$$J_\alpha(u) = \max\left\{\frac{J_1(u)}{\alpha}, \frac{J_2(u)}{1-\alpha}\right\}, \quad 0 < \alpha < 1,$$

представляющий собой свертку Гермейера исходных функционалов. Как известно, параметры Θ_α, для которых выполняется

$$\min_{\Theta} J_\alpha(\Theta) = J_\alpha(\Theta_\alpha),$$

являются оптимальными по Парето решениями многокритериальной задачи. Необходимые условия оптимальности в рассматриваемой задаче формулируются следующим образом [1].

Теорема.1 Оптимальными по Парето управлениями в многокритериальной задаче являются обобщенные H_2-оптимальные управления с параметрами $\Theta_\alpha(t) = \mathrm{Z}(t)Y^{-1}(t)$, где матрицы $Y(t) = Y^T(t) \geq 0$ и $Z(t)$ находятся из решения задачи $\inf \gamma^2$ при ограничениях, определяемых линейными матричными неравенствами:

$$\dot{Y} + AY + YA^{\mathrm{T}} + B_u Z + Z^{\mathrm{T}} B_u^{\mathrm{T}} + B_v B_v^T = 0, \quad Y(0) = R,$$
$$\begin{pmatrix} Y & YC_1^{\mathrm{T}} + Z^{\mathrm{T}} D_1^{\mathrm{T}} \\ C_1 Y + D_1 Z & \gamma^2 \alpha^2 I \end{pmatrix} \geq 0, \quad \begin{pmatrix} Y & YC_2^{\mathrm{T}} + Z^{\mathrm{T}} D_2^{\mathrm{T}} \\ C_2 Y + D_2 Z & \gamma^2 (1-\alpha)^2 I \end{pmatrix} \geq 0. \qquad (2)$$

3. Задачи оптимального управления механическими системами

3.1. Задача виброзащиты

В качестве первого примера рассмотрим механическую систему с двумя степенями свободы, показанную на рисунке 1 и представляющую собой упругий объект, который моделируется двумя материальными точками 2 и 3, связанными между собой линейными упругим и диссипативным элементами. Этот объект связан упругим и диссипативным элементами и управляемым элементом (называемым далее виброизолятором) с другим телом 1, которое является подвижным основанием. Динамика данной механической системы (в безразмерных переменных и параметрах) описывается дифференциальными уравнениями

$$\ddot{x}_1 = -2\beta\dot{x}_1 + \beta\dot{x}_2 - 2x_1 + x_2 + v + u,$$
$$\ddot{x}_2 = -\beta(\dot{x}_2 - \dot{x}_1) - x_2 + x_1 + v,$$

где x_1 и x_2 – координаты материальных точек 2 и 3 относительно подвижного основания, u – усилие, создаваемое виброизолятором при его деформации (т. е. при смещении точки 2 относительно точки 1), v – с точностью до знака ускорение основания (материальной точки 1), $\beta = 0.1$ – заданный положительный параметр демпфирования. Задача виброизоляции состоит в поиске нестационарного управления $u = \theta_1(t)x_1 + \theta_2(t)x_2 + \theta_3(t)\dot{x}_1 + \theta_4(t)\dot{x}_2$, обеспечивающего как наименьшую возможную деформацию механической системы, так и минимальную силу, противодействующую смещению упругого объекта относительно основания. С этой целью выберем выходы

$$z_1 = (x_1, x_2 - x_1)^{\mathrm{T}}, \qquad z_2 = -x_1 - \beta\dot{x}_1 + u,$$

и в качестве критериев возьмем обобщенные H_2 нормы относительно этих выходов.

Зададим на интервале времени длительностью $T = 20$ равномерную сетку с шагом $h = 0.05$ и положим $R = \mathrm{diag}(10,10,1,1)$. На рисунке 1 изображено множество Парето, точка A соответствует случаю $u = 0$.

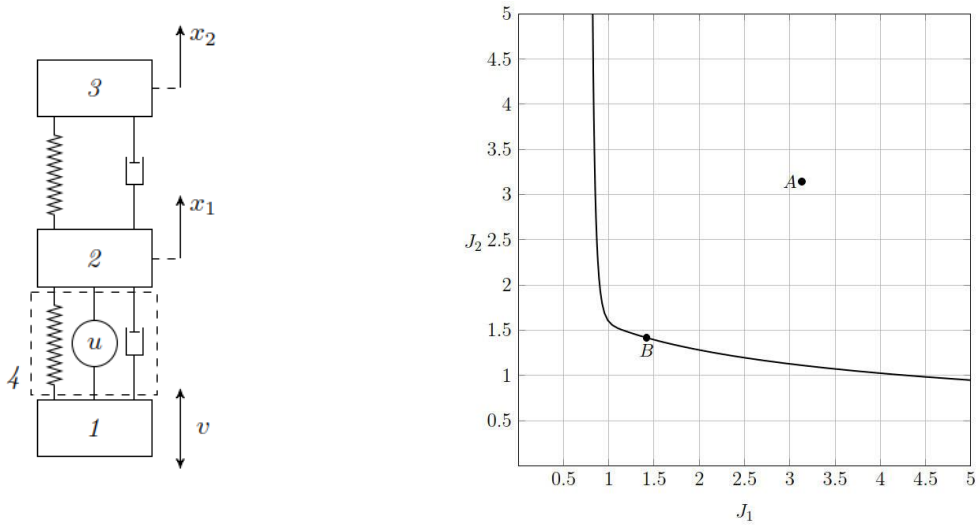


Рис. 1. Схематическое изображение механической системы с двумя степенями свободы (слева) и множество Парето на плоскости критериев (справа)

3.2. Управление движением ротора в электромагнитных подшипниках

Движение жесткого вертикально вращающегося ротора в радиальных электромагнитных подшипниках, представляющих собой четыре пары электромагнитов, описывается в безразмерной форме следующей системой дифференциальных уравнений [2]:

$$\begin{aligned} \ddot{\alpha} &= \mu_2 u_2 - \mu_1 u_1 - \rho_0 \dot{\beta} + w_1, \quad \ddot{x} = u_3 + u_4 + w_3, \\ \ddot{\beta} &= \mu_1 u_3 - \mu_2 u_4 + \rho_0 \dot{\alpha} + w_2, \quad \ddot{y} = u_1 + u_2 + w_4, \end{aligned} \tag{2}$$

где x и y - координаты центра масс ротора, α и β - углы поворота ротора относительно осей x и y соответственно, $w_1, \dots, w_4$ - внешние силы, действующие на ротор, и $u_1, \dots, u_4$ - управления, порожденные электромагнитными силами, возникающими в подшипниках. Безразмерные параметры, входящие в систему, имеют вид: $\mu_1 = ml_0l_1/J$, $\mu_2 = ml_0l_2/J$, $\rho_0 = J_z\omega\sqrt{m\delta_0}/J$, где l_1 и l_2 - расстояния от центра масс до верхнего и нижнего подшипников соответственно, $l_0 = l_1 + l_2$, J и J_z - главные моменты инерции ротора, ω - заданная угловая частота вращения ротора относительно оси z, δ_0 - зазор между телом и электромагнитами [2].

Задача гашения колебаний ротора, вращающегося в электромагнитных подшипниках, состоит в поиске стационарного управления, обеспечивающего как наименьшее возможное смещения ротора в верхних и нижних электромагнитных подшипниках, так и минимальные значения управлений. С этой целью выразим смещения ротора в подшипниках через координаты центра масс и углы поворотов:

$$x_u = x + \nu_1\beta, \qquad x_l = x - \nu_2\beta, \qquad y_u = y - \nu_1\alpha, \qquad y_l = y + \nu_2\alpha, \tag{3}$$

и выберем целевые выходы как

$$z_1 = (x_u, x_l, y_u, y_l)^{\mathrm{T}}, \qquad z_2 = (u_1, u_2, u_3, u_4)^{\mathrm{T}},$$

здесь $\nu_1 = l_1/l_0$, $\nu_2 = l_2/l_0$. В качестве критериев возьмем обобщенные H_2 нормы относительно этих выходов, считая, что соответствующие нормы вычисляются по формулам:

$$|z_1| = \max\{|x_u|, |x_l|, |y_u|, |y_l|\}, \qquad |z_2| = \max\{|u_1|, |u_2|, |u_3|, |u_4|\}.$$

Решение поставленной задачи можно искать как в классе управлений в форме линейной обратной связи по состоянию, так называемые централизованные управления, так и в классе децентрализованных управлений, зависящих только от переменных x_u, x_l, y_u, y_l:

$$\begin{aligned} u_1 &= -\nu_2(ay_u + b\dot{y}_u), \quad u_2 = -\nu_1(ay_l + b\dot{y}_l), \\ u_3 &= -\nu_2(ax_u + b\dot{x}_u), \quad u_4 = -\nu_1(ax_l + b\dot{x}_l). \end{aligned}$$

Задача поиска централизованного регулятора решается с использованием линейных матричных неравенств на основе теоремы из предыдущего пункта, а задача поиска децентрализованного регулятора решена аналитически:

$$a = \frac{1-\alpha}{\nu_1 \alpha}\sqrt{\frac{\nu_1^2+\mu_0^2}{3(\nu_2^2+\mu_0^2)}}, \qquad b = \frac{2^{1/2}}{3^{1/4}}\frac{(\nu_1^2+\mu_0^2)^{1/4}(\nu_2^2+\mu_0^2)^{1/4}}{\sqrt{\nu_1\mu_0(\nu_2^2+\mu_0)}}\sqrt{\frac{1-\alpha}{\alpha}}, \qquad 0<\alpha<1,$$

где $\mu_0 = ml_1l_2/J$.

В качестве иллюстрации приведем оптимальные по Парето множества в плоскости критериев (J_1, J_2) для следующих значений параметров: $J = 0.74$кг $\cdot$ м^2, $J_z = 1.7 \cdot 10^{-2}$кг $\cdot$ м^2, $l_1 = 0.3$м, $l_2 = 0.2$м, $m = 13.7$кг, $\delta_0 = 5 \cdot 10^{-4}$м, $\omega = 314$ рад/с, $\mu_0 = 1.11$, $\nu_1 = 0.6$ и $\nu_2 = 0.4$. На рисунке 2 сплошной линией изображена кривая, соответствующая случаю централизованного регулятора, а штриховой линией - децентрализованного регулятора. Из представленного рисунка видно, что оптимальные по Парето децентрализованные управления лишь немного уступают централизованным управлениям.

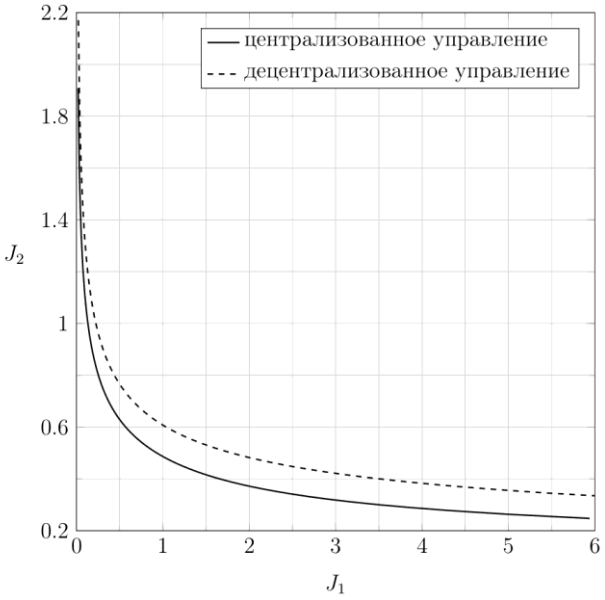


Рис. 2. Множество Парето на плоскости критериев: сплошная линия – централизованное управление, штриховая линия – децентрализованное управление

3.3. Управление движением сферического робота

В качестве последнего примера, иллюстрирующего изложенный ранее подход, рассмотрим задачу обеспечения движения центра сферического робота вдоль заранее заданной траектории. Сферический робот представляет собой две тонкостенные концентрические сферические оболочки. Внешняя оболочка имеет массу M и радиус R, а момент инерции относительно любой оси, проходящей через ее центр, равен J. Аналогичные параметры для внутренней оболочки равны r, m и j соответственно. Робот приводится в движение за счет вращения оболочек относительно друг друга.

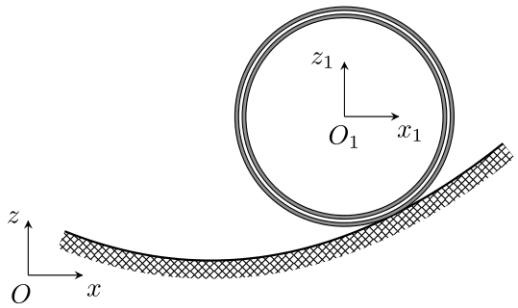


Рис. 3. Схематическое изображение сферического робота

Уравнения движения робота в безразмерных переменных в случае, отсутствия проскальзывания и верчения относительно вектора нормали к опорной поверхности, имеют вид [3]:

$$\dot{\xi} = [\Omega \times n], \qquad (1+\gamma)\dot{\Omega} + \gamma\, n\, \langle \dot{\Omega} \cdot n \rangle = Q - \gamma[n \times k] \tag{4}$$

где $\xi = (x, y, z)$ - координаты центра масс робота, Ω - угловую скорость робота, n - единичный вектор нормали к опорной поверхности, $k = (0, 0, 1)$ – вектор, противоположный направлению силы тяжести, Q – вектор управления и $\gamma = R^2(M + 6m)/J$.

Теперь рассмотрим вопрос об организации движения вдоль заданной траектории, которую описывает центр сферической оболочки. Будем считать, что траектория движения описывается параметрическим уравнением $\xi_0 = \Xi(t)$, где $\Xi(t)$ – дважды непрерывно дифференцируемая функция. Требуется найти вектор управления Q, обеспечивающий приближение робота к заданной траектории. Если в начальный момент времени $t = 0$ выполняется соотношение $\Omega(0) = [n \times \dot{\xi}(0)]$, то вектор Q находится из второго уравнения (4):

$$Q = Q_0 = (1+\gamma)\dot{\Omega}_0 + \gamma\, n\, \langle \dot{\Omega}_0 \cdot n \rangle + \gamma[n \times k]. \tag{5}$$

Пусть теперь $\Omega(0) \neq \left[n \times \dot{\xi}_0\right]$, в этом случае обеспечить приближение к заданной траектории можно с помощью дополнительного корректирующего управления, а именно

$$Q = Q_0 + \Theta(\zeta - \zeta_0), \qquad \zeta = \left(\xi^\top,\ \dot{\xi}^\top\right)^\top. \tag{6}$$

где матрица коэффициентов обратной связи Θ, вообще говоря, может быть функцией времени. Однако, для использования соотношения (5) требуется знать вектор нормали n, что не всегда возможно, поэтому заменим закон управления (5) на более простой и не требующий знания вектора нормали:

$$Q = (1+\gamma)[k \times \dot{\xi}_0] + \Theta(\zeta - \zeta_0). \tag{7}$$

Для определения параметров обратной связи Θ введем в рассмотрение два функционала:

$$J_1(\Theta) = \sup_{\zeta(0)\neq 0} \frac{\sup_{t\geq 0}|\zeta - \zeta_0|}{\sqrt{\zeta^\top(0)\, R\, \zeta(0)}}, \qquad J_2(\Theta) = \sup_{\zeta(0)\neq 0} \frac{\sup_{t\geq 0}|\Theta(\zeta - \zeta_0)|}{\sqrt{\zeta^\top(0)\, R\, \zeta(0)}}.$$

где $R = R^\top > 0$ – заданная матрица. Нетрудно видеть, что первый функционал характеризует максимальное отклонение траектории движения робота от заданной траектории ξ_0 вне зависимости от начальных условий и отличий вектора нормали от вектора k, а второй – максимальное значение корректирующих управлений.

Таким образом, для определения матрицы обратной связи Θ требуется для системы (4) найти оптимальные по Парето решения двухкритериальной задачи оптимального управления с функционалами J_1 и J_2.

При проведении числительных экспериментов примем следующие значения параметров $\kappa = 0.8$, $\gamma = 10.5$. Будем считать, что робот движется по плоскости $z = 0$, а требуемая траектория движения центра масс робота является окружностью

$$x_0(t) = \rho\, \cos\alpha\, t\,, \qquad y_0(t) = \rho\, \sin\alpha\, t\,, \qquad z_0(t) = \rho.$$

Начальные условия зададим следующим образом:

$$\xi(0) = (\rho, 0, 1), \qquad \dot{\xi}(0) = (0, \alpha\,\rho, 0), \qquad \Omega(0) = (-\alpha\,\rho\, +\, 0.1, -0.2, 0).$$

На рис. 4 при значениях параметров $\rho = 2$ и $\alpha = 0.5$ представлены результаты численного моделирования приближения сферического робота к движению вдоль окружности: линией черного цвета показана желаемая траектория, а красной линией – траектория движения центра масс сферического робота. Как можно видеть, построенное управление обеспечивает решение поставленной задачи, то есть траектория движения центра робота приближается к заданной окружности.

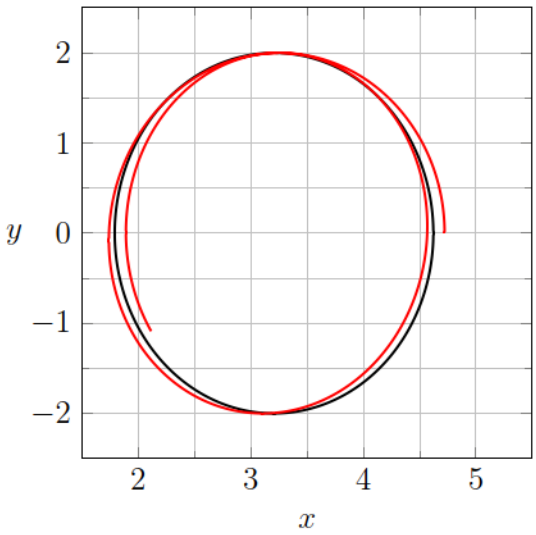


Рис. 4. Приближение центра масс сферического робота к заданной траектории

Литература

1. Баландин Д.В., Коган М.М. Оптимальное по Парето обобщенное H_2-управление и задачи виброзащиты // АиТ. 2017. № 8. С. 76-90.
2. Баландин Д.В., Бирюков Р.С., Коган М.М. Оптимальное по Парето управление ротором в электромагнитных подшипниках // Известия РАН. Теория и системы управления. 2018. № 6. С. 131-142.
3. Balandin D.V., Biryukov R.S., Tuzikov A.M. Path Tracking Control of a Spherical Robot with Pendulum-Like Driver // Communications in Computer and Information Science. 2024. V. 1914, p. 176-189.

НЕКОТОРЫЕ ЗАДАЧИ И АЛГОРИТМЫ НАВИГАЦИИ АВТОНОМНЫХ ТРАНСПОРТНЫХ СРЕДСТВ[1*]

Борисов Н.А., Егоров К.С.

Нижегородский государственный университет им. Н.И. Лобачевского

В данной работе проводится обзор современных решений в области автономных транспортных средств и технологий их программирования, а также обсуждается опыт преподавания робототехники в институте информационных технологий, математики и механики ННГУ им. Н.И. Лобачевского.

Ключевые слова: робототехника, беспилотные транспортные средства, алгоритмы картографирования.

1. Введение

В современном мире автоматизация технологических процессов в промышленности и повседневной жизни играет всё возрастающую роль. В первую очередь это связано со стремлением заменить низкоквалифицированный, монотонный ручной труд различными автоматическими машинами с целью как удешевить весь процесс производства или оказания различных услуг, так и сделать его более эффективным в целом. Также сказывается и всё увеличивающийся недостаток квалифицированных кадров рабочих специальностей, поэтому иногда вопрос автоматизации производства равнозначен вопросу выживаемости самого предприятия [3,4,5].

Одна из сфер, в которой автоматизация процессов за последнее десятилетие увеличивает свою долю – это сфера транспорта. Причём, если рассматривать беспилотные автомобили на дорогах общего пользования - то в данном случае, конечно, рано ещё говорить об их повсеместном внедрении, тогда как транспорт, работающий в особых условиях - такой, который работает, во-первых, вне дорог общего пользования с интенсивным движением; во-вторых, на сравнительно небольших расстояниях и малых скоростях – приобретает всё большую степень автономности.

Для выполнения задач планирования движения на местности необходимо как-то получить о ней информацию. Описание операционной области называется картой местности и представляет собой граф, вершины которого – некоторые точки на поверхности, а рёбра - пути, по которым можно попасть из одной точки в другую. Планирование же маршрута может осуществляться с помощью любого алгоритма обхода графа.

Сбор информации о местности может также выполняться разными способами - они различаются по типу применяемых сенсоров. Это могут быть лидары (лазерные дальномеры), ультразвуковые дальномеры, а также камеры – в случае с ними мы говорим уже о технологиях анализа изображений. Перечисленные сенсоры обычно используются в том случае, когда внести какие-либо изменения в операционную среду невозможно или крайне нежелательно (пример - роботы-планетоходы, домашние роботы). Методы навигации в таком случае называются “неитрузивными”, к ним относится SLAM в разных его вариациях.

Методы же с использованием искусственных меток называются, наоборот, “интрузивными”. Разработано множество алгоритмов такого типа – начиная от простого следования по линии, которое используется, например, при обучении школьников или студентов основам робототехники, а также в автономных транспортных средствах под названием AGV (Automated Guided Vehicle). Это специальные грузовые (реже пассажирские) транспортные средства, осуществляющие перевозку грузов или пассажиров по фиксированному маршруту по физически зафиксированному пути.

[1*] Работа выполнена при поддержке научно-образовательного математического центра «Математика технологий будущего» (проект № 075-02-2024-1439).

Есть более сложные методы с использованием искусственных меток, такие, как, например, система FROG(англ. Free Range on Grid). К сожалению, эта система проприетарная, поэтому в этой работе рассматривалась лишь как идея для создания своего аналога такой системы.

У интрузивных и неинтрузивных методов есть как свои преимущества, так и недостатки. К преимуществам неинтрузивных методов относится, безусловно, отсутствие необходимости в какой либо дополнительной инфраструктуре. Однако же, есть такие существенные недостатки, как низкая точность некоторых алгоритмов SLAM, чувствительность датчиков к помехам. В случае использования более точных реализаций метода SLAM существенно возрастает требовательность к вычислительным ресурсам, что ведёт к удорожанию оборудования.

У интрузивных методов же очевидным недостатком является необходимость в дополнительной инфраструктуре. С другой стороны, это обстоятельство является и преимуществом – датчики, которые используются в данных методах, во-первых, дешевле, а во вторых менее подвержены разнообразным помехам.

Полученные компетенции были использованы при разработке и внедрении курса «Основы робототехники», читаемы сейчас для магистров 1 курса по направлению подготовки «Программная инженерия» очной формы обучения.

2. Обзор существующих решений в области автономной навигации

Задача автономной навигации – это задача определения роботом своего местоположения на местности относительно цели движения и построение оптимального (по разным критериям) маршрута до этой цели. Такие методы делятся на две главных подгруппы – “интрузивные” и “неинтрузивные”, т.е. использующие и не использующие различные искусственные метки на местности соответственно. Существует множество технических реализаций алгоритмов из обоих семейств, различающихся по используемому оборудованию, сложности и требовательности к вычислительным ресурсам.

У обоих семейств алгоритмов есть свои преимущества и недостатки - у “интрузивных” методов теоретически более высокая надежность в связи с использованием более простых датчиков, а у “неинтрузивных” – наоборот.

К “интрузивным” методам можно отнести разнообразные алгоритмы езды по линии, считывание QR-кодов и т.д. В семействе “неинтрузивных” методов были рассмотрены разнообразные реализации алгоритма SLAM (англ. Simultaneous Localisation and Mapping). Это было рассмотрено в работах [1,2].

SLAM (англ. Simultaneous Localisation and Mapping) – это метод одновременной локализации и картографирования, т.е. на заранее неизвестной местности робот, используя данный метод, и определяет своё местоположение с использованием различных сенсоров, и одновременно с этим строит карту местности или обновляет её участки. Алгоритм подходит для использования разными роботами – наземными, квадрокоптерами, подводными роботами и т.д.

С помощью “органов чувств” робота (дальномеров, одометров и т.д.) вычисляется расстояние от агента до препятствий в окружении и, таким образом, составляется 2D или 3D карта местности.

Математическая модель робота может быть сформулирована следующим образом [2]:

$U_{\{1\dots t\}} = \{U_1, U_2, \dots, U_t\}$, где $U_{\{1\dots t\}}$ – управление роботом в момент времени t;

$Z_{\{1\dots t\}} = \{Z_1, Z_2, \dots, Z_t\}$, где $Z_{\{1\dots t\}}$ – обозреваемая роботом окружающая среда в момент времени t;

$X_{\{1\dots t\}} = \{X_1, X_2, \dots, X_t\}$, где $X_{\{1\dots t\}}$ – местоположение робота в момент времени t.

С учётом наличия карты местности m функция траектории движения робота может быть записана в виде $P\left(X_{\{1\dots t\}}, m \middle| Z_{\{1\dots t\}}, U_{\{1\dots t\}}\right)$.

Существует множество реализаций SLAM:

- Feature-based SLAM: использует легко идентифицируемые элементы в среде и с учётом местоположения этих ориентиров строят внутреннее представление о состоянии среды, в которой робот находится. Исторически первый и наиболее используемый по сей день метод.

- Graph-based SLAM: карта представляется в виде графа, вершины которого - позы робота в разное время, а рёбра – пространственные ограничения. По этому графу определяется наиболее вероятная поза с учётом рёбер графа.
- Grid-based SLAM: среда, в которой находится робот, разбивается на сетку точек, каждая из которых может принимать три значения - занята препятствием, свободна или не исследована. Наиболее простой в представлении метод.
- Topological SLAM – топологическое описание карты, в котором легко различимые значимые места заносятся в узлы графа.
- Semantic SLAM – семантическое представление карты местности, главная идея заключается в сопоставлении объектов местности с семантическими концепциями в составленной карте.

3. Учебный курс по робототехнике для студентов магистратуры

Учебная программа по робототехнике в институте ИТММ ННГУ им. Н.И.Лобачевского, предназначенная для студентов магистратуры, обучающихся по специальности «Программная инженерия», преследует следующие цели:

1. Знакомство с современными подходами в программировании промышленных и бытовых роботов.
2. Комплексный обзор методов и алгоритмов, использующихся в робототехнике.

В рамках первой цели рассматривается использование экосистемы ROS (англ. Robot Operating System) – специализированной надстройки над ОС Linux/Windows для распределённой обработки данных, которая существенно упрощает разработку сложных приложений для управления роботами и/или системой роботов. Это достигается введением таких абстракций, как node и topic – программная абстракция для запуска отдельно взятой программы, отвечающей за какой-то аспект функционирования робота и области памяти, хранящей данные соответственно.

В рамках второй цели студентам предлагаются к рассмотрению такие задачи, как реализация автоматического управления электронным устройством, реализация средств контроля и диспетчеризации. Примеры задач - автономное движение и навигация, системы управления автономными транспортными средствами.

Актуальность и полезность данного курса заключается в начальной подготовке специалистов-разработчиков робототехнических систем. Курс активно дорабатывается по результатам освоения материала студентами и современными тенденциями и потребностями на рынке труда и в сфере робототехники. В частности, планируется миграция на систему ROS2, являющуюся дальнейшим развитием ROS, а также включение в курс рассмотрение таких вопросов, как задачи компьютерного зрения, использование искусственного интеллекта в анализе поступающих данных и поддержке принятия решений.

Литература

1. Gao, X., & Zhang, T. (2021). Introduction to Visual SLAM. Springer Singapore. https://doi.org/10.1007/978-981-16-4939-4.
2. Pedrosa, E., L. Reis, C. M. D. Silva and H. S. Ferreira. Autonomous Navigation with Simultaneous Localization and Mapping in/outdoor. 2020.
3. Gmapping [Электронный ресурс] URL: http://wiki.ros.org/gmapping, свободный ресурс – Загл. с экрана. Язык англ. Дата обращения: 14.08.2020.
4. RTAB-Map, Real-Time Appearance-Based Mapping [Электронный ресурс] URL: http://introlab.github.io/rtabmap/, свободный ресурс – Загл. с экрана. Язык англ. Дата обращения: 22.06.2020.
5. Жигалов К.Ю., Клочкова Е.Н. Решение задачи построения оптимального маршрута движения транспорта по пересеченной местности с использованием ГИС // Московский экономический журнал. 2020, №1. С. 71-81. DOI 10.24411/2413-046X-2020-10039.

6. Abd Algfoor Z., Sunar M. S., Kolivand H. A Comprehensive Study on Pathfinding Techniques for Robotics and Video Games // International Journal of Computer Games Technology. 2015. Article ID 736138. https://doi.org/10.1155/2015/736138.
7. https://fab.cba.mit.edu/classes/865.21/topics/path_planning/robotic.html - построение маршрута движения робота.
8. https://pythonist.ru/algoritm-a-star-i-ego-realizacziya-na-python/ - алгоритм A*.

ИССЛЕДОВАНИЕ ЭФФЕКТИВНОСТИ РАСПАРАЛЛЕЛИВАНИЯ ИТЕРАЦИОННОГО АЛГОРИТМА РЕШЕНИЯ НЕЛИНЕЙНОЙ СИСТЕМЫ ПЕРЕНОСА ИЗЛУЧЕНИЯ И СТАТИСТИЧЕСКОГО РАВНОВЕСИЯ

А.А. Бусалов[1], А.В. Калинин[1,2], А.А. Тюхтина[1,2]

[1]Нижегородский университет им. Н.И. Лобачевского
[2]Институт прикладной физики РАН

В данной работе приведен алгоритм распараллеливания при численном решении нелинейной системы переноса излучения и статистического равновесия с использованием стандарта OpenMP. В рамках работы разработано два подхода к распараллеливанию: по угловым и по пространственным переменным. Приведены исследования эффективности реализованных алгоритмов. Внедрение распараллеливания на общей памяти позволило существенно сократить календарное время расчетов.

Ключевые слова: нелинейное уравнение переноса, параллельные вычисления, эффективность распараллеливания, стандарт OpenMP.

1. Введение

Численное моделирование процессов переноса излучения и взаимодействия излучения с веществом является исключительно трудной задачей, что обусловлено большой размерностью сеточных уравнений и спецификой моделируемого процесса [1-3]. В работах [4,5] был теоретически обоснован и численно исследован линеаризующий итерационный алгоритм решения нелинейной задачи теории переноса излучения и статистического равновесия в диффузионном и кинетическом приближениях. Однако, время проведения расчетов, особенно в оптически плотных и разноплотных средах по разработанному алгоритму с подробной дискретизацией фазового пространства было велико из-за большой размерности сеточных уравнений. В связи с этим внедрение распараллеливания в существующий алгоритм, с целью сокращения времени счета, является актуальной задачей.

В рамках данной работы рассмотрены вопросы, связанные с внедрением и тестированием модели распараллеливания на общей памяти с использованием стандарта OpenMP в разработанный ранее линеаризующий итерационный алгоритм.

2. Постановка задачи

Нелинейная система интегро-дифференциальных уравнений переноса излучения и статистического равновесия в пространственно многомерном случае, соответствующая модели двухуровневого атома в рамках предположения о полном перераспределения излучения по частоте имеет вид [1,2]:

$$\frac{1}{c}\frac{\partial\varphi(x,\nu,\omega,t)}{\partial t}+\left(\omega,\nabla\right)\varphi(x,\nu,\omega,t)+h\nu_{12}\frac{\kappa(\nu)}{4\pi}\cdot$$
$$\cdot\left[B_{12}C_1(x,t)-B_{21}C_2(x,t)\right]\varphi(x,\nu,\omega,t)=h\nu_{12}\frac{\kappa(\nu)}{4\pi}A_{21}C_2(x,t), \quad (1)$$

$$\left[C_{12}n_e(x)+B_{12}\int\limits_I\int\limits_\Omega\frac{\kappa(\nu)}{4\pi}\varphi(x,\nu,\omega,t)d\omega d\nu\right]C_1(x,t)=$$
$$=\left[A_{21}+C_{21}n_e(x)+B_{21}\int\limits_I\int\limits_\Omega\frac{\kappa(\nu)}{4\pi}\varphi(x,\nu,\omega,t)d\omega d\nu\right]C_2(x,t), \quad (2)$$

$$C_1(t,x)+C_2(t,x)=f(x) \quad (3)$$

для определения интенсивности излучения $\varphi(x,\nu,\omega,t)$ и концентраций $C_1(x,t)$ и $C_2(x,t)$ атомов, находящихся в основном и в возбужденном состояниях соответственно.

Здесь $x=\{x_1,x_2,x_3\}\in G\subset R^3, \omega\in\Omega=\{\omega=\{\omega_1,\omega_2,\omega_3\}\in R^3:\omega_1^2+\omega_2^2+\omega_3^2=1\}$, $\nu\in I=[0,\nu_0]$; $t=[0,T], T>0; G$ – выпуклое ограниченное множество с гладкой границей ∂G, имеющей всюду внешнюю нормаль $n(x)\left(x\in\partial G\right)$. Система (1) - (3) дополняется начальными и граничными условиями:

$$\varphi(x,\nu,\omega,t)=0, x\in\partial G,\left(\omega,n(x)\right)<0,\omega\in\Omega,\nu\in I,t\in[0,T], \quad (4)$$

$$\varphi(x,\nu,\omega,0)=\varphi_0(x,\nu,\omega). \quad (5)$$

3. Особенности программной реализации и исследования эффективности распараллеливания

Внедрение OpenMP распараллеливания потребовало внесения существенных изменений в структуру данных программы решения уравнения переноса в диффузионном и кинетическом приближениях в пространственно многомерном случае. Двумерная декомпозиция характеризуется двумя переменными: Nx – число OpenMP разбиений по переменной X (по столбцам пространственной сетки); Ny – число OpenMP разбиений по переменной Y (по строкам пространственной сетки). На рисунке 1 показана геометрическая декомпозиция двумерной области (красным выделены границы OpenMP фрагментов).

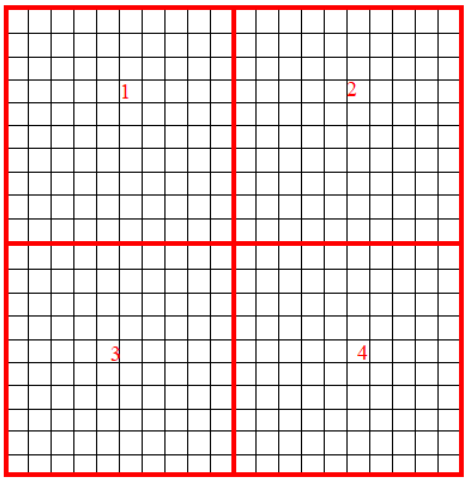


Рис. 1. Геометрическая декомпозиция двумерной области
(красным выделены границы OpenMP фрагментов)

Данная схема практически не имеет ограничений по масштабируемости (ограничение только в числе строк и столбцов), в отличие от схемы распараллеливания по угловым переменным. Кроме того, данная схема является универсальной для кинетического и диффузионного приближений, так как в диффузионном приближении нет зависимости от угловых переменных.

Для тестирования была выбрана задача о распространении излучения в разнопплотной среде. Постановка подробно описана в работе [4]. Исследование эффективности распараллеливания проводилось методом дробления при фиксированном числе пространственных точек.

В таблице 1 приведены временные характеристики расчетов и показатели эффективности, рассчитанные по формуле:

$$E=\frac{t_1}{t_n\cdot n}\cdot 100\%, \quad (6)$$

где n -–число процессоров, t_1 – время счета на одном процессоре, t_n – время счета на n процессорах.

Таблица 1. Результаты тестирования задачи методом дробления

Число точек	Число процессоров	Время счета, сек	Эффективность,%
100000	1	24.67	-
	2	12.58	95
	4	7.21	86
	8	3.94	78

По результатам, приведенным в таблице 1 видно, что получена достаточно хорошая масштабируемость при эффективности распараллеливания выше 70%. Кроме того, внедрение OpenMP распараллеливания позволило существенно сократить время расчетов и повысить точность (появилась возможность более детальной дискретизации фазового пространства)

Литература

1. Иванов В.В. Перенос лучистой энергии в атмосферах звезд и планет. М.: Гостехтеоритздат, 1956.
2. Михалас Д. Звездные атмосферы. М.: Мир, 1982.
3. Гермогенова Т.А. Локальные свойства решений уравнения переноса. М.: Наука, 1966.
4. Kalinin A.V., Tyukhtina A.A., Busalov A.A. Linearizing Algorithm for Solving a Nonlinear Initial Boundary Value Problem of Radiation Transfer in Spatially Multidimensional Domains // D. Balandin et al. (Eds.): MMST 2023, CCIS 1914. Springer, 2023. p. 44-69.
5. Kalinin, A.V., Tyukhtina, A.A., Busalov, A.A.: An iterative method for solving a nonlinear system of the theory of radiation transfer and statistical equilibrium in a plane-parallel layer. In: Balandin, D., et al. (eds.) MMST 2022. CCIS, vol. 1750, p. 106-120. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-24145-1 9.

ГИБРИДНЫЙ АЛГОРИТМ РЕШЕНИЯ ЗАДАЧИ РАЗМЕЩЕНИЯ КОМПОНЕНТ ИНТЕГРАЛЬНЫХ СХЕМ

М.А. Быкова, А.А. Урмаев, А.И. Яушева, Д.В. Пачурина, Д.И. Попугаев

Нижегородский государственный университет им. Н.И. Лобачевского

В работе рассматривается задача размещения компонент интегральных схем. Решение задачи традиционно разбивается на два больших этапа – этап глобального размещения и этап детального размещения. В работе рассматривается этап детального размещения. Представлена математическая постановка задачи. Предлагается гибридный алгоритм решения этапа детального размещения.

Ключевые слова: размещение элементов интегральных схем, глобальное размещение, детальное размещение, гибридный алгоритм.

1. Введение

Физическое проектирование электронных систем включает следующие этапы: компоновка электронной схемы, планирование пространства изделия, размещение компонентов в пространстве изделия, трассировку электронных цепей. Практические задачи физического проектирования нередко имеют большие порядки. Так, например, число элементов современных интегральных схем с субмикронными топологическими нормами может достигать 10^9.

Размещение – это процесс определения расположения схемотехнических устройств на определенной поверхности. Это важный этап в процессе проектирования интегральных схем, поскольку он влияет на реализуемость трасс цепей, производительность итогового изделия, распределение тепла по поверхности кристалла и, в меньшей степени, на энергопотребление конструкции.

В данной работе представлено содержательное описание задачи размещения, ее математическая постановка, описан разработанный гибридный алгоритм решения задачи.

2. Содержательное описание задачи

В задаче размещения компонент электронных схем имеются: размещаемые элементы прямоугольной формы с их размерами (требуемое количество посадочных мест по горизонтали и вертикали); схема, описывающая связи между элементами, которые реализуются в виде проводников; монтажное пространство прямоугольной формы с заданными размерами (количество посадочных мест по горизонтали и вертикали). В задаче требуется для каждого элемента указать его монтажную позицию. Допустимость размещения определяется ограниченной вместимостью монтажных позиций.

Размещение – сложная с точки зрения вычислений задача. Даже простой случай размещения схемы только с модулями единичного размера и 2-контактными сетями вдоль прямой линии для минимизации общей длины трасс цепи является NP-полным [1]. Проблема размещения современных интегральных схем гораздо сложнее.

Сложность, объём исходных данных и число неизвестных в задачах размещения таковы, что этап размещения разбивается на несколько управляемых этапов. Основными этапами являются этап глобального размещения и этап детального размещения. В работе также предлагается добавить этап локальной оптимизации.

Первый этап заключается в поиске решения «огрубленной» задачи размещения – отказ от учета ряда ограничений и некоторых исходных данных, позволяет значительно упростить задачу и обеспечивает поиск приемлемого решения за разумное время. Как правило отказываются от необходимости размещать элементы строго в посадочные места, а также игнорируются факты

взаимного перекрытия размещенных компонент интегральной схемы. Такой этап получил название глобального размещения.

Второй этап заключается в корректировке решения задачи глобального размещения. Так как с точки зрения общего распределения элементов по кристаллу решение в целом имеет приемлемое качество, то это дает возможность сконцентрироваться на локальных изменениях и корректировках решения. Такой подход обеспечивает возможность декомпозиции общей задачи размещения на серию локальных задач, которые по объёму исходных данных и числу неизвестных значительно меньше исходной проблеме. Комплекс задач, решаемых на этом этапе, называют задачами детального размещения.

Третий этап заключается в локальной оптимизации решения, полученного в результате работы первых двух этапов. Алгоритм локальной оптимизации пытается еще улучшить найденное размещение с точки зрения оценки длин будущих трасс, а также разместить элементы, которые не удалось разместить в результате работы этапа детального размещения.

В данной работе исследование сконцентрировано на втором и третьем этапах решения задачи.

3. Математическая постановка задачи размещения

Как говорилось выше, задача размещения элементов интегральной схемы решается в несколько этапов. Основными этапами являются этап глобального размещения и этап детального размещения. В работе предполагается, что этап глобального размещения уже реализован, и получены начальные (глобальные) координаты элементов схемы. Приведем математическую постановку для задачи детального размещения.

Гиперграфом $G(V, E)$ будем моделировать электронную схему, где $V = \{v_1, \dots, v_n\}$ – множество элементов схемы; $E \subseteq 2^V$ – связи между элементами (как электрическая цепь связывает несколько элементов схемы, так и её математическая модель – гиперребро связывает несколько соответствующих вершин графа).

Каждый элемент v_i характеризуется собственным размером $w_i = (w_{i1}, w_{i2}) \in Z^2$, $i = \overline{1,n}$. Размеры определяют количество посадочных мест кристалла, занимаемых элементом по горизонтали и вертикали.

Пусть монтажное пространство имеет размеры (W, H). Пусть на монтажном пространстве имеется m посадочных мест $j \in \{1, \dots, m\}$, каждое из которых характеризуется местоположением в двумерном пространстве кристалла $p_j = (p_{j1}, p_{j2}) \in Z^2$.

Вектор, определяющий размещение вершин на кристалле $\vec{x} = (x_1, \dots, x_n)$, будет решением задачи. Отметим, что элемент вектора x_i определяет положение левого верхнего угла элемента v_i, в то время как сам элемент занимает прямоугольную область согласно своим размерам.

$$x_i \in \{1, \dots, m\}, i = \overline{1,n} \tag{1}$$

При этом значение p_{x_i} соответствует позиции размещения элемента v_i.

Сформулируем требование, чтобы никакая пара размещенных элементов не пересекалась:

$$\begin{gathered} p_{x_i1} + w_{i1} \le p_{x_j1} \text{ или } p_{x_j1} + w_{j1} \le p_{x_i1} \\ p_{x_i2} + w_{i2} \le p_{x_j2} \text{ или } p_{x_j2} + w_{j2} \le p_{x_i2} \\ p_{x_i1} + w_{i1} \le \mathrm{W} \\ p_{x_i2} + w_{i2} \le \mathrm{H} \\ i \ne j, \;\; i,j = \overline{1,n}. \end{gathered} \tag{2}$$

Система ограничений (1), (2) учитывает исполнение требований размещения всех компонент строго в собственные посадочные места и размещения в рамках посадочного места не более одного компонента.

Таким образом, система ограничений (1), (2) обеспечивает все существенные требования детального размещения.

В основу функции цели предлагается положить вычислительно простые оценки длин HPWL, основанные на вычислении полупериметра ограничивающего прямоугольника каждой цепи: $l(x, e) = (\max\limits_{i \in e}\{x_{i1}\} - \min\limits_{i \in e}\{x_{i1}\}) + (\max\limits_{i \in e}\{x_{i2}\} - \min\limits_{i \in e}\{x_{i2}\})$.

Основная цель в задаче глобального размещения – минимизация оценки длины трасс цепей, которую можно записать в виде:

$$F(x) = \sum_{e \in E} l(x, e) \to min \quad (3)$$

Задачу (1)-(3) поиска размещения $\vec{x}$ будем называть задачей детального размещения элементов интегральной схемы.

4. Алгоритм решения задачи размещения

Общий алгоритм решения задачи состоит из следующих этапов:

1. Решение задачи глобального размещения;
2. Решение задачи детального размещения;
3. Локальная оптимизация полученного решения.

Опишем методы, предлагаемые для реализации каждого этапа алгоритма.

4.1. Многоуровневый алгоритм решения задачи глобального размещения

Для решения задачи глобального алгоритма предлагается использовать многоуровневый алгоритм [2].

Центральной идеей многоуровневого метода является последовательное редуцирование размерности задачи путем обобщения информации о задаче и включения в редуцированный вариант наиболее существенных данных, решение редуцированной задачи и последовательное восстановление полученного решения до размерности исходной задачи. При этом на каждом этапе восстановления выполняется улучшение восстанавливаемого решения с учетом добавления раннее исключенных из рассмотрения данных. Многоуровневые методы с успехом применяются для решения задач разбиения графа [3], коммивояжера [4] и многих других.

Таким образом, вместо поиска решения на исходном графе происходит построение иерархии порожденных графов, в который каждый следующий строится на основе предыдущего и имеет меньшие размеры относительно родительского. Последний граф имеет наименьшую размерность, и для него выполняется поиск начального решения. Начальное решение строится на основе модифицированного алгоритма силовой укладки графа [5]. На основе полученного решения происходит восстановление решения на исходном графе. Редуцированный граф поэтапно восстанавливается, при этом на каждом шаге восстановления происходит локальная оптимизация решения также на основе модифицированного алгоритма силовой укладки графа.

4.2. Гибридный алгоритм решения задачи детального размещения

Суть гибридного алгоритма состоит в разбиении области монтажного пространства на регионы, на каждом из которых отдельно решается задача детального размещения, а затем полученные результаты объединяются.

Для решения задачи в каждом регионе можно применить любой алгоритм детального размещения. В работе предлагается последовательное применение радиального алгоритма и алгоритма тетрис, описанных ниже. Особенность радиального алгоритма состоит в том, что он равномерно распределяет элементы по области, а алгоритм тетрис наоборот группирует элементы, располагая их более плотно. В результате работы этих алгоритмов могут остаться неразмещенные элементы.

После объединения результатов, полученных в каждом из регионов, запускается процедура доразмещения неразмещенных ранее элементов в образовавшиеся промежутки.

На схеме (см. рис. 1) показаны исходные данные задачи размещения.

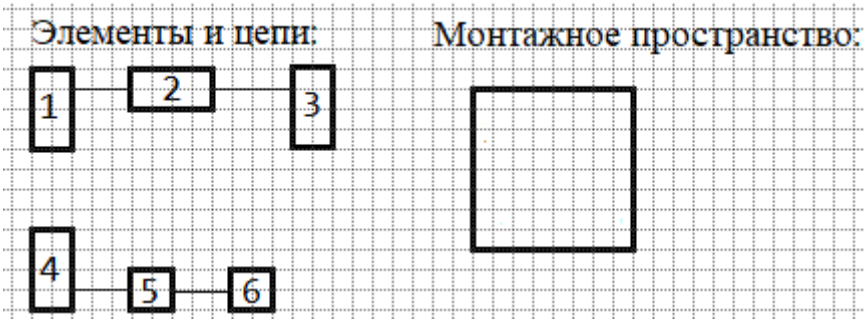


Рис. 1. Исходные данные задачи

На рис. 2 показаны основные этапы работы алгоритма.

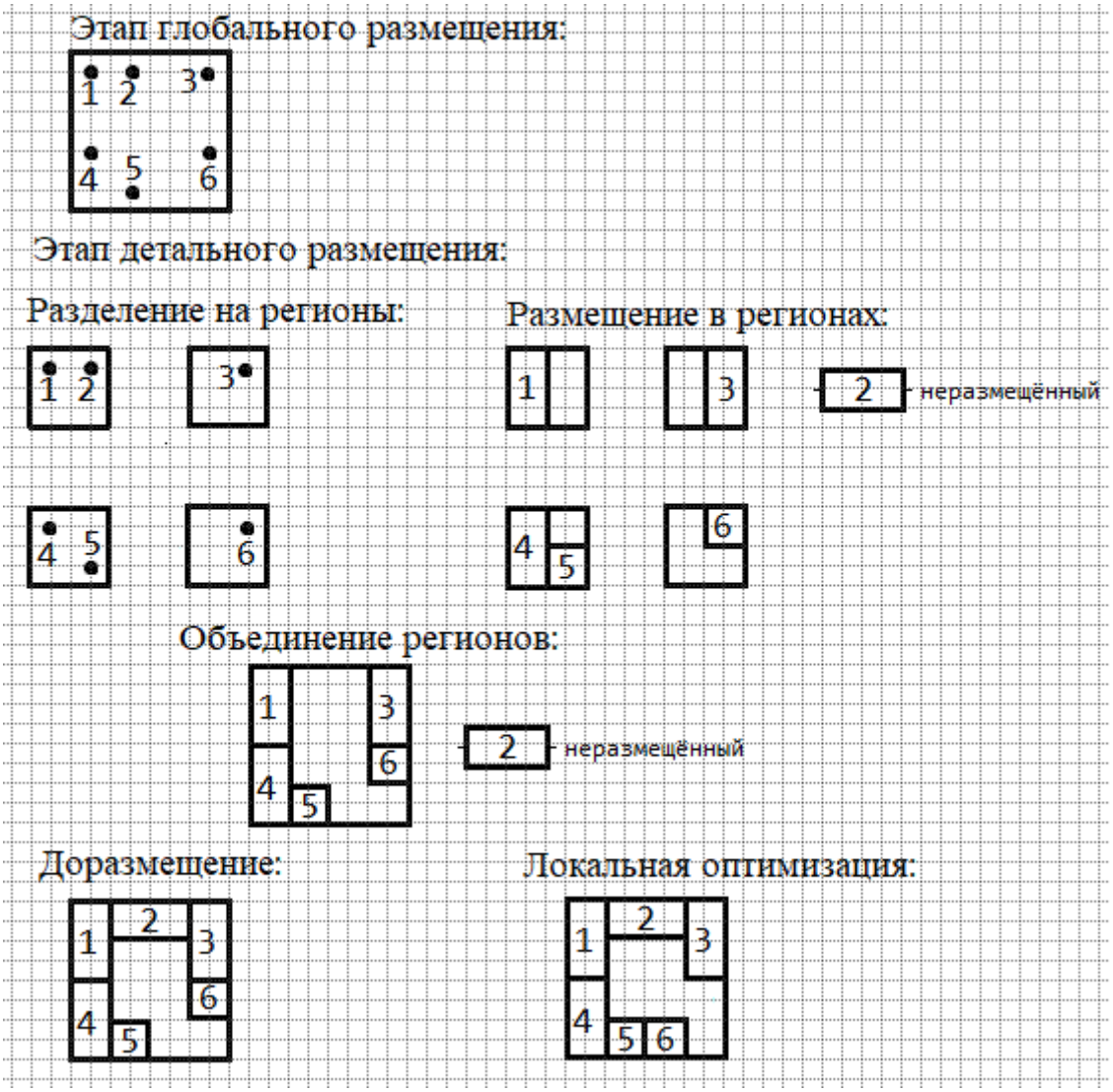


Рис. 2. Этапы работы алгоритма размещения

Так как вычисления в каждом из регионов независимы, гибридный алгоритм допускает возможность распараллеливания, что потенциально может позволить сократить время его работы.

4.3. Радиальный алгоритм детального размещения

Алгоритм осуществляет размещение элементов на кристалле, основываясь на целевых координатах, полученных на этапе глобального размещения. Алгоритм состоит из следующих этапов:

1. Устанавливается радиус отклонения от целевых координат; начальное отклонение равно нулю;
2. На основе целевых координат и текущего радиуса отклонения генерируются все возможные координаты для размещения элемента;
3. Каждая сгенерированная координата последовательно проверяется на допустимость: если текущий элемент можно разместить на кристалле по указанной координате с учетом его размеров, то элемент размещается. Алгоритм переходит на пятый этап. Если подходящей области не найдено, алгоритм переходит к четвертому этапу;
4. Если для текущего элемента не удалось найти место с заданным радиусом отклонения, радиус увеличивается, и алгоритм возвращается ко второму этапу;
5. Если радиус отклонения достигает максимального допустимого значения и подходящих координат для размещения элемента не найдено, элемент остаётся неразмещённым, и за это начисляется штраф;
6. Если все элементы обработаны, алгоритм завершает работу и рассчитывает значения критериев задачи. В противном случае, он переходит к следующему элементу и повторяет цикл, начиная с первого этапа.

4.4. Алгоритм детального размещения Тетрис

Алгоритм Тетрис, который будет описан далее, основан на алгоритме из статьи [6] и его модификации, представленной в [7].

Алгоритм может работать в двух режимах: горизонтальном и вертикальном. В горизонтальном режиме элементы сортируются по возрастанию их x-координаты, при этом первоначальные места-кандидаты определяются по столбцу с нулевой x-координатой. В вертикальном режиме сортировка осуществляется по возрастанию y-координаты, а места-кандидаты задаются по столбцу с нулевой y-координатой. Кроме того, при сортировке можно учитывать размеры элементов (по одной из координат или их общие размеры).

Размещение элементов происходит в порядке, определённом их сортировкой. Ниже описан алгоритм для горизонтального размещения.

1. Устанавливается маркер для отслеживания статуса размещения элемента.
2. Вычисляется расстояния до возможных мест для размещения элемента. После этого места-кандидаты сортируются по возрастанию полученного расстояния.
3. Места-кандидаты проверяются последовательно на возможность размещения в них элемента. Если в прямоугольной области, подходящей под топологию элемента, (от левого верхнего угла (сгенерированные координаты) до правого нижнего угла (сгенерированные координаты + топология элемента)) все места на кристалле свободны и эта область не выходит за пределы кремниевой подложки, то соответствующие ячейки в булевой матрице помечаются как занятые, и элемент считается размещённым на подложке. После этого алгоритм переходит на следующий этап. В случае отсутствия подходящей области, элемент остаётся не размещённым.
4. После завершения попыток размещения элемента, производится пересчет мест-кандидатов для размещения последующих элементов.

4.5. Алгоритм локальной оптимизации

Алгоритм локальной оптимизации пытается еще улучшить полученное решение. Идея алгоритма основана на алгоритме FastDP [8].

Алгоритм состоит из следующих шагов:

1. Определение оптимального региона
    a. Для каждого элемента найти цепи, связанные с элементом;

b. Вычислить ограничивающий прямоугольник, включающий все элементы цепей, найденных на предыдущем шаге, но не включающий текущий элемент;
c. Определить оптимальный регион для текущего элемента как геометрический центр полученного прямоугольника;

2. Поиск свободного места и принятие решение о перемещении элемента
   a. Выбрать позиции оптимального региона, в которые можно пометить текущий элемент с учетом его размеров;
   b. Оценить длину трасс цепей с учетом нового положения элемента. Если длина цепей улучшилась, то элемент занимает место.

5. Заключение

В работе рассмотрена актуальная задача размещения элементов интегральных схем, описан разработанный алгоритм ее решения.

Литература

1. Гэри М., Джонсон Д. Вычислительные машины и труднорешаемые задачи. М.: Мир, 1982. – 416 с.
2. Д.И. Батищев, Н.В. Старостин, А.В. Филимонов. Многоуровневый алгоритм решения задачи компоновки интегральных схем. Системы управления и информационные технологии. - 2007. - Т. 29. - № 3. - С. 48-52.
3. C. Walshaw and M. Cross. Mesh Partitioning: a Multilevel Balancing and Refinement Algorithm. SIAM J. Sci. Comput., 22(1):63–80, 2000. (originally published as Univ. Greenwich Tech. Rep.98/IM/35).
4. Starostin, Nikolay and Klyuev, Ilya. (2014). Iteration Multilevel Method for the Travelling Salesman Problem. Communications in Computer and Information Science. 466. 477-482. 10.1007/978-3-319-11854-3_41.
5. C. Walshaw. A multilevel algorithm for force-directed graph drawing. Journal of Graph Algorithms and Applications, 7(3):253–285, 2003.
6. D. Hill, Method and System for High Speed Detailed Placement of Cells Within an Integrated Circuit Design, U.S. Patent No. 6,370,673. April 9, 2002.
7. A. Khatkhate, C. Li, A. R. Agnihotri, M. C. Yildiz, S. Ono, C. K. Koh, and P. H. Madden, Recursive bisection based mixed block placement, in Proc. ACM Int. Symp. on Physical Design, pp. 84–89, April 2004.
8. M. Pan, N. Viswanathan, and C. Chu, An efficient and effective detailed placement algorithm, in Proc. IEEE/ACM Int. Conf. on Computer-Aided Design, pp. 48–55, November 2005.

ОПТИМИЗАЦИЯ ФУНКЦИЙ УМНОЖЕНИЯ ПЛОТНОЙ ЛЕНТОЧНОЙ МАТРИЦЫ НА ВЕКТОР[1*]

А.А. Воденеева, К. И. Ковалев, Е.А. Козинов, А.Ю. Пирова, А.В. Устинов, В.Д. Волокитин, И.Б. Мееров

Нижегородский государственный университет им. Н.И. Лобачевского

Библиотеки базовых операций линейной алгебры (BLAS) необходимы для работы большого числа научных и инженерных пакетов, решающих задачи численного моделирования. OpenBLAS – одна из широко используемых реализаций стандарта BLAS, оптимизированная под различные архитектуры и типы процессоров. Несмотря на высокие показатели производительности для большинства функций библиотеки, были выделены функции, нуждающиеся в дополнительной оптимизации на процессорах x86. В работе предложен новый алгоритм векторизации функций умножения ленточной матрицы на вектор. Вычислительные эксперименты показали, что предлагаемый алгоритм позволяет в 1,5-2 раза увеличить производительность таких функций для матриц с узкой шириной ленты.

Ключевые слова: BLAS, умножение матрицы на вектор, векторизация.

1. Введение

Во многих вычислительных задачах механики, физики и других научных областях процесс вычислений сводится к многократному решению задач линейной алгебры: решению систем линейных уравнений, SVD-разложению, поиску собственных чисел и собственных векторов и др. В основе всех этих задач лежат базовые операции над векторами и матрицами: умножение матрицы на вектор, умножение двух матриц, скалярное произведение векторов и др. От эффективности их реализации зависит производительность больших пакетов для численного моделирования. В 1979 г. был разработан стандарт BLAS (Basic Linear Algebra Subprograms) – стандартный интерфейс библиотек для базовых операций линейной алгебры [4], а также разработана эталонная реализация refblas. Все функции в стандарте BLAS разделены на три группы: BLAS первого уровня – операции между векторами (сложение векторов, скалярное произведение векторов и др.), BLAS второго уровня – операции между векторами и матрицами (умножение матрицы на вектор и др.), BLAS третьего уровня – операции между матрицами (умножение матриц, решение СЛАУ с треугольной матрицей и др.). В настоящее время разработано большое число реализаций стандарта, адаптированных для различных архитектур: Intel oneAPI MKL, Goto BLAS, Netlib BLAS, ATLAS (процессоры x86), ACML (процессоры AMD), cuBLAS, CLBlast (GPU) и др. Среди открытых библиотек большой популярностью в научном сообществе пользуется библиотека OpenBLAS [2], которая содержит низкоуровневые оптимизированные реализации функций для различных типов процессоров, в том числе, для процессоров Intel, AMD и новых процессоров архитектуры RISC-V [5].

В большинстве библиотек линейной алгебры оптимизация функции под архитектуру подразумевает использование векторных SIMD инструкций. В связи с постоянным развитием аппаратного обеспечения, распространением процессоров новых типов, создание и поддержание высокой производительности библиотек для функций линейной алгебры остается актуальной задачей. Отметим, что сейчас многими научными и индустриальными группами решается задача переноса существующего программного обеспечения на процессоры архитектуры RISC-V, в том числе, библиотеки линейной алгебры OpenBLAS [6, 7].

Конечной целью работы является разработка оптимизированных реализаций для нескольких базовых алгоритмов линейной алгебры для процессоров архитектур x86 и RISC-V. В качестве ба-

[1*] Работа выполнена при поддержке программы академического лидерства ННГУ "Приоритет-2030"

зовой реализации на x86 выбрана библиотека OpenBLAS. В данной работе решается задача оптимизации функций умножения ленточной матрицы на вектор. Матрицы такого типа возникают при численном решении некоторых дифференциальных уравнений в частных производных, например, при решении задачи Дирихле для уравнения Пуассона итерационными методами. В работе предлагается новый алгоритм умножения ленточной матрицы на вектор с использованием SIMD инструкций.

Работа построена следующим образом. В разделе 2 описывается новый векторный алгоритм умножения ленточной матрицы на вектор. В разделе 3 приводятся результаты вычислительных экспериментов. В заключении приводятся выводы и планы дальнейшей работы.

2. Оптимизация алгоритма умножения ленточной матрицы на вектор

На этапе предварительного тестирования OpenBLAS были выбраны три функции для дальнейшей оптимизации:

gbmv – умножение матрицы на вектор для ленточной матрицы общего вида ($y = \alpha \times op(A)x + \beta y$),

sbmv – умножение симметричной ленточной матрицы на вектор ($y = \alpha Ax + \beta y$),

tbmv – умножение треугольной ленточной матрицы на вектор ($x = op(A)x$),

здесь A – матрица, $op(A) = A$ или $op(A) = A^T$, x, y – вектора, α, β – скаляры.

В базовой версии алгоритма произведение $op(A)x$ вычисляется в цикле по строкам матрицы, на каждой итерации которого строка матрицы векторной операцией умножается на фрагмент вектора x (рис. 1, а). В библиотеке OpenBLAS для этого вызываются операции BLAS первого уровня – AXPY (сложение векторов, $y = ax + y$, где x, y – вектора, a – скаляр) или DOT (скалярное произведение векторов, $r = \sum_{i=0}^{n-1} x_i y_i$, где x, y – вектора), которые векторизованы для различных архитектур и наборов векторных инструкций. Однако, в случае малого числа диагоналей (например, для типа double меньше 8 диагоналей) в строке матрицы меньше элементов, чем максимально возможно загрузить в векторный регистр, и для функций AXPY и DOT вызываются скалярные версии. Для возможности использования векторных инструкций на матрицах с малым числом диагоналей было предложено изменить порядок обхода матрицы: вместо обхода по строкам использовать обход по диагоналям, поскольку в диагонали можно выделить блоки элементов, помещающихся в векторный регистр (рис. 1, б). В этом случае векторные операции сложения и умножения выполняются между элементами диагонали и элементами векторов x, y.

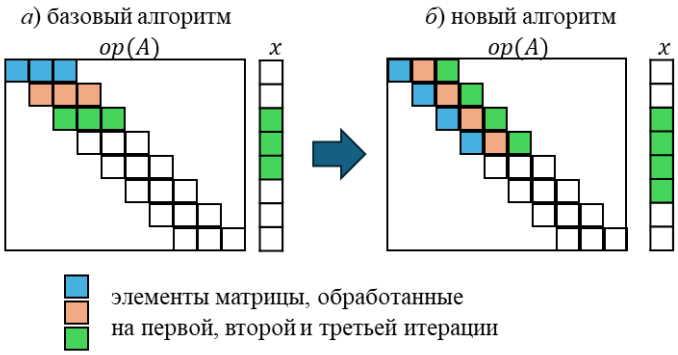


Рис. 1. Способ обхода матрицы для функций gbmv, tbmv, sbmv

3. Вычислительные эксперименты

3.1 Методика тестирования

Тестирование проводилось на узле кластера «Лобачевский» со следующими характеристиками: Intel(R) Xeon(R) Silver 4310T CPU, частота 2.30GHz, память 64 Gb, L2 кэш 1.2 Мб; операционная система CentOS 7.0, компилятор gcc-9.5.0. В качестве базовой реализации использовалась библиотека OpenBLAS v. 0.3.28, для проверки корректности и производительности использовался пакет BLAS-Tester [3]. Для реализации собственных алгоритмов использовался набор

векторных инструкций AVX-512. Результаты для функции tbmv приведены для не транспонированных нижне-треугольных матриц с не единичной главной диагональю.

3.2 Результаты вычислительных экспериментов

На рис. 2 показана производительность функции tbmv в референсной реализации и после оптимизации при разном числе диагоналей. Из графика видно, что при числе диагоналей меньше 8 оптимизированная версия опережает исходную в 1.26–2.11 раз (чем меньше число диагоналей, тем больше выигрыш). При дальнейшем увеличении числа диагоналей исходная реализация незначительно, до 6%, опережает оптимизированную. Это соответствует ожиданиям, поскольку при вызове функции AXPY все ненулевые элементы строки будут загружены в векторные регистры.

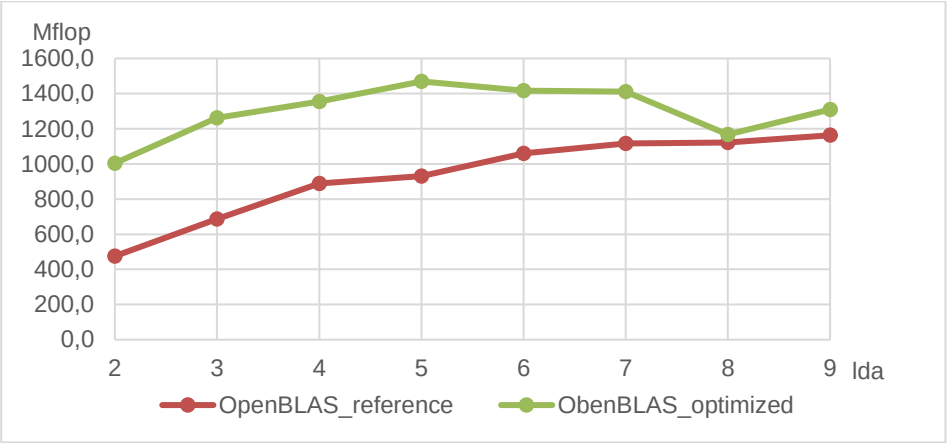


Рис. 2. Производительность функции tbmv до и после оптимизации
в зависимости от числа диагоналей.
Размер матрицы – 1 млн. строк, lda – число диагоналей

На рисунке 3 показана производительность функции tbmv при увеличении размера матрицы и фиксированном числе диагоналей. Из графика видно, что производительность референсной и оптимизированной версий при изменении размера матрицы изменяется незначительно, и, в среднем, оптимизированная версия опережает исходную в 1.6 раз.

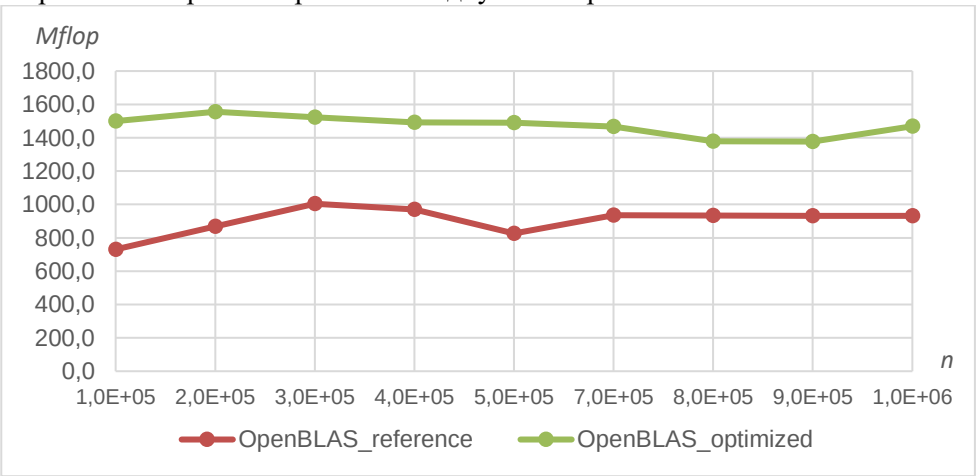


Рис. 3. Производительность функции tbmv до и после оптимизации в зависимости от размера матрицы.
n – размер матрицы, число диагоналей – 5

Заключение

В работе рассматривается задача оптимизации функций умножения ленточной матрицы на вектор. Матрицы такого вида встречаются, в частности, при численном решении уравнения

Пуассона. В качестве базовой реализации была выбрана библиотека с открытым исходным кодом OpenBLAS, широко используемая в научном сообществе. Оптимизация была выполнена с использованием векторных инструкций AVX-512 для процессоров архитектуры x86, и позволила увеличить производительность рассматриваемых функций до 1,5 – 2 раз для матриц с узкой шириной ленты. В дальнейшем предполагается перенос оптимизированных алгоритмов для процессоров архитектуры RISC-V.

В целом, оптимизация базовых функций линейной алгебры является актуальной задачей, поскольку при появлении новых процессоров и вычислительных архитектур требуется поддерживать высокую производительность базовых математических библиотек, повсеместно использующихся в научном сообществе и индустрии.

Литература

1. Intel Math Kernel Library Reference Manual URL: https://www.intel.com/content/www/us/en/developer/tools/oneapi/onemkl.html (дата обращения: 01.09.2024).
2. OpenBLAS. URL: https://github.com/OpenMathLib/OpenBLAS (дата обращения: 01.09.2024).
3. BLAS-Tester. URL: https://github.com/xianyi/BLAS-Tester (дата обращения: 01.09.2024).
4. Lawson C. L., Hanson R. J., Kincaid D. R., Krogh F. Basic linear algebra subprograms for FORTRAN usage. ACM Transactions on Mathematical Software, vol. 5, issue 3, 1979, p. 308-323. DOI: 10.1145/355841.355847.
5. Furber S.B. VLSI RISC architecture and organization. 1st edn. CRC Press. 1989.
6. Зайцева К.А., Пузикова В.В., Соколов А.Д. О проблемах использования библиотеки OpenBLAS в продуктовом коде на RISC-V // Труды Института системного программирования РАН. 2023. Том. 35. № 5. С. 91-106.
7. Fibich C. et al. Evaluation of Open-Source Linear Algebra Libraries targeting ARM and RISC-V Architectures // 2020 15th Conference on Computer Science and Information Systems (FedCSIS). – IEEE, 2020. – С. 663-672.

ОБРАБОТКА ЭКСПЕРИМЕНТОВ СОУДАРЕНИЯ КАПЛИ С ПОДЛОЖКОЙ И БИНАРНЫЙ КЛАССИФИКАТОР РАЗРУШЕНИЯ КАПЛИ

М.Д. Вульф[1], Д.С. Жариков[2], Д.С. Коломенский[1]

[1]Сколковский институт науки и технологии, Центр системного проектирования
[2]Московский авиационный институт, национальный исследовательский университет

В ходе работы были созданы классификаторы для точного определения взаимодействия суспензии и подложки, что важно для капельной 3D-биопечати. Описана установка для нанесения материала, регулирующая скорость и размер капель. Полученные данные экспериментов использовались для построения моделей машинного обучения, прогнозирующих тип соударения капель с подложками. Установка для разбрызгивания капель позволяет контролировать процесс с высокой точностью. Ручная разметка видео и обработка данных позволили построить модели машинного обучения, предсказывающие тип соударения капли и подложки. Важные переменные были выявлены с помощью интерпретации коэффициентов линейной модели и инструментов CatBoost.

Ключевые слова: Генератор капель, столкновение капли, суспензии, капельная 3D-биопечать, машинное обучение

1. Введение

Капельная 3D-биопечать представляет собой одну из наиболее перспективных технологий в биомедицине, позволяя создавать сложные структуры для различных медицинских нужд. Эта технология обеспечивает возможность разработки индивидуализированных моделей для каждого пациента, что существенно увеличивает эффективность и точность лечебных процессов. Хотя взаимодействие капли однофазной жидкости со стенкой хорошо изучено и может быть предсказано с помощью безразмерных чисел, жидкости с крупными частицами не подчиняются этим закономерностям. Для их исследования необходимо провести экспериментальные работы, и целесообразно применить методы машинного обучения, что и является новшеством данной исследования.

Практическая значимость работы заключается в том, что разработанные ML-классификаторы, определяющие тип соударения суспензии с подложкой, могут быть использованы в будущем в капельной 3D-биопечати.

В рамках исследования проведены эксперименты по нанесению суспензий на различные подложки. Видеозаписи соударения капель с подложками были вручную размечены и проанализированы на предмет возможных ошибок в классификации. Полученные данные подверглись анализу для выявления потенциальных взаимосвязей. С помощью различных логических условий был разработан целевой показатель разбрызгивания – Breaking up. На основе этих параметров были созданы модели машинного обучения для прогнозирования типов соударения капель с разнообразными подложками.

2. Описание установки и экспериментов

2.1. Описание установки

В данной работе представлена испытательная установка, которая включает в себя блок генератора капель, подложку, микроконтроллер, источник света, две высокоскоростные камеры – фронтальную и верхнюю, а также настольный ПК, размещенные на раме стендовой испытатель-

ной установки [1]. ПК и микроконтроллер выполняют функции синхронизации генератора капель с камерами и источником света, а также обеспечивают регистрацию температуры и влажности окружающего воздуха. Кроме того, установка позволяет получать капли необходимого размера.

2.2. Описание экспериментов

Для сравнения результатов ударов по различным поверхностям был использован 52,5%-ный водный раствор глицерина с частицами полиэтилена. В зависимости от смачиваемости подложки результаты воздействия капель суспензии на стенки варьировались от растекания до отскакивания. Испытания с одним и тем же субстратом и идентичными частицами применялись для сопоставления результатов воздействия различных жидкостей.

Эти тесты продемонстрировали, что тип результата удара также зависит от вязкости жидкости и поверхностного натяжения, изменяясь от растекания до разбрызгивания. Испытания с тем же субстратом и дистиллированной водой в качестве суспензии позволили сравнить результаты воздействия суспензий с различными размерами частиц полиэтилена. Было установлено, что форма пятна и распределение частиц в пространстве чувствительны к размеру частиц.

3. Разметка данных и моделирование

3.1. Записи экспериментов

В качестве данных для разметки в данной работе были использованы видеоматериалы соударения суспензий и подложек. Всего было просмотрено 395 видео, на 4 из которых отсутствовал момент соударения капель и поверхностей. При анализе использовались не только видеоматериалы, снятые сбоку, но также и фото падения капель. Процесс соударения капли с поверхностью был условно разделен на несколько этапов, каждый из которых представлен ниже.

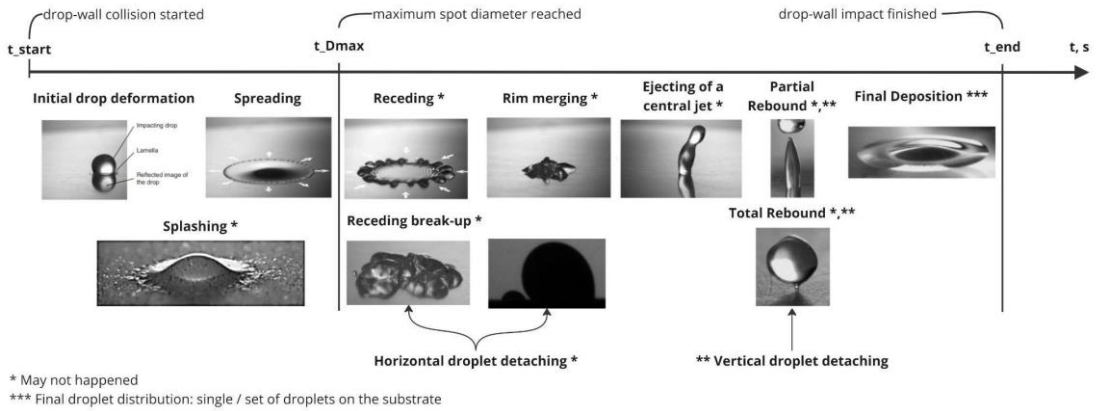


Рис. 1. Схема возможных состояний капли при соприкосновении с подложкой

С помощью различных значений признаков и их комбинаций были составлены различные целевые показатели. В данной работе исследовались случаи, при которых происходило отделение мелких капель в процессе стягивания после соударения с поверхностью. Соотношение бинарных меток в обучающей и тестовой выборке представлено ниже (таблица 1).

Таблица 1. Баланс классов

	Количество “0”	Количество “1”
Обучающая выборка	218	79
Тестовая выборка	55	20

3.2. Моделирование

В процессе моделирования результатов соударения были применены различные модели бинарной классификации, однако для анализа выбраны две из них: логистическая регрессия и градиентный бустинг CatBoost. Линейная модель предоставляет статистически значимую интерпретацию признаков, в то время как CatBoost демонстрирует наивысшее качество по сравнению с другими распространенными классификаторами, такими как SVM, RandomForestClassifier, XGBoost, LogisticRegression и KN-классификатор. Для прогнозирования использовались безразмерные параметры: число Вебера, число Рейнольдса, угол наклона подложки, отношение плотности частиц к плотности жидкости, смачиваемость подложки, нелинейная комбинация чисел Вебера и Рейнольдса, отношение диаметра частиц к диаметру капли, а также шероховатость подложки. Метрики качества представлены ниже (таблица 2).

Таблица 2. Метрики качества

	F1	ROC AUC
Logistic Regression	0.76	0.82
CatBoost	0.81	0.86

Все признаки линейной модели, кроме отношения плотности частиц к плотности жидкости и шероховатости показывают статистическую значимость на уровне значимости 0.05. Для нелинейной модели наиболее значимыми признаками оказались число Рейнольдса, а также смачиваемость (рис. 2).

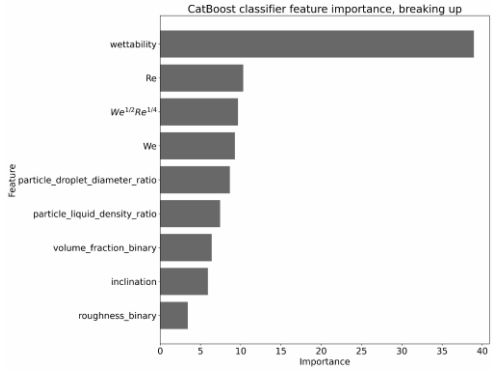


Рис. 2. Значимость признаков модели CatBoost

4. Заключение

В работе была описана установка для разбрызгивания капель, которая позволяет контролировать процесс разбрызгивания капель с достаточно высокой точностью. Также была рассмотрена ручная разметка видео соударения капель с различными подложками и обработка размеченных данных. Полученные результаты и проведенный EDA позволили построить модели машинного обучения, который предсказывают возможный тип соударения капли и подложки с высокой точностью и стабильностью. Были показаны также наиболее важные и значимые переменные в процессе прогнозирования с помощью интерпретации коэффициентов линейной модели и встроенных инструментов библиотеки CatBoost.

Литература

1. A Novel Droplet Generator and a Bench – Testing Rig Enabling Research on Suspension Droplet Wall Impingement and Particle Deposition // MDPI URL: https://www.mdpi.com/2079-6412/13/1/129 (дата обращения 01.04.2024).
2. Теория и практика машинного обучения : учебное пособие / В. В. Воронина, А. В. Михеев, Н. Г. Ярушкина, К. В. Святов., Ульяновск : УлГТУ, 2017. С. 154-167.
3. Машинное обучение часть 1 / А.М.Миронов, М.: Московский Государственный Университет, 2017. 84 с.
4. Практическая статистика для специалистов Data Science: Пер. с англ. / П. Брюс, Э. Брюс. – СПб.: БХВ-Петербург, 2018. – 304 с.
5. Гельман Э., Хилл Дж., Вехтари А. Регрессия: теория и практика. С примерами на R и Stan / пер. с анг. В. С. Яценкова. – М.: ДМК Пресс, 2022. – 748 с.

РЕСУРСНАЯ СИСТЕМА МАССОВОГО ОБСЛУЖИВАНИЯ С УПРАВЛЕНИЕМ ПЕРЕРАСПРЕДЕЛЕНИЕМ РЕСУРСА ДЛЯ АНАЛИЗА НАРЕЗКИ БЕСПРОВОДНОЙ СЕТИ[1*]

И.Е.З. Гебриал[1], К.А. Леонтьева[1], И.А. Кочеткова[1,2]

[1]Российский университет дружбы народов имени Патриса Лумумбы
[2]Федеральный исследовательский центр «Информатика и управление» РАН

В беспроводных сетях пятого и шестого поколения технология нарезки сети используется для создания нескольких виртуальных сегментов, изолированных друг от друга. Под изоляцией при этом понимается не нарушение качества обслуживания пользователей одного сегмента при увеличении объема ресурса другого сегмента. В работе задача перераспределения ресурса между сегментами сети анализируется при помощи управляемой ресурсной системы массового обслуживания и марковского процесса принятия решений. Ресурс между сегментами делится по частоте, а между пользователями в каждом из сегментов – по времени. Управление состоит в выборе нового перераспределения частот при поступлении сигнала контроллера, осуществляющего мониторинг нарезки беспроводной сети.

Ключевые слова: 5G, 6G, нарезка сети, ресурсная управляемая система, марковский процесс принятия решений.

1. Введение

Одной из главных целей сетей пятого поколения (5G) является предоставление высококачественных услуг и обеспечение быстрой передачи данных для пользователей [1]. Для достижения этой цели были разработаны различные технологии, в том числе и нарезка сети [2, 3]. Она позволяет разделить физическую инфраструктуру на несколько виртуальных сегментов, каждый из которых соответствует определенному сценарию использования 5G, такому как расширенная мобильная широкополосная связь, массовая связь между машинами, а также сверхнадежная связь с низкой задержкой.

В связи с важностью обеспечения максимального удовлетворения потребностей пользователей и минимизации затрат возникает задача эффективного распределения ресурсов между этими сегментами [4, 5]. В этом случае возникает проблема перераспределения ресурсов. Для решения этой проблемы используются различные подходы, такие как марковские процессы принятия решений, теория массового обслуживания [6, 7], машинное обучение [4], нейронные сети [8].

В данной работе рассматривается модель перераспределения ресурса между двумя сегментами. Одной из ее главных особенностей является динамическая временная нарезка сети, которая обеспечивает изоляцию между сегментами. Модель построена для работы без очередей, поскольку современные технические системы стремятся минимизировать их наличие и задержки. Это означает, что в таких системах проблема потери и блокировки запросов становится приоритетной. Кроме того, за счет включения механизмов изоляции и блокировки модель обеспечивает соблюдение требований и условий, определенных в соглашении об уровне обслуживания для различных сегментов.

[1*] Работа выполнена в рамках проекта № 025319-2-000 Системы грантовой поддержки научных проектов РУДН.

2. Системная модель

Рассмотрим базового оператора, который является поставщиком ресурса, и двух виртуальных операторов – поставщиков услуг и четырех пользователей: A, B, C и D (рис. 1). Будем считать, что пользователи A и B относятся к первому виртуальному оператору, а пользователи C и D – ко второму. Распределение ресурса в первую очередь осуществляется по частоте, поэтому общий доступный диапазон частот F распределяется между двумя операторами, образуя поддиапазоны F_1 и F_2 соответственно. Кроме того, каждый пользователь каждого оператора занимает определенную долю временного кадра (фрейма), обеспечивая выполнение заранее заданных условий.

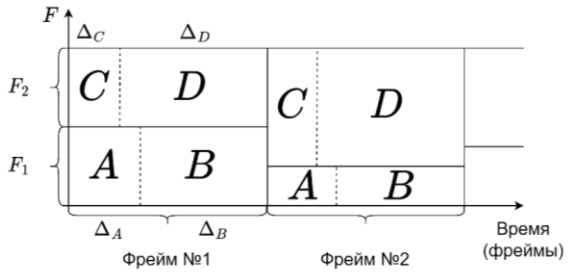


Рис. 1. Системная модель

Общий диапазон частот разделен на равные интервалы длины f_0 и общее число таких интервалов определяется величиной $M = F/f_0$. Планирование перераспределения ресурса осуществляется в рамках одного фрейма, который полностью разделяется между всеми пользователями сегмента, $\sum_{j=1}^{n_i} y_{ij}^t = 1,\ i = 1,2$. Мощность P одинаковая для всех пользователей. Таким образом, требование к ресурсу определяется следующей тройкой параметров:

$$\boldsymbol{y}_{ij}(m, n_1, n_2) = \left(y_{ij}^f = F_i(m), y_{ij}^t = \frac{z_{ij}}{\sum_{k=1}^{n_i} z_{ik}}, y_{ij}^P = P\right), z_{ij} = \left(F_i(m)\log_2\left(1 + \frac{PG_tG_P}{PL_{ij}(N_0 + I_0)}\right)\right)^{-1}$$

где количество частотных интервалов первого сегмента m и количеством сессий на i-м сегменте $n_i,\ i = 1,2$.

Описанная система может быть представлена в виде управляемой ресурсной системы массового обслуживания. В системе имеется два входящих пуассоновских потока запросов с интенсивностями λ_1 и λ_2, соответствующих двум операторам. Средний объем передаваемого эластичного трафика при обслуживании пользователей i-го оператора распределен по экспоненциальному закону с параметром μ_i.

Для описания системы введены два порога скорости передачи данных:

- d_i – минимальная гарантируемая скорость передачи данных, которую необходимо поддерживать для каждого сегмента,
- b_i – скорость передачи данных, при которой сохраняется изоляция между сегментами.

При поступлении сигналов от контроллера происходит проверка необходимости перераспределения ресурса. Интервалы между поступлениями таких сигналов также распределены по экспоненциальному закону с параметром δ.

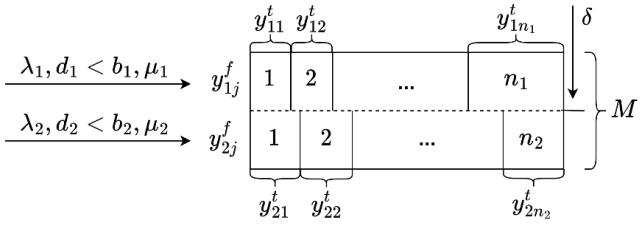


Рис. 2. Управляемая система

3. Управляемая система массового обслуживания

3.1. Марковский процесс принятия решений

Марковский процесс принятия решений состоит из четырех компонент: пространства состояний, множества действий, матрицы интенсивностей и функции вознаграждения $\left(\mathcal{X}, \mathcal{A}(\boldsymbol{x}), \mathbf{Q}_a, R(\boldsymbol{x}'|\boldsymbol{x}, a)\right)$. Управление состоит в выборе нового значения m. При определении первой и второй компонент марковского процесса предполагалось, что все пользователи расположены на минимально возможном расстоянии от базовой станции, а фрейм разделен поровну между всеми пользователями каждого сегмента. Тогда максимальное число пользователей, удовлетворяющих этим условиям:

$$N_i(m) = \frac{F_i(m)}{d_i}\log_2\left(1 + \frac{PG_tG_p}{PL_0(N_0 + I_0)}\right).$$

Пространство состояний выглядит следующим образом:

$$\mathcal{X} = \{\boldsymbol{x} = (m, n_1, n_2)\colon 0 \le n_1 \le N_1(m);\ 0 \le n_2 \le N_2(m);\ 0 \le m < M\}.$$

Максимальная скорость передачи данных определяется формулой:

$$b_i^{\max}(n_i) = M_i(n_i) f_0 \frac{1}{n_i}\log_2\left(1 + \frac{PG_tG_P}{PL_0(N_0 + I_0)}\right),$$

где $M_1(n_1) = \dfrac{b_1 n_1}{f_0 \log_2\left(1 + \frac{PG_tG_p}{PL_0(N_0+I_0)}\right)}$, $\quad M_2(n_2) = M - \dfrac{b_2 n_2}{f_0 \log_2\left(1 + \frac{PG_tG_p}{PL_0(N_0+I_0)}\right)}$.

Множество возможных действий $\mathcal{A} = \{0, \cdots, M\}$, $\mathcal{A}(m, n_1, n_2) \subseteq \mathcal{A}$ запишем в виде системы:

$$\mathcal{A}(m, n_1, n_2) = \begin{cases} \emptyset, & (m, n_1, n_2) \in \mathcal{X}\colon b_1^{max}(n_1) \le b_1, b_2^{max}(n_2) \le b_2 \\ \{m, \cdots, M - M_2(n_2)\}, & (m, n_1, n_2) \in \mathcal{X}\colon b_1^{max}(n_1) \le b_1, b_2^{max}(n_2) > b_2 \\ \{M_1(n_1), \cdots, m\}, & (m, n_1, n_2) \in \mathcal{X}\colon b_1^{max}(n_1) > b_1, b_2^{max}(n_2) \le b_2 \\ \{M_1(n_1), \cdots, M - M_2(n_2)\}, & (m, n_1, n_2) \in \mathcal{X}\colon b_1^{max}(n_1) > b_1, b_2^{max}(n_2) > b_2 \end{cases}$$

Множество вариантов занятия ресурса имеет вид:

$$\mathcal{Y}_0 = \left\{(m, n_1, n_2) \in \mathcal{X}\colon B_{1j}(m, n_1, PL_{1j}) \ge d_1, j = 1, \dots, n_1, B_{2k}(m, n_2, PL_{2k}) \ge d_2, k = 1, \dots, n_2\right\};$$

$$\mathcal{Y}_1 = \left\{(m, n_1, n_2) \in \mathcal{X}\colon B_{1j}(m, n_1, PL_{1j}) \ge b_1, j = 1, \dots, n_1, B_{2k}(m, n_2, PL_{2k}) \ge d_2, k = 1, \dots, n_2\right\};$$

$$\mathcal{Y}_2 = \left\{(m, n_1, n_2) \in \mathcal{X}\colon B_{1j}(m, n_1, PL_{1j}) \ge d_1, j = 1, \dots, n_1, B_{2k}(m, n_2, PL_{2k}) \ge b_2, k = 1, \dots, n_2\right\};$$

где $B_{ij}(m) = F_i(m) y_{ij}^t C_i$, $C_i = \log_2\left(1 + \frac{PG_tG_P}{PL_{ij}(N_0+I_0)}\right)$, $i = 1,2$.

3.2. Интенсивности переходов

Пространство состояний и множество действий в модели описывают оптимальный сценарий с точки зрения максимальной пропускной способности соты. Однако в реальных условиях пользователи распределены на различных, зачастую случайных, расстояниях от базовой станции и испытывают различное затухание сигнала.

Определим интенсивности переходов в системе:

$$q(\boldsymbol{x}'|\boldsymbol{x}, a) = \begin{cases} \lambda_1 \cdot p_1(m, n_1, n_2), & \boldsymbol{x}' = (m, n_1 + 1, n_2), \\ \lambda_2 \cdot p_2(m, n_1, n_2), & \boldsymbol{x}' = (m, n_1, n_2 + 1), \\ \delta \cdot g_a(m, n_1, n_2), & \boldsymbol{x}' = (a, n_1, n_2), \\ n_1\mu_1 E(\xi_{1j}^b), & \boldsymbol{x}' = (m, n_1 - 1, n_2), \boldsymbol{x}\colon n_1 > 0 \\ n_2\mu_2 E(\xi_{2j}^b), & \boldsymbol{x}' = (m, n_1, n_2 - 1), \boldsymbol{x}\colon n_2 > 0 \end{cases}$$

Вероятности того, что при поступлении нового запроса минимальная скорость передачи данных на i-м сегменте не будет нарушена, имеют вид:

$$p_i(m, n_1, n_2) = \frac{\prod_{j=1}^{n_1+1} \Pr\left(\xi_{1j}^b(m) \geq d_1\right) \prod_{j=1}^{n_2} \Pr\left(\xi_{2j}^b(m) \geq d_1\right)}{\Pr\left(\mathcal{Y}_0(m, n_1, n_2)\right)}$$

При поступлении управляющего сигнала возможны три сценария перераспределения ресурса. Первый сценарий: новый объем ресурса первого сегмента меньше текущего, т.е. $a < m$. В таком случае вероятность не нарушения изоляции первого сегмента примет вид:

$$g_a(m > a, n_1, n_2) = \frac{\Pr\left(\xi_{2k}^b(a) \geq d_2, \xi_{1j}^b(m) \geq d_1\right)}{\Pr\left(\mathcal{Y}_0(m, n_1, n_2)\right)}, \qquad j = \overline{1, n_1}, \qquad k = \overline{1, n_2}$$

Второй сценарий: новый объем ресурса первого сегмента больше текущего $a > m$. Тогда важно, чтобы не нарушилась изоляция второго сегмента. Получаем следующую вероятность:

$$g_a(m < a, n_1, n_2) = \frac{\Pr\left(\xi_{1j}^b(a) \geq d_1, \xi_{2k}^b(m) \geq d_2\right)}{\Pr\left(\mathcal{Y}_0(m, n_1, n_2)\right)}, \qquad j = \overline{1, n_1}, \qquad k = \overline{1, n_2}$$

Третий сценарий: новый объем ресурса первого сегмента равен текущему $a = m$. Вероятность имеет вид $g_a(m = a, n_1, n_2) = 1 - g_a(m > a, n_1, n_2) - g_a(m < a, n_1, n_2)$.

4. Заключение

В представленных тезисах построена управляемая ресурсная система массового обслуживания для модели перераспределения радиоресурса при нарезке беспроводной сети 5G. Марковский процесс принятия решений был детализирован с учетом характеристик распределения частот и временных интервалов между сегментами. Он включает в себя описание пространства состояний системы, множества действий, а также интенсивностей переходов. В докладе будет представлен вывод функции распределения скорости передачи трафика и сравнительный анализ разных функций вознаграждения.

Литература

1. Moltchanov D., Sopin E., Begishev V., et al. A tutorial on mathematical modeling of 5G/6G millimeter wave and terahertz cellular systems // IEEE Communications Surveys & Tutorials. 2022. Vol. 24, № 2. P. 1072-1116. DOI: 10.1109/COMST.2022.3156207.
2. Pesando L., Fischer J.K., Shariati B., et al. Standardization of the 5th Generation Fixed Network for enabling End-to-End Network Slicing and Quality-Assured Services // IEEE Communications Standards Magazine. 2022. Vol. 6, № 4. P. 96-103. DOI: 10.1109/MCOMSTD.0002.2100097.
3. Filali A., Nour B., Cherkaoui S., et al. Communication and Computation O-RAN resource slicing for URLLC services using deep reinforcement learning // IEEE Communications Standards Magazine. 2023. Vol. 7, № 1. P. 66-73. DOI: 10.1109/MCOMSTD.0002.2100078.
4. Yarkina N., Gaydamaka A., Moltchanov D., et al. Performance assessment of an ITU-T compliant machine learning enhancements for 5 G RAN network slicing // IEEE Transactions on Mobile Computing. 2024. Vol. 23, № 1. P. 719-736. DOI: 10.1109/TMC.2022.3228286.
5. Kochetkova I., Vlaskina A., Burtseva S., et al. Analyzing the effectiveness of dynamic network slicing procedure in 5G network by queuing and simulation models // Lecture notes in computer science. 2020. P. 71-85. DOI: 10.1007/978-3-030-65726-0_7.
6. Adou Y., Markova E., Gaidamaka Y. Modeling and Analyzing Preemption-Based service prioritization in 5G Networks slicing Framework // Future Internet. 2022. Vol. 14, № 10. P. 299. DOI: 10.3390/fi14100299.
7. Kochetkova I. Leonteva K., Ghebrial I., et al. Controllable Queuing System with Elastic Traffic and Signals for Resource Capacity Planning in 5G Network Slicing // Future Internet. 2023. Vol. 16, № 1. P. 18. DOI: 10.3390/fi16010018.
8. Efrosinin D., Vishnevsky V., Stepanova N. Optimal scheduling in general Multi-Queue system by combining simulation and neural network techniques // Sensors. 2023. Vol. 23, № 12. P. 5479. DOI: 10.3390/s23125479.

КРАТКОСРОЧНЫЙ ПРОГНОЗ СКОРОСТИ ВЕТРА НА ОСНОВЕ ИСКУССТВЕННЫХ НЕЙРОННЫХ СЕТЕЙ С ПРИМЕНЕНИЕМ МЕТОДА ДЕКОМПОЗИЦИИ НА ВАРИАЦИОННЫЕ МОДЫ

И.В. Дель, А.В. Старченко

Национальный исследовательский Томский государственный университет

Шум во входных данных, обусловленный различными внешними факторами, часто снижает точность построенных по ним прогнозов, как следствие, влияет на производительность и качество математической модели. Для точного краткосрочного прогнозирования скорости ветра по измеренным историческим значениям метеорологических параметров за предыдущие часы, представлен гибридный метод, который использует искусственные нейронные сети (ИНС) в сочетании с фильтрацией входного сигнала методом декомпозиции на вариационные моды (Variational Mode Decomposition – VMD). Предлагаемый метод значительно уменьшает ошибку прогнозирования временных рядов ИНС.

Ключевые слова: нейронные сети, временные ряды, VMD, прогноз скорости ветра.

1. Введение

Прогнозирование скорости ветра с использованием нейронных сетей становится все более актуальным в условиях изменения климата и увеличения частоты экстремальных погодных явлений. Кроме того, краткосрочный прогноз локальной скорости ветра чрезвычайно важен для обеспечения безопасной и эффективной работы ветровых электрических станций. Современные методы машинного обучения, включая нейронные сети, способны обрабатывать большие объемы данных и выявлять сложные зависимости, что позволяет значительно повысить точность прогнозов.

Целью работы является реализация и применение гибридного метода для краткосрочного прогнозирования скорости ветра, который использует ИНС в сочетании с предварительной фильтрацией входного сигнала методом декомпозиции на вариационные моды (VMD). По измеренным стационарной метеостанцией историческим значениям скорости ветра за предыдущие часы, разложенных на моды с помощью VMD, необходимо с помощью ИНС предсказать скорость ветра через 1, 3 и 6 часов.

2. Данные и методы исследования

Исходные метеорологические данные для исследования были получены с ультразвуковой метеостанции (УЗМ) Базового экспериментального комплекса (БЭК) ИОА СО РАН за период времени с 01.01.2021 по 31.10.2022.

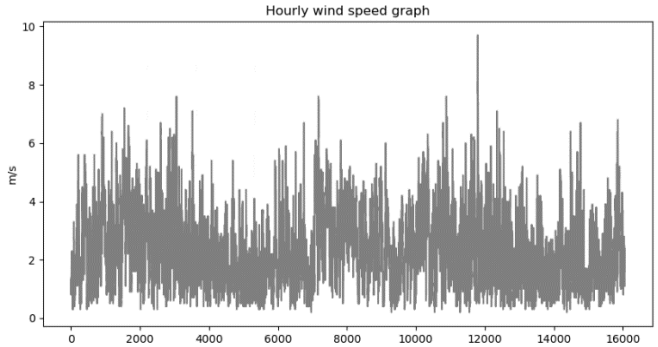


Рис. 1. Изменение скорости ветра в районе УЗМ БЭК с 01.01.2021 по 31.10.2022

Размер набора данных составляет 16056 ежечасных наблюдений. Изменение скорости ветра в районе УЗМ БЭК показано на рис. 1. Максимальная скорость ветра за весь рассматриваемый период составила 9.7 м/с, среднее значение скорости ветра – 2.3 м/с.

Выбранные данные были разделены на три выборки: обучающую выборку, соответствующую 70% данных от основной выборки, валидационную выборку, соответствующую 20% и тестовую выборку, соответствующую 10%, чтобы можно было проверить способность сети к прогнозированию после фазы обучения.

2.1. Метод декомпозиции на вариационные моды

В [1] предложен нерекурсивный метод декомпозиции входного сигнала на вариационные моды (VMD), в котором моды извлекаются одновременно. Метод ищет ансамбль мод и соответствующие им центральные частоты таким образом, чтобы моды после сглаживания в совокупности воспроизводили входной сигнал [1]. В данной работе этот метод применяется к входному набору данных (ежечасно измеренным значениям приземной скорости ветра) перед использованием модели ИНС для прогнозирования скорости ветра.

Метод VMD может разложить любой значимый входной сигнал $y(t)$ на дискретное количество составляющих (k-моды), где требуется, чтобы каждая мода (подсигналы) u_k была в основном компактной вокруг центральной частоты ω_k, определенной в процессе декомпозиции. Каждая мода рассматривается как амплитудно-модулированный и частотно-модулированный (АМ-ЧМ) сигнал.

Ограниченная вариационная задача для нахождения всех мод определяется как [1]:

$$\min_{u_k,\omega_k}\left\{\sum_k\left\|\partial_t\left[\left(\delta(t)+\frac{j}{\pi t}\right)*u_k(t)\right]e^{-j\omega_k t}\right\|_2^2\right\} \text{ при условии } \sum_k u_k = y,$$

где y обозначает сигнал для разложения; $\{u_k\}=\{u_1, u_2, \ldots, u_k\}$ – k-я мода; $\{\omega_k\}=\{\omega_1, \omega_2, \ldots, \omega_k\}$ – центральная частота k-й моды; δ – распределение Дирака; t – время; k – количество мод; * обозначает свертку. Более высокие значения индекса k соответствуют модам с компонентами более низкой частоты.

2.2. Архитектура нейросетевой модели

Архитектура построенной искусственной нейронной сети для краткосрочного прогнозирования скорости ветра состоит из последовательных трех слоев: входного, одного скрытого и выходного полносвязных слоев. Увеличение количества скрытых слоев не привело к улучшению производительности и качества прогнозирования ИНС.

В рассматриваемой нейронной сети оптимальный поиск весовых коэффициентов осуществляется методом адаптивной инерции с коррекцией весовых коэффициентов путем добавления L2-регуляризации AdamW (вариант стохастического градиентного спуска) [2]. В качестве функции потерь использовалась функция Хьюбера [3] – комбинация среднеквадратичной ошибки при малых погрешностях и средней абсолютной ошибки при больших:

$$Huber = \begin{cases} \frac{1}{2}\left(X_{pred} - X_{true}\right)^2, \text{ если } \left|X_{pred} - X_{true}\right| \le \delta; \\ \delta\left|X_{pred} - X_{true}\right| - \frac{1}{2}\delta^2, \text{ если } \left|X_{pred} - X_{true}\right| > \delta, \end{cases}$$

где X_{pred} – предсказанное выходное значение, X_{true} – реальное выходное значение, а $\delta = 2$ м/с [3].

Для оценки качества прогнозирования временных рядов с помощью построенной нейронной сети использовались: средняя абсолютная ошибка *MAE*, коэффициент детерминации R^2 и индекс согласия *IOA* между предсказанными значениями и наблюдениями скорости ветра [4]:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|d_i - y_i\right|, R^2 = 1 - \frac{\sum_{i=1}^{n}\left(d_i - y_i\right)^2}{\sum_{i=1}^{n}\left(d_i - \bar{d}\right)^2}, IOA = 1 - \frac{\sum_{i=1}^{n}\left(d_i - y_i\right)^2}{\sum_{i=1}^{n}\left(\left|y_i - \bar{d}\right| + \left|d_i - \bar{d}\right|\right)^2},$$

где y_i – предсказанное выходное значение, d_i – реальное выходное значение, а $\bar{d}$ – среднее значение, n – количество использованных для сравнения значений.

3. Результаты

В таблице 1 приведены значения рассмотренных метрик для прогноза скорости ветра через 1, 3 и 6 часов.

Таблица 1. Значения метрик с применением метода VMD

	MAE	$\boldsymbol{R^2}$	***IOA***
через 1 час	0.054 м/с	0.996	0.999
через 3 часа	0.079 м/с	0.992	0.998
через 6 часов	0.129 м/с	0.978	0.994

Также приведем значения метрик для прогноза скорости ветра без применения технологии VMD (таблица 2).

Таблица 2. Значения метрик без применения метода VMD

	MAE	$\boldsymbol{R^2}$	***IOA***
через 1 час	0.449 м/с	0.68	0.908
через 3 часа	0.631 м/с	0.449	0.817
через 6 часов	0.777 м/с	0.245	0.688

При сравнении полученных значений метрик, можно сделать вывод, что предлагаемый метод значительно уменьшает ошибку прогнозирования ИНС. Из таблиц также видно, что точность прогнозирования в каждом случае (таблица 1 и таблица 2) снижается с увеличением периода прогнозирования, причем при использовании разработанного гибридного метода прогнозирования, характер снижения гораздо меньше.

Работа выполнена при поддержке Министерства науки и высшего образования РФ (соглашение № 075-02-2024-1437).

Литература

1. Dragomiretskiy K., Zosso D. Variational mode decomposition // IEEE Transactions on Signal Processing. 2014. V. 62 (3). P. 531 – 544.
2. Loshchilov I., Hutter F. Fixing weight decay regularization in adam // arXiv preprint arXiv:1711.05101. 2017.
3. Huber P.J. Robust Estimation of a Location Parameter // Ann. Math. Statist. 1964. V. 35, No. 1. P. 73-101.
4. Дель И.В., Старченко А.В. Использование методов нейросетевого моделирования для прогнозирования качества атмосферного воздуха // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2023. № 65. С. 15–24. DOI: 10.17223/19988605/65/2.

ПЕРСПЕКТИВЫ КЛАССИФИКАЦИИ НОРМАТИВНЫХ ТРЕБОВАНИЙ К ПРЕДПРИНИМАТЕЛЯМ ПОСРЕДСТВОМ ПРИМЕНЕНИЯ КЛАССИЧЕСКИХ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ

А.Г. Дёгтев[1], В.В. Разин[2]

[1]Московский физико-технический институт (национальный исследовательский университет),
[2]Нижегородский государственный университет имени Н.И. Лобачевского

В статье рассматриваются перспективы применения классических методов машинного обучения для классификации нормативных требований к предпринимателям. Актуальность исследования обусловлена необходимостью оптимизации процессов соблюдения законодательства и повышения безопасности при осуществлении предпринимательской деятельности. В ходе работы проведен анализ существующих подходов к классификации нормативных актов, выявлены ключевые проблемы и ограничения традиционных методов. Рассмотрены отдельные классические алгоритмы машинного обучения, способные автоматизировать процесс классификации. Результаты экспериментов показывают, что использование таких методов позволяет значительно повысить скорость и точность обработки информации, что, в свою очередь, способствует более эффективному управлению рисками и соблюдению нормативных требований. Статья подчеркивает важность дальнейших исследований в этой области для разработки и внедрения классической модели машинного обучения в целях информирования предпринимателей о нормативных требованиях.

Ключевые слова: нормативные требования, машинное обучение, бизнес, классификация, юридические тексты, дерево решений, метод k-ближайший соседей, случайный лес, линейный классификатор, machine learning, classification, legal-tech, obligatory requirements.

1. Введение

В настоящее время в Российской Федерации активно ведется цифровизация государственного управления. Сфера государственного регулирования не является исключением. По данным федеральной государственной информационной системы «Реестр обязательных требований» в России действует более 527 тыс. нормативный требований федерального уровня [1].

За несоблюдение нормативных требований предпринимателям грозят предписания, штрафы и административное приостановление деятельности. В этой связи важно обеспечить информирование предпринимателей о нормативных требованиях в отношении их вида деятельности.

Актуальность исследования обусловлена тем, что до 2025 года в России должен быть осуществлен переход к риск-ориентированному подходу, следовательно, проверки будут проводиться только в отношении предпринимателей, по которым имеется информация о возможном нарушении [2]. Плановые проверки проводиться не будут. В этой связи необходимо создать инструмент для ознакомления предпринимателей с исчерпывающим составом обязательных требований.

Цель исследования заключается в анализе существующих методов классификации нормативных требований по видам бизнеса и определение наиболее оптимального для дальнейшей разработки сервиса классификации.

Новизна работы заключается в способе предобработке данных и разработке модели машинного обучения для классификации нормативных требований.

Анализ литературы не позволил найти в свободном доступе статьи, связанные с классификацией нормативных требований, как в России, так и в международной практике. В большинстве случаев исследователи сосредоточены на классификации исковых заявлений в целях прогнозирования результатов судебного разбирательства. В этих целях используются такие подходы, как

линейный классфикатор (Boella G. et al.[3]), наивный байес (Hausladen et al., 2020 [4]), метод k−ближайших соседей (Pudaruth et al. [5]), случайный лес (Chen H. et al. [6]), метод опорных векторов (Sengupta, 2022 [7]), решающее дерево (Thompson [8]).

Сводные данные о результатах обзора литературы представлены в таблице 1.

Таблица 1. Результаты применения моделей машинного обучения для классификации юридических текстов

Название модели	Назначение	Авторы статей	Язык	Значение F1-меры
Метод k−ближайших соседей	использовался для классификации юридических текстов с разбиением документов на равные сегменты и голосованием по большинству	Pudaruth et al., 2018	Английский	83,5%
Случайный лес	интерпретируемые модели машинного обучения применялись для классификации юридических текстов	Chen H. et al, 2022	Английский	83,2%
Линейные классификаторы	были применены для классификации юридических текстов на основе TF-IDF признаков	Boella et al., 2011	Итальянский	76%
Метод опорных векторов	классификация юридических документов по отрасли права	Sengupta, 2022	Хинди	75%
Наивный байес	применяется как один из контролируемых алгоритмов для классификации юридических текстов	Hausladen et al., 2020	Английский	65%
Дерево решений	классификация юридических документов по отрасли права	Thompson, 2001	Английский	64%

Таким образом, по результатам обзора литературы наиболее эффективными подходами являются метод k-ближайших соседей, случайный лес и линейный классификатор.

2. Данные

В качестве данных используются отдельные нормативные требования из базы нормативных требований федеральной государственной информационной системы «Реестр обязательных требований» [1]. Нормативные требования представляют собой короткие юридические тексты, в среднем состоящие из 22 слов. Распределение нормативных требований по количеству слов представлено на рис 1.

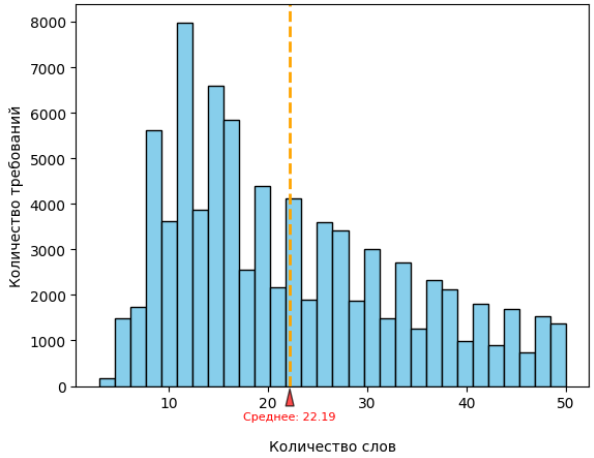


Рис. 1. Распределение нормативных требований по количеству слов

При этом 7% нормативных требований состоят из 100 и более слов. Такие требования, согласно методическим рекомендациям по выделению нормативных требований в нормативных правовых актах, являются некорректными, так как в действительности содержат несколько нормативных требований, которые должны быть разделены. Поэтому для обучения модели эти нормативные требования не будут использоваться.

Количество слов нормативного требования зависит от сферы регулирования, однако слишком длинные нормативные требования затрудняют понимание и толкование нормы. Согласно исследованию, проведенному Национальным исследовательским университетом «Высшая школа экономики», предложения в юридических текстах не должны быть длиннее 50 слов. Больший объем затрудняет толкование норм [9]. Распределение нормативных требований по группам по количеству слов представлено на рис. 2.

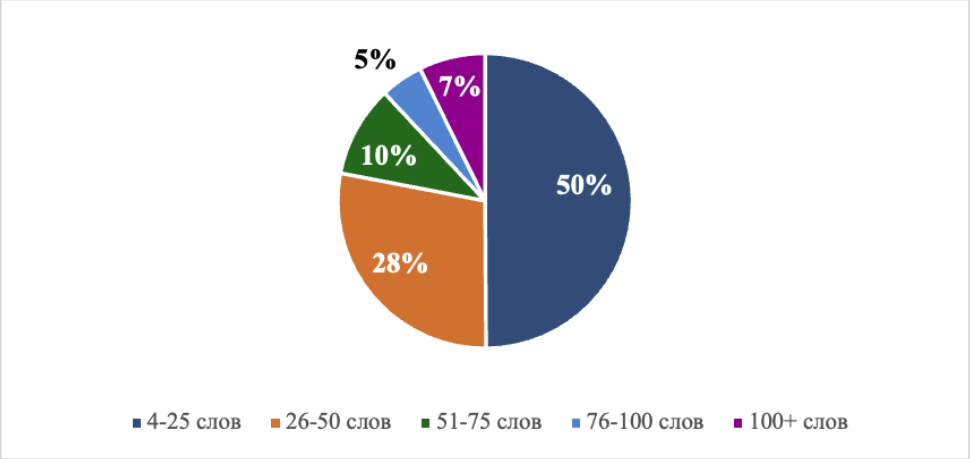


Рис. 2. Распределение нормативных требований по группам по количеству слов

Общий массив нормативных требований состоит из 11 561 483 слов, 84% которых являются существительными. Части речи, составляющие нормативные требования представлены на рис. 3.

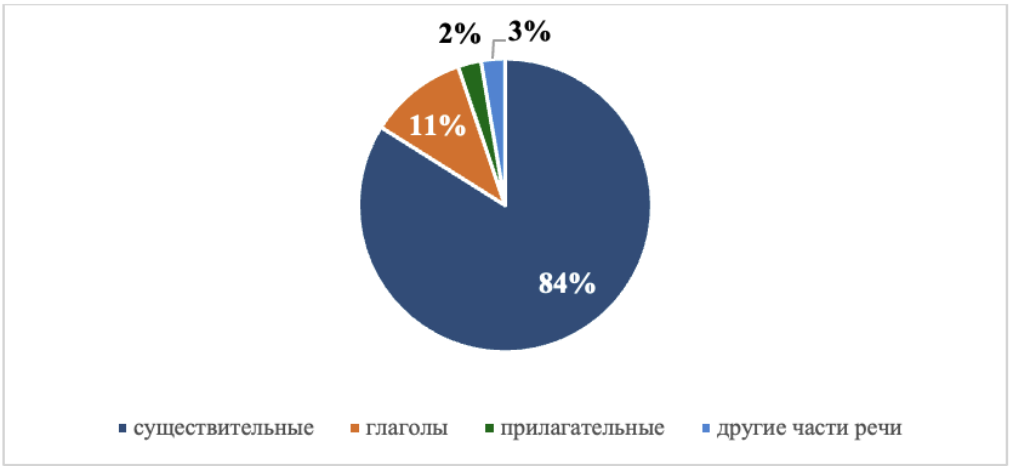


Рис. 3. Части речи, составляющие нормативные требования

Ключевой особенностью нормативного требования является наличие глаголов (категорий состояния) долженствования: должен, обязан, необходимо и т.д. Таким образом, подчеркивается обязанность контролируемого лица соблюдать нормативное требование. В противном случае нормативное требование будет расценено как рекомендация, необязательная для выполнения. Облако слов, построенное на основе базы нормативных требований представлено на рис. 4.

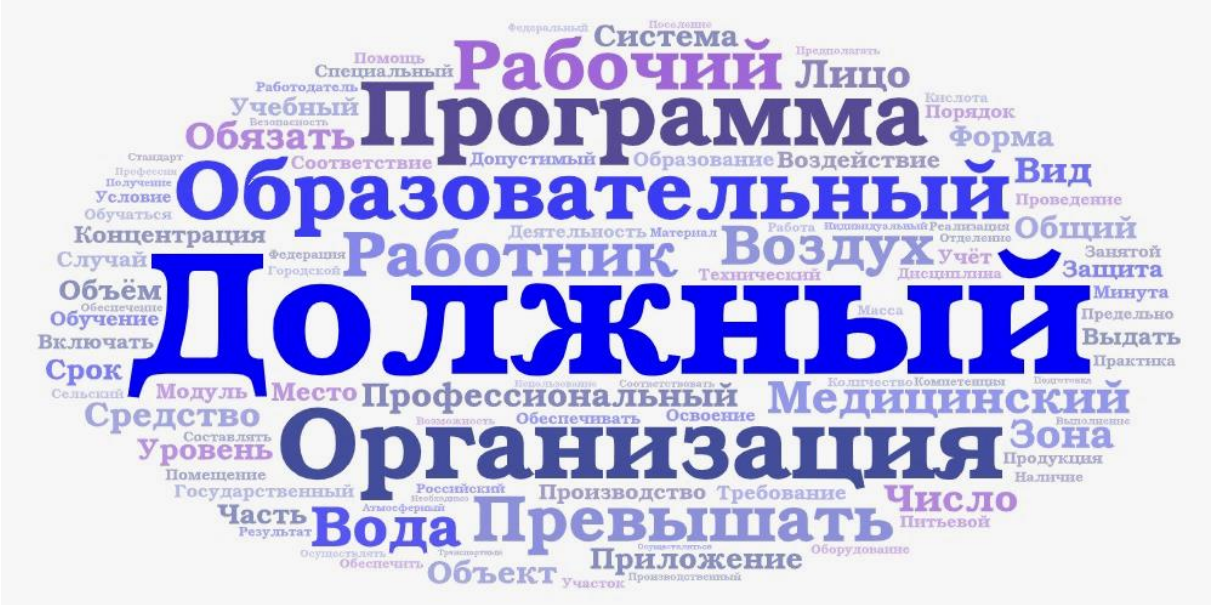


Рис. 4. Облако слов, построенное на основе базы нормативных требований

Всего для обучения модели использовано 1904 требования по 4 категориям бизнеса (кафе, свиноферма, ферма по выращиванию зерновых культур, химчистка).

3. Метрики

В качестве метрики эффективности моделей используется F1-мера – среднее гармоническое значение точности и полноты, что позволяет оценить эффективность модели в контексте правильного прогнозирования положительных объектов. Данная мера отражает долю корректно классифицированных положительных примеров среди всех предсказанных положительных объектов, а также среди всех фактических положительных объектов [10].

В целях определения наиболее эффективного подхода к классификации нормативных требований по F1-мере данные нормативных требований обработаны посредством 6 алгоритмов машинного обучения.

4. Результаты

В целях предобработки текстов использовался метод очистки текста от лишних символов (новых строк, пробелов, точек и точек с запятой), а также токенизация, удаление пунктуации, лемматизация и удаление стоп-слов.

В результате наибольшую эффективность при решении задачи классификации показал метод опорных векторов, метод k-ближайших соседей и случайный лес.

Таблица 2. Результаты апробации классических моделей на наборе данных

Модель	F1-мера
Метод опорных векторов	99,7%
Метод k-ближайших соседей	99,2%
Случайный лес	99,1%
Логистическая регрессия	98,9%
Решающее дерево	98,4%
Гауссовский наивный байесовский классификатор	94,2%

Результаты исследования показали, что метод опорных векторов достиг наивысшего показателя F1-меры, составив 99,7%. Это можно объяснить несколькими факторами. Во-первых, метод опорных векторов эффективно работает с линейно разделяемыми данными, что позволяет

ему создавать четкие границы решений. Во-вторых, метод максимизации марджина (разделительной зоны) между классами делает метод опорных векторов устойчивым к шуму и выбросам, что особенно важно для текстовых данных, содержащих разнообразные формулировки и терминологию. Наконец, метод опорных векторов поддерживает использование различных ядерных функций, что позволяет модели адаптироваться к нелинейным зависимостям, хотя в данном исследовании использовалось линейное ядро.

Метод k-ближайших соседей занял второе место с показателем F1-меры 99,2%. Метод k-ближайших соседей является простым и гибким методом, который адаптивно учитывает расстояния между образцами. Его способность обрабатывать сложные границы решений делает его эффективным для задач классификации текстовых данных. Однако, метод k-ближайших соседей может быть чувствителен к масштабу признаков и требует значительных вычислительных ресурсов при работе с большими наборами данных.

Случайный лес показал F1-меру 99,1%. Этот метод основан на ансамбле решающих деревьев, что позволяет ему снижать риск переобучения и улучшать обобщающую способность модели. Случайный лес также эффективно обрабатывает нелинейные зависимости и устойчив к выбросам, что делает его подходящим для задач классификации текстовых данных.

5. Заключение и выводы

В рамках настоящего исследования были рассмотрены классические методы машинного обучения для классификации нормативных требований к предпринимателям. Целью исследования было оценить эффективность различных алгоритмов в задаче классификации текстовых данных, представляющих собой нормативные требования, установленные законодательством для предпринимателей.

Результаты исследования показывают, что классические методы машинного обучения могут быть эффективно применены для классификации нормативных требований к предпринимателям. Метод опорных векторов и метод k-ближайших соседей продемонстрировали наивысший показатель F1-меры, что делает их предпочтительными для практических применений. Случайный лес также показал высокий показатель F1-меры и устойчивость, что делает его подходящим для задач, требующих высокой обобщающей способности.

Дальнейшие исследования могут быть направлены на улучшение предобработки текстовых данных, использование более сложных моделей, таких как нейронные сети [11-17], и интеграцию дополнительных источников данных для повышения показателя F1-меры классификации [18]. Кроме того, можно рассмотреть применение ансамблевых методов [19, 20], объединяющих несколько моделей, для дальнейшего улучшения результатов.

Эти результаты открывают новые возможности для автоматизации процессов классификации и анализа нормативных требований, что может значительно упростить работу предпринимателей и регуляторных органов.

Литература

1. Federal State Information System 'Register of Mandatory Requirements' [Электронный ресурс] – URL: https://ot.gov.ru/mr (дата обращения 10.09.2024).
2. Подпункт «к» пункта 10 Перечня поручений Президента РФ от 30.03.2024 № Пр-616 http://www.kremlin.ru/acts/assignments/orders/73759/print.
3. Boella G. et al. Conditional labelling for abstract argumentation // Springer Berlin Heidelberg, 2012. – С. 232-248.
4. Hausladen C. I., Schubert M. H., Ash E. Text classification of ideological direction in judicial opinions //International Review of Law and Economics. – 2020. – Т. 62. – С. 105903.
5. Pudaruth S., Soyjaudah K. M. S., Gunputh R. P. An innovative multi-segment strategy for the classification of legal judgments using the k-nearest neighbour classifier //Complex & Intelligent Systems. – 2018. – Т. 4. – С. 1-10.
6. Chen H. et al. A comparative study of automated legal text classification using random forests and deep learning //Information Processing & Management. – 2022. – Т. 59. – №. 2. – С. 102798.

7. Sengupta S., Dave V. Predicting applicable law sections from judicial case reports using legislative text analysis with machine learning //Journal of Computational Social Science. – 2022. – Т. 5. – №. 1. – С. 503-516.
8. Thompson P. Automatic categorization of case law //Proceedings of the 8th international conference on Artificial intelligence and law. – 2001. – С. 70-77.
9. Григорьева Н. Л. и др. Сложность российских законов. Опыт синтаксического анализа // Федеральное государственное автономное образовательное учреждение высшего образования" Национальный исследовательский университет" Высшая школа экономики". – 2020. – С. 312-312.
10. Боженко В. В., Клюканов В. К. Применение алгоритмов машинного обучения в задачах классификации и кластеризации //Обработка, передача и защита информации в компьютерных системах'22. – 2022. – С. 28-33.
11. Ruiz M. E., Srinivasan P. Hierarchical neural networks for text categorization //Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval. – 1999. – С. 281-282.
12. Chalkidis I. et al. LEGAL-BERT: The muppets straight out of law school //arXiv preprint arXiv:2010.02559. – 2020.
13. Leitner E., Rehm G., Moreno-Schneider J. Fine-grained named entity recognition in legal documents // International Conference on Semantic Systems. – Cham: Springer International Publishing, 2019. – С. 272-287.
14. Li G., Wang Z., Ma Y. Combining domain knowledge extraction with graph long short-term memory for learning classification of chinese legal documents // IEEE Access. – 2019. – Т. 7. – С. 616-627.
15. Devlin J. et al. Bert: Pre-training of deep bidirectional transformers for language understanding // arXiv preprint arXiv:1810.04805. – 2018.
16. Tagarelli A., Simeri A. Unsupervised law article mining based on deep pre-trained language representation models with application to the Italian civil code // Artificial Intelligence and Law. – 2022. – Т. 30. – №. 3. – С. 417-473.
17. Nguyen T. S. et al. Recurrent neural network-based models for recognizing requisite and effectuation parts in legal texts // Artificial Intelligence and Law. – 2018. – Т. 26. – С. 169-199.
18. HaCohen-Kerner Y., Miller D., Yigal Y. The influence of preprocessing on text classification using a bag-of-words representation // PloS one. – 2020. – Т. 15. – №. 5. – С. 525.
19. Chen H. et al. A comparative study of automated legal text classification using random forests and deep learning // Information Processing & Management. – 2022. – Т. 59. – №. 2. – С. 798.
20. Rabelo J. et al. COLIEE 2020: methods for legal document retrieval and entailment // Springer International Publishing, 2021. – С. 196-210.

ЧИСЛЕННО-АНАЛИТИЧЕСКИЙ МЕТОД РАСЧЕТА ОСАДКИ СИСТЕМЫ «СООРУЖЕНИЕ-ГРУНТ»[1*]

Н.С. Дюкина, А.В. Домрачева

Нижегородский государственный университет им. Н.И. Лобачевского

Предложен численно-аналитический способ расчета осадки системы «сооружение-грунт» от поля сил тяжести. Метод заключается в задании аналитически определенного поля напряжений для более масштабной по сравнению с сооружением подобласти грунта и последующей корректировке поля напряжений во всей расчетной области с применением имеющейся в программном комплексе «Логос Прочность» процедуры динамической релаксации. Дисбаланс энергии системы на начальном этапе численного расчета будет локализован в небольшой окрестности сооружения. Приводятся результаты численных экспериментов, демонстрирующих эффективность предложенного метода для точного определения осадки системы.

Ключевые слова: осадка, грунтовое основание, численное моделирование, напряженно-деформированное состояние.

1. Введение

В строительной отрасли при проектировании новых сооружений большое значение имеет определение деформаций грунтов под действием внешних сил. Численный расчет заглубленных сооружений на сейсмические воздействия проводится с учетом поля силы тяжести и грунтового основания и требуют предварительного определения осадки системы «сооружение-грунт». Включение инерционного поля силы тяжести в расчет приводит к значительным по времени и амплитуде колебаниям расчетной области, для гашения которых в инженерных программных комплексах используются различные методы. Актуально развитие экономичных методов расчета предварительной осадки, позволяющих повысить скорость и точность расчета.

2. Численно-аналитический метод расчета

Авторы предлагают при численном определении НДС системы «сооружение-грунт» использовать аналитическое решение, полученное для однородного массива грунтового основания в условиях одноосной деформации [1]. При отсутствии напряжений на дневной поверхности грунта напряжения в i-том слое конечных элементов выражаются [2]:

$$\sigma_{zz}^{i} = -\rho_i g\left(L - \sum_{k=i}^{n} h_k - z\right) - \sum_{k=1}^{i-1} \rho_k g h_k, \quad \sigma_{xx}^{i} = \sigma_{yy}^{i} = \sigma_{zz}^{i} \frac{\nu_i}{1-\nu_i}$$

где L – толщина всего грунтового массива, h_k – толщина k-го слоя ячеек, ρ_k – плотность в k-том слое, ρ_0 – плотность на поверхности грунта, E_i, ν_i – упругие постоянные i-го слоя.

В этом случае дисбаланс энергии системы на начальном этапе численного расчета будет определяться наличием сооружения, корректировка НДС потребуется только в небольшой окрестности сооружения.

3. Результаты расчета

Для демонстрации метода рассмотрен однородный грунтовый массив –прямоугольный параллелепипед с основанием 15×15м и высотой 5м – с жестко заделанной нижней гранью и проскальзыванием по вертикали на боковых гранях. Использованы 8-узловые конечные элементы размерами 1×1×1м. Грунт – глина с параметрами ρ = 1500 кг/м^3, ν = 0,32, E = 15 МПа. Массив

[1*] Работа выполнена при поддержке гранта РНФ № 23-29-00103.

грунта находится в поле силы тяжести в условиях одноосной деформации. Расчет осадки проводился в ПП «Логос Прочность» в трех вариантах: с применением и без применения процедуры динамической релаксации, с заданием и без задания аналитически определенного начального поля напряжений. Статическое решение, получаемое итерационно с помощью процедуры динамической релаксации, определяется задаваемой в процедуре погрешностью. Результаты подтвердили предположение о том, что задание начальных условий увеличивает эффективность расчетов с помощью уменьшения начального отклонения, влияющего на вычислительную погрешность. Полученные результаты распространены на случай моделирования расчетной области, включающей грунтовое основание и сооружение. Полученное в результате расчета распределение сжимающих напряжений σ_{zz} в области «сооружение-грунт» показано на рис.1, графики зависимостей величины осадки от времени в трех точках, показанных на рис.1, приведены на рис.2.

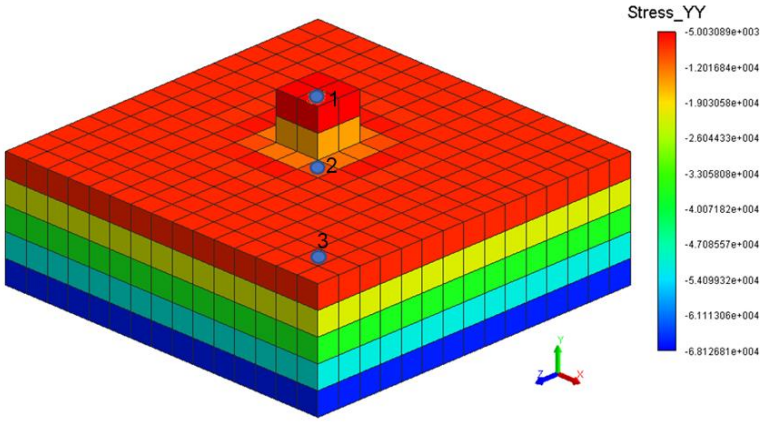


Рис. 1. Поле сжимающих напряжений в расчетной области «сооружение-грунт»

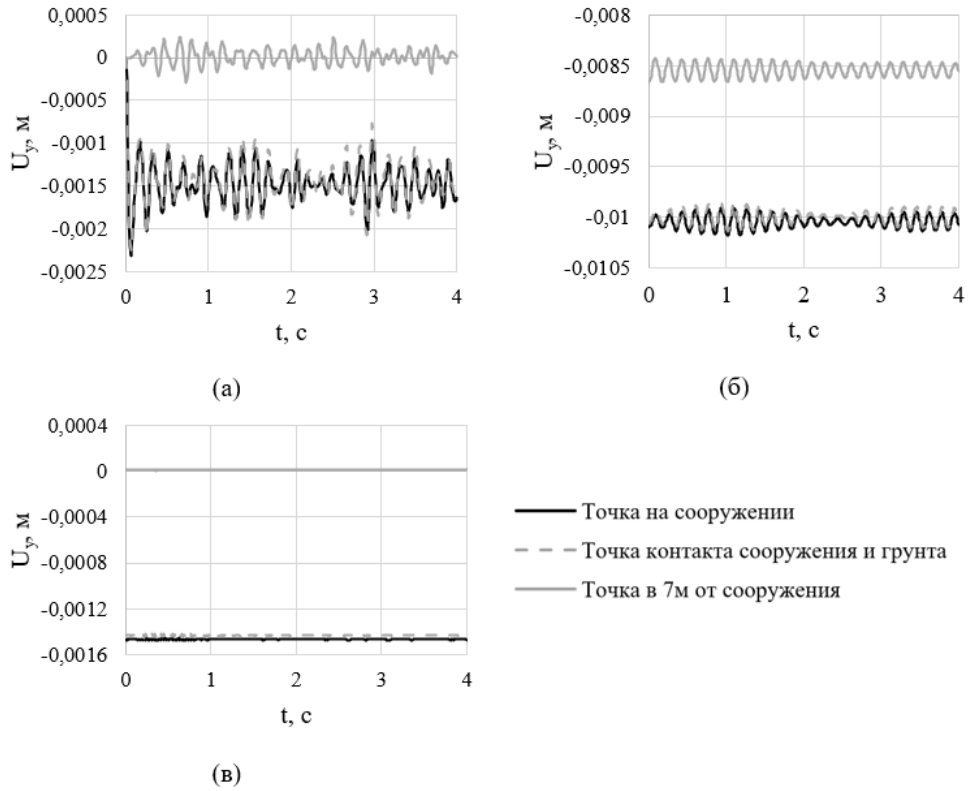


Рис. 2. Зависимости осадки от времени при расчете без начального поля напряжений и процедуры динамической релаксации (а), без начального поля напряжений, с процедурой динамической релаксации (б), с начальным полем напряжений и процедурой динамической релаксации (в)

Как видно из рис.2, задание начального поля напряжений совместно с применением процедуры динамической релаксации приводит к почти полному прекращению колебаний расчетной

области. Применение предложенного метода при решении задач осадки поможет не только сократить время расчетов, но и получить более точные результаты, так как погрешность от колебаний грунта будет снижена в разы.

Литература

1. Ландау Л.Д., Лифшиц Е.М. Теория упругости. М.: Наука, 1987. 248 с.
2. Дюкина Н.С., Кибец А.И., Жестков М.Н. Анализ сейсмических колебаний заглубленных сооружений с учетом трансверсально-изотропного основания // Проблемы прочности и пластичности. 2013. Т. 75. № 1. С. 40-46.

ОЦЕНКА КАЧЕСТВА ПОЛИТИКИ АКТИВАЦИИ КАНАЛОВ, ПОЛУЧЕННОЙ С ПОМОЩЬЮ ОБУЧЕНИЯ С ПОДКРЕПЛЕНИЕМ[1*]

А.А. Живцова, К.Е. Самуйлов

Российский университет дружбы народов им. Патриса Лумумбы

Мобильные сети пятого поколения (5G A.) предусматривают использование многошаговой полудуплексной передачи в сети радиодоступа. В такой конфигурации для обеспечения абонентов требуемыми показателями качества необходимо особое внимание уделять задаче планирования передачи по каналам сети. Формулируя эту задачу с помощью Марковского процесса принятия решений со счетным множеством состояний, можно численно найти оптимальную политику управления с помощью методов обучения с подкреплением. Однако, возникает необходимость оценить погрешность решения, полученного с помощью численных алгоритмов. Для этого в данной работе проводится сравнение целевого показателя численно найденной оптимальной политики с результатами других известных политик.

Ключевые слова: многошаговые беспроводные сети, активация каналов, планирование передачи, обучение с подкреплением, управляемая сеть массового обслуживания, полудуплекс, задержка.

1. Введение

В спецификации мобильной связи 5G Advanced в 2024 году международная стандартизирующая организация 3GPP зафиксировала необходимость поддержки многошаговой беспроводной ретрансляции [1]. Известно, что при использовании миллиметрового диапазона частот в многошаговой сети предпочтительным является полудуплексный режим передачи, ограничивающий одновременную передачу по каналам [2]. Для организации передачи данных в такой сети необходимо в начале каждого временного интервала планирования применять правило (или политику) активации каналов, определяющее каналы, по которым в течение интервала планирования будет осуществляться передача. Предполагается, что в сетях 5G Advanced эта функция может быть реализована с помощью методов машинного обучения [3].

Заметим, что задача организации передачи в многошаговой беспроводной полудуплексной сети исследуется с 90-х годов прошлого века с помощью моделей в виде сетей массового обслуживания в дискретном времени. Полученные в ранних работах результаты в основном предполагали введение некоторой политики активации и аналитическое исследование ее области стабильности (или пропускной способности) [4]. При этом акцент ставился на достижение оптимальности по пропускной способности, снижение вычислительной сложности и уменьшение задержки, однако последнее свойство исследовалось только в асимптотических предположениях [5]. Развитие методов обучения с подкреплением, позволяющих находить политику управления в системах со счетным множеством состояний, позволило использовать известную модель многошаговой беспроводной сети для получения политики активации, оптимальной с точки зрения задержки и пропускной способности [6].

Отметим, что в виду того, что методы обучения с подкреплением основаны на аппроксимации, помимо доказательства возможности получения теоретически оптимальной политики активации, интересна оценка качества полученного решения. Также, ввиду отсутствия явного вида политики, полученной с помощью методов обучения с подкреплением, качество этой политики можно оценить имитационно сравнивая с разработанными ранее политиками. Такому сравнению посвящена данная работа.

[1*] Работа выполнена в рамках проекта № 021937-2-000 Системы грантовой поддержки научных проектов РУДН.

2. Математическая модель

Будем моделировать беспроводную многошаговую сеть в виде сети массового обслуживания в дискретном времени. Для этого предположим, что перед каждым каналом связи сформированы очереди неограниченного объема, в которых находятся пакеты, ожидающие передачи по данному каналу. Пусть в сети задано F фиксированных и уникальных маршрутов пакетов. Пакеты, следующие по одному маршруту, назовем потоком и будем считать, что перед каждым каналом имеются отдельные очереди для всех проходящих по нему потоков. Пусть маршрут f состоит из H_f каналов, пронумерованных от 1 до H_f. Тогда, все очереди в сети можно однозначно определить парой (f, h), в которой f – это поток, а h – это порядковый номер канала в потоке. Всего $N = \sum_{f=1}^{F} H_f$ очередей. На каждом такте t к первому каналу в пути каждого потока f поступает $a_f(t)$ новых пакетов, $a_f(t) \sim Poiss(\lambda_f)$. Обслуживание очереди моделирует передачу пакетов по каналу. Индикатор обслуживания очереди (f, h) на такте t обозначим через $i_{f,h}(t) \in \{0,1\}$. При наличии обслуживания очередь (f, h) покидает группа пакетов размером не более $c_{f,h}$. Обозначим длину очереди (f, h) на такте t через $Q_{f,h}(t)$ и выразим динамику этой величины соотношением (1).

$$\begin{cases} Q_{f,1}(t+1) = Q_{f,1}(t) + a_f(t) - i_{f,1}(t) min\big(Q_{f,1}(t),\, c_{f,1}\big), \;\; f = 1, \dots, F \\ Q_{f,h}(t+1) = Q_{f,h}(t) + i_{f,h-1}(t) min\big(Q_{f,h-1}(t),\, c_{f,h-1}\big) - i_{f,h}(t) min\big(Q_{f,h}(t),\, c_{f,h}\big), \\ h = 2, \dots, H_f, f = 1, \dots, F. \end{cases} \quad (1)$$

Определим Марковский процесс принятия решений следующим образом. Пусть состоянием системы на такте t является вектор длин очередей $\boldsymbol{Q}(t) = (Q_{f,h}(t))$. Тогда пространство состояний имеет вид $\mathcal{S} = \mathbb{Z}_+^N$. Пусть действие на такте t определяется через вектор индикаторов обслуживания $\boldsymbol{i}(t) = (i_{f,h}(t))$. Отметим, что пространство действий $\mathcal{A}$ конечно и состоит из заранее заданного набора векторов $\boldsymbol{i}(t)$. В данном случае будем считать, что $\mathcal{A}$ сформировано таким образом, что ни один узел сети не может использовать более одного канала одновременно. Для минимизации средней задержки в сети определим функцию награды на такте t следующим образом

$$R(t) = -\sum_{f=1}^{F} \sum_{h=1}^{H_f} \big(Q_{f,h}(t) - i_{f,h}(t) \min\big(Q_{f,h}(t),\, c_{f,h}\big)\big)$$

и отметим, что награда (а точнее штраф) на такте имеет значение числа пакетов в системе, оставшихся после проведения обслуживания выбранных очередей. Таким образом, введенная функция награды зависит от состояния системы на такте $\boldsymbol{Q}(t)$ и выбранного на такте действия $\boldsymbol{i}(t)$. Заметим, что объектом поиска является распределение $\pi\big(\boldsymbol{i}(t) \big| \boldsymbol{Q}(t)\big)$, такое что средняя дисконтированная награда максимальна, то есть $\mathrm{M}G = \mathrm{M} \sum_{\tau=0}^{\infty} \gamma^{\tau} R(\tau) \to max$, если $\boldsymbol{i}(t) \sim \pi$.

Распределение π называется политикой и $\gamma \in [0,1)$. Заметим, что через теорему Литтла можно показать, что минимизация средней дисконтированной награды ведет к минимизации средней задержки в сети. Далее для определенности отметим нижним индексом π все величины, полученные при условии, что действие выбирается в соответствии с политикой π, то есть $\boldsymbol{i}(t) \sim \pi$. Для состояния $s \in \mathcal{S}$ введем оценочную функцию состояния $V_\pi(s) = M(G_\pi \mid \boldsymbol{Q}(0) = s)$ равную средней дисконтированной награде по траекториям, начинающимся в состоянии s и продолжающимся в соответствии с политикой π.

3. Решение с помощью обучения с подкреплением

Для поиска политики активации будем использовать метод Proximal Policy Optimization (PPO) [7], являющейся on-policy алгоритмом, основанным на Policy Gradient теореме, позволяющей находить градиент средней дисконтированной награды G_π с помощью градиента политики π, другой заданной политики и advantage функции, выражающей соотношение между политиками. Пусть π^{η_k} и V^{θ_k} – нейронные сети с параметрами η_k и θ_k, аппроксимирующие политику и оценочную функцию состояний соответственно. Параметры нейронных сетей изменяются на каждой итерации k алгоритма обучения.

Предположим, что горизонт планирования конечен и равен T, определим дисконтированную награду с момента t как $G_{\pi^{\eta_k}}(t) = \sum_{\tau=t}^{T} \gamma^{\tau} R(\tau)$ и функцию оценки состояния s как $v_{\pi^{\eta_k}}(s) = M(G_{\pi^{\eta_k}}(t) | \boldsymbol{Q}(t) = s)$. Тогда считаем, что $V^{\theta_k}(s) \approx v_{\pi^{\eta_k}}(s)$. Поэтому функция потерь L_θ для сети V^{θ_k} имеет вид

$$L_\theta = \frac{1}{T} \sum_{t=0}^{T} (V^{\theta_k}(\boldsymbol{Q}_\pi(t)) - G_\pi(t))^2. \quad (2)$$

Функция потерь для сети π^{η_k} основывается на Policy Gradient теореме, поэтому для ее расчета используется две политики π и π^{η_k} (подробнее в алгоритме). С помощью π происходит генерация траекторий и расчет градиента средней дисконтированной награды политики π^{η_k}, на которой и проводится оптимизация. Для дальнейшего изложения сначала введем advantage функцию

$$A^{\theta_k}(t) = R_\pi(t) + \gamma V^{\theta_k}(\boldsymbol{Q}_\pi(t+1)) - V^{\theta_k}(\boldsymbol{Q}_\pi(t)) \quad (3)$$

и функцию среза

$$g_\epsilon\left(A^{\theta_k}(t)\right) = \begin{cases} 1+\epsilon, A^{\theta_k}(t) \geq 0, \\ 1-\epsilon, A^{\theta_k}(t) < 0. \end{cases} \quad (4)$$

Тогда функция потерь сети π^{η_k} имеет вид

$$L_\eta = -\frac{1}{T} \sum_{t=0}^{T} min\left[\frac{\pi^{\eta_k}(\boldsymbol{i}_\pi(t)|\boldsymbol{Q}_\pi(t))}{\pi\,(\boldsymbol{i}_\pi(t)|\boldsymbol{Q}_\pi(t))} A^{\theta_k}(t),\ g_\epsilon\left(A^{\theta_k}(t)\right)\right]. \quad (5)$$

При этом политика π фиксирована.

Зададим гиперпараметры N, M – параметры числа итераций, α_1, α_2 -- скорости обучения, ϵ – параметр функции среза g_ϵ. Представим алгоритм обучения.

```
Для каждого n=0, …, N
    Фиксируем π = π^{η_nM}
    Для каждого m=1, …, M
        k=nM+m
        Генерируем траекторию Q_π(t), i_π(t) и R_π(t) длины T
        Определяем величины G_π(t) ∀ t=0,…T
        Определяем advantage функцию A^{θ_k}(t) (3)
        Считаем функции потерь (2) и (5)
        Делаем шаг градиентного спуска
            η_{k+1} ← η_k − α_1 ∇_η L_η,
            θ_{k+1} ← θ_k − α_2 ∇_θ L_θ
```

Рис. 1. Алгоритм получения политики активации

4. Оценка эффективности полученного решения

Руководствуясь работой [8], мы реализуем сеть, изображенную на рисунке 2 и зададим максимальные объемы передающейся пачки следующим образом $c_{2,1} = c_{4,1} = 19$, $c_{1,2} = c_{3,1} = 23$, $c_{1,1} = c_{2,2} = 34$. Также, базируясь на работе [8], где было проведено сравнение показателя задержки для различных динамических политик активации каналов, для оценки качества решения, полученного с помощью обучения с подкреплением (далее политики π_{ppo}), мы используем гибридную $\alpha\beta$ политику $\pi_{\alpha\beta}$ [9], так как она обеспечивает наилучшую задержку в сети. Отметим, что в работе [6] политика, полученная с помощью обучения с подкреплением (алгоритм DDPG), сравнивается с политикой GreedyMS, которая, как показано в [8], на много уступает выбранной $\alpha\beta$ политике. Определим $\alpha\beta$ политику уравнениями (6)-(10).

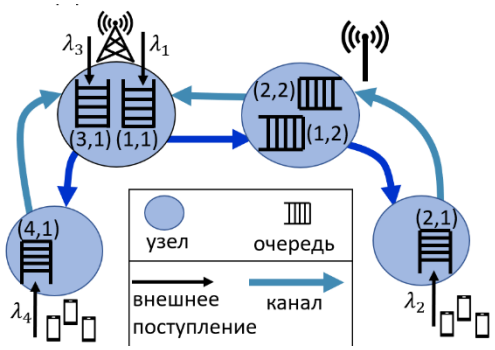


Рис. 2. Топология рассматриваемой сети

$$u_f(t) = \left(\sum_{n=1}^{H_f} (Q_{f,n}(t))^{\alpha+1} \right)^{\frac{\beta-\alpha}{\alpha+1}} \tag{6}$$

$$y_{f,h}(t) = \chi \frac{\sum_{k=h}^{H_f} \frac{1}{c_{f,k}}}{\sum_{k=1}^{H_f} \sum_{m=k}^{H_f} \frac{1}{c_{f,m}}} \tag{7}$$

$$Z_{f,h}(t) = \begin{cases} Q_{f,h}(t), & Q_{f,h}(t) < y_{f,h}(t) \\ \left(Q_{f,h}(t) - y_{f,h}(t)\right)^{\alpha} + y_{f,h}(t), & Q_{f,h}(t) \geq y_{f,h}(t) \end{cases} \tag{8}$$

$$w_{(f,h)}(t) = \begin{cases} Q_{f,H_h}(t) c_{f,H_f} u_f(t), \\ (Q_{f,h}(t) - Q_{f,h+1}(t)) c_{f,h} u_f(t), \quad h = 1 \ \ldots H_f - 1 \end{cases} \tag{9}$$

$$\pi_{\alpha\beta} : \boldsymbol{i}(t) \in argmax\, \boldsymbol{i} \cdot \boldsymbol{w}(t), \quad \boldsymbol{w}(t) = \left(w_{f,h}(t)\right) \tag{10}$$

В качестве π^{η_k} будем использовать полносвязную нейронную сеть, состоящую из 12 слоев с функцией активации гипербалического тангенса на всех слоях кроме последнего (там применяется функция (softmax). Входом в сеть служит состояние системы $\boldsymbol{Q}(t)$ а выходом является распределение действий при условии этого состояния $\pi^{\eta_k}(\cdot\,|s)$. В качестве V^{θ_k} используется полносвязная сеть из 8 слоев с функцией активации LeakyReLU(0.2). Входом в сеть также является состояние системы $\boldsymbol{Q}(t)$ а выходом служит предполагаемая средняя дисконтированная награда по траекториям, начинающимся в этом состоянии. При этом размер скрытых слоев в сетях π^{η_k} и V^{θ_k} равен 512. Произведя подбор гиперпараметров, мы остановились на использовании следующих значений: $T = 50$ – максимальная длина траектории, $N = 30$, $M = 850$ – параметры числа итераций, $\alpha_1 = 10^{-5}, \alpha_2 = 10^{-6}$ – скорости обучения, $\gamma = 0.9$ – коэффициент дисконтирования, $\epsilon = 0.05$ – коэффициент среза функции потерь. Для гибридной $\alpha\beta$ политики $\alpha = 0.5$, $\beta = 10$, $\chi = 50$. Изобразим на рисунках 3 и 4 распределение задержек, полученное при применении политик π_{ppo} и $\pi_{\alpha\beta}$.

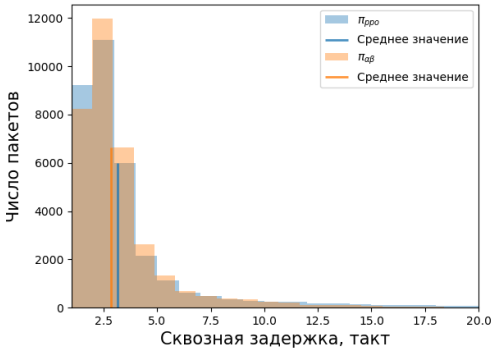


Рис. 3. Распределение задержки

$\lambda_1 = \lambda_3 = 1.34, \lambda_2 = \lambda_4 = 0.33$

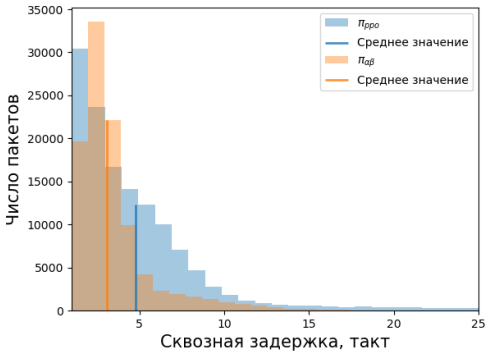


Рис. 4. Распределение задержки

$\lambda_1 = \lambda_3 = 4.02,\ \lambda_2 = \lambda_4 = 1$

5. Заключение

В данной работе была рассмотрена полудуплексная сеть многошаговой беспроводной ретрансляции. Для нее с помощью метода обучения с подкреплением на основе критерия минимизации задержки получена политика активации каналов. Задержки, вызванные применением данной политики, численно сравнены с задержками, полученными применением известной гибридной $\alpha\beta$ политики. Приведенная оценка показала, что при низкой нагрузке системы применение методов обучения с подкреплением позволяет добиться задержки, близкой к задержке известной политики, однако при более высокой нагрузке полученная средняя задержка почти в два раза превышает среднюю задержку известной политики.

Литература

1. 3GPP. Service requirements for the 5G system: Technical Specification (TS) / 3GPP. 06/2024. No. 22.261. Version 19.7.0.
2. 3GPP. Study on Integrated Access and Backhaul : Technical Report (TR) / 3GPP. 12/2018. No. 38.874. Version 16.0.0.
3. 3GPP. Study on 5G System Support for AI/ML-based Services: Technical Specification (TS) / 3GPP. 12/2022. No. 23.700-80. Version 18.0.0.
4. Georgiadis L., Neely M., Tassiulas L. Resource Allocation and Cross-Layer Control in Wireless Networks // Foundations and Trends in Networking. 2006. Jan. Vol. 1. DOI: 10.1561/1300000001.
5. Ji B., Joo C., Shroff N. B. Delay-based Back-Pressure scheduling in multi-hop wireless networks // 2011 Proceedings IEEE INFOCOM. 2011. P. 2579–2587. DOI: 10.1109/INFCOM.2011.5935084.
6. Gupta M., Rao A., Visotsky E., Ghosh A., Andrews J. G. Learning Link Schedules in Self-Backhauled Millimeter Wave Cellular Networks // IEEE Transactions on Wireless Communications. 2020. Vol. 19, no. 12. P. 8024–8038. DOI: 10.1109/TWC.2020.3018955.
7. Schulman J., Wolski F., Dhariwal P., Radford A., Klimov O. Proximal Policy Optimization Algorithms//ArXiv.2017..abs/1707.06347. URL:https://api.semanticscholar.org/CorpusID:28695052.
8. Zhivtsova A., Beschastnyi V., Koucheryavy Y., Samouylov K. A Survey of Delay-Oriented Dynamic Link Scheduling Policies for 5G/6G Integrated Access and Backhaul Systems // IEEE Access. 2024. Vol. 12. P. 118565–118586. – DOI: 10.1109/ACCESS.2024.3446569.
9. Venkataramanan V. J., Lin X., Ying L., Shakkottai S. On Scheduling for Minimizing End-to-End Buffer Usage over Multihop Wireless Networks // 2010 Proceedings IEEE INFOCOM. 2010. P. 1–9. DOI: 10.1109/INFCOM.2010.5462117.

ГЕНЕТИЧЕСКИЙ АЛГОРИТМ ДЛЯ ПОСТРОЕНИЯ РАСПИСАНИЯ НА МНОГОЯДЕРНОМ ПРОЦЕССОРЕ С УЧЕТОМ ВЗАИМНОГО ВЛИЯНИЯ РАБОТ[1*]

Ю.В. Захарова, М.Ю. Сахно

Институт математики им. С.Л. Соболева СО РАН, Омский филиал

В работе рассматривается задача планирования работ на многоядерном процессоре с учетом их взаимного влияния. Особенностью постановки является то, что работы при совместном выполнении замедляют друг друга по некоторому заданному правилу за счет использования общего ресурса – шины данных. Исследуются свойства этой задачи и её допустимых решений, предлагается адаптивный генетический алгоритм для её решения. Особенностью алгоритма является то, что для выбора одного оператора скрещивания из нескольких используется метод обучения с подкреплением. Проводится вычислительный эксперимент, результаты которого сравниваются с результатами известного жадного алгоритма и коммерческого пакета.

Ключевые слова: генетический алгоритм, многоядерный процессор, построение расписаний

1. Введение

Рассматривается задача составления расписания выполнения n работ (программных модулей) на m ядрах процессора при условии, что работы взаимно влияют друг на друга за счет потребления ресурсов (таких как кэш или шина данных). Необходимо минимизировать общее время завершения всех работ. В [1] доказана NP-трудность этой задачи. В настоящей работе предполагается, что взаимное влияние работ друг на друга происходит из-за конкуренции за шину данных [1, 8, 9]. Шина данных – это часть системной шины, которая служит для передачи данных между компонентами компьютера. Также предполагается, что шина данных делится между работами по закону, представленному в [1].

Для задач составления расписаний, возникающих в многопроцессорных компьютерных системах, актуальным направлением является разработка конструктивных жадных алгоритмов для онлайн-постановок и метаэвристик для оффлайн-постановок, среди которых есть класс эволюционных алгоритмов, хорошо зарекомендовавших себя при решении задач составления расписаний с ресурсными ограничениями (см., например, [2, 3]). В настоящей работе предлагается генетический алгоритм для рассматриваемой задачи и проводится вычислительный эксперимент, где результаты генетического алгоритма сравниваются с результатами жадного алгоритма, предложенного ранее для этой же задачи в [1], и результатами коммерческого пакета CPLEX.

2. Генетический алгоритм

Генетический алгоритм имитирует процесс биологической эволюции популяции особей, где в качестве особей выступают пробные решения задачи. Стандартная схема генетического алгоритма предполагает построение начальной популяции, а затем на каждой итерации за счет операторов воспроизведения, скрещивания и мутации на основе текущей популяции строится следующее поколение (новая популяция) [7].

В реализации предлагаемого генетического алгоритма GA для рассматриваемой задачи в качестве операторов воспроизведения (селекции) использовались ранговая и турнирная селекции [4], в качестве операторов скрещивания (кроссинговера) – одноточечный кроссинговер, кроссинговер с частичным отображением, циклический кроссинговер и порядковый кроссинговер [5], а

[1*] Исследование выполнено за счет гранта Российского научного фонда № 22-71-10015.

в качестве операторов мутации – мутации обмена и сдвига. Кроме этого в генетическом алгоритме присутствуют следующие параметры: размер популяции k, количество элит в популяции e, вероятность применения оператора кроссинговера P_c и вероятность применения оператора мутации P_m, для турнирной селекции – размер турнира. В качестве кодировки решений задачи мы используем перестановки длины n, соответствующей числу работ.

Настройка параметров алгоритма выполнялась автоматически с использованием пакета IRACE [6].

Была исследована сходимость предложенного алгоритма и определена оценка среднего числа итераций до первого достижения оптимума для случая, когда поиск направлен на нахождение оптимального расписания среди тех, где ядра максимально загружаются имеющимися работами и не простаивают, если ещё имеются невыполненные работы.

Теорема 1. Генетический алгоритм GA в среднем не более чем за

$$O\left((n-1)\left(\frac{kn(k+1)(n-1)}{2P_m(1-P_c)}\right)^{(n-1)}\right)$$

итераций впервые достигает оптимум задачи построения расписания на многоядерном процессоре с учетом взаимного влияния работ.

Доказательство. Вычислим вероятность того, что из особи ξ за *n-1* итерацию получится особь ξ^*, представляющая оптимальное решение задачи. Вероятность особи быть выбранной в операторе ранговой селекции равна $\frac{2}{k(k+1)}$. Вероятность того, что при мутации сдвига один из генов переместится на позицию, совпадающую с позицией из оптимального решения, равна $P_m \frac{1}{n(n-1)}$, при этом кроссинговер должен отсутствовать. Таким образом, вероятность получить из особи ξ особь ξ^* за *n-1* итерацию равна $\left(\frac{2(1-P_c)P_m}{k(k+1)n(n-1)}\right)^{(n-1)}$, а среднее число итераций до достижения этого события не превосходит $(n-1)\left(\frac{kn(k+1)(n-1)}{2P_m(1-P_c)}\right)^{(n-1)}$. Теорема доказана.

Аналогичные теоремы справедливы и при использовании турнирной селекции и мутации обмена.

3. Вычислительный эксперимент

Генетический алгоритм был протестирован на двух сериях тестовых примеров: для первой серии было сгенерировано 30 примеров с 50 работами и 8 ядрами, для второй серии – 30 примеров со 100 работами и 24 ядрами. В качестве работ были взяты процедуры из библиотеки Intel MKL и процент потребления ими шины данных был измерен методом, описанным в [1]. Отметим, что в первой серии никакие 8 работ не замедляли друг друга, т.к. процент потребления шины данных каждой работой менее 12%.

В таблице 1 представлены результаты сравнения решений, найденных генетическим алгоритмом, с решениями жадного алгоритма, предложенного ранее. Жадный алгоритм выбирает работы для выполнения так, чтобы они как можно меньше замедляли друг друга, при этом ядра не должны простаивать (т.е. каждое ядро должно выполнять по одной работе в каждый момент времени, пока есть работы для выполнения). Во всех примерах генетический алгоритм демонстрирует лучшее качество решений. Среднее преимущество по серии из 50 работ составляет 6.1%, а по серии из 100 работ – 19.44%.

В таблице 2 представлены результаты сравнения решений генетического алгоритма с нижней границей, которая вычисляется как максимум из длины самой длинной работы и суммы длительностей всех работ (при условии, что работа выполняется без замедления), деленная на количество ядер. Для серии с 50 работами алгоритм отклоняется от нижней границы не более чем на 0.3%, в то время как для серии со 100 работами минимальное отклонение составляет 45.03%. Это связано с тем, что при вычислении нижней границы не учитываются возможные замедления работ друг друга.

Таблица 1. Минимальное, среднее и максимальное отклонение решений жадного алгоритма от решений генетического алгоритма

	min	**avg**	**max**
50 работ, 8 ядер	1.97%	6.1%	12.97%
100 работ, 24 ядра	13.77%	19.44%	27.83%

Таблица 2. Минимальное, среднее и максимальное отклонение решений генетического алгоритма от нижней границы

	min	**avg**	**max**
50 работ, 8 ядер	0.13%	0.18%	0.26%
100 работ, 24 ядра	45.03%	53.17%	61.83%

При малом числе работ ($n = 10$) генетический алгоритм также сравнивался с коммерческим пакетом CPLEX и показал сравнимые результаты.

Литература

1. Eremeev A.V., Malakhov A.A., Sakhno M.A., Sosnovskaya M.Y. Multi-core Processor Scheduling with Respect to Data Bus Bandwidth // Proceedings of International Conference Optimization and Applications, OPTIMA 2020. CCIS. 2020. Vol. 1340. Springer, Cham. P. 55-69. DOI: 10.1007/978-3-030-65739-0_5.
2. Slowik A., Kwasnicka H. Evolutionary Algorithms and Their Applications to Engineering Problems // Neural Comput & Applic. 2020. Vol. 32. P. 12363–12379.
3. Gao K., Huang Y., Sadollah A. A Review of Energy-Efficient Scheduling in Intelligent Pro-duction Systems // Complex Intell. Syst. 2020. Vol. 6. P. 237-249.
4. Gen M., Lin L. Genetic Algorithms and Their Applications // Springer Handbook of Engi-neering Statistics. 2023. DOI: 10.1007/978-1-4471-7503-2_33.
5. Kellegoz T., Toklu B., Wilson J. Comparing Efficiencies of Genetic Crossover Operators for One Machine Total Weighted Tardiness Problem // Applied Mathematics and Computation. 2008. Vol. 199. № 2. P. 590-598. DOI: 10.1016/j.amc.2007.10.013.
6. Lopez-Ibanez M., Dubois-Lacoste J., Perez Caceres L., Birattari M., Stutzle T. The Irace Package: Iterated Racing for Automatic Algorithm Configuration // Operations Research Per-spectives. 2016. Vol. 3. P. 43-58. DOI: 10.1016/j.orp.2016.09.002.
7. Reeves C. Genetic Algorithms for the Operations Researcher // INFORMS Journal on Computing.1997. Vol. 9. № 3. P. 231-250.
8. Merkel A., Stoess J., Bellosa F. Resource-Conscious Scheduling for Energy Efficiency on Multi-core Processors // Proc. 5th European Conference on Computer Systems, EuroSys'10. 2010. P. 153–166. DOI: 10.1145/1755913.1755930.
9. Zhuravlev S., Blagodurov S., Fedorova A. Addressing Shared Resource Contention in Multi-core Processors via Scheduling // Proc. 15th International Conference on Architectural Sup-port for Programming Languages and Operating Systems, ASPLOS'10. 2010. P. 129-142. DOI: 10.1145/1735970.1736036.

МОДЕЛИРОВАНИЕ АЛГОРИТМА ДВИЖЕНИЯ РОБОТА В УСЛОВИЯХ СКЛАДСКИХ ПОМЕЩЕНИЙ

Д.А. Збруев, Н.А. Борисов

Нижегородский государственный университет им. Н.И. Лобачевского

Описан комплекс управления складом, состоящий из компьютера, управляющего складом, и робота-погрузчика. Задача комплекса – обеспечить движение погрузчика из текущего положения к ячейкам с необходимыми товарами по кратчайшему маршруту при условии возможного наличия препятствий на складе. Кроме того, робот должен в автоматическом режиме брать товар из выбранных ячеек

Ключевые слова: роботизация склада, робот-погрузчик, поиск кратчайшего пути.

1. Введение

В настоящее время идет активная автоматизация больших складов, позволяющая сократить время на выполнение различных операций и исключить человеческий фактор. В связи с этим происходит активное внедрение беспилотных погрузчиков, в частности подобные аппараты выпускаются фирмами Rocla и TOYOTA [1].

Однако у этих аппаратов есть ряд серьезных недостатков, главный из которых - ограниченное число маршрутов (не более 3-х), что не позволяет полноценно использовать данные аппараты на больших складах. Также эти погрузчики могут брать не более одного ящика за поездку, что при интенсивном товарообороте может приводить к очень большим задержкам. Кроме того, учитывая курс на импортозамещение, создание комплекса, не использующего иностранные средства связи (а в иностранные погрузчики используют GPS), является очень востребованным.

В связи с этим было принято решение создать принципиально новую действующую модель автоматизированного комплекса управления складом, предназначенного для перевозки грузов, состоящего из системы управления складом и робота-погрузчика, способного передвигаться по неограниченному числу маршрутов и брать до 2-х ящиков за поездку, чтобы данный комплекс можно было применять, в том числе на больших складах. Комплекс должен использовать российское программное обеспечение, а также предоставлять возможность специальному роботу-погрузчику перемещаться в любую точку склада.

2. Разработка и исследование программного комплекса

2.1. Описание схемы взаимодействия узлов комплекса

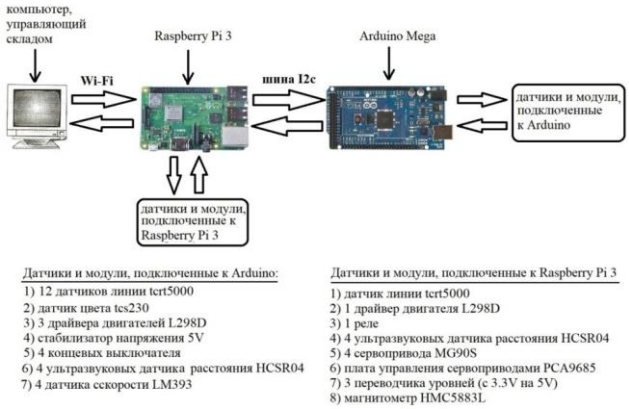


Рис. 1. Схема взаимодействия узлов комплекса управления складом

Оператор склада, используя специальное приложение, написанное на python, выбирает товары (по артикулам), которые необходимо взять со склада. Далее программа строит кратчайший маршрут из текущего положения робота к ячейкам с необходимыми товарами, после чего маршрут преобразуется в массив команд для робота, который передается по Wi-Fi на микроконтроллер Raspberry Pi 3, установленный на робота. На Raspberry Pi 3 производится проверка принятого массива, после чего по Wi-Fi отправляется сообщение об успешном приеме, далее производится передача массива команд на микроконтроллер Arduino Mega. Управление роботом передается Arduino Mega, и робот начинает движение по складу.

Необходимо отметить, что 2 микроконтроллера на роботе-погрузчике выполняют разные задачи. Raspberry Pi 3 обеспечивает связь по Wi-Fi с компьютером, управляющим складом, а также осуществляет общее управление роботом и полностью «принимает управление» при возникновении нештатных ситуаций (например, появление препятствия на пути движения робота), так как для объезда препятствия необходима большая вычислительная мощность, чем та, которой располагает Arduino Mega. К Arduino Mega при этом подключена большая часть датчиков, плат управления моторами и концевых выключателей (благо большое количество пинов позволяет это сделать). Главная же задача Arduino Mega - управлять перемещением робота при движении по построенному маршруту.

После того, как массив команд был принят Arduino Mega, робот начинает движение под управлением этого микроконтроллера. Если в процессе движения робот не встретил препятствий, то после окончания поездки с Arduino на Raspberry Pi 3 будет отправлено сообщение о том, что поездка прошла успешно (это сообщение будет ретранслировано на компьютер, управляющий складом). Если же робот встретил препятствие, робот отправляет сообщение о том, что на участке маршрута между перекрестками с номерами n1 и n2 робот встретил препятствие. Далее Raspberry Pi 3 принимает управление, и робот пытается объехать препятствие. Если препятствие удается объехать, то Raspberry Pi 3 запоминает участок маршрута, где находится препятствие, и дает команду Arduino на продолжение движения по ранее переданному маршруту.

После успешного завершения поездки на компьютер, управляющий складом, будет отправлено сообщение о том, что между перекрестками с номерами n1 и n2 находится препятствие, но робот сумел его объехать. Если же объехать препятствие не получилось, отправляется сообщение о том, что между перекрестками с номерами n1 и n2 находится препятствие и робот не может его объехать. Соответствующее сообщение отображается на экране компьютера, управляющего складом. Далее соответствующее ребро будет удалено из графа, описывающего склад, после чего автоматически будет произведен перерасчет маршрута. Если к указанным ячейкам можно подъехать, то программа преобразует новый маршрут в массив команд, после чего отправит роботу по Wi-Fi, о чем будет сообщено оператору склада.

2.2. Описание принципа движения робота по складу

Для того, чтобы робот мог перемещаться по складу к нужным ячейкам, был применен следующий принцип: на полу склада нанесены линии белого цвета, вдоль которых едет робот, используя специальные датчики, а именно датчики линии, распознающие белый и чёрный цвета (на самом деле, цвет линии можно сделать любым, но для упрощения восприятия и отладки были выбраны именно такие цвета).

В местах пересечения линий (далее будем называть их перекрёстками) нанесены специальные метки. В днище погрузчика установлен датчик, который считывает эти метки при въезде на перекрёсток.

Таким способом робот определяет свое местоположение на складе. Маршрут представляет собой набор определённых команд. Рассмотрим это более подробно на примере:

На рис. 2 изображен простейший склад (вид сверху). Допустим, погрузчику необходимо из исходного положения добраться до зачеркнутой крестом ячейки, взять оттуда товар и вернуться в исходную точку. Для удобства все перекрестки пронумерованы.

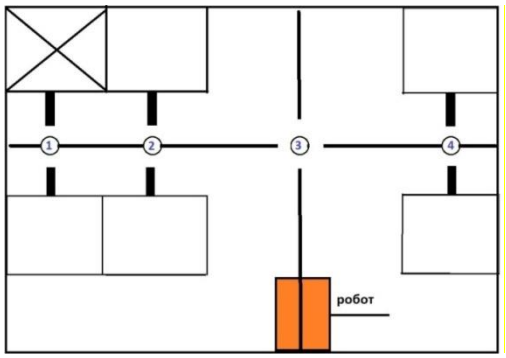


Рис. 2. Складское помещение (вид сверху)

В данном примере маршрут будет выглядеть следующим образом:

1. начать движение вперед,
2. на следующем перекрестке (обозначен цифрой 3) повернуть влево, продолжить движение вперед,
3. через один перекресток (обозначен цифрой 1, перекресток 2 робот должен просто проехать) остановиться, повернуть вправо и взять товар с n-го яруса, где n={1, 2},
4. повернуть вправо, начать движение вперед,
5. через один перекресток (обозначен цифрой 3) повернуть вправо, продолжить движение вперед,
6. при достижении исходной точки остановиться.

Если представить этот склад в виде связного графа (рис. 3), где вершинами графа являются перекрестки, а ребрами - дороги (на рисунке отмечены жирными черными линиями), то склад будет иметь вид:

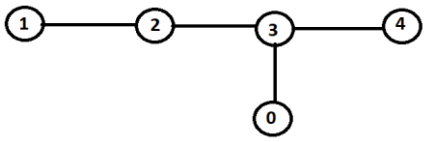


Рис. 3. Связный граф, описывающий складское помещение на рис. 2

Соответственно, маршрут можно записать так:

0, 3, 2, 1, 2, 3, 0, где 0 - исходное положение робота.

Для построения наиболее оптимального маршрута используем алгоритм A-star [3]. Данный алгоритм выбран, так как он обладает свойством оптимальности, т.е. он гарантирует получение лучшего из возможных результатов (этим свойством не обладают “жадные” алгоритмы). Кроме того, алгоритм обладает высоким быстродействием.

2.3. Описание эвристической функции оценки расстояния

Для работы алгоритма A-star необходимо иметь возможность рассчитать значение F для любой вершины. Расчет производится по формуле:

F = G + L

где G (в общем случае) - вес пути от стартовой вершины до рассматриваемой вершины, L - оценка веса пути от рассматриваемой вершины до конечной.

$$L = |x1 - x2| + |y1 - y2|$$

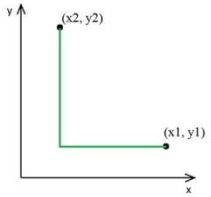


Рис. 4. Формула расчета манхэттенского расстояния

Поскольку все перекрестки на складах имеют прямые углы, то для расчета L используется манхэттенское расстояние, которое рассчитывается по формуле (рис. 4).

2.4. Анализ эффективности программы при различных размерах смоделированного склада

После реализации алгоритма A-star были проведены эксперименты, в ходе которых проверялась правильность работы алгоритма, и измерялось время работы алгоритма в секундах при различных размерах склада. В таблице 1 представлены результаты этих экспериментов.

Таблица 1. Результаты экспериментов

Геометрические размеры склада в ячейках	Количество ячеек на складе	Среднее время построения маршрута, с	Геометрические размеры склада в ячейках	Количество ячеек на складе	Среднее время построения маршрута, с
40x10	400	0.0	300x250	75000	3.075
70x40	2800	0.03125	400x300	120000	10.1
100x60	6000	0.1	500x400	200000	21.7617
140x90	12600	0.2	600x500	300000	39.125
170x120	20400	0.4	700x600	420000	60.25
210x150	31500	0.653125	1000x800	800000	118.5
250x200	50000	1.5			

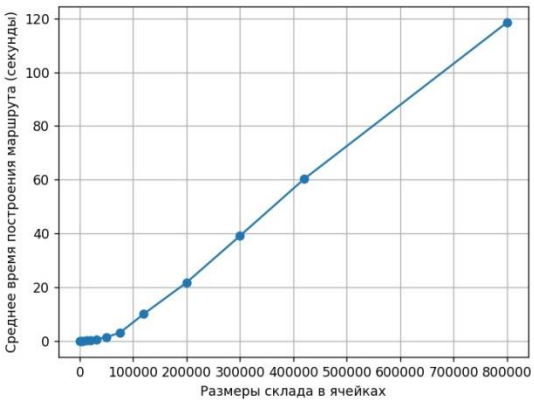


Рис. 5. График зависимости максимального времени построения маршрута от количества ячеек на складе

В ходе экспериментов было установлено, что алгоритм A-star строит оптимальный маршрут для движения робота по складу, причем, если размер склада меньше 40000 ячеек, то маршрут будет построен в среднем менее чем за 1 секунду. Но на больших складах (больше 100000 ячеек) время построения маршрута превышает 5 секунд. Для ускорения процесса необходимо написать алгоритм A-star на языке C++ (в данный момент вся программа написана на python). Также необходимо написать эвристическую функцию оценки расстояния на языке C++. Кроме того, можно создать двумерный массив расстояний между различными перекрестками. Этот массив будет автоматически рассчитываться при создании склада, храниться в базе данных и при запуске программы считываться из базы данных. Эти оптимизации будут реализованы в следующей версии программы.

2.5. Построение массива команд

Итак, при помощи алгоритма A-star маршрут между двумя перекрёстками построен. Напомню, что маршрут - это массив номеров перекрёстков, через которые необходимо проехать

роботу, чтобы добраться до пункта назначения. Но робот, перемещаясь по складу, не может считывать номера перекрёстков, так как система считывания номеров перекрёстков сильно усложнила бы систему и сделала её менее надёжной. Следовательно, необходимо превратить маршрут (то есть массив номеров перекрёстков) в массив инструкций, понятных роботу. Одна инструкция для робота состоит из двух чисел:

1. Номер команды, которую должен выполнить робот.
2. Число перекрёстков, через которое эта команда должна быть выполнена. Далее этот параметр будет называться «счетчик».

Такая система позволяет отвязаться от номеров перекрёстков, так как робот не может узнать номер конкретного перекрёстка, но может с точностью 100% посчитать их количество на своем пути.

2.6. Объезд препятствия

После обнаружения препятствия первая задача робота – оценить ширину препятствия и ширину коридоров, находящихся справа и слева от препятствия [2]. Для этого робот сканирует пространство обеими передними лидарами, при этом измеряется расстояние до объектов при текущем угле поворота. Далее производится оценка ширины препятствия и двух коридоров. Рассмотрим этот процесс на примере оценки ширины правого коридора (рис. 6).

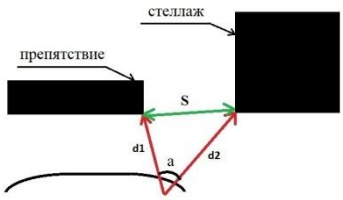


Рис. 6. Условные обозначения для расчета ширины проезда

Для нахождения искомой ширины S необходимо воспользоваться теоремой косинусов:

$$S^2 = {d_1}^2 + {d_2}^2 - 2 \cdot d_1 \cdot d_2 \cdot \cos(a)$$

Необходимо пояснить, что при ширине препятствия больше 22 см точность измерения ширины препятствия и коридоров с помощью лидаров существенно падает. Это связано с особенностями самих лидаров: по сути, это сервопривод, на котором установлен датчик HCSR04 (рис. 7) [4].



Рис. 7. Лидар

В дальнейшем планируется установка более совершенных лидаров. Поэтому, если ширина препятствия больше 22 см, используется другой метод оценки ширины коридоров. Идея этого метода заключается в том, что измерение ширины препятствия и коридоров проводится не с помощью лидаров, а с помощью датчиков скорости вращения колёс, которые обладают очень высокой точностью при измерении расстояния. Реализован этот процесс следующим образом: после проведения измерения ширины препятствия при помощи лидаров, если полученное значение больше 22 см, то робот поворачивает на 90° вправо, далее производится измерение ширины правого коридора.

Общая блок-схема алгоритма объезда препятствий изображена на рис. 8.

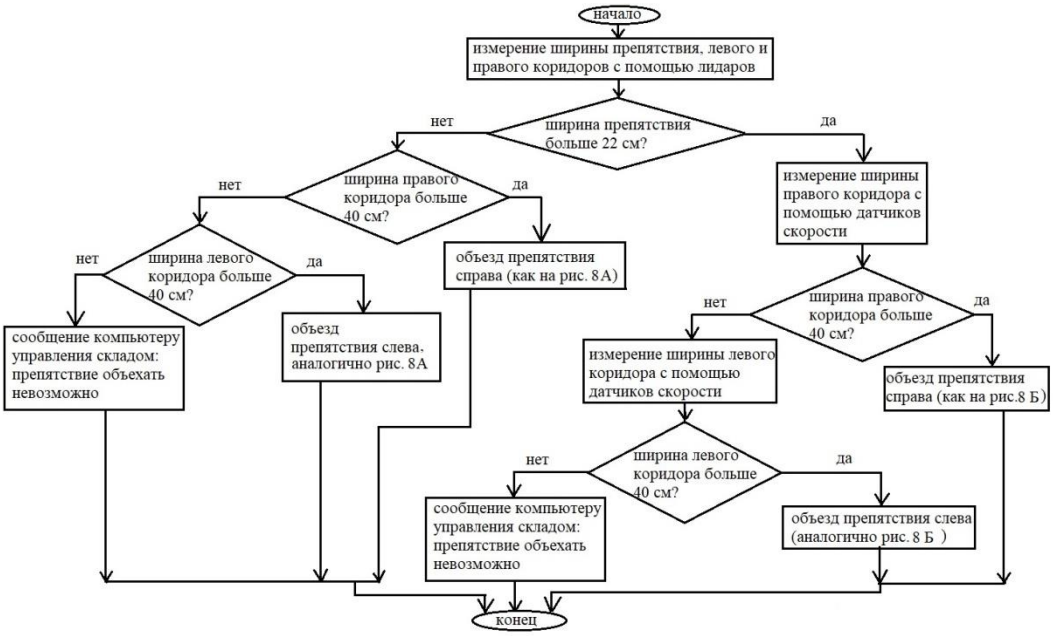


Рис. 8. Блок-схема алгоритма объезда препятствий

3. Заключение

В данной работе описан общий принцип работы комплекса управления складом. Написано программное обеспечение, выполнена настройка системы таким образом, что робот берет грузы из заданных ячеек с надёжностью 100%. Реализована и проверена система обмена данными между роботом и графическим приложением. Реализована и испытана система объезда препятствий. Код проекта, написанный на Python [5] можно найти по ссылке https://github.com/denis915959/robot.

Литература

1. Движение робота к точке с заданными координатами / Хабр (habr.com) [электронный документ]. URL: https://habr.com/ru/post/277829/ (дата обращения: 27.09.2024).
2. SciELO - Brazil - An approach to avoid obstacles in mobile robot navigation: the tangential escape An approach to avoid obstacles in mobile robot navigation: the tangential escape [электронный документ]. URL: https://www.scielo.br/j/ca/a/VCrdFB5Pmt8TgF7CKrNS4Bg/?lang=en (дата обращения: 27.09.2024).
3. Поисковый алгоритм A Star: что это и как эффективно его использовать? | CoderNet [электронный документ]. URL: https://codernet.ru/articles/web/poiskovyij_algoritm_a_star_chto_eto_i_kak_effektivno_ego_ispolzovat/?ysclid=lcapb8q112185226572 (дата обращения: 27.09.2024).
4. Ультразвуковой дальномер HC-SR04: описание, подключение, схема, характеристики [электронный документ]. URL: https://amperka.ru/product/hc-sr04-ultrasonic-sensor-distance-module (дата обращения: 27.09.2024).
5. Дронов В.А., Прохоренок Н.А. Python 3 и PyQt 5. Разработка приложений – СПб.: БХВ-Петербург, 2017. – 832 с.

ГРАФЫ ТРИГОНАЛЬНЫХ КРИВЫХ И ЖЕСТКИЕ ИЗОТОПИИ ОСОБЫХ ВЕЩЕСТВЕННЫХ АЛГЕБРАИЧЕСКИХ КРИВЫХ БИСТЕПЕНИ (4,3) НА ГИПЕРБОЛОИДЕ[1*]

В.И. Звонилов

Нижегородский государственный университет им. Н.И. Лобачевского

В работе даны доказательства сформулированных ранее автором утверждений о единственности компонент связности 16-ти классов вещественных алгебраических кривых бистепени (4,3), имеющих единственную невырожденную двойную точку или точку возврата. Основным техническим средством являются графы вещественных тригональных кривых на поверхностях Хирцебруха.

Ключевые слова: пространства вещественных алгебраических кривых, кривые на квадриках, тригональные кривые.

1. Определения

Пусть $[x_0{:}\,x_1]$, $[y_0{:}\,y_1]$ – однородные координаты на гиперболоиде $P^1 \times P^1$. *Вещественная алгебраическая кривая бистепени* (m_1, m_2) на гиперболоиде определяется уравнением $F(x_0, x_1; y_0, y_1) = 0$, где F – вещественный однородный многочлен степеней m_1 и m_2 по x_0, x_1 и y_0, y_1.

Обозначим через $C_{m,n}$ пространство вещественных алгебраических кривых бистепени (m, n) на гиперболоиде.

Вещественная схема вещественной алгебраической кривой A (на плоскости, на поверхности) – это схема взаимного расположения компонент её вещественной части $\mathbb{R}A$ (вещественных ветвей - для особой кривой). *Овалом* называется компонента, стягиваемая на плоскости, на поверхности.

Вещественная алгебраическая кривая A принадлежит *типу I*, если множество $\mathbb{C}\,\widetilde{A} \setminus \mathbb{R}\,\widetilde{A}$ несвязно, где $\widetilde{A}$ – нормализация кривой A, и *типу II*, если оно связно. Если A принадлежит типу I, то ориентация множества $\mathbb{R}\,\widetilde{A}$ как края замыкания одной из двух половин множества $\mathbb{C}\,\widetilde{A} \setminus \mathbb{R}\,\widetilde{A}$ называется *комплексной*. *Комплексная схема* кривой есть вещественная схема + тип + комплексные ориентации (для типа I).

Вещественная схема называется *неопределённой*, если существуют кривые обоего типа с этой схемой.

Жесткой изотопией вещественных алгебраических кривых некоторого класса называется путь в пространстве кривых этого класса.

2. Поверхности Хирцебруха, преобразования Нагаты, тригональные кривые

Поверхность Хирцебруха $\Sigma_k, k \geq 0$, является пространством линейного расслоения $q{:}\,\Sigma_k \to P^1$, т.е. вещественной рациональной линейчатой поверхностью с исключительным вещественным сечением E_k, $E_k^2 = -k$. Поверхность $\Sigma_0 = P^1 \times P^1$ является гиперболоидом, в котором в качестве исключительного сечения можно взять любую кривую бистепени $(0,1)$. В пространстве $C_{m,n}$ при изучении кривых с единственной особенностью исключительным сечением считается кривая бистепени $(0,1)$, проходящая через особую точку кривой.

Положительным (отрицательным) преобразованием Нагаты называется послойное бирациональное преобразование $\Sigma_k \to \Sigma_{k+1}$ (соответственно, $\Sigma_k \to \Sigma_{k-1}$), состоящее из раздутия

[1*] Работа поддержана Минобрнауки РФ, проект FSWR-2023-0034.

точки $p \in E_k$ (соответственно, $p \notin E_k$) и последующего стягивания в точку собственного прообраза слоя $q^{-1}(p)$.

Положительное (отрицательное) преобразование Нагаты переводит E_k в E_{k+1} (соответственно, в E_{k-1}).

Алгебраическая кривая $C \subset \Sigma$ называется *тригональной*, если сужение $q|_C$ является трёхлистным разветвлённым накрытием. Кривая называется *собственной*, если она не пересекается с исключительным сечением.

Положительное и отрицательное преобразования Нагаты устанавливают взаимосвязь между (тригональными) кривыми бистепени $(m, 3)$ на гиперболоиде и собственными тригональными кривыми: кривой $C \subset \Sigma_0$ отвечает её образ $N(C) \subset \Sigma_m$ при положительном преобразовании Нагаты, раздувающем точки пересечения кривой C с кривой $E_0 \subset \Sigma_0 = P^1 \times P^1$. Ясно, что при этом жёстко изотопным кривым отвечают жёстко изотопные.

3. Граф тригональной кривой

В некоторой аффинной карте на Σ_k исключительное сечение E_k и собственная тригональная кривая C определяются уравнениями $y = \infty$ и $y^3 + b(x)y + w(x) = 0$, где b и w – сечения расслоения q, которые можно рассматривать как однородные многочлены степеней $2k$ и $3k$. Рациональная функция $j = j_C = \frac{4b^3}{\Delta} = 1 - \frac{27w^2}{\Delta}$ называется *j-инвариантом* кривой C.

Пусть круг D – фактор-пространство прямой $\mathbb{C}P^1$ по комплексному сопряжению, а pr: $\mathbb{C}P^1 \to D$ – проекция. В образе j-инварианта $j_C: \mathbb{C}P^1 \to \mathbb{C}P^1 = \mathbb{C} \cup \{\infty\}$ вещественной тригональной кривой $C \subset \Sigma_k$ раскрасим $\mathbb{R}P^1$ следующим образом: 0, 1, и ∞ – это, соответственно, чёрная, белая и крест-вершина; $(\infty, 0)$, $(0,1)$ и $(1, \infty)$ – это, соответственно, сплошное, жирное и штриховое ребро. Прообраз $\Gamma_C = \mathrm{pr}(j_C^{-1}(\mathbb{R}P^1))$, наделённый соответствующей раскраской, называется *графом* тригональной кривой C.

4. Основной результат

В работе автора утверждалось, что неособая кривая бистепени (4,3) на гиперболоиде определяется с точностью до жесткой изотопии своей комплексной схемой. Для доказательства этого использовался подход, предложенный в [1] для получения жесткой изотопической классификации плоских вещественных квинтик и основанный на теореме 1 работы [2], указывающей все компоненты связности пространства кривых бистепени (4,3), имеющих единственную невырожденную двойную точку или точку возврата. В настоящей работе указанная теорема 1 доказывается другим способом – с использованием графов вещественных тригональных кривых на поверхности Хирцебруха Σ_3, полученных из особых кривых бистепени (4,3). Доказательство этой теоремы было начато автором в [1]; в [2] был указан пробел в этом доказательстве, который устранён в настоящей работе.

Теорема. *Для каждой из 16-ти вещественных схем, указанных на рис.* 1, *жёсткий изотопический класс особой вещественной алгебраической кривой бистепени* (4,3) *с этой схемой однозначно определяются комплексной схемой этой кривой.*

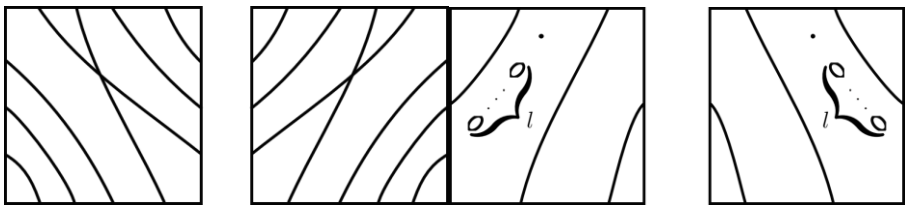


Рис. 1. Схемы особых кривых бистепени
(4,3) на гиперболоиде, $0 \le l \le 5$ (при $l = 3$ схема неопределённая)

Литература

1. Degtyarev A., Itenberg I., Kharlamov V. Real Enriques surfaces. Lecture Notes in Math., 2000, vol. 1746 , Springer-Verlag, 259 p.
2. Звонилов В.И. Жесткая изотопическая классификация вещественных алгебраических кривых бистепени (4,3) на гиперболоиде // Вестник Сыктывкарского университета, 1999, Сер. 1, вып. 3, С. 81-88.
3. Звонилов В. И. Жесткая изотопическая классификация вещественных алгебраических кривых бистепени (4,3) на гиперболоиде. Приложение // Вестник Сыктывкарского университета, 2003, Сер. 1, вып. 5, С. 239-242.

РЕАЛИЗАЦИЯ АЛГОРИТМОВ ПОДГОТОВКИ НАЧАЛЬНОГО СОСТОЯНИЯ СИСТЕМЫ КУБИТОВ С УЧЕТОМ ОГРАНИЧЕНИЙ СОВРЕМЕННЫХ КВАНТОВЫХ КОМПЬЮТЕРОВ

А.Д. Ивлев

Нижегородский государственный университет им. Н.И. Лобачевского

Подготовка произвольного состояния системы кубитов является важной и актуальной задачей квантовых вычислений. Её важность обуславливается тем, что многие квантовые алгоритмы требуют предварительной загрузки классических данных на квантовые устройства. В данной работе рассмотрены некоторые актуальные подходы и их алгоритмическая трансляция в язык QASM с учетом топологии и базисных наборов гейтов современных квантовых компьютеров. Проведены эксперименты на симуляторах идеальных и реальных квантовых компьютеров, по результатам которых проведено сравнение теоретических оценок точности и размеров квантовых схем и их практической реализации.

Ключевые слова: квантовые вычисления, подготовка квантового состояния, QASM

1. Введение

Подготовка произвольного состояния системы кубитов является важной и актуальной задачей квантовых вычислений. Это обуславливается тем, что многие квантовые алгоритмы требуют предварительной загрузки классических данных на квантовые устройства, например квантовые нейронные сети [1], контроль результатов с помощью гейта SWAP [2] или решение систем линейных уравнений [3].

Задается нормированный вектор состояния системы кубитов $|\psi\rangle \in \mathbb{C}^{2^n}$, где n – количество кубитов. Необходимо построить квантовую схему, которая задаёт унитарное преобразование U, которое переводит исходное состояние $|0\rangle^{\otimes n}$ в желаемое $|\psi\rangle$ ($U|0\rangle^{\otimes n} = |\psi\rangle$).

Известно, что размер квантовых схем подготовки n-кубитного квантового состояния оценивается $O(2^n)$, при использовании однокубитных гейтов и гейта CNOT в качестве универсального базового набора [4]. Но при этом теоретическая глубина схем может достигать $O(n^2)$ при использовании дополнительных кубитов [5].

В данной работе рассмотрены некоторые актуальные подходы, описанные в статьях [5, 6] и их алгоритмическая трансляция в язык QASM [7] с учетом топологии и базисных наборов гейтов, на примере архитектуры квантовых компьютеров IBM. На основе экспериментов была оценена точность алгоритмов и проведено сравнение теоретических оценок размеров и глубины идеальных квантовых схем и их практической реализации.

2. Подготовка квантового состояния с помощью управляемых гейтов поворота

В данном разделе будет рассмотрен модифицированный алгоритм подготовки квантового состояния с помощью управляемых гейтов [5]. Данный алгоритм не использует дополнительных кубитов, а размер и глубина его квантовой схемы оцениваются O(2^n).

Основная идея алгоритма заключается в том, что мы представляем все комплексные числа вектора состояния в показательном виде, рассматриваем подзадачу построения вектора из действительных чисел, и далее – подзадачу наложения фазы на полученный вектор.

Первая подзадача сводится к распределению 1 (соответствует начальному состоянию, где все кубиты в состоянии $|0\rangle$) по всему вектору в необходимых пропорциях. Разделим первую часть алгоритма на n итераций ($i = \overline{1,n}$), на каждой из которых мы будем рассматривать только первые 2^i значений вектора. То есть на каждой итерации мы будем задействовать только первые

i кубитов, причём старший из них изначально будет находиться в состоянии 0. Задача итерации распределить значения из первой половины на весь под вектор из 2^{i} значений.

Разобьём 2^{i} значений на пары по их позициям $(j, 2^{i-1}+j), j = \overline{0, 2^{i-1}-1}$. Можно заметить, что данные значения описываю состояния, которые отличаются только значением старшего, для данной итерации, кубита. Следовательно мы можем выделить данные состояния, используя множественный контроль [8], и применить гейт $Ry(\varphi)$ к каждой паре по отдельности. То есть мы свели задачу итерации к переходу пар из состояния $(c, 0)$ к (a, b).

Пример первых двух итерации построения вектора модулей приведен ниже (рис. 1).

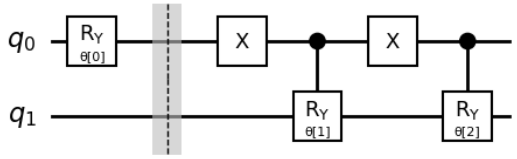


Рис. 1. Квантовая схема первых двух шагов построения вектора модулей

Для завершения основной части алгоритма нам осталось определить углы φ для применения гейтов $Ry(\varphi)$. Рассмотрев данный процесс для одной пары, получим что $a = c * \cos(\varphi/2), b = c * \sin(\varphi/2)$. Тогда $\varphi = 2\arccos(a/c)$ при $c \neq 0$ ($\varphi = 0$ при $c = 0$). При $i =$ n мы знаем a, b для каждой пары, так как это известные нам значения желаемого вектора состояний $|\psi\rangle$. Поэтому для определения углов мы будем идти от последней итерации к первой "сворачивая" вектор к нулевому состоянию, используя соотношение $c = \sqrt{a^2 + b^2}$.

Вторая подзадача в общем случае может быть решена с использованием гейта $P(\varphi)$, но ввиду того, что данный гейт редко входит в базовый набор гейтов реальных квантовых систем, будем использоваться гейт $Rz(\varphi)$, что повлечёт за собой накопление глобальной фазы, но данная проблема будет рассмотрена отдельно.

В исходной статье решение предполагает наложение углов по аналогии с конструированием вектора модулей, последовательно применяя после гейтов $Ry(\varphi)$ гейты $Rz(\psi)$ с множественным контролем. В моём решении предлагается модифицировать последнюю итерацию алгоритма применяя гейты $Ry(\varphi)$ и $Rz(\psi)$ одновременно.

Углы фазы ψ в отличии от исходной статьи не требуется вычислять, так как они нам известны из представления вектора состояния в показательном виде $(r_k \mathrm{e}^{i\psi_k})$. Следовательно, на последней итерации $(c, 0) \rightarrow (a\mathrm{e}^{i\psi_1},\ b\mathrm{e}^{i\psi_2})$, то есть мы можем не задействовать дополнительные гейты CNOT для поворота на фазу, так как второе значение в паре $(c, 0)$ равно 0. Пример одного шага последней итерации приведен на рис. 2.

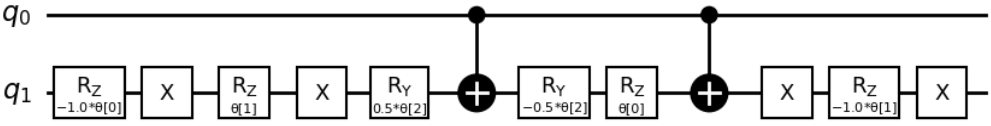


Рис. 2. Блок квантовой схемы, выполняющий поворот на заданную фазу

Данный алгоритм является теоретически точным, то есть его реальная точность ограничена лишь точностью выполнения операций на классическом компьютере при подсчете углов и на квантовом компьютере при исполнении схемы.

3. Подготовка квантового состояния с помощью концепции разделяй и властвуй

Предложенный в статье [5] алгоритм, представляет собой оптимизацию предыдущего алгоритма, которая за счет использования $2^n - n - 1$ дополнительных кубитов приводит асимптотику глубины схемы к $O(n^2)$, но сохраняя её размер $\mathrm{O}(2^n)$.

Идея заключается в развертке алгоритма в ширину вместо длины за счет древовидной структуры алгоритма и гейтов CSWAP. Такая организация позволяет выполнять квантовые гейты параллельно. Пусть $\mathrm{N} = 2^n$, тогда алгоритм построения схемы может быть описан как на рис. 3.

input : $N-1$ dimensional vector angle = `gen_angles` ($abs(x)$)
input : $N-1$ dimensional vector angle_z = `gen_angles_z` ($phase(x)$)
output : Quantum circuit to load x in the amplitudes of a quantum system entangled with ancillary qubits

```
circuit = quantum circuit with N − 1 qubits q[0], . . . , q[N − 2]
for k ← 0 to N − 2 do
    R_y(angle[k], q[k])
for k ← 0 to N − 2 do
    R_z(angle_z[k], q[k])
actual = parent(N − 2)
while actual ≥ 0 do
    left_index = left(actual)
    right_index = right(actual)
    while right_index < N − 1 do
        cswap(q[actual], q[left_index], q[right_index])
        left_index = left(left_index)
        right_index = left(right_index)
    actual = actual - 1
```

Рис. 3. Подготовка квантового состояния с помощью концепции разделяй и властвуй [5]

Углы для гейтов Ry вычисляются также, как в предыдущем случае, но для Rz требуются дополнительные вычисления [5]. На рис. 4 повернутое дерево углов и схема, созданная с помощью стратегии разделяй и властвуй для подготовки состояния 3-х кубитов. Результат храниться в кубитах q0, q1 и q3, а прочие кубиты являются дополнительными.

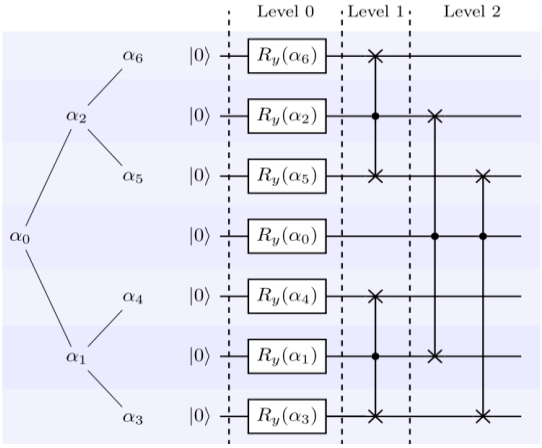


Рис. 4. Повернутое дерево углов и схема, созданная с помощью стратегии разделяй и властвуй для подготовки состояния 3-х кубитов [5]

4. Подготовка квантового состояния в виде аппроксимации унитарного оператора

Любая квантовая схема представляет собой некоторый унитарный оператор. Соответственно получив вид данного оператора для схемы, описанной в разделе 2, можно аппроксимировать его.

Первый столбец данного оператора всегда будет вектором состояния, которое мы хотим построить. Что естественным образом получается из нашей постановки задачи в виде $U|0\rangle^{\otimes N} = |\psi\rangle$, где $|\psi\rangle$ – желаемое состояние. Соответственно для решения данной задачи в общем случае достаточно брать унитарный оператор U с первым столбцом $|\psi\rangle$.

Для этого решим более общую задачу аппроксимации произвольного унитарного оператора U размера $2^n \times 2^n$ некоторой конечной квантовой схемой на основе некоторого конечного набора базовых гейтов. Так как базовые наборы гейтов для разных квантовых компьютеров часто отличаются, то, представляя один универсальный набор гейтов через другой с помощью данного алгоритма, мы можем обеспечить переносимость квантовых схем на разные квантовые процессоры, что также будет использоваться в данной работе.

Идея алгоритма заключается в том, что мы заготавливаем несколько параметризованных шаблонных схем. И пытаемся оптимизировать параметры, порядок применения и количество повторений данных шаблонов, чтобы полученная в результате матрица схемы минимально отличалась от желаемой. Так как мы хотим аппроксимировать произвольный унитарный оператор, то мы изначально ничего не знаем о его виде, поэтому, для универсальности, за базовый шаблон будем брать некоторые аналоги схемы 2 (рис. 5).

Следовательно, предложенный универсальный метод может быть не оптимальным и количество повторений шаблонных схем для достижения высокой точности может достигать 2^n, но как правило не более. Подбор оптимальных параметров будет происходить, как решение задачи оптимизации: $\min_p ||U - M(p)||$, где M(p) – матрица параметризованной схемы с набором параметров p – соответственно, и используется норма Фробениуса.

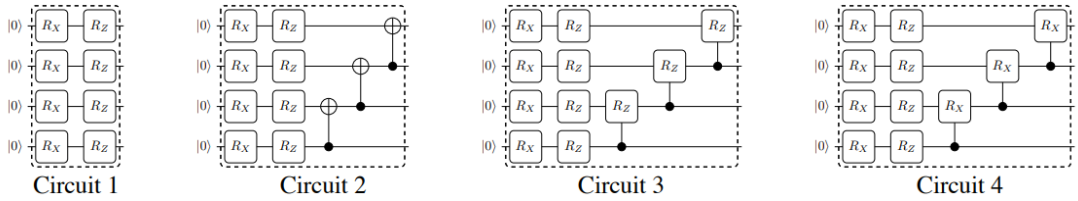


Рис. 5. Типовые примеры шаблонных схем [9]

5. Влияние ограничений современных квантовых компьютеров на построение алгоритмов

На примере квантовых компьютеров, находящиеся в открытом доступе через сервис IBM Quantum, рассмотрим влияние их ограничений на построение алгоритмов. В данный момент базовым набором гейтов для всех доступных для общего пользования квантовых машин является набор ECR, I, Rz, SX, X. Он отличен от используемых теоретиками, но он также представляет собой универсальную систему. Аппроксимируем базовые гейты, используемые для построения алгоритмов подготовки начального состояния, с помощью алгоритма аппроксимации унитарного оператора.

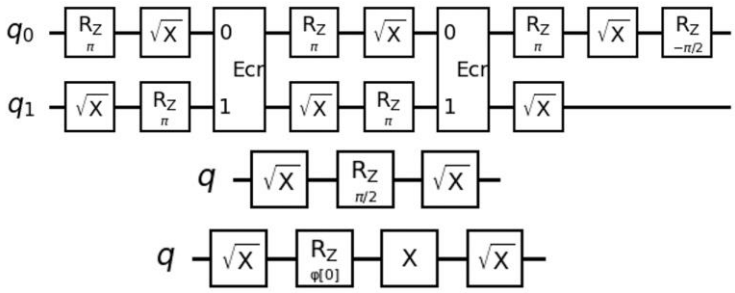


Рис. 6. Аппроксимация гейтов CNOT (с точностью до глобальной фазы $e^{i\frac{\pi}{4}}$), H и Ry(φ)

Несложно заметить, что такая замена увеличивает в разы количество используемых гейтов. Например, количество двухкубитных гейтов увеличится ровно в 2 раза.

Следующим ограничением является топология квантового компьютера. В отличие от идеальной модели связь кубитов представляет собой неполный ориентированный граф, что накладывает ограничения на применение двухкубитных гейтов. Так как любые 2 кубита связаны, либо напрямую, либо по цепочке, то, если кубиты не являются соседями представим исходный CNOT эквивалентной последовательностью гейтов между соседними кубитами [10], но такой подход увеличивает сложность применения CNOT от O(1) к O(p), где p – длина цепочки кубитов. Если связь есть, только в одну сторону, то мы можем «перевернуть» CNOT с помощь гейтов H примененных к обоим кубитам до и после обратного CNOT.

Также на реальных системах присутствуют ошибки выполнения гейтов. Но в данной работе не рассматриваются оптимизационные алгоритмы и алгоритмы компенсации ошибок, чтобы отследить точность работы подготовки начального состояния в реализации приближенной к теоретической.

Также может возникнуть проблема наличия глобальной фазы. Она обычно не доставляет проблем, так как не влияет на результат измерения, но всё же есть некоторые алгоритмы, такие что при возникновении глобальной фазы внутри которых, мы можем получить результат отличный от желаемого, например алгоритм подготовки начального состояния.

Применение гейта X с r управляющими кубитами (CnNOT), сконструированного через CnRz(π) [8], будет прибавлять глобальную фазу $\frac{\pi}{2^{r+1}}$. Соответственно применив два раза данный гейт с одинаковыми параметрами, мы не изменим модуль состояния, но увеличим глобальную фазу на $\frac{\pi}{2^r}$.

В нашем случае алгоритм подготовки начального состояния с помощью гейтов поворота не подвержен данной ошибке при $n > 1$, так как на i итерации r = i, причём гейт CnNOT применяется 2^{i+1} раз, следовательно глобальная фаза на итерации равна $2^{i+1}\sum_{k=1}^{i}\frac{\pi}{2^{k+1}} = \pi 2^i \sum_{k=1}^{i}\frac{1}{2^k} = \pi\left(2^i - 1\right)$. Угол глобальной фазы кратен π, следовательно он равен 0 (при $i = 0$) или π, тогда требуется его погасить последовательностью гейтов XZYXZY применённых к любому кубиту. Его оптимизация с помощью дополнительных кубитов унаследует данное свойство, а алгоритм подготовки начального состояния в виде аппроксимации не подвержен глобальной фазе, если шаблонная схема подобрана с учетом базисных гейтов и её количество повторений не меньше 2^n.

6. Результаты экспериментов

Построение и симуляция исполнения квантовых схем реализована с помощью библиотеки qiskit. Далее под идеальным симулятором будет пониматься AerSimulator, а под симулятором реального квантового компьютера FakeSherbrooke, который отображает поведение системы ibm_sherbrooke.

Решение задачи оптимизации для аппроксимации унитарного оператора производиться численно с использованием библиотеки scipy, с помощью метода, вызываемого по умолчанию функцией minimize. За базовый шаблон взят аналог схемы 2 (рис. 5) в базисных гейтах квантового компьютера FakeSherbrooke. а количество её повторений будем считать фиксированным равным 2^n. Тогда количество параметров схемы равно $3n2^n$.

Прежде рассмотрим размеры и глубину квантовых схем в зависимости от количества кубитов. Результаты представлены в табл. 1, где также указано количество двухкубитных гейтов ECR. Параметр ширины схемы не рассматривается, так как он вычисляется явно и равен n для алгоритма с управляемыми гейтами поворота и аппроксимирующего алгоритма, и $2^n - 1$ для алгоритма, реализующего стратегию «разделяй и властвуй».

Таблица 1. Сравнение размера полученных схем, количества гейтов ECR и глубины схем

Количество кубитов		1	2	3	4	5	6	7	8
Управляемые повороты	размер, ECR, глубина	17, 0, 17	148, 8, 93	2204, 200, 1247	23476, 2184, 12087	184152, 16648, 92893	1240956, 110600, 621351	7820778, 692744, 3900877	47883542, 4230664, 23838361
Разделяй и властвуй	размер, ECR, глубина	5, 0, 5	212, 20, 117	1141, 110, 426	6505, 793, 1433	13086, 1579, 1827	30273, 3699, 3074	82370, 10509, 5609	-
Аппроксимация	размер, ECR, глубина	14, 0, 14	95, 3, 49	266, 14, 112	853, 45, 337	2205, 124, 751	5397, 315, 1645	12749, 762, 3563	29381, 1785, 7657

Алгоритм подготовки квантового состояния с помощью управляемых гейтов при действиях для обхода ограничений приобрёл сложность $O(n^n)$ за счет того, что он плохо ложиться на линейную топологию.

Алгоритм «разделяй и властвуй» оказался более устойчивый к топологии, так как большее количество двухкубитных гейтов были применены между соседними кубитами. Но в отличие от теоретических оценок глубины схемы $O(n^2)$ он перешел к $O(2^n)$, а размер схемы к $O(n2^n)$, и имеет схожую асимптотику с алгоритмом в виде аппроксимации унитарного оператора, но с большими константами. Также уже при $n = 8$ алгоритму не хватило дополнительных кубитов.

За счёт того, что шаблонная схема была заготовлена сразу с учётом ограничений квантовых компьютеров IBM, то алгоритм подготовки состояния в виде аппроксимации унитарного оператора достигает меньшего количества используемых гейтов, сохраняя оценку размера схемы $O(n2^n)$и её глубины $O(2^n)$. Но главную сложность данного подхода это решение классической задачи оптимизации для подбора $3n2^n$ параметров.

Сравним точность данных алгоритмов с использованием идеального симулятора и симулятора реального квантового компьютера в зависимости от количества кубитов. Зададим количество измерений квантовых схем в 20000, так как данное значение является максимальным при запуске на реальных квантовых компьютерах IBM. Исходя из размеров квантовых схем и количества измерений, эксперименты были проведены для подготовки состояний систем до 3-х кубитов (табл. 2). Для каждого размера схемы проведены испытаний на 10 произвольных векторах. Результат представляется, как средняя норма разницы исходных и полученных векторов.

Таблица 2. Погрешность работы алгоритмов
(идеальный симулятор, симулятор реального квантового компьютера)

Количество кубитов	1	2	3
Управляемые повороты	0.00255, 0.00735	0.00617, 0.08812	0.00841, 0.44880
Разделяй и властвуй	0.00382, 0.00898	0.00489, 0.11497	0.00852, 0.29804
Аппроксимация	0.00294, 0.00956	0.00727, 0.02402	0.01213, 0.25297

Точность исполнения схемы на симуляторе идеального квантового компьютера обуславливается, только погрешностью выполнения классических вычислений и количеством измерений. При исполнении на симуляторе реального квантового компьютера добавляется погрешность исполнения квантовых гейтов и, так как она относительно велика, то вносит существенный вклад из-за размеров схем. Соответственно алгоритм подготовки состояния в виде аппроксимации унитарного оператора показывает несколько лучшие результаты, но если посмотреть на абсолютные значения результатов, то ошибка является существенной. Например, на рис. 7 изображены результаты, полученные алгоритмом “разделяй и властвуй” при подготовке состояния 3-х кубитов (ошибка 0.24173) и 4-х кубитов (ошибка 0.38991).

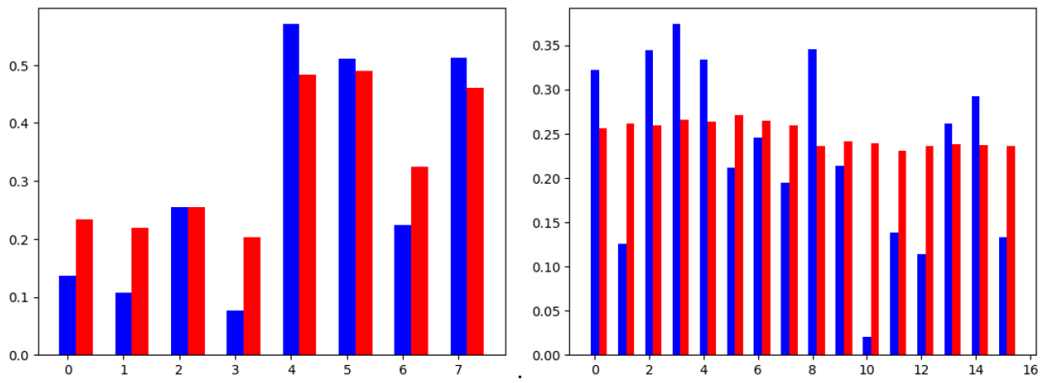


Рис. 7. Заданный вектор состояния (синий) и полученный на симуляторе реального квантового компьютера (красный) алгоритмом “разделяй и властвуй”

Если при ошибке около 0.25 результат ещё различим, но с погрешностями, то из-за особенностей квантового компьютера при исполнении больших схем результат может стримиться к равновероятному состоянию и ошибка будет расти.

6. Заключение

В данной работе были рассмотрены три возможных алгоритма подготовки начального состояния. Их алгоритмическая реализация во всех случаях показала смещение оценки асимптотики размера схемы и её глубины в худшую сторону. Лучший результат показал алгоритм подготовки состояния в виде аппроксимации унитарного оператора с $O(n2^n)$ и $O(2^n)$ соответственно, но он требует ресурсоёмкой предварительной классической подготовки $3n2^n$ параметров схемы.

Все алгоритмы показали низкую точность при алгоритмической реализации на архитектуру современного квантового компьютера ibm_sherbrooke. И на данный момент из-за высокой асимптотики практической реализации в явном виде они пригодны лишь для подготовки состояний систем из не более чем 3-х кубитов. Возможно, применяя различные оптимизационные методы и алгоритмы компенсации ошибок, удастся увеличить это число, но без новых подходов, которые снизят асимптотику размера схемы или её глубины на практике, не стоит ожидать значительного улучшения.

Литература

1. Hur T., Kim L., Park D. K. Quantum convolutional neural network for classical data classification // arXiv preprint arXiv:2108.00661. – 2021.
2. Barenco A. et al. Stabilisation of Quantum Computations by Symmetrisation // arXiv preprint quant-ph/9604028. – 1996.
3. Harrow A.W., Hassidim A., Lloyd S. Quantum algorithm for linear systems of equations // Physical review letters. – 2009. – Т. 103. – №. 15. – С. 150–502.
4. Sun X. et al. Asymptotically Optimal Circuit Depth for Quantum State Preparation and General Unitary Synthesis // arXiv preprint arXiv:2108.06150v3. – 2021.
5. Araujo I. F. et al. A divide-and-conquer algorithm for quantum state preparation // Scientific reports. – 2021. – Т. 11. – №. 1. – С. 6329.
6. Sim S., Johnson P. D., Aspuru-Guzik A. Expressibility and entangling capability of parameterized quantum circuits for hybrid quantum-classical algorithms // Advanced Quantum Technologies. – 2019. – Т. 2. – №. 12.
7. Cross A.W. et al. Open quantum assembly language // arXiv preprint arXiv:1707.03429. – 2017.
8. Barenco A. et al. Elementary gates for quantum computation //Physical review A. – 1995. – Т. 52. – №. 5. – С. 3457.
9. Shende V.V., Bullock S.S., Markov I.L. Synthesis of quantum logic circuits // Proceedings of the 2005 Asia and South Pacific Design Automation Conference. – 2005. – С. 272–275.

АНАЛИЗ ПРОИЗВОДИТЕЛЬНОСТИ OPENBLAS НА ПРОЦЕССОРАХ С АРХИТЕКТУРОЙ RISC-V И X86[1*]

Е.А. Козинов, А.Ю. Пирова, В.Д. Волокитин, Е.А. Панова, А.В. Линев, И.Б. Мееров

Нижегородский государственный университет им. Н.И. Лобачевского

Базовые операции линейной алгебры используются во многих научных приложениях и являются ключевыми вычислительно-трудоемкими операциями. В рамках работы анализируется производительность широко используемых библиотек оптимизированных функций линейной алгебры OpenBLAS и Intel oneMKL на процессорах с архитектурами RISC-V и x86. Предлагается методика тестирования, позволяющая сравнить производительность библиотек для большого числа тестовых случаев и различных объемов данных. На основе проведенных вычислительных экспериментов делаются выводы о производительности OpenBLAS в целом, а также выявляются функции, нуждающиеся в дополнительной оптимизации.

Ключевые слова: OpenBLAS, MKL, базовые функции линейной алгебры, RISC-V.

1. Введение

В последние несколько лет появились и получили большое распространение процессоры с открытой и свободно распространяемой архитектурой RISC-V [1-3]. Еще несколько лет назад такие процессоры были представлены только маломощными прототипами с невысокой производительностью. Вместе с тем, уже сейчас существуют промышленные образцы, сравнимые по производительности с Raspberry Pi 4 [4, 5], а также процессоры, подходящие для высокопроизводительных вычислений [6, 7]. Появление современных и производительных процессоров требует комплексного развития программных средств: операционных систем, компиляторов, математических библиотек, а также подходов к анализу и оптимизации производительности [8-10].

Одной из важных групп математических операций, которые входят в вычислительное ядро многих прикладных научных приложений, библиотек для машинного обучения, математического моделирования, являются операции с векторами и плотными матрицами. В 1979 г. был предложен стандарт интерфейса библиотек для базовых операций линейной алгебры, который получил название BLAS (Basic Linear Algebra Subprograms) [11]. В настоящее время разработано большое число открытых и коммерческих библиотек с реализациями этого стандарта для различных архитектур, в том числе OpenBLAS, Goto BLAS, Netlib BLAS, ATLAS, cuBLAS, Intel oneMKL BLAS и др. Среди реализаций, широко используемых для процессоров архитектуры x86, можно выделить BLAS из коммерческой библиотеки Intel oneMKL [12], оптимизированный для процессоров Intel, а также открытую библиотеку OpenBLAS [13,14], которая содержит оптимизированные реализации для различных типов процессоров и поддерживается большим сообществом разработчиков.

В рамках данной работы проведен сравнительный анализ производительности библиотек OpenBLAS и Intel oneMKL на процессорах архитектуры x86, а также протестирована производительность библиотеки OpenBLAS на доступных процессорах с архитектурой RISC-V. В результате анализа выделена группа функций библиотеки OpenBLAS, имеющих потенциал для низкоуровневой оптимизации на процессорах архитектур RISC-V и x86.

Статья имеет следующую структуру. В разделе 2 приводится краткое описание методики проведения вычислительных экспериментов. В разделе 3 описываются результаты вычислительных экспериментов. В заключении приводятся выводы и определяются направления дальнейших исследований.

[1*] Работа выполнена при поддержке программы академического лидерства ННГУ "Приоритет-2030"

2. Методика проведения вычислительных экспериментов

Анализ производительности больших математических библиотек является нетривиальной задачей. Как правило, библиотеки содержат большое количество функций, сами функции могут содержать разные реализации в зависимости от размера исходных данных и целевой архитектуры, а также несколько параллельных реализаций. Как результат, замер производительности ведет к построению таблиц большой размерности, в которых затруднительно обобщить результаты. Для упрощения понимания и обобщения собранных данных мы предложили и использовали методику, представленную ниже.

В рамках работы производился замер производительности всех функций линейной алгебры с матрично-векторными и матричными операциями, описанными в стандарте BLAS (BLAS второго и третьего уровня). Для детального анализа производительности каждой функции были выделены три класса запускаемых тестов, исходя из объема входных данных: «small», «medium», «large». В классе «small» объем исходных данных подбирался так, чтобы все данные попадали в кэш L1. В классе «medium» объем исходных данных подбирался так, чтобы все данные не попадали в кэш L1, но при этом не выходили за кэш L3. В классе «large» объем исходных данных подбирался так, чтобы объем исходных данных превосходил кэш L3 и, при этом, данные полностью помещались в оперативную память. В каждом классе тестов мы меняли размер исходных данных в зависимости от платформы, на которой производились вычислительные эксперименты.

Для замера производительности использовалась библиотека BLAS-Tester [14, 15]. Тесты повторялись несколько раз, замерялись время работы и производительность в MFlop. Так как на замер производительности могут влиять внешние факторы, среди всех результатов в рамках одного класса отбрасывались 20% замеров с наименьшим значением MFlop и 5% замеров с наибольшим значением MFlop. Оставшиеся замеры усреднялись.

Эксперименты проводились как для последовательных реализаций, так и для параллельных. В разделе 3 в результатах вычислительных экспериментов приведены данные для двух случаев: последовательный запуск и запуск на всех доступных вычислительных ядрах процессора.

В первой серии экспериментов BLAS-Tester собирался на платформе с процессорами архитектуры x86 с библиотеками OpenBLAS и Intel oneMKL. Выполненные эксперименты позволили оценить производительность OpenBLAS и найти несколько функций, нуждающихся в низкоуровневой оптимизации. Во второй серии экспериментов BLAS-Tester и OpenBLAS собирались для платформы с процессором архитектуры RISC-V, что позволило сделать выводы о производительности доступных нам процессоров с архитектурой RISC-V.

3. Результаты вычислительных экспериментов

Вычислительные эксперименты проводились на нескольких разделах суперкомпьютера Лобачевский. В качестве тестовой платформы с процессорами архитектуры x86 использовался сервер с двумя процессорами Intel Xeon 4310T (всего 2 x 10 ядер Ice Lake, тактовая частота до 3.4 GHz) и 64 GB оперативной памяти, операционной системой CentOS 7, компилятором GCC 9.5. В качестве тестовой платформы с процессорами архитектуры RISC-V использовался раздел, построенный на основе плат Lichee Pi 4A [4] с процессорами TH1520 RISC-V (всего 4 ядра THead C910, тактовая частота до 1.8GHz, RISC-V ISA RV64GCV0p7), 16 GB оперативной памяти. На платах установлена операционная система Debian 12. На первой тестовой платформе тестировались библиотеки OpenBLAS v.0.3.28 и Intel oneMKL версии 2023.0.0 с помощью пакета BLAS-Tester [15]. На второй тестовой платформе – только библиотека OpenBLAS v.0.3.28. Компиляция OpenBLAS и BLAS-Tester на RISC-V осуществлялась кросс-компилятором GCC 10.4.0 с поддержкой векторных инструкций RVV 0.7.1 (Xuantie-900 linux-5.10.4 glibc gcc Toolchain V2.8.1 B-20240115).

Вычислительные эксперименты проводились согласно методике, описанной в разделе 2. Размеры входных данных для тестовых классов определялись таблицей 1. Тестирование проводилось для вычислений в одинарной и двойной точности чисел с плавающей запятой.

Таблица 1. Допустимый объем данных для разных тестовых классов, в зависимости от вычислительной платформы

Тестовый класс	Платформа x86	Платформа RISC-V
large	> 50 МБ и < 64 ГБ	> 4 МБ и < 16 ГБ
medium	> 2 МБ и < 15 МБ	> 256 КБ и < 1 МБ
small	< 0.5 МБ	< 128 КБ

В таблице 2 представлена часть собранных результатов вычислительных экспериментов для нескольких функций[1]. В первом столбце указано имя тестируемой функции. Во втором столбце приведен тестовый класс. Третий столбец содержит информацию о числе используемы потоков, при этом в скобках указано число потоков для процессоров RISC-V. Столбцы 4–5 содержат значения усредненной производительности функции в MFlop. В столбце 7 сравнивается производительность OpenBLAS и Intel oneMKL. В последнем столбце показано отставание процессоров RISC-V от x86 в рамках используемых платформ.

Таблица 2. Производительность некоторых функций линейной алгебры для библиотек OpenBLAS и Intel oneMKL, тип данных double (данные приведены в MFlop)

1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Функция	**Тестовый класс**	**Число потоков**	**OpenBLAS x86**	**Intel oneMKL x86**	**OpenBLAS RISC-V**	**OpenBLAS / Intel oneMKL**	**RISC-V / OpenBLAS**
gemm	large	1	72191,3	76284,3	2870,8	0,95	0,04
gemm	large	20(4)	247135,2	250516,3	6816,2	0,99	0,03
gemv	large	1	3627,2	3258,7	538,6	1,11	0,15
gemv	large	20(4)	7491,6	7233,9	987,5	1,04	0,13
sbmv	large	1	569,0	995,2	26,1	**0,57**	0,05
sbmv	large	20(4)	562,0	1052,5	17,8	**0,53**	0,03
trsm	large	1	53771,9	80173,0	2458,7	**0,67**	0,05
trsm	large	20(4)	174787,7	268473,0	7014,6	**0,65**	0,04
gemm	medium	1	40407,8	44052,9	2845,4	0,92	0,07
gemm	medium	20(4)	130896,9	87760,2	4674,0	1,49	0,04
gemv	medium	1	3256,4	3184,2	640,7	1,02	0,20
gemv	medium	20(4)	3880,2	3125,2	647,4	1,24	0,17
sbmv	medium	1	733,0	1426,1	535,4	**0,51**	0,73
sbmv	medium	20(4)	738,8	1267,5	435,1	**0,58**	0,59
trsm	medium	1	13518,0	39211,6	2164,6	**0,34**	0,16
trsm	medium	20(4)	53934,9	63398,2	5084,2	**0,85**	0,09
gemm	small	1	17075,9	17580,7	2337,1	0,97	0,14
gemm	small	20(4)	18671,1	7523,9	2420,9	2,48	0,13
gemv	small	1	2727,2	2490,8	598,7	1,09	0,22
gemv	small	20(4)	1689,1	677,3	648,5	2,49	0,38
sbmv	small	1	389,0	690,4	317,5	**0,56**	0,82
sbmv	small	20(4)	407,9	913,8	152,4	**0,45**	0,37
trsm	small	1	3254,2	7762,9	1397,5	**0,42**	0,43
trsm	small	20(4)	5228,4	2998,0	1720,8	1,74	0,33

Анализ результатов вычислительных экспериментов показал, что в библиотеке OpenBLAS большая часть функций работает с производительностью, близкой к Intel oneMKL. Примерами

[1] Результаты замера всех функций не приводятся из-за большого объема данных.

таких функций являются gemm и gemv (умножение двух матриц и умножение матрицы на вектор соответственно) из таблицы 2. В однопоточном режиме производительность данных функций схожа на всех классах. В многопоточном варианте Intel oneMKL в случае малого объема исходных данных может отставать. Подобный эффект, скорее всего, связан с используемыми в Intel oneMKL механизмами распараллеливания и NUMA архитектурой тестовой платформы.

В процессе проведения экспериментов было выявлено пять функций из OpenBLAS, на которых наблюдается заметное отставание по производительности от Intel oneMKL (около 30–50%), а значит, имеющих потенциал для низкоуровневой оптимизации. Это функции BLAS второго уровня gbmv, tbmv, sbmv, tbsv и функция BLAS третьего уровня trsm. Результаты экспериментов для функций sbmy и trsm приведены в таблице 2. Данные функции демонстрируют почти двухкратное отставание на тестовых классах large и medium.

Сравнивая результаты производительности на процессорах RISC-V и (столбец 8 таблицы 2), можно сделать вывод о том, что уже сейчас широкодоступные маломощные процессоры с архитектурой RISC-V отстают от достаточно современного сервера архитектуры x86 в 25-30 раз на вычислительно сложных задачах (compute-bound) (gemm), и всего в 7 раз на задачах, ограниченных пропускной способностью памяти (memory-bound) (gemv). Так как функции gbmv, tbmv, sbmv, tbsv так же относятся к memory-bound задачам, то нам кажется, что эти функции обладают потенциалом для оптимизации. Поскольку вычислительные алгоритмы рассматриваемых функций одинаковы для платформ x86 и RISC-V, то общее улучшение производительности для x86 за счет изменения вычислительного алгоритма может ускорить и работу на RISC-V.

Заключение

В рамках работы предложена методика и проведены масштабные вычислительные эксперименты по сравнению производительности двух широко используемых библиотек базовых математических функций линейной алгебры – OpenBLAS и Intel oneMKL. Исходя из результатов проведенных вычислительных экспериментов выявлено, что на большинстве математических функций линейной алгебры в однопоточном режиме OpenBLAS сравним с Intel oneMKL, в случае использования многопоточной реализации – OpenBLAS уступает Intel oneMKL. В результате тестирования выбрано пять функций, на которых OpenBLAS заметно уступает Intel oneMKL, а значит, имеется потенциал для их оптимизации. В качестве дальнейшего исследования планируется низкоуровневая оптимизация выявленных функций для процессоров с архитектурой x86 и RISC-V.

Литература

1. Asanović, K., Patterson, D.A. Instruction sets should be free: The case for RISC-V. // EECS Department. University of California. Berkeley. UCB/EECS-2014-146 (2014).
2. Waterman A., Asanović K. The RISC-V Instruction set manual, volume I: user-level ISA, Document version 20190608-Base-Ratified. // RISC-V Foundation. 2019.
3. Furber S.B. VLSI RISC architecture and organization. 1st edn. CRC Press. 1989.
4. Lichee PI. URL: https://sipeed.com/licheepi4a (дата обращения: 24.09.2024).
5. SpacemiT x60. URL: https://www.spacemit.com/en/spacemit-x60-core/ (дата обращения: 24.09.2024).
6. Davis J.D. RISC-V in Europe: the road to an open source HPC stack. URL: https://www.european-processor-initiative.eu/wp-content/uploads/2022/03/EPI-@-HPCUser-Forum.pdf (дата обращения: 24.09.2024).
7. First International workshop on RISC-V for HPC. URL: https://riscv.epcc.ed.ac.uk/ community/ isc23-workshop/ (дата обращения: 24.09.2024).

ВЛОЖЕНИЕ ФИЛЬТРОВАННЫХ ДЕФОРМАЦИЙ ГРАДУИРОВАННЫХ НЕАЛЬТЕРНИРУЮЩИХ ГАМИЛЬТОНОВЫХ АЛГЕБР ЛИ В ОБЩУЮ АЛГЕБРУ ЛИ КАРТАНОВСКОГО ТИПА[1*]

А.В. Кондратьева, М.И. Кузнецов

Нижегородский государственный университет им. Н.И. Лобачевского

В докладе представлены результаты проверки условий теоремы о минимальном вложении фильтрованных деформаций $\boldsymbol{\mathcal{L}}$ градуированных неальтернирующих гамильтоновых алгебр Ли $\boldsymbol{L}$ над полем характеристики 2 в алгебру Ли специальных дифференцирований (векторных полей) $\boldsymbol{W(F)}$ над алгеброй разделенных степеней $\boldsymbol{O(F)}$ пространства переменных $\boldsymbol{E}$ размерности $\boldsymbol{n}$.

Ключевые слова: неальтенирующие гамильтоновы алгебры Ли, фильтрованные деформации, теорема вложения.

Основные результаты

Согласно теореме вложения [1] для транзитивной алгебры Ли $\mathcal{L}$ с отмеченной подалгеброй $\mathcal{L}_0$ существует минимальное вложение $\tau\colon (\mathcal{L}, \mathcal{L}_0) \to \big(W(F'), W(F')_{(0)}\big)$. Пусть $\mathcal{L}$ – фильтрованная алгебра Ли, $\mathcal{L} = \mathcal{L}_{-1} \supset \mathcal{L}_0 \supset \mathcal{L}_1 \supset \cdots$ с ассоциированной градуированной алгеброй $L = L_{-1} + L_0 + L_1 + \cdots$, $L_{(0)} = L_0 + L_1 + \cdots$. Известно, что $F(L, L_0) \le F'$. Достаточные условия совпадения флагов содержатся в теореме, которая является частным случаем теоремы, доказанной в [1, 2].

Теорема 1. Пусть $L = L_{-1} + L_0 + \cdots$, – транзитивная градуированная алгебра Ли. Если

(i) L_{-1} – неприводимый L_0-модуль,

(ii) $H^1(L_0, L_{-1}) = 0$,

(iii) $mtp\left(L_{-1}, H_1(L_{-1})\right) \le m\left(F\left(L, L_{(0)}\right)\right) - n$, $n = dim\, L_{-1}$,

то для любой фильтрованной деформации $\mathcal{L}$ алгебры Ли L справедливы утверждения:

a) $F(\mathcal{L}, \mathcal{L}_0) = F\left(L, L_{(0)}\right)$,

b) $Der\, \mathcal{L} \cong N_{\overline{W}(F)}(\tau(\mathcal{L}))$.

Здесь $\tau\colon \mathcal{L} \to W(F)$ – минимальное вложение, $\overline{W}(F)$ – р-замыкание $W(F)$ в $Der\, O(F)$, $mtp(Q, V)$ – кратность вхождения L_0-модуля Q в композиционный ряд L_0-модуля V.

Авторы доказывают, что условия теоремы выполняются при некоторых ограничениях для фильтрованных алгебр Ли $\mathcal{L}$ с ассоциированной градуированной алгеброй, изоморфной градуированной неальтернирующей гамильтоновой алгебре $L, P(F,\ \omega_0)^{(1)} \subseteq L \subseteq \tilde{P}(F,\ \omega_0)$ (см. [3 – 5]).

Доказательство основано на результатах, приведенных в [6].

Пусть $V = L_{-1}$, $o(V)$ – ортогональная алгебра Ли, соответствующая форме ω_0, $so(V) = o(V) \cap sl(V)$. Подалгебру L_0 можно рассматривать как подалгебру в $o(V)$.

Теорема 2. Пусть G – *подалгебра в* $o(V)$, $o(V)^{(1)} \subseteq G$. Если $n \ne 3$ или $n = 3$, $G \not\subseteq so(V)$, то $H^1(G, V) = 0$. Если $n = 3$, $G \subseteq so(V)$, то $\dim H^1(G, V) = 2$.

Пусть E^0 – подпространство изотропных векторов в $E = L_{-1}^*$ относительно формы, двойственной форме ω_0 на пространстве L_{-1}.

Теорема 3. Кратность L_0-модуля L_{-1} в L_0-модуле $H_1(L_{-1})$ не превосходит $m(F) - n$, где $m(F) = m_1 + \cdots + m_n$, при условии $n \ne 4$ или $n = 4$, $E_1 \not\subset E^0$.

[1*] Работа выполнена при финансовой поддержке Министерства науки и высшего образования Российской Федерации (проект FSWR-2023-0034).

Таким образом, при указанных в теоремах 2 и 3 ограничениях, выполнены условия теоремы 1. В частности, для фильтрованных деформаций $\mathcal{L}$ градуированных неальтернирующих гамильтоновых алгебр Ли L флаги $F(\mathcal{L}, \mathcal{L}_0)$ и $F\left(L, L_{(0)}\right)$ совпадают.

Литература

1. Кузнецов М.И. Усеченные индуцированные модули над транзитивными алгебрами Ли характеристики p // Изв. АН СССР. Сер.матем. 1989. Т.53, № 3. С. 557-589.
2. Кузнецов М.И. Теорема вложения для транзитивных фильтрованных алгебр Ли харак-теристики p // Изв. вузов. Матем. 1991. №.10. С. 43-45.
3. Кузнецов М.И., Кондратьева А.В., Чебочко Н.Г. О гамильтоновых алгебрах Ли харак-теристики 2 // Матем. журнал. 2016. Т.16, № 2. С. 54-65.
4. Kondrateva A.V. Non-alternating Hamiltonian Lie algebras in three variables // Lobachevskii Journal of Mathematics. 2021. V.42, No. 12. P. 2841-2853.
5. Кондратьева А.В., Кузнецов М.И., Неальтернирующие гамильтоновы формы над ал-геброй разделенных степеней в характеристике 2 // Изв. вузов. Матем. 2023. №.6. С. 95-100.
6. Кондратьева А.В., Кузнецов М.И., Фильтрованные деформации градуированных не-альтернирующих гамильтоновых алгебр Ли // Изв. вузов. Матем. 2024. №.9. (в печати).

ФОРМИРОВАНИЕ СПЕЦИАЛИЗИРОВАННЫХ ПОПУЛЯЦИЙ В РЕКУРРЕНТНЫХ НЕЙРОННЫХ СЕТЯХ, ОБУЧЕННЫХ С ПОДКРЕПЛЕНИЕМ ЗАДАЧЕ КОНТЕКСТНО-ЗАВИСИМОГО ПРИНЯТИЯ РЕШЕНИЯ[1*]

Р.А. Кононов, Н.А. Поспелов, О.В. Масленников, В.И. Некоркин

Институт прикладной физики им. А.В. Гапонова-Грехова РАН,
Нижегородский государственный университет им. Н. И. Лобачевского,
Институт перспективных исследований мозга МГУ имени М.В. Ломоносова

В докладе рассматривается динамика рекуррентной нейронной сети с функцией активации ReLU, обученной с подкреплением задаче двухальтернативного выбора с контекстом. Описаны два типа аттракторов, кодирующих решения сети и признаки задачи: квазипериодические траектории и неподвижные точки. Выделены популяции нейронов, активность которых соответствует динамике в окрестности данных аттракторов, и исследована эволюция этих популяций в процессе обучения.

Ключевые слова: нелинейная динамика, аттрактор, нейронная сеть, когнитивная задача, обучение с подкреплением.

1. Введение

Современные исследования когнитивных функций нередко опираются на модели, полученные при помощи машинного обучения из эмпирических данных. [1,2,3] Проблема подобного подхода в том, что полученная путём численной оптимизации нейронная сеть – чёрный ящик, требующий дальнейшего исследования [4,5].

Часто исследование динамики полученных ИНС ограничивается анализом численных траекторий активности сети [6] и полагается на понижение размерности траектории, даже в случае достаточно просто заданных систем, например, вида:

$$\boldsymbol{x}_{n+1} = \boldsymbol{ReLU}(\, \boldsymbol{x}_n \boldsymbol{W}^h + \boldsymbol{I}_n \boldsymbol{W}^I \,), \qquad (1)$$

$$ReLU(x) = \begin{cases} x, & x > 0 \\ 0, & x \le 0 \end{cases}$$

В докладе анализируется динамический механизм [8], формируемый в нейронной сети вида (1) при обучении её классической когнитивной задаче двухальтернативного выбора с контекстом [7] с точки зрения устройства фазового пространства. Для обучения сети используется алгоритм обучения с подкреплением, который полагается на разреженную во времени награду, что позволяет обучать сеть на основе сигналов, структурно повторяющих награду, получаемую животными в натурных экспериментах в этой области.

Отдельное испытание состоит из трёх этапов, первый из которых служит для разделения испытаний, на втором подаются две пары зашумленных стимулов и сигнал контекста, выделяющий одну из этих пар. От сети требуется сравнить стимулы в этой паре и выбрать наибольший, на третьем этапе стимулы отсутствуют, а от агента требуется удержать решение в памяти и дать ответ в конце испытания. Разница между стимулами в одной паре обозначим как coh, тогда каждое испытание конфигурируется контекстом и парой $(coh_A, coh_B) \in [0,1]$. Награда – это отдельный скалярный сигнал, она равна нулю всюду, кроме последнего момента испытания. Если ответ правильный, в этот момент она равна единице.

[1*] Работа выполнена при поддержке гранта Российского научного фонда № 23-72-10088.

2. Результаты

Ансамбль сетей с различной инициализацией весов обучен вышеописанной задаче. Критерий остановки обучения: доля верных ответов сети > 0.95, где под точностью подразумевается вероятность выбора сетью правильного ответа. Конкретный метод обучения: проксимальное обновление стратегии [9].

Стимулы внутри одного этапа – постоянные сигналы с аддитивным шумом. Полагая их фиксированными параметрами, можно изучить автономную динамику внутри каждого из этапов. Сформулирован критерий, позволяющий найти неподвижные точки и квазипериодические аттракторы в сети, полагаясь на структуру связей активных нейронов. Это позволяет выделить нейронные популяции, стоящие за объектами в фазовом пространстве, кодирующими входные стимулы или принятое решение.

Изучена структура популяций и её эволюция в процессе обучения для всех сетей ансамбля. Популяции в разных сетях эволюционируют схожим образом относительно точности сети. Кроме того, изучено распределение типов аттракторов относительно параметров испытания. Эволюция данных популяций в процессе машинного обучения продемонстрирована на рис.1$а$. Зарождение пары популяций, кодирующих решения сети, согласуется с появлением у сети нетривиальной стратегии решения задачи. Несмотря на то, что стимулы одного знака coh кодируются одной популяцией, тип кодирующего аттрактора может меняться в зависимости от этого параметра. На рис.1$б$ показано, что один и тот же знак может кодироваться как неподвижной точкой, так и квазипериодическими аттракторами с разной частотой. Установлена сложная зависимость между coh стимулов и типом аттрактора, кодирующего её на соответствующем этапе испытания. Показано, что квазипериодический аттрактор чаще кодирует пары стимулов в испытаниях с малой по модулю coh (рис.1$в$).

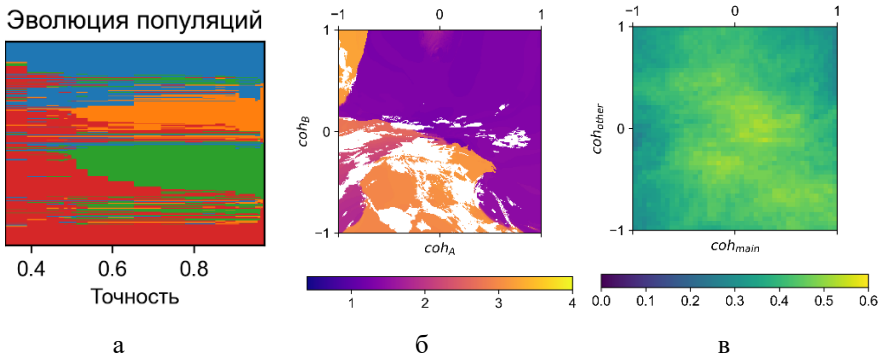


Рис. 1. Эволюция популяций нейросети ($а$) относительно точности решения задачи. Зависимость частоты кодирующего аттрактора на плоскости параметров испытания для одной сети, белым выделены неподвижные точки ($б$) и распределение квазипериодических аттракторов на этой плоскости для ансамбля сетей ($в$)

Литература

1. Saurabh Vyas et al. Computation through neural population dynamics // Annual Review of Neuroscience. 2020. 43 С. 249–275.
2. Richards, B.A., Lillicrap, T.P., Beaudoin, P. et al. A deep learning framework for neuroscience // Nat Neurosci. 2019. 22, p. 1761-1770.
3. Omri Barak, Recurrent neural networks as versatile tools of neuroscience research // Current Opinion in Neurobiology. 2017. V.46. p. 1-6.
4. Масленников О.В., Пугавко М.М., Щапин Д.С., Некоркин В.И., Нелинейная динамика и машинное обучение рекуррентных спайковых нейронных сетей // 2022. УФН 192 1089–1109
5. Pugavko, M.M., Maslennikov, O.V. & Nekorkin, V.I. Multitask computation through dynamics in recurrent spiking neural networks // Sci Rep. 2023. 13, 3997.
6. Yang, Guangyu Robert et al, Artificial Neural Networks for Neuroscientists: A Primer // Neuron. V. 107, № 6, P. 1048 – 1070.

7. Valerio Mante et al, Context-dependent computation by recurrent dynamics in prefrontal cortex // Nature. 2013. 503.7474 , p. 78–84.
8. Кононов Р.А., Масленников О.В., Некоркин В.И. Динамика рекуррентных нейронных сетей с кусочно-линейной функцией активации в задаче контекстно-зависимого принятия решения // Известия вузов. ПНД (принята к публикации).
9. Schulman, John, et al, Proximal policy optimization algorithms // arXiv preprint arXiv:1707.06347 (2017).

СТРУКТУРА И ДИНАМИКА СПАЙКОВЫХ НЕЙРОННЫХ СЕТЕЙ В ОБРАБОТКЕ СТОХАСТИЧЕСКИХ ВХОДНЫХ СИГНАЛОВ[1*]

Н.А. Кудряшкин[1,2], Р.А. Кононов[1,2], В.С. Тиселько[2,3], О.В. Масленников[1,2]

[1]Нижегородский государственный университет им. Н.И. Лобачевского,
[2]Институт прикладной физики им. А.В. Гапонова-Грехова РАН,
[3]Лаборатория сложных сетей, Центр нейрофизики и нейроморфных технологий

Спайковые или импульсные нейронные сети являются предметом исследования в области вычислительной нейронауки как модели биологических сетей; в то же время, в связи с развитием нейроморфных вычислений, большой интерес представляет динамика таких сетей, настроенных методами машинного обучения. В докладе представлены результаты исследования динамики и структуры рекуррентных сетей нейронов накопление-сброс, которые обучены одной из двух задач: фильтрации или интегрирования стохастического входного сигнала. Показано формирование нетривиальных структур связей в процессе обучения с учителем, проведён сравнительный анализ коллективной динамики в сетях, обученных разным задачам.

Ключевые слова: спайковые нейроны, рекуррентные нейронные сети, машинное обучение, нелинейная динамика, сложные сети.

1. Введение

В настоящей работе проведён анализ динамики рекуррентной нейронной сети [1] спайковых нейронов, заданной уравнением (1):

$$Y[t] = W_{out}\, f(W_{in}X[t] + W_{rec}\, Y[t-1]), \quad (1)$$

где $Y[t], Y[t-1]$ – векторы выхода с сети в моменты времени t и $t-1$ соответственно, W_{out} – матрица весов выходного линейного слоя, $f(X[t],\ Y[t-1])$ – функция слоя нейронов с рекуррентными связями, W_{in} – матрица весов входного линейного слоя, W_{rec} – матрица весов рекуррентного слоя, $X[t]$ – вектор входного сигнала в момент времени t.

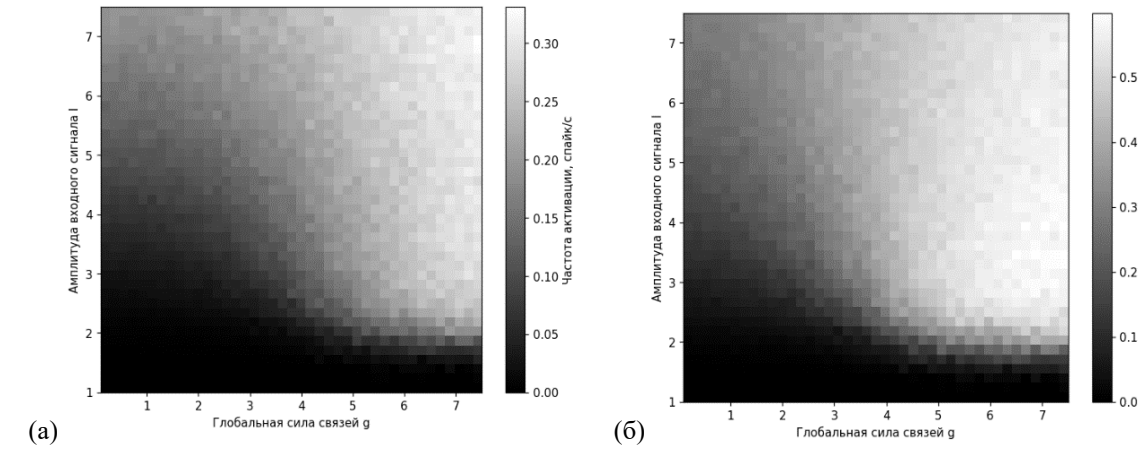
В ходе работы на первом этапе исследована динамика необученной сети при подаче на вход постоянного сигнала. В качестве характеристик динамики использованы средняя частота активации и доля активных (выдающих импульс хотя бы раз за время испытания) нейронов. На Рис. 1а, Рис. 1б представлена зависимость данных величин от глобальной силы связей между нейронами g и уровня сигнала I при фиксированном значении параметра внутренней динамики – характерного времени $\tau = 2.0$.

В нижней области каждого из рисунка 1(а) и 1(б), отмеченной чёрным, сеть находится в режиме покоя – она не реагирует на входной сигнал, внутренняя динамика отсутствует. Противоположным случаем является режим чрезмерной активности – наличие в сети большого количества активных нейронов, выдающих импульсы в каждый момент времени. С точки зрения обработки информации наиболее интересным является пограничный режим. Поэтому рабочим диапазоном гиперпараметров был выбран $g = 3.0,\ I = [2.0;\ 4.0]$.

2. Постановка задач

В задаче фильтрации на сеть подавалась последовательность случайных уровней стохастического сигнала. Соотношение сигнал/шум составляло 0.25, при этом шум был задан равномерным распределением на интервале [-0.5, 0.5). Целевым выходом являлся незашумлённый сигнал.

[1*] Работа выполнена при поддержке гранта Российского научного фонда № 23-72-10088.



(а) (б)

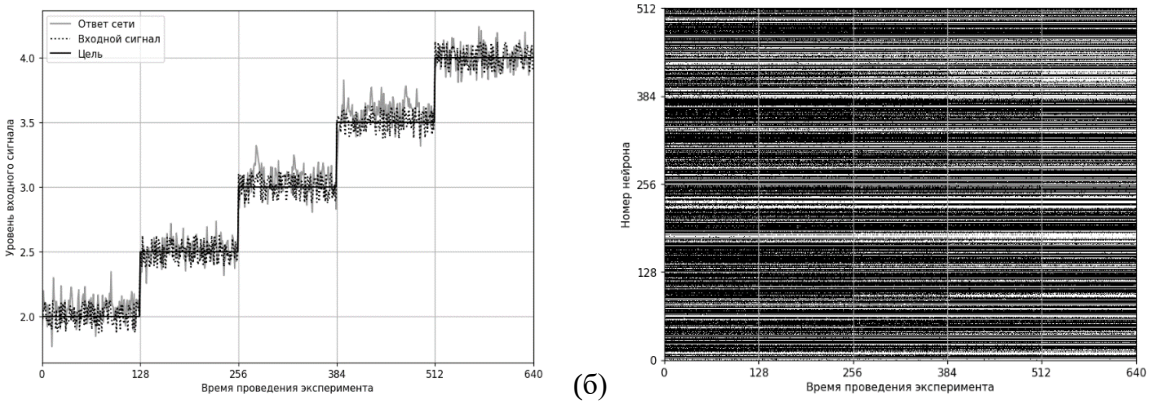
Рис. 1. Динамика необученной сети из 64 нейронов. Средняя частота активации по всей сети (а), доля активных нейронов (б)

В задаче интегрирования [2], которая лежит в основе навигации в окружающей среде биологических организмов, сеть принимала случайный сигнал (скорость когнитивного агента), значения которого имели нормальное распределение с нулевым средним и эмпирически подбираемой для разных параметров сети дисперсией. Скорость оставалась неизменной в течение заданного промежутка времени. В качестве целевого выхода приняты значения координаты, полученные интегрированием скорости с заданным временным шагом *dt*.

3. Обучение и результаты

Обучение, т.е. настройка весовых коэффициентов сети, осуществлено при помощи пакета SpikingJelly, основанного на PyTorch [3, 4]. При этом ключевыми функциональными элементами сети являются спайковые нейроны типа накопление-сброс с утечкой (LIF – нейроны) [5, 6].

После задания гиперпараметров на границе активности было проведено обучение сети с учителем, по методу обратного распространения ошибки [7, 8] с использованием суррогатных градиентов. В качестве суррогатной функции использована функция арктангенса. Для оптимизации весов применён метод адаптивной оценки моментов (Adam). Для расчёта ошибки использована функция средних квадратов. При обучении задаче фильтрации на вход сети подавался стохастический сигнал постепенно нарастающего уровня. Обучение завершалось после 100 эпох.

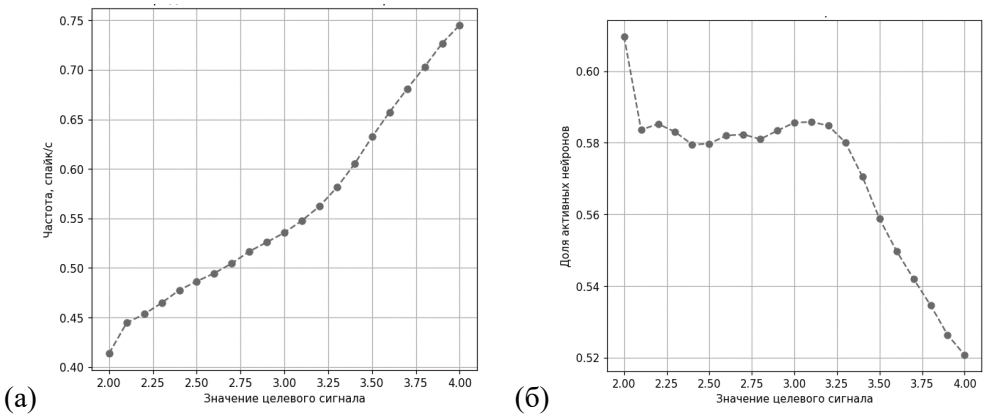


(а) (б)

Рис. 2. Сравнение сигналов (а) и спайковая диаграмма при решении задачи фильтрации (б)

Результаты работы сети из 512 нейронов при решении задачи фильтрации представлены на Рис. 2. На Рис. 2а изображён пример целевого и входного сигнала, а также соответствующий им выход сети. Видно, что сеть различает уровни входа и воспроизводит их близко к уровню целевого сигнала. На Рис. 2б изображена диаграмма спайков, характеризующая внутреннюю динамику сети при обработке задачи.

Проведена оценка средней частоты активных нейронов и их количества в обученной сети в зависимости от уровня целевого сигнала. Эти характеристики получены для 60 различных реализаций входного сигнала, а затем усреднены. Результаты изображены на Рис. 3.



(а) (б)

Рис. 3. Средняя частота (а) и доля активных нейронов (б)

Анализ графиков позволяет установить факт кодирования уровня выходного сигнала частотой активности нейронов в сети. При этом связь между точностью выхода сети и уровнем входного сигнала отсутствует.

4. Заключение

Таким образом, в работе исследована динамика сети спайковых нейронов, выбран оптимальный для обучения режим коллективной активности, и проведено машинное обучение двум целевым задачам. Установлена зависимость между качеством обучения и гиперпараметрами сети, таким как размер сети и сила связи, а также параметром собственной динамики нейронов (характерным временным масштабом) и характеристиками входных сигналов.

Литература

1. Maslennikov, O.V., Nekorkin, V.I. Stimulus-induced sequential activity in supervisely trained recurrent networks of firing rate neurons. Nonlinear Dyn 101, 1093–1103 (2020). https://doi.org/10.1007/s11071-020-05787-0.
2. Cueva, Christopher & Wei, Xue-Xin. (2018). Emergence of grid-like representations by training recurrent neural networks to perform spatial localization. https://doi.org/10.48550/arXiv.1803.07770.
3. SpikingJelly documentation. URL: https://spikingjelly.readthedocs.io/zh-cn/0.0.0.0.14/index.html.
4. Wei Fang et al., SpikingJelly: An open-source machine learning infrastructure platform for spike-based intelligence. Sci.Adv.9, eadi1480 (2023). https://doi.org/10.1126/sciadv.adi1480.
5. Abbott L.F. (1999). Lapicque's introduction of the integrate-and-fire model neuron (1907). Brain Research Bulletin. 50 (5–6): 303–4. doi:10.1016/S0361-9230(99)00161-6
6. Dmitrichev A.S., Kasatkin D.V., Klinshov V.V., Kirillov S.Yu., Maslennikov O.V., Shchapin D.S., Nekorkin V.I. Izvestiya VUZ. AND, vol. 26, no. 4, 2018. https://doi.org/10.18500/0869-6632-2018-26-4-5-58.
7. Rumelhart D.E., Hinton G.E., and Williams R.J. Learning internal representations by error propagation. In Parallel distributed processing, vol. 1, pp. 318–362, Cambridge, MA: MIT Press, 1986.
8. Масленников О.В., Пугавко М.М., Щапин Д.С., Некоркин В.И. Нелинейная динамика и машинное обучение рекуррентных спайковых нейронных сетей. УФН 192 1089–1109 (2022) https://doi.org/10.3367/UFNe.2021.08.039042.

ОПРЕДЕЛЕНИЕ УПРУГИХ СВОЙСТВ КОМПОЗИТНЫХ МАТЕРИАЛОВ МЕТОДОМ РЕШЕНИЯ ОБРАТНОЙ ЗАДАЧИ КОЛЕБАНИЙ ПЛАСТИН НА ВИБРАЦИОННОМ СТЕНДЕ[1*]

С.А. Лавренков, А.В. Васюков

Московский физико-технический институт

Данная работа посвящена обратной задаче поиска упругих параметров многослойных композитов по данным из экспериментов с пластинами из исследуемого материала на вибрационном стенде. Прямая задача поиска амплитудно-частотной характеристики колебаний пластины по заданным упругим параметрам сформулирована как серия краевых задач для уравнения движения тонкого образца в вибрационном стенде, решаемых методом конечных элементов. Обратная задача сведена к минимизации функционала ошибки между численной и экспериментально полученной АЧХ колебаний. В работе представлены численные решения для различных вариантов укладки слоёв в ламинате, производится анализ решений, направленный на улучшение экспериментальной методики.

Ключевые слова: неразрушающий контроль, композитные материалы, обратная задача, автоматическое дифференцирование.

1. Введение

Композитные материалы широко используются в современном машиностроении, однако в некоторых областях применяются образцы композитов, находящихся в статусе экспериментальных и изготавливаемых мелкими сериями. Между этими сериями упругие параметры материала могут значительно изменяться, в связи с чем возникает необходимость разработки метода определения этих параметров, который позволил бы более точно прогнозировать свойства изготавливаемой детали в процессе производства. При этом, в силу ограниченности партий материала, необходимо, чтобы метод являлся неразрушающим. Вибрационные испытания довольно часто проводятся при изготовлении механических конструкций, поэтому необходимое оборудование – вибрационные стенды и датчики широко распространены. В силу приведённых выше фактов, данная работа фокусируется на разработке неразрушающего метода получения упругих модулей материалов по частотному отклику колебаний образца из исследуемого материала на вибрационном стенде.

Подходы, близкие к описываемому в данной работе, могут быть найдены в нескольких источниках. В работе [1] рассматриваются схожие математическая модель и экспериментальная установка, но в ней особое внимание уделяется нелинейности колебаний и для получения частотного отклика используется лазерный виброметр. Родственная обратная задача о поиске упругих параметров бесконечной пластины, состоящей из нескольких изотропных слоёв с разными свойствами и толщинами, рассматривается в [2].

Используемый в настоящей работе метод был впервые описан в статье [3]. Данная работа дополняет этот метод, позволяя применить его к более широкому кругу задач.

2. Методология исследования

Рассматриваются прямая и обратная задачи колебаний пластины на вибрационном стенде. Экспериментальная установка, позволяющая получить частотный отклик колебаний, состоит из вибростенда (генератора синусоидальных колебаний с изменяющейся частотой), пластины из исследуемого материала и двух синхронизированных механических акселерометров, один из ко-

[1*] Работа выполнена при финансовой поддержке Российского научного фонда (код проекта 22-11-00142).

торых закреплён на пластине. Исходя из этого, прямая задача формулируется следующим образом: при заданных параметрах геометрии пластины, упругих модулях, её плотности, параметрах акселерометра (размеры и масса) получить на заданном диапазоне частот значения частотного отклика. Тогда, обратная задача: по заданным геометрии пластины, её плотности, параметрам акселерометра, а также частотному отклику колебаний в некотором диапазоне найти такие упругие параметры материала, при которых результат моделирования частотного отклика совпадает с экспериментальным.

2.1. Прямая задача

В рамках гипотезы Кирхгофа-Лява и классической теории ламинатов система уравнений упругости сводится к системе из трёх уравнений на компоненты вектора перемещения двумерной срединной плоскости пластины в трёхмерном пространстве. Данная система выписана в книге [4]. Для решения задачи о колебаниях производится переход в частотную область, в связи с чем решение перестаёт зависеть от времени, а в системе уравнений возникает параметр – частота колебаний. Краевая задача для данной системы формулируется с граничными условиями Дирихле на зажатом в вибростенде конце пластины и нулевыми граничными условиями Неймана на оставшейся границе. Решение серии таких краевых задач с разным параметром частоты колебаний позволяет получить частотный отклик колебаний, т.е. решить прямую задачу.

Системы уравнений дискретизуются с помощью метода конечных элементов, затем полученные системы линейных уравнений решаются солвером для разреженных матриц UMFPACK. Также для ускорения вычислений используется распараллеливание по частоте колебаний с помощью OpenMP, в силу независимости краевых задач внутри серии. Алгоритм решения прямой задачи реализован с помощью библиотеки jax для языка Python, позволяющей использовать автоматическое дифференцирование солвера прямой задачи.

2.2. Обратная задача

Обратная задача формулируется как задача поиска минимума функционала ошибки между численным решением и экспериментальным частотным откликом:

$$\min_{\theta} L(\theta) = \min_{\theta} \left[\frac{1}{N} \sum_{n=1}^{N} \left(\log|F(\omega_n, \theta)| - \log\left|F_{\exp}(\omega_n)\right| \right) \right]$$

где θ — вектор неизвестных упругих модулей, F_{exp} — экспериментально полученный частотный отклик, F — численное решение, $\{\omega_n\}_{n=1}^{N}$ — дискретный набор частот.

Минимизация функционала производится численно, методом дифференциальной эволюции и алгоритмом simplicial homology global optimisation для глобальной оптимизации и trust-region методом второго порядка для локальной оптимизации. При этом для вычисления градиентов и матрицы Гессе используется автоматическое дифференцирования солвера прямой задачи.

3. Результаты

Решения обратной задачи были получены для синтетических данных: по заданным упругим модулям решалась прямая задача, затем полученный частотный отклик подавался на вход алгоритма решения обратной задачи. Рассматривались композитные прямоугольные пластины с размерами 100 мм х 20 мм с различными толщиной и числом слоёв, в вибростенде был зажат один из торцов размера 20 мм. Были проведены расчёты для различных типов слоёв: рассматривались однонаправленные слои с выделенным направлением армирующих нитей, а также слои с полотняным переплетением нитей, при этом упругие модули подбирались так, чтобы соответствовать композитам из углеродного волокна и эпоксидной смолы [5]. Укладки слоёв (направление каждого слоя в ламинате относительно осей пластины) рассматривались как симметричные (например (0°, 30°, 0°)), так и несимметричные (например (0°, 30°, 60°)).

По результатам моделирования было выявлено, что постановки с большим количеством слоёв требуют большего времени работы алгоритма обратной задачи, а также снижают точность

решения по сравнению с тонкими пластинами. Также было обнаружено, что для решения обратной задачи для тонкой пластины с малым числом слоёв требования к ширине диапазона частот могут быть снижены, что позволяет упростить эксперимент, рассматривая более узкий интервал частот без потери точности и скорости решения обратной задачи.

Литература

1. Li H., Xue P., Zhang T., and Wen B. Nonlinear vibration study of fiber-reinforced compo-site thin plate with strain dependent property based on strain energy density function method // Mechanics of Advanced Materials and Structures. 2020. Vol. 27, No. 9. P. 761-773.
2. Romanov V.G., Weng C.I, Chen T.C. An inverse problem related to layered elastic plate // Journal of Inverse and Ill posed Problems. 2000. Vol. 8, No. 4. P. 421-447.
3. Aksenov V., Vasyukov A., Beklemysheva K. Acquiring elastic properties of thin composite structure from vibrational testing data // Journal of Inverse and Ill-posed Problems. 2024. Vol. 32, No. 3. P. 467-484.
4. Reddy J.N. Mechanics of Laminated Composite Plates and Shells. Boca Raton, FL: CRC Press, 2003.
5. Rahmani H., Najaf S. H. Mahmoudi, Ashori A., and Golriz M. Elastic Properties of Carbon Fibre-Reinforced Epoxy Composites // Polymers and Polymer Composites. 2015. Vol. 23, No. 7. P. 475-482.

РЕШЕНИЕ ЗАДАЧИ СЕМАНТИЧЕСКОЙ СЕГМЕНТАЦИИ ДЛЯ РАБОТЫ ВИДЕО МОДУЛЯ БЕСПИЛОТНОГО РОБОТА[1*]

Р.Е. Ломовский[1], М.И. Кумсков[1], В.М. Буданов[2]

[1]Механико-математический факультет МГУ имени М.В. Ломоносова,
[2]НИИ Механики МГУ имени М.В. Ломоносова

В статье рассматривается семантическая сегментация видео для модуля восприятия беспилотного робота с использованием датчиков LiDAR и камер. Основное внимание уделяется алгоритму U-Net и различным архитектурам предобученных энкодеров, позволяющим извлекать признаки из изображений. Обучение проводилось в два этапа с разным количеством классов, а также использовалось квантование в процессе обучения для уменьшения размера модели и повышения производительности. Представлены результаты экспериментов на датасете RUGD, оцененные по метрике IoU (Intersection over Union). Полученные результаты подтверждают, что выбор архитектуры энкодера влияет на качество сегментации, что критично для работы беспилотного робота.

Ключевые слова: semantic segmentation, U-Net, quantization aware training, беспилотный ровер, модуль восприятия

1. Введение

Перед нашим коллективом стоит задача создания программного обеспечения для управления колесным роботом. Для получения информации об окружающей среде и состоянии самого робота предполагается использование большого количества датчиков. В частности, для визуального ориентирования, выявления препятствий, оценки состояния сцены предполагается использование LiDAR и камеры. По отдельности эти два датчика имеют ряд недостатков, которые можно нивелировать их совместным использованием. Видео с камер может дать нам полезную семантическую информацию, которую можно совместить с лидарным облаком точек для более подробного описания состояния окружающей среды. Автоматическому извлечению семантической информации из видео будет посвящена данная статья.

2. Постановка задачи беспилотного робота

Подойти к задаче создания ПО для беспилотного робота довольно трудно. Предлагается рассмотреть ее как частный случай задачи **беспилотного автомобиля** (англ. self-driving-car) и использовать наработанный алгоритм, преобразовав его с учетом нашей специфики. Задача управления беспилотным автомобилем состоит из следующих блоков:

- Localization – корректировка положения объекта в пространстве,
- Perception – восприятие того, что происходит вокруг робота (именно в этом блоке предполагается использовать камеры и сегментировать препятствия),
- Prediction – предсказания траекторий движения объектов сцены,
- Planning – планирование траектории движения робота,
- Control – перевод высокоуровневых инструкций в низкоуровневые (поверни колесо, ускорься, остановись и т.д.),
- HD maps – получение карты сцены с высокой точностью, (не актуально для нашей задачи, получение такой карты не предполагается).

С точки зрения обмена данными, эти блоки соотносятся следующим образом:

[1*] Работа выполнена при частичной поддержке проекта 23-Ш01-11 «Исследование интеллектуальной робототехнической системы, оснащенной ветроэнергетической установкой» НОШ МГУ «Космос».

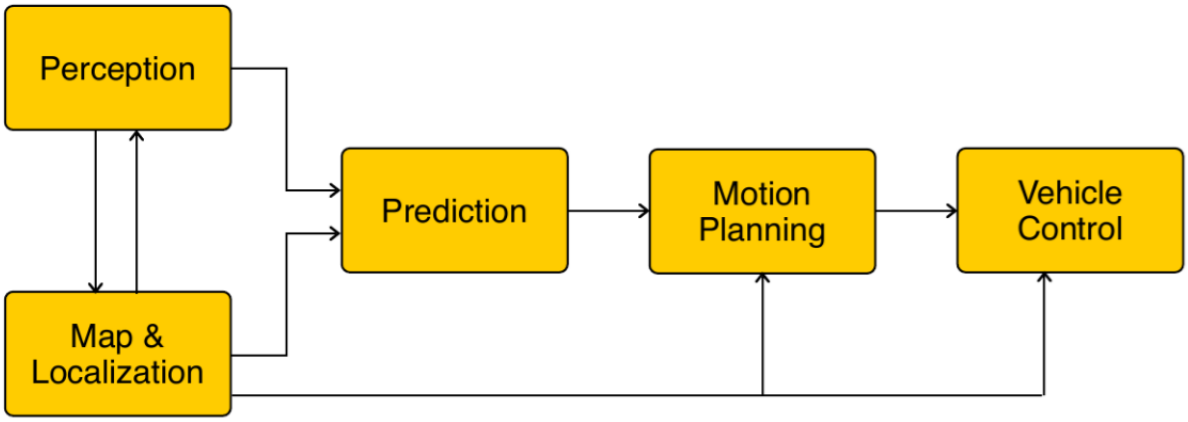


Рис. 1.

Данная статья посвящена решению задачи восприятия (Perception), остановимся на ней подробнее.

3. Модуль восприятия. LiDAR и семантическая сегментация видео

В условиях нашей задачи, предполагается использование двух средств получения визуальной информации о состоянии окружающей среды:

- камера, установленная на передней части робота,
- LiDAR.

3.1. LiDAR

LiDAR (англ. Light Detection and Ranging «обнаружение и определение дальности с помощью света») – лазерный локатор, использующий технологию испускания лазером волн оптического диапазона с дальнейшей регистрацией лазерных импульсов, которые были рассеяны объектами). Примерная схема устройства LiDAR:

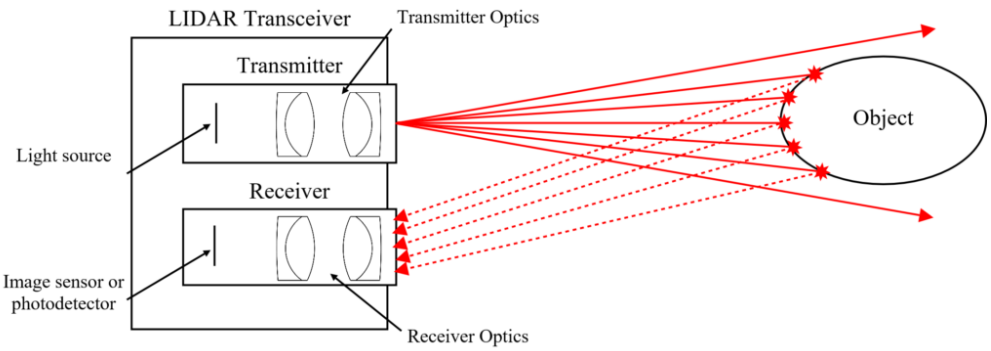


Рис. 2.

Не вдаваясь в схему работы данного устройства, будем считать, что с его помощью в некоторые дискретные моменты времени (порядка 10 раз в секунду) мы будем получать координаты точек в системе координат, связанной с роботом, в которых произошло отражение лидарных лучей. Такой набор точек в некоторый момент времени называется point-cloud (англ. «облако точек»). Такая информация крайне полезна в задаче оценки состояния среды, но имеет ряд недостатков, вот некоторые из них:

- лидарные лучи не дают информацию о цвете объектов от которых они отражаются (некоторые объекты могут иметь схожую форму в виде облака точек, без информации о цвете их может быть трудно различить);
- лидарные лучи могут отражаться от дождя, тумана или снега, что может сильно «зашумить» картину сцены.

Так или иначе, непонимание лидаром структуры препятствий - потенциально большая проблема при оценке состояния среды (например, лидар может не видеть разницы между преграждающими дорогу камнями и травой: первое препятствие придется объехать, второе можно пройти). На помощь в данном вопросе приходит камера, на основе показаний которой можно сделать вывод о семантической структуре объектов, произведя семантическую сегментацию.

3.2. Задача семантической сегментации

Семантичексая сегментация [1] (англ. semantic segmentation) – присвоение каждому пикселю изображения метки определенного класса.

Самым современным и распространенным подходом решения задачи семантической сегментации является подход глубоких нейронных сетей различных архитектур, наиболее распространенными из которых являются:

- Unet [2] – архитектура, изначально созданная для сегментации медицинских изображений, нашла широкое применение во многих областях;
- DRN-A-50 (26, 40) [3] – группа архитектур, описанная в статье Dilated Residual Networks;
- HRNet [4].

3.3. U-Net

Данная архитектура представляет собой 4 последовательных уменьшения разрешения изображения в 2 раза с помощью операции **max-pooling 2 × 2**, после чего следуют 4 увеличения разрешения также в 2 раза с помощью операции **up-conv 2 × 2**, что позволяет извлекать из изображения признаки на разных уровнях разрешающей способности.

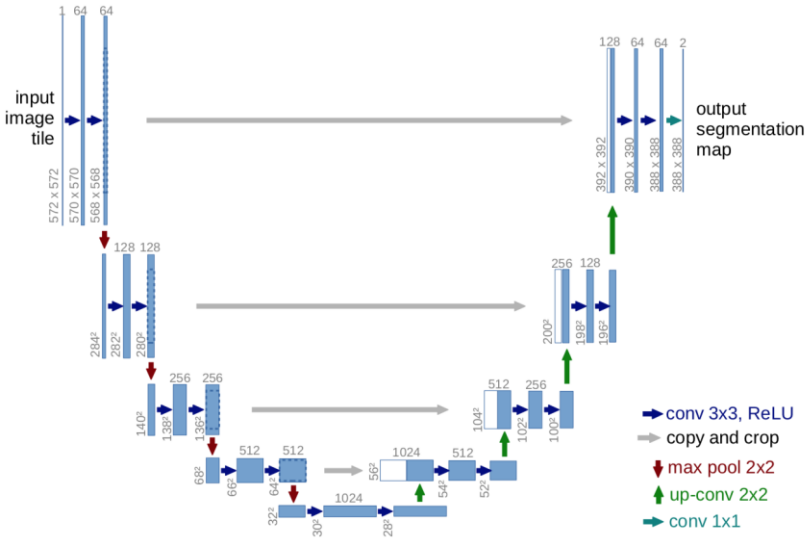


Рис. 3.

3.4. Предобученные энкодеры

Работа с изображениями в пиксельном пространстве может быть затруднительной по многим причинам, перед подачей на сегментатор стоит извлечь из него карту признаков (англ. feature map) - тензор, в каждой ячейке которого будет закодирован какой-то предобученный признак.

Для извлечения этих признаков используются предобученные нейронные сети различных архитектур, такая сеть называется **encoder** [5]. В целях улучшения качества итоговой модели или улучшения ее производительности следует перебрать несколько архитектур энкодеров, об этом подробнее в следующем разделе.

4. Способы оптимизации работы модели

Тот факт, что сегментатор должен работать автономно, накладывает некоторые ограничения на размер модели и тип данных, в котором хранятся веса. О первой проблеме следует позаботиться на этапе обучения. Если с количеством обучающих параметров U-Net[2] ничего не поделаешь, то итоговый вес модели можно уменьшить с помощью подбора архитектуры энкодера: произвести обучение для нескольких архитектур. В случае, если они покажут сопоставимое качество – выбрать ту, которая «меньше» (об этом подробнее в разделе «вычислительные эксперименты»).

Вторая проблема решается за счет применения квантования [6] – преобразования параметров модели из представления с плавающей точкой в представление с более низкой точностью, например, с использованием 8-битных целых чисел.

5. Вычислительные эксперименты

5.1. Обучение сегментатора с маленьким числом классов

Эксперименты проводились с использованием фреймворка segmentation_models.pytorch [7] на языке программирования Python. Для обучения использовался открытый датасет RUGD [8] сегментированных фотографий, снятых с ровера, размеченных на классы: dirt, tree, sand, water, rock , log (классы используемые в обучении).

Обучалась связка U-Net + encoder, с энкодерами следующих архитектур:

- imm-mobilenetv3_small_minimal_100 [9] – 0.43 млн. обучаемых параметров,
- efficientnet-b0 [10] – 4 млн. обучаемых параметров,
- vgg11 [11] – 9 млн. обучаемых параметров,
- resnet18 [12] – 11 млн. обучаемых параметров,
- resnext50_32x4d [13] – 22 млн. обучаемых параметров.

Качество модели измерялось по метрике IoU [14] (Intersection over Union), считая попиксельные пересечения и объединения ненулевых классов (0 – класс void, недетектируемый) в предсказанной и ground truth масках.

По результатам проведенных вычислительных экспериментов были получены следующие результаты.

Таблица. 1.

encoder	кол-во обучаемых параметров	dice_loss	IoU - score
timm-mobilenetv3_sall_minimal_100	0.43 млн.	0.07892	0.8679
efficientnet-b0	4 млн.	0.06945	0.8817
vgg11	9 млн.	0.0752	0.873
resnet18	11 млн.	0.07515	0.8734
se_resnext50_32x4d	22 млн.	0.06947	0.8818

Приведем примеры работы сегментатора с энкодером timm-mobilenetv3_sall_minimal_100 (класс ’tree’).

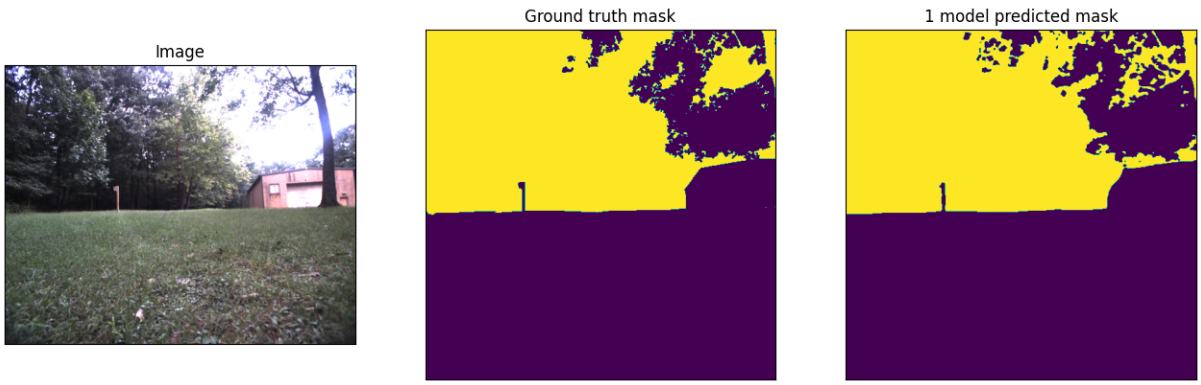


Рис. 4.

5.2. Обучение сегментатора с расширенным количеством классов

В дальнейших экспериментах мы пошли по направлению увеличения количества полезных классов с целью лучшей интерпретации сцены: в обучающий датасет были добавлены классы grass, pole, sky, vehicle, container/generic-object, asphalt, gravel, building, mulch, rock-bed, bicycle, person, fence, bush, sign, bridge, concrete, picnic, table.

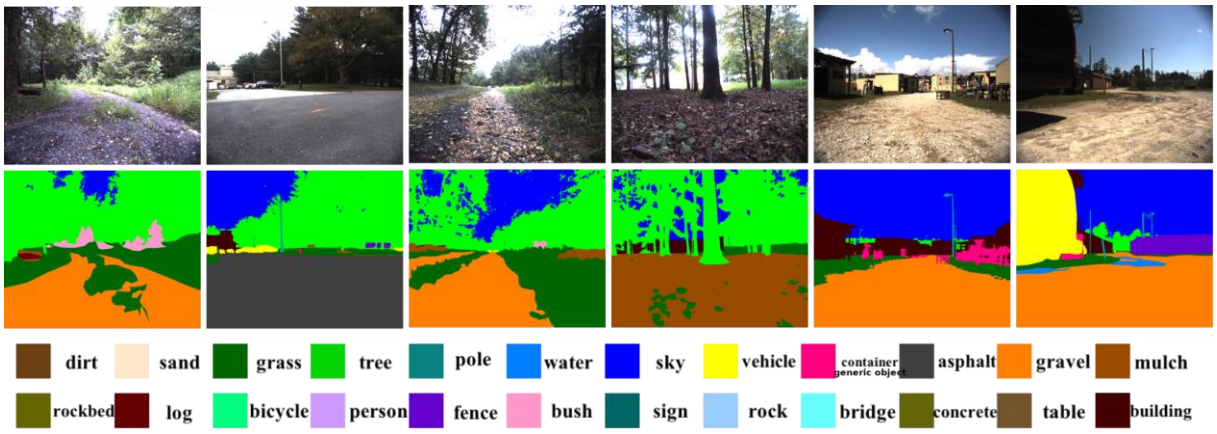


Рис. 5.

После обучения представленных архитектур, качество модели по метрике IoU ожидаемо снизилось, но осталось приемлемым.

Таблица. 2.

encoder(на 24 класса)	кол-во обучаемых параметров	IoU - score
timm-mobilenetv3_small_minimal_100	0.43 млн.	0.7815
efficientnet-b0	4 млн.	0.7771
vgg11	9 млн.	0.7911

По итогу обученная на большем количестве классов модель показывает достаточное качество для локализации объекта в пространстве.

5.3. Применение квантования во время обучения

Для увеличения вычислительной эффективности модели на CPU было решено применить **квантование в процессе обучения** [16] с помощью фреймворка **NNCF** [15] (Neural Network Compression Framework). Квантование уже обученной модели часто приводит к существенной

потере ее качества, квантование в процессе обучения помогает избежать существенного ухудшения метрик, поскольку в процессе обучения модель имеет возможность «подсмотреть»за новым типом представления данных (весов модели). Применение **QAT** [16] оказалось удачным: метрики заново обученных моделей практически не изменились (в районе 0.87 по IoU в зависимости от архитектуры энкодера). Подход признан удачным и будет использоваться в дальнейшем для существенного уменьшения размеров модели и приведения ее весов к более удобному 8-битовому представлению.

Литература

1. Jonathan Long, Evan Shelhamer, Trevor Darrell. Fully Convolutional Networks for Semantic Segmentation. arXiv, 2015. URL: https://arxiv.org/pdf/1411.4038.
2. Olaf Ronneberger, Philipp Fischer, Thomas Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. arXiv, 2015. URL: https://arxiv.org/abs/1505.04597.
3. YSaeideh Yousefzadeh, Hamidreza Pourreza, Hamidreza Mahyar.A Triplet-loss Dilated Residual Network for High-Resolution Representation Learning in Image Retrieval. arXiv, 2023. URL: https://arxiv.org/pdf/2303.08398.
4. Shun Takagi, Li Xiong, Fumiyuki Kato, Yang Cao, Masatoshi Yoshikawa. HRNet: Differentially Private Hierarchical and Multi-Resolution Network for Human Mobility Data Synthesization. arXiv, 2024. URL: https://arxiv.org/abs/2405.08043.
5. Yann LeCun, Yoshua Bengio & Geoffrey Hinton. Deep learning. Nature. 2015.
6. Jiwei Yang, Xu Shen, Jun Xing, Xinmei Tian, Houqiang Li, Bing Deng, Jianqiang Huang, Xiansheng Hua. Quantization Networks. 2019.
7. Segmentation_models.pytorch. – URL: https://github.com/qubvel-org/segmentation_models.pytorch.
8. RUGD Team. RUGD Dataset for Semantic Segmentation, 2019. URL: http://rugd.vision/, дата публикации: 1 сентября 2019.
9. Andrew G. Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, Hartwig Adam. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. arXiv, 2017. URL:https://arxiv.org/abs/1704.04861.
10. Mingxing Tan, Quoc V. Le. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks.arXiv, 2019. URL: https://arxiv.org/abs/1905.11946.
11. Karen Simonyan, Andrew Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv, 2014. URL: https://arxiv.org/abs/1409.1556.
12. Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun. Deep Residual Learning for Image Recognition.arXiv, 2015. URL:https://arxiv.org/abs/1512.03385.
13. Saining Xie, Ross Girshick, Piotr Dollár, Zhuowen Tu, Kaiming He. Aggregated Residual Transformations for Deep Neural Networks. arXiv, 2016. URL: https://arxiv.org/abs/1611.05431.
14. Hamid Rezatofighi, Nathan Tsoi, JunYoung Gwak, Amir Sadeghian, Ian Reid, Silvio Savarese. Generalized Intersection over Union: A Metric and A Loss for Bounding Box Regression. arXiv, 2015. URL: https://arxiv.org/abs/1902.09630.
15. Neural Network Compression Framework. URL: https://github.com/openvinotoolkit/nncf.
16. Eunhyeok Park, Sungjoo Yoo, Peter Vajda. Quantization Networks. 2019. – URL: https://openaccess.thecvf.com/content_CVPR_2019/html/Yang_Quantization_Networks_CVPR_2019_paper.html.

ЛИЕВО-НИЛЬПОТЕНТНЫЕ ЦЕНТРАЛЬНО СУЩЕСТВЕННЫЕ КОЛЬЦА[1*2*]

О.В. Любимцев[1], А.А. Туганбаев[2]

[1]Нижегородский государственный университет им. Н.И. Лобачевского,
[2]Национальный исследовательский университет МЭИ

Кольцо R называется центрально существенным, если оно либо коммутативно, либо для любого его нецентрального элемента a существуют такие ненулевые центральные элементы x и y, что $ax = y$. В работе изучаются центрально существенные кольца, которые нильпотентны как кольца Ли (центрально существенные кольца не всегда нильпотентны как кольца Ли, так как известно, что они не обязаны быть PI кольцами). Доказана лиево-нильпотентность артиновых справа центрально существенных колец с конечно порожденной присоединенной группой; полусовершенных центрально существенных колец с периодической группой единиц и нильпотентным радикалом Джекобсона; цепных справа центрально существенных колец.

Ключевые слова: центрально существенное кольцо, лиево-нильпотентное кольцо, присоединенная группа кольца.

1. Введение

Все кольца, рассматриваемые в работе, являются ассоциативными, не обязательно с единицей. Кольцо R называется *центрально существенным*, если оно либо коммутативно, либо для любого его нецентрального элемента a существуют такие ненулевые центральные элементы x и y, что $ax = y$. Если R – центрально существенное кольцо с ненулевой единицей, то существование указанных центральных элементов гарантировано для любого ненулевого элемента $a \in R$. Основные результаты о центрально существенных кольцах приведены в монографии [1].

Хорошо известно, что каждое ассоциативное кольцо R можно рассматривать как кольцо Ли относительно лиевого умножения $[r, s] = rs - sr$ для всех $r, s \in R$. Для элементов $r_1, r_2, \ldots, r_n$. из R, Ли-коммутатор $[r_1, r_2, \ldots, r_n]$ определяется индуктивно: $[r_1, r_2, \ldots, r_n] = [[r_1, r_2, \ldots, r_{n-1}], r_n]$ для всех $n \neq 2$. Кольцо R называется *лиево-нильпотентным* (*Lie nilpotent*), если существует такое положительное целое n, что $[r_1, r_2, \ldots, r_{n+1}] = 0$ для всех $r_1, r_2, \ldots, r_{n+1}$ из R. Наименьшее n с таким свойством есть *класс лиево-нильпотентности* R. Через $R^{(2)}$ обозначается идеал, порожденный всеми элементами $[r, s]$, и, индуктивно, $R^{(n)}$ – идеал, порожденный всеми $[a, b]$, где $a \in R^{(n-1)}$, $b \in R$. Кольцо R называется *сильно лиево-нильпотентным* (*strongly Lie nilpotent*), если $R^{(m)} = 0$ для некоторого m. Для пары элементов $r, s \in R$ мы пишем $[r, {}_n s]$ для $[r, s, \ldots, s]$, где s берется ровно n раз. Кольцо R называется *лиево-энгелевым*, если для всех $x, y \in R$, существует такое натуральное $n = n(x, y)$, что $[x, {}_n y] = 0$.

Множество всех элементов кольца R образует полугруппу относительно операции присоединенного умножения $r \circ s = r + s + rs$ для всех $r, s \in R$. Множество

$R^\circ = \{r \in R \mid r \circ s = 0 = s \circ r$ *для некоторого* $s \in R\}$

образует группу, которая называется *присоединенной группой* кольца R. Если R имеет 1, то $1 + R^\circ$ совпадает с группой единиц $U(R)$ кольца R и отображение $r \to 1 + r$ для $r \in R^\circ$ является изоморфизмом R° на $U(R)$.

Приведем необходимые обозначения и определения. Мы обозначаем $C(R)$ и $J(R)$ центр и радикал Джекобсона кольца R, соответственно. Хорошо известно, что радикал Джекобсона артинова справа кольца является нильпотентным. Наименьшее n, такое, что $J(R)^n = 0$, называется *индексом нильпотентности* $J(R)$. Кольцо R называется *локальным*, если $R/J(R)$ – тело. Модуль

[1*] Исследование О.В. Любимцева выполнено при поддержке Минобрнауки РФ, проект FSWR-2023-0034.
[2*] Исследование А.А.Туганбаева выполнено при поддержке гранта Российского научного фонда № 22-11-00052.

M называется *цепным*, если множество всех подмодулей в *M* линейно упорядочено по включению. Кольцо *R*, цепное как правый *R*-модуль, называется *цепным справа*. Кольцо *R* называется *полусовершенным*, если факторкольцо *R/J(R)* артиново и все идемпотенты *R/J(R)* могут быть подняты по модулю *J(R)* до идемпотентов в *R*. Кольцо называется *прямо неразложимым*, если это кольцо нельзя представить в виде прямой суммы двух ненулевых идеалов. Группа называется *локально нильпотентной*, если каждая ее конечно порожденная подгруппа нильпотентна. Кольцо *R* называется *локально конечным*, если каждое конечное подмножество в *R* мультипликативно порождает конечную полугруппу.

2. Основные результаты

Следующая лемма является аналогом 1.1.5 из [1] для колец без единицы.

Лемма 1. Пусть *R* – центрально существенное кольцо, $r, x_1, \ldots, x_n, y_1, \ldots, y_n \in R$ и s – неделитель нуля в *R*. Тогда из

$$\begin{cases} x_1y_1 + x_2y_2 + \cdots + x_ny_n = s \\ x_1ry_1 + x_2ry_2 + \cdots + x_nyr_n = 0 \end{cases}$$

следует, что $r = 0$.

Ясно, что если центрально существенное кольцо разложимо в прямую сумму идеалов, то изучение строения этого кольца сводится к изучению строения каждого идеала. Так как идемпотенты центрально существенного кольца центральны (см. 1.1.4 из [1]), то при изучении центрально существенных полусовершенных колец достаточно рассмотреть прямо неразложимые кольца.

Лемма 2. Пусть *R* – прямо неразложимое артиново справа центрально существенное кольцо. Тогда *R* либо нильпотентное кольцо, либо такое унитальное кольцо, что *R/J(R)* – поле.

Предложение 3. Пусть *R* – конечное прямо неразложимое центрально существенное кольцо, которое не нильпотентно. Тогда группа единиц *U(R)* нильпотентна.

С учетом леммы 1, леммы 2 и предложения 3, имеет место следующая

Теорема 4. Артиново справа центрально существенное кольцо *R* с конечно порожденной присоединенной группой R° является лиево-нильпотентным кольцом класса нильпотентности не выше *n*, где *n* – индекс нильпотентности радикала Джекобсона *J(R)* кольца *R*.

Следствие 5. Если *R* – артиново справа центрально существенное кольцо с конечно порожденной группой *U(R)*, то $R = C(R) + J(R)$.

Следствие 6. Пусть *R* – артиново справа не нильпотентное центрально существенное кольцо с конечно порожденной группой R°. Тогда выполнены следующие условия:

- *R* – артиново слева;
- *R* удовлетворяет условию минимальности для двусторонних идеалов.

Для полусовершенных центрально существенных колец с периодической группой единиц *U(R)* имеем

Теорема 7. Пусть *R* – полусовершенное центрально существенное кольцо с периодической группой единиц *U(R)*. Тогда

- *U(R)* – локально нильпотентная группа;
- $R = C(R) + J(R)$;
- *R* – локально конечное лиево-энгелево кольцо.

Если при этом радикал Джекобсона *J(R)* кольца *R* нильпотентен (в частности, если *R* артиново справа), то группа *U(R)* нильпотентна и *R* – сильно лиево-нильпотентное кольцо.

Следствие 8. Для полусовершенного центрально существенного кольца *R* с периодической группой единиц *U(R)* следующие утверждения эквивалентны:

- *R* – артиново справа;
- *R* – артиново слева;
- *R* удовлетворяет условию минимальности для двусторонних идеалов.

Как показывает следующая теорема, цепные справа центрально существенные кольца с единицей лиево-нильпотентны.

Теорема 9. Пусть R – артиново справа, цепное справа центрально существенное кольцо с единицей и радикалом Джекобсона индекса нильпотентности n. Тогда R является лиево-нильпотентным кольцом класса нильпотентности $\leq n$.

3. Замечания и открытые вопросы

Замечание 10. Существуют центрально существенные кольца, которые не являются PI кольцами, см. теорему 3.4.5 из [1]. Следовательно, существуют не лиево-нильпотентные центрально существенные кольца.

Известно, что центрально существенная групповая алгебра FG конечной группы G над полем F характеристики 0 или $p > 0$, где p не делит порядок группы G, коммутативна; см. замечание 3.2.2 из [1] Следовательно, интерес представляют только групповые алгебры простой характеристики, делящей порядок группы G.

Замечание 11. Если G – конечная группа, то центрально существенная групповая алгебра FG над полем F лиево-нильпотентна.

Замечание 12. Если R – артиново справа центрально существенное кольцо с циклической группой единиц $U(R)$, то R коммутативно.

Замечание 13. Пусть R – конечное центрально существенное кольцо с единицей и $J(R)^2 = 0$. Тогда R коммутативно.

Открытый вопрос. Верно ли, что любое артиново справа центрально существенное кольцо лиево-нильпотентно?

Литература

1. Туганбаев А.А. Центрально существенные кольца. Москва: Флинта, 2022, 160 с. ISBN 978-5-9765-5093-3.

РАСПОЗНАВАНИЕ ПРЕПЯТСТВИЙ С ИСПОЛЬЗОВАНИЕМ СЕМАНТИЧЕСКОЙ СЕГМЕНТАЦИИ ПРИ ДВИЖЕНИИ АВТОНОМНОГО МОБИЛЬНОГО РОБОТА[1*]

В.В. Маносин[1], М.И. Кумсков[1], В.М. Буданов[2]

[1] Механико-математический факультет МГУ имени М.В. Ломоносова,
[2] Научно-исследовательский институт механики МГУ имени М.В. Ломоносова

В данной работе рассмотрена задача навигации автономного мобильного робота по незнакомой пересеченной местности с использованием современных методов семантической сегментации. Более конкретно исследуется задача поиска и распознавания препятствий, возникающих по ходу движения робота. Проводится обзор и сравнение современных нейросетевых архитектур, предназначенных для семантической сегментации, таких как FCN, SegNet, U-Net, применительно к поставленной задаче. Вычислительные эксперименты осуществлялись на Robot Unstructured Ground Driving (RUGD) Dataset с измененным числом классов. С помощью архитектуры U-Net на данном наборе данных удалось достичь качества сегментации 94.84% по метрике IoU.

Ключевые слова: мобильный робот, распознавание препятствий, нейронные сети, семантическая сегментация.

1. Введение

1.1. Мобильный робот

Мобильный робот (ровер) представляет собой высокотехнологичное устройство, специально разработанное и предназначенное для автономного перемещения при различных сценариях. Высокая маневренность ровера обеспечивается использованием колесного шасси. Однако необходимо отметить, что колесная система является одновременно и недостатком, так как позволяет преодолевать лишь небольшие неровности ландшафта. Информация об окружающей среде поступает с помощью высококачественной камеры и датчика LiDAR (Light Detection and Ranging), установленных на мобильном роботе. Сочетание данных с этих двух устройств позволяет как идентифицировать объекты, так и с высокой точностью определять расстояние до них.

2. Задача распознавания препятствий

2.1. Описание задачи

Задача распознавания препятствий направлена на создание алгоритмов, способных выявлять преграды, мешающие продвижению ровера, в окружающей среде. Эти преграды могут быть как статическими (камни, деревья, здания), так и динамическими (люди, животные). Главная цель распознавания препятствий заключается в том, чтобы предоставить роверу информацию о рельефе окружающей местности. С помощью полученной информации автономный робот может эффективно ориентироваться в окружающей среде, корректировать свой маршрут и избегать столкновений с различными потенциально опасными объектами.

[1*] Работа выполнена при частичной поддержке проекта 23-Ш01-11 «Исследование интеллектуальной робототехнической системы, оснащенной ветроэнергетической установкой» НОШ МГУ «Космос».

2.2. Семантическая сегментация в задаче распознавания препятствий

Рассмотрим более подробно способ решения задачи распознавания препятствий с помощью семантической сегментации изображения. Семантическая сегментация – это задача компьютерного зрения, которая заключается в классификации каждого пикселя изображения, то есть отнесении его к определенному классу. В отличие от обычной классификации, где изображение зачастую помечено всего одной меткой, семантическая сегментация обеспечивает более подробную информацию о содержимом изображения, разделяя объекты попиксельно.

Решение задачи семантической сегментации подразумевает создание детализированной попиксельной цветовой маски изображения, где каждому пикселю присваивается определённый цвет в зависимости от того, к какому типу объектов он относится. В частности, объекты, идентифицируемые как преграды, будут отмечены специфическим цветом, что позволит как определять их местоположение, так и понимать их форму и размеры. Также существует возможность разделения большого класса преград на более узкие (деревья, кусты, камни, здания, ямы) для более точного распознавания объектов. Однако решение такой задачи может быть связано с усложнением модели и существенным увеличением вычислений, что не всегда допустимо в условиях ограниченности вычислительных ресурсов, поэтому данная задача требует дальнейших исследований.

2.3. Сочетание камеры и LiDAR

Камера и LiDAR могут быть связаны для решения задачи распознавания препятствий с помощью семантической сегментации изображения. Эти два датчика отлично дополняют друг друга: камера предоставляет информацию о цвете и текстуре объектов, что полезно для идентификации и сегментации различных классов препятствий. В то время как с помощью LiDAR можно определить расстояние до объектов и построить их 3D-модель, позволяя более точно идентифицировать местоположение препятствий и их форму. Также объединение информации разной природы способствует улучшению точности предсказания за счет компенсации недостатков друг друга. Например, в условиях низкой освещенности камера выдает нечеткую зашумленную картинку, из которой может быть сложно извлечь какие-либо признаки о препятствиях, в свою очередь, LiDAR работает в любое время суток и при любой освещенности без ограничений. Таким образом, модели, использующие комбинированные данные, могут быть более эффективны для решения задачи распознавания препятствий.

2.4. Использование нейронных сетей в задаче распознавания препятствий

В настоящее время методы глубокого обучения, представляющие собой мощные архитектуры нейронных сетей, стали основой для решения множества задач в области компьютерного зрения, в частности, задачи семантической сегментации изображения. Эти подходы обеспечивают значительные улучшения в качестве и точности сегментации по сравнению с традиционными методами, позволяя эффективно справляться со сложными изображениями с большим количеством объектов на них.

3. Архитектуры нейронных сетей для семантической сегментации

3.1. Fully Convolution Network

FCN, или Fully Convolutional Network, представляет собой архитектуру полностью сверточной нейронной сети, которая впервые продемонстрировала возможность успешного применения таких нейросетей для решения задачи семантической сегментации изображения, значительно превосходя по метрикам классические методы [1]. Основной принцип работы FCN заключается в переосмыслении классических классификационных архитектур путём адаптации их структуры, а именно замены полносвязных слоев на слои свертки, что позволило нейросети обрабатывать входные данные произвольного размера.

Большинство современных архитектур для сегментации изображений разработаны на основе концепции Fully Convolutional Networks. То есть они используют исключительно свёрточные слои, которые позволяют извлекать признаки из изображений, а также слои пулинга, уменьшающие пространственное разрешение для упрощения вычислений и выделения наиболее значимых характеристик. В дополнение к этому используется апсемплинг, необходимый для увеличения размерности изображения и восстановления его исходного разрешения для построения корректной сегментационной маски.

3.2. U-Net

U-Net – это архитектура сверточной нейронной сети, разработанная для решения задачи семантической сегментации биомедицинских изображений [2]. U-Net в настоящее время считается одной из самых популярных и широко используемых моделей для семантической сегментации благодаря своей способности достигать хороших результатов при сравнительно небольшом объеме вычислительных ресурсов и минимальных требованиях к размерам обучающих наборов данных. Это делает её универсальным инструментом для решения многих задач семантической сегментации далеко за пределами биомедицинских изображений.

Архитектура U-Net состоит из двух основных компонент: кодировщика и декодировщика, связанных между собой. Кодировщик выполняет функцию сжатия входного изображения и извлечения признаков. Процесс начинается с применения двух сверток размера 3x3, после каждой из которых используется функция активации ReLU. Затем применяется операция пулинга, которая снижает пространственные размеры входных данных, что уменьшает количество вычислений и выделяет наиболее существенные признаки изображения. Данная последовательность операций (свертка + ReLU + свертка + ReLU + пулинг) повторяется четыре раза, каждый раз уменьшая размерность и увеличивая количество каналов, что обеспечивает извлечение всё более сложных признаков из исходного изображения.

Декодировщик осуществляет обратное преобразование сжатых данных, однако в оригинальной архитектуре не восстанавливает размер изображения до исходного. Он также использует две последовательные свертки 3x3 с активацией ReLU после каждого применения. В дополнение к этому применяется транспонированная свертка, увеличивающая пространственные размеры входных данных. Как и в кодировщике, данный блок операций (свертка + ReLU + свертка + ReLU + транспонированная свертка) также повторяется четыре раза.

Финальный блок модели включает в себя две свертки 3x3 с активацией ReLU после каждой свертки. Для окончательной интерпретации выходных данных используется свертка размером 1x1, которая отвечает за сопоставление каждого вектора признаков с желаемым количеством классов, что позволяет получить желаемую сегментационную маску.

Одной из ключевых особенностей архитектуры U-Net является использование skip connections [3]. Они обеспечивают объединение слоев кодировщика с соответствующими слоями декодировщика. Эти соединения позволяют использовать как низкоуровневые, так и высокоуровневые признаки, что значительно повышает качество сегментации. Благодаря этому подходу, модель способна восстанавливать мелкие детали объектов, которые могли бы быть утеряны во время прохода через слои кодировщика.

3.3. SegNet

SegNet – это архитектура сверточной нейронной сети, разработанная для семантической сегментации дорожных сцен и сцен в помещениях [4]. Основной акцент при проектировании этой архитектуры был сделан на обеспечение высокой производительности нейронной сети как с точки зрения времени вычислений, так и с точки зрения экономии используемой памяти. Это особенно важно, поскольку в реальных приложениях, таких как автономные автомобили, требуется проводить вычисления в реальном времени. Также обычно вычислительные ресурсы на мобильных устройствах существенно ограничены, что является дополнительным плюсом в пользу SegNet.

Архитектура SegNet, как и U-Net, состоит из кодировщика и декодировщика. Кодировщик в SegNet состоит из 13 последовательных применений свертки, после каждой из которых применяется процедура нормализации, которая улучшает сходимость и стабильность обучения модели, а также функция активации ReLU, вводящая нелинейность в модель. Переход от одного блока сверток к другому осуществляется с помощью операции пулинга, что позволяет уменьшать размерность и выделять наиболее значимые признаки изображения. Для инициализации процесса обучения SegNet можно использовать заранее подготовленные веса, обученные на больших наборах данных, например, на известном наборе данных ImageNet [5]. Важной особенностью SegNet является то, что после каждого блока свертки запоминаются индексы, полученные в результате пулинга, которые затем передаются соответствующему блоку в декодировщике. Это позволяет декодировщику эффективно восстанавливать информацию, потерянную в процессе сжатия данных.

Ключевым отличием SegNet от архитектуры U-Net является метод использования информации о пулинге. Вместо того, чтобы передавать полные карты признаков, как это делает U-Net, SegNet использует только индексы пулинга, что позволяет существенно сократить объем памяти, необходимый для обучения модели и ее реализации. Каждый уровень декодировщика соотносится с соответствующим уровнем кодировщика, и таким образом декодировщик также состоит из 13 блоков сверток с процедурой нормализации и функцией активации ReLU. Переход между блоками декодировщика реализуется посредством апсемплинга, что позволяет восстанавливать разрешение изображения, постепенно увеличивая размеры выходных данных на каждом уровне. Последним слоем применяется многопеременная логистическая регрессия, которая представляет результаты в форме сегментационных масок, позволяя выделять различные классы объектов на изображении.

4. Вычислительные эксперименты

4.1. Формирование набора данных

Для проведения сравнения представленных архитектур необходимо сначала сформировать набор данных, на котором будут обучаться модели. В дальнейшем на этом же наборе будет замеряться время работы уже обученных нейросетей. Для этих целей был использован RUGD Dataset, ориентированный на семантическое понимание неструктурированной внешней среды с целью применения в приложениях автономной навигации по сложным и труднопроходимым участкам (например, лесные массивы) [6].

Набор данных состоит из видеопоследовательностей, которые были сняты с камеры, установленной на мобильном колесном роботе. Данный робот был спроектирован так, чтобы обладать компактными размерами, что позволяет ему маневрировать в загроможденной среде и успешно проходить через узкие пространства. При этом он также обладает достаточной прочностью и устойчивостью, чтобы преодолевать сложную местность, состоящую из разных типов рельефа и различных природных препятствий.

Плотные попиксельные аннотации, предоставляемые с данным набором данных, были созданы для каждого пятого кадра видеоряда, что позволяет обеспечить высокую детализацию информации о каждом элементе среды. В общей сложности в аннотациях можно встретить 24 смысловые категории, которые описывают окружающую среду и объекты в ней.

Для удобства анализа смысловые категории были разделены на две основные группы. В первую группу входят те категории, которые не представляют значительных препятствий для перемещения робота. Эти категории, такие как void, dirt, sand, grass, sky, asphalt, gravel, mulch и rock_bed, характеризуют доступные для проезда участки, где условия не оказывают серьезного влияния на способность ровера передвигаться. Во вторую группу попадают категории, представляющие собой препятствия, которые могут затруднять или полностью блокировать движение ровера. К этим категориям относятся tree, pole, water, vehicle, container, building, log, bicycle, person, fence, bush, sign, rock, bridge и picnic_table. Эти аннотации позволяют моделям обучаться различать проходимые и непроходимые участки, что крайне важно для разработки эффективных алгоритмов автономной навигации и обеспечения безопасности роботов при их перемещении по сложным ландшафтам.

4.2. Эксперименты с U-Net

Проведем вычислительные эксперименты для U-Net с предобученными на ImageNet энкодерами, такими как Resnet34 [3] и различными модификациями Mobilenetv3_small [7], (время работы замеряется как среднее время генерации масок для батчей с 32 изображениями в облачной среде Kaggle).

Таблица 1. Результаты экспериментов

Предобученный энкодер	Число параметров	Время работы	IoU
Resnet34	21 млн.	0.30 с.	94.00 %
Mobilenetv3_small_100	0.93 млн.	0.28 с.	94.84 %
Mobilenetv3_small_minimal_100	0.43 млн.	0.25 с.	93.80 %

Предсказания, полученные с помощью Resnet34 + U-Net достаточно точны, что подтверждается метрикой IoU. Однако из-за большого числа параметров обученная модель занимает много места в хранилище. Объем модели можно уменьшить за счет использования энкодера, предназначенного для мобильных устройств.

Результаты сегментации с использованием Mobilenetv3_small_100 + U-Net получаются ничуть не хуже, а даже немного лучше предыдущей архитектуры (94.84 % IoU против 94.00 % IoU). Число параметров же уменьшается больше, чем в 22.5 раза, что существенно экономит пространство на жестком диске, занимаемое обученной моделью. Также немного уменьшается время работы (на 0.02 секунды).

Последняя архитектура Mobilenetv3_small_minimal_100 + U-Net показывает худшее из 3 моделей качество сегментации, всего 93.8 % IoU, однако отставание от лучшей модели не так велико (1.04 %), выигрыш же в числе параметров почти в 49 раз по сравнению с самой тяжелой моделью. Также получено ускорение на 0.05 секунд по сравнению с Resnet34 или на 0.03 секунды по сравнению с mobilenetv3_small_100.

5. Выводы

Таким образом, архитектуры семантической сегментации полностью применимы к решению задачи распознавания препятствий при движении ровера. Полученные результаты в 94.84 % по метрике IoU с высокой скоростью работы гарантируют качественное построение сегментационных масок. Это в дальнейшем позволит строить точную карту местности, необходимую для полноценной навигации мобильного автономного робота. Для построения карты местности возможно использование сочетания таких методов как SLAM и глубокого обучения с подкреплением [8], однако это выходит за рамки представленного доклада.

Литература

1. Jonathan Long, Evan Shelhamer, Trevor Darrell Fully Convolutional Networks for Semantic Segmentation. URL: arxiv.org/abs/1411.4038 (дата обращения: 14.06.2024).
2. Olaf Ronneberger, Philipp Fischer, Thomas Brox U–Net Convolutional Networks for Biomedical Image Segmentation. URL: arxiv.org/abs/1505.04597 (дата обращения: 23.07.2024).
3. Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun Deep Residual Learning for Image Recognition. URL: arxiv.org/abs/1512.03385 (дата обращения: 24.07.2024).
4. Vijay Badrinarayanan, Alex Kendall, Roberto Cipolla SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. URL: arxiv.org/abs/1511.00561 (дата обращения: 02.08.2024).

5. Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, Li Fei-Fei ImageNet Large Scale Visual Recognition Challenge. URL: arxiv.org/abs/1409.0575 (дата обращения: 02.08.2024).
6. Maggie Wigness, Sungmin Eum, John G. Rogers III, David Han, Heesung Kwon A RUGD Dataset for Autonomous Navigation and Visual Perception in Unstructured Outdoor Environments. URL: rugd.vision/pdfs/RUGD_IROS2019.pdf (дата обращения: 21.08.2024).
7. Andrew Howard, Mark Sandler, Grace Chu, Liang-Chieh Chen, Bo Chen, Mingxing Tan, Weijun Wang, Yukun Zhu, Ruoming Pang, Vijay Vasudevan, Quoc V. Le, Hartwig Adam Searching for MobileNetV3. URL: arxiv.org/abs/1905.02244 (дата обращения: 04.09.2024).
8. Seunghyeop Nam, Changseok Woo, Sinkyu Kang, Tuan Anh Nguyen, Dugki Min SLAM_DRLnav: A SLAM-Enhanced Deep Reinforcement Learning Navigation Framework for Indoor Self-driving. URL: ieeexplore.ieee.org/document/10181042 (дата обращения: 15.09.2024).

УСЛОВИЯ ПРЕДСТАВЛЕНИЯ ПРОИЗВОЛЬНОЙ ЛИНЕЙНОЙ СИСТЕМЫ В КАНОНИЧЕСКОЙ УПРАВЛЯЕМОЙ ФОРМЕ

А.В. Мухин

Нижегородский государственный университет им. Н.И. Лобачевского

Рассматривается задача приведения произвольной линейной непрерывной стационарной системы с необратимыми матрицами входа и выхода к канонической управляемой форме. Определены условия, выражаемые в виде рангового равенства относительно матриц системы, при которых соответствующий базис существует. Главное достоинство представления системы в канонической управляемой форме – простое решение задач стабилизации по измеряемому выходу. Рассмотрены вопросы стабилизации линейных систем, приводимых к канонической управляемой форме. Получено решение в виде линейного матричного неравенства для случаев, когда произведение матриц выхода и входа имеет полный ранг.

Ключевые слова: каноническая управляема форма, статический регулятор по выходу, стабилизация, линейные матричные неравенства.

1. Введение

В зависимости от выбранного базиса одна и та же линейная система может иметь разное представление. При этом свойства самой системы являются инвариантными, т.е. сохраняются при переходе от одного базиса к другому. Существует, вообще говоря, бесконечное число различных базисов, применяющихся для тех или иных целей. Одна из них – декомпозиция пространства состояний на непересекающиеся подпространства. Добиться такого представления можно с помощью, так называемых канонических управляемых (наблюдаемых) форм. Если система приведена к такой форме, то пространство состояний раскладывается в прямую сумму управляемого (наблюдаемого) и неуправляемого (ненаблюдаемого) подпространств. Причем если подпространство устойчивых состояний принадлежит подпространству неуправляемых (ненаблюдаемых) состояний, то система стабилизируема. Алгоритм приведения линейной системы с полной обратной связью к канонической управляемой (наблюдаемой) форме можно найти в [1], с неполной обратной связью при дополнительном условии – в [2]. Заметим, что именно стабилизация систем в условиях неполной обратной связи, когда в измерении задействованы не все переменные, имеет большое практическое значение. В общем случае задача является NP-трудной [3]. Одним из подходов к ее решению является поиск частных случаев, разрешимых в рамках выпуклой оптимизации. Обзор некоторых выпуклых частных случаев приведен в [4]. Методом обнаружения таких случаев является как раз метод линейного преобразования базиса системы. В качестве примеров можно указать работы [5, 6]. Если произвольную систему с неполной обратной связью можно представить в канонической управляемой форме, то решение задачи стабилизации гарантируется в рамках выпуклой оптимизации с помощью техники линейных матричных неравенств.

Статья посвящена поиску условий приведения произвольной линейной системы к канонической управляемой форме. Рассмотрены вопросы стабилизации линейных систем, записанных в канонической управляемой форме.

2. Предварительные сведения

Дана произвольная линейная непрерывная стационарная система:

$$\begin{gathered}\dot{\tilde{x}} = \tilde{A}\tilde{x} + \tilde{B}u,\ \tilde{x}(0) = \tilde{x}_0, \\ y = \tilde{C}\tilde{x},\end{gathered} \tag{1}$$

где $\tilde{x} \in R^n$ – состояние; $u \in R^m$ – вход; $y \in R^p$ – измеряемый выход; $\tilde{A} \in R^{n\times n}$ – обратимая матрица системы; $\tilde{B} \in R^{n\times m}$ – матрица входа; $\tilde{C} \in R^{p\times n}$ – матрица выхода.

Произвольность системы означает, что она может быть как полностью, так и частично управляемой и наблюдаемой. Без потери общности полагаем, что размерности матриц $\tilde{B}$ и $\tilde{C}$ неизбыточны:

$$rank\left(\tilde{B}\right) = m < n,$$
$$rank\left(\tilde{C}\right) = p < n,$$

и $p \geq m$. Нас интересуют условия, при которых систему (1) можно привести к канонической управляемой форме. Идея приведения линейной системы к каноническим управляемым, наблюдаемым формам принадлежит Р. Калману. Алгоритм приведения системы с полной обратной связью, т.е. когда одна из матриц $\tilde{B}$ или $\tilde{C}$ обратима, изложен в [1]. Если система не является полностью управляемой, то в качестве матрицы преобразования следует взять линейно-независимые столбцы матрицы управляемости и дополнить их до полного базиса. После линейного преобразования система принимает вид:

$$\dot{x} = \begin{pmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{pmatrix} x + \begin{pmatrix} B_{11} \\ 0 \end{pmatrix} u, \qquad x(0) = x_0,$$

где $A_{11}, B_{11} \in R^{m\times m}$, m – количество линейно-независимых столбцов матрицы управляемости.

Пространство состояний системы раскладывается в прямую сумму управляемого и неуправляемого подпространств. Причем подпространство управляемых состояний совпадает с подпространством полностью управляемой подсистемы, определяемой парой (A_{11}, B_{11}) . Если неуправляемое подпространство, определяемое блоком A_{22} устойчиво, то система стабилизируема. Алгоритм приведения системы к канонической форме Калмана, когда матрицы C и B одновременно необратимы, изложен в [2]. Главное условие приведения – неполная управляемость и наблюдаемость, т.е. ранг матрицы управляемости меньше n и ранг ядра матрицы наблюдаемости отличен от нуля. В общем случае, когда матрицы C и B одновременно необратимы и никаких требований относительно управляемости и наблюдаемости не задано, этот алгоритм неприменим.

3. Необходимые и достаточные условия приведения произвольной линейной системы к канонической управляемой форме

Разобьем матрицы (1) на блоки следующим образом:

$$\tilde{A} = \begin{pmatrix} \tilde{A}_{11} & \tilde{A}_{12} \\ \tilde{A}_{21} & \tilde{A}_{22} \end{pmatrix},$$
$$\tilde{B} = \begin{pmatrix} \tilde{B}_{11} \\ \tilde{B}_{21} \end{pmatrix},$$
$$\tilde{C} = (\tilde{C}_{11} \quad \tilde{C}_{12}),$$

где $\tilde{A}_{11} \in R^{p\times p}$; $\tilde{B}_{11} \in R^{p\times m}$; $\tilde{C}_{11} \in R^{p\times p}$.

Воспользуемся методом линейных преобразований базиса. Выполним линейное преобразование базиса системы (1) посредством обратимой матрицы $T \in R^{n\times n}$. В новом базисе система (1) примет вид:

$$\dot{x} = T\tilde{A}T^{-1}x + T\tilde{B}u = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} x + \begin{pmatrix} B_{11} \\ B_{21} \end{pmatrix} u, \; x(0) = x_0, \tag{2}$$
$$y = \tilde{C}T^{-1}x = \tilde{C}x = (C_{11} \quad C_{12})x.$$

Для того чтобы система (2) удовлетворяла канонической управляемой форме необходимо и достаточно, чтобы выполнялись следующие равенства:

$$C_{12} = 0_{p\times(n-p)},$$
$$B_{21} = 0_{(n-p)\times m}, \tag{3}$$
$$A_{21} = 0_{(n-p)\times p},$$

Если это так, то система (2) примет вид:

$$\dot{x} = \begin{pmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{pmatrix} x + \begin{pmatrix} B_{11} \\ 0 \end{pmatrix} u, \qquad x(0) = x_0,$$
$$y = (C_{11} \quad 0)x.$$

Разобьём матрицу преобразования на два блока следующим образом:

$$T = \begin{pmatrix} T_p \\ T_{n-p} \end{pmatrix},$$

где $T_p \in R^{p \times n}$.

Докажем лемму:

Лемма 1. *Для существования базиса, в котором имеют место равенства* (3) *необходимо и достаточно, чтобы выполнялось ранговое условие:*

$$rank \begin{pmatrix} \tilde{C} \\ \tilde{B}^T \\ \tilde{C}\tilde{A}^T \end{pmatrix} = p. \quad (4)$$

Доказательство. Пусть равенства (3) выполняются. Запишем уравнение преобразования матрицы выхода, учитывая первое равенство (3):

$$\tilde{C}T^{-1} = (C_{11} \quad 0).$$

Так как $rank\left(\tilde{C}\right) = p$, то матрица C_{11} обратима. Значит, $T_p = C_{11}^{-1}\tilde{C}$. Пусть $T^{-1} = (Z_p \quad Z_{n-p})$, где $Z_p \in R^{n \times p}$. Запишем условие обратимости матрицы преобразования:

$$TT^{-1} = \begin{pmatrix} C_{11}^{-1}\tilde{C} \\ T_{n-p} \end{pmatrix}(Z_p \quad Z_{n-p}) = \begin{pmatrix} I_p & 0 \\ 0 & I_{n-p} \end{pmatrix}$$

В результате получаем равенства:

$$C_{11}^{-1}\tilde{C}Z_p = I_p,$$
$$C_{11}^{-1}\tilde{C}Z_{n-p} = 0_{p \times (n-p)},$$
$$T_{n-p}Z_p = 0_{(n-p) \times p},$$
$$T_{n-p}Z_{n-p} = I_{n-p}.$$

Откуда сразу находим блоки Z_p, Z_{n-p} и T_{n-p}:

$$Z_p = \tilde{C}^T(\tilde{C}\tilde{C}^T)^{-1}C_{11},$$
$$Z_{n-p} = \mathcal{N}_{\tilde{C}},$$
$$T_{n-p} = (\mathcal{N}_{\tilde{C}}^T\mathcal{N}_{\tilde{C}})^{-1}\mathcal{N}_{\tilde{C}}^T,$$

где $\mathcal{N}_{\tilde{C}}$ – ядро матрицы $\tilde{C}$.

Запишем матрицы T и T^{-1}:

$$T = \begin{pmatrix} T_p \\ T_{n-p} \end{pmatrix} = \begin{pmatrix} C_{11}^{-1}\tilde{C} \\ (\mathcal{N}_{\tilde{C}}^T\mathcal{N}_{\tilde{C}})^{-1}\mathcal{N}_{\tilde{C}}^T \end{pmatrix},$$
$$T^{-1} = (\tilde{C}^T(\tilde{C}\tilde{C}^T)^{-1}C_{11} \quad \mathcal{N}_{\tilde{C}}).$$

Так как столбцы $\mathcal{N}_{\tilde{C}}$ образуют ядро матрицы $\tilde{C}$, то запишем первое условие:

$$\tilde{C}\mathcal{N}_{\tilde{C}} = 0_{p \times (n-p)}. \quad (5)$$

Тогда и $C_{11}^{-1}\tilde{C}\left((\mathcal{N}_{\tilde{C}}^T\mathcal{N}_{\tilde{C}})^{-1}\mathcal{N}_{\tilde{C}}^T\right)^T = 0$. Следовательно, строки $(\mathcal{N}_{\tilde{C}}^T\mathcal{N}_{\tilde{C}})^{-1}\mathcal{N}_{\tilde{C}}^T$ дополняют строки $C_{11}^{-1}\tilde{C}$ до базиса, т.е. матрица T обратима. Из уравнения преобразования матрицы $\tilde{B}$ с учетом второго равенства (3)

$$T\tilde{B} = \begin{pmatrix} C_{11}^{-1}\tilde{C} \\ (\mathcal{N}_{\tilde{C}}^T\mathcal{N}_{\tilde{C}})^{-1}\mathcal{N}_{\tilde{C}}^T \end{pmatrix}\tilde{B} = \begin{pmatrix} B_{11} \\ 0 \end{pmatrix},$$

следует, что

$$(\mathcal{N}_{\tilde{C}}^T\mathcal{N}_{\tilde{C}})^{-1}\mathcal{N}_{\tilde{C}}^T\tilde{B} = 0_{(n-p) \times m}.$$

Отбросив обратимую матрицу слева, получим второе условие:

$$\tilde{B}^T\mathcal{N}_{\tilde{C}} = 0_{m \times (n-p)}. \quad (6)$$

Верхний блок матрицы B находится по формуле $B_{11} = C_{11}^{-1}\tilde{C}\tilde{B}$. Далее, выполнив перемножение $T\tilde{A}T^{-1}$, найдем блок A_{21}:

$$A_{21} = (\mathcal{N}_{\tilde{C}}^T\mathcal{N}_{\tilde{C}})^{-1}\mathcal{N}_{\tilde{C}}^T\tilde{A}\tilde{C}^T(\tilde{C}\tilde{C}^T)^{-1}C_{11}.$$

С учетом последнего равенства (3), отбросив обратимые матрицы $(\mathcal{N}_{\tilde{C}}^T\mathcal{N}_{\tilde{C}})^{-1}$ и $(\tilde{C}\tilde{C}^T)^{-1}C_{11}$, получаем еще одно условие:

$$\mathcal{N}_{\tilde{C}}^T\tilde{A}\tilde{C}^T = 0_{(n-p) \times p} \leftrightarrow \tilde{C}\tilde{A}^T\mathcal{N}_{\tilde{C}} = 0_{p \times (n-p)}. \quad (7)$$

Объединив условия (5)-(7), запишем уравнение:

$$\begin{pmatrix} \tilde{C} \\ \tilde{B}^T \\ \tilde{C}\tilde{A}^T \end{pmatrix} \mathcal{N}_{\tilde{C}} = 0_{(2p+m)\times(n-p)}.$$

Это уравнение разрешимо тогда и только тогда, когда выполняется условие (4). Необходимость (4) доказана. Пусть теперь условие (4) выполняется. Тогда выполняются и равенства (5)-(7). Из равенства (5) следует первое равенство, а из (6) и (7) – два последних равенства (3) соответственно. Лемма доказана.

Таким образом, для приведения произвольной линейной системы к канонической управляемой форме необходимо и достаточно, чтобы строки матриц $\tilde{C}$, $\tilde{B}^T$ и $\tilde{C}\tilde{A}^T$ были линейно-зависимыми. Пусть условие (4) выполняется. Применим статический закон управления по измеряемому выходу:

$$u = Ky,$$

где $K \in R^{m\times p}$ – матрица регулятора.

Тогда матрица замкнутой системы при $C_{11} = I_p$ примет следующий вид:

$$A_c = \begin{pmatrix} \tilde{C}\tilde{A}\tilde{C}^T(\tilde{C}\tilde{C}^T)^{-1} + \tilde{C}\tilde{B}K & \tilde{C}\tilde{A}\mathcal{N}_{\tilde{C}} \\ 0 & (\mathcal{N}_{\tilde{C}}^T\mathcal{N}_{\tilde{C}})^{-1}\mathcal{N}_{\tilde{C}}^T\tilde{A}\mathcal{N}_{\tilde{C}} \end{pmatrix}. \quad (8)$$

Получаем систему в канонической управляемой форме как в случае с обратимой матрицей выхода. Однако вряд ли можно рассчитывать на то, что условие (4) можно встретить в практических линейных системах, так как оно выглядит ограничительно. Поэтому рассмотрим вместо него другое, более общее условие:

$$rank\begin{pmatrix} \tilde{C} \\ \tilde{B}^T \end{pmatrix} = p. \quad (9)$$

Из (9) следует, что $rank\left(\tilde{C}\tilde{B}\right) = \min\{m, p\} = m$. Предположим, что в системе (1) выбран такой базис, в котором образ матрицы выхода имеет вид $\tilde{C} = (I_p \quad 0)$. Выполним линейное преобразование так, чтобы выполнялось второе условие (3) и при этом матрица выхода не изменилась. Запишем уравнения преобразований:

$$\begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}\begin{pmatrix} \tilde{B}_{11} \\ \tilde{B}_{21} \end{pmatrix} = \begin{pmatrix} B_{11} \\ 0_{(n-p)\times m} \end{pmatrix}, \quad (10)$$

$$(I_p \quad 0_{p\times(n-p)})\begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}^{-1} = (I_p \quad 0_{p\times(n-p)}).$$

Пусть $\begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}^{-1} = \begin{pmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{pmatrix}$. Тогда из последнего равенства следует, что $R_{12} = 0_{p\times(n-p)}$. Тогда и $S_{12} = 0_{p\times(n-p)}$. Из первого уравнения (10) получаем:

$$S_{21}\tilde{B}_{11} + S_{22}\tilde{B}_{21} = 0_{(n-p)\times m}.$$

Так как матрица обратима, то ее диагональные блоки также обратимы. Умножим обе части последнего равенства слева на S_{22}^{-1} и введем обозначение $S_{22}^{-1}S_{21} = Z$. В результате получим:

$$\tilde{B}_{21} = -Z\tilde{B}_{11}. \quad (11)$$

Так как $\tilde{B}_{11} = \tilde{C}\tilde{B}$, то $rank\ (B_{11}) = m$. Тогда решением уравнения (11) является матрица

$$Z = -\tilde{B}_{21}\left(\tilde{B}_{11}^T\tilde{B}_{11}\right)^{-1}\tilde{B}_{11}^T.$$

Положим, что $S_{11} = I_p$ и $S_{22} = I_{n-p}$. Убеждаемся, что матрица преобразования

$$S = \begin{pmatrix} I_p & 0 \\ Z & I_{n-p} \end{pmatrix}$$

обеспечивает равенства (10). Обратная матрица будет равна

$$S^{-1} = \begin{pmatrix} I_p & 0 \\ -Z & I_{n-p} \end{pmatrix}$$

Выполнив преобразование системы с помощью матрицы S, получим матрицу замкнутой системы в новом базисе:

$$A_c = \begin{pmatrix} \tilde{A}_{11} + \tilde{A}_{12}Z + \tilde{B}_{11}K & \tilde{A}_{12} \\ Z\tilde{A}_{11} + \tilde{A}_{21} - \left(\tilde{A}_{22} + Z\tilde{A}_{12}\right)Z & \tilde{A}_{22} + Z\tilde{A}_{12} \end{pmatrix}, \quad (12)$$

где $Z = -\tilde{B}_{21}\left(\tilde{B}_{11}^T\tilde{B}_{11}\right)^{-1}\tilde{B}_{11}^T$.

Таким образом, если выполняется более слабое условие (9), то матрицу замкнутой системы можно привести к виду (12), в котором неизвестная матрица K локализована в одном ее блоке.

4. Применение к решению задач стабилизации

Пусть условие (4) выполнено. Обратимся сначала к матрице (8). Пространство состояний системы с такой матрицей раскладывается в прямую сумму двух подпространств, одно из которых есть пересечение управляемого и наблюдаемого подпространств, а второе – пересечение неуправляемого и ненаблюдаемого подпространств. Если исходная система стабилизируема и детектируема, то пара $\left(\tilde{C}\tilde{A}\tilde{C}^T(\tilde{C}\tilde{C}^T)^{-1}, \tilde{C}\tilde{B}\right)$ полностью управляема. Тогда, если пересечение неуправляемого и ненаблюдаемого подпространств, определяемое блоком $(\mathcal{N}_{\tilde{C}}^T\mathcal{N}_{\tilde{C}})^{-1}\mathcal{N}_{\tilde{C}}^T\tilde{A}\mathcal{N}_{\tilde{C}}$ устойчиво, то система стабилизируема. Сформулируем утверждение:

Утверждение 1. *Если условие* (4) *выполнено, то для существования статического регулятора по выходу, обеспечивающего асимптотическую устойчивость замкнутой системы необходимо и достаточно, чтобы пара* $\left(\tilde{C}\tilde{A}\tilde{C}^T(\tilde{C}\tilde{C}^T)^{-1}, \tilde{C}\tilde{B}\right)$ *была стабилизируемой и блок* $(\mathcal{N}_{\tilde{C}}^T\mathcal{N}_{\tilde{C}})^{-1}\mathcal{N}_{\tilde{C}}^T\tilde{A}\mathcal{N}_{\tilde{C}}$ *гурвицев.*

Пусть теперь выполняется ослабленное условие (9). Рассмотрим задачу стабилизации матрицы (12) в предположении, что блок

$$\tilde{A}_{22} + Z\tilde{A}_{12} = \tilde{A}_{22} - \tilde{B}_{21}\left(\tilde{B}_{11}^T\tilde{B}_{11}\right)^{-1}\tilde{B}_{11}^T\tilde{A}_{12}$$

гурвицев. Заметим, этого можно добиться путем выбора соответствующего базиса. Тогда для устойчивости (12) необходимо и достаточно, чтобы существовала матрица функции Ляпунова $Y = \begin{pmatrix} Y_{11} & Y_{12} \\ Y_{12}^T & Y_{22} \end{pmatrix} \succ 0$, обеспечивающая разрешимость билинейного матричного неравенства [7]:

$$\begin{pmatrix} \tilde{A}_{11} + \tilde{A}_{12}Z + \tilde{B}_{11}K & \tilde{A}_{12} \\ Z\tilde{A}_{11} + \tilde{A}_{21} - \left(\tilde{A}_{22} + Z\tilde{A}_{12}\right)Z & \tilde{A}_{22} + Z\tilde{A}_{12} \end{pmatrix}\begin{pmatrix} Y_{11} & Y_{12} \\ Y_{12}^T & Y_{22} \end{pmatrix} + (*)^T \prec 0.$$

В соответствии с известной в теории управления леммой исключения [7], это неравенство разрешимо тогда и только тогда, когда разрешима система:

$$\mathcal{N}_{B^T}^T\left(\begin{pmatrix} \tilde{A}_{11} + \tilde{A}_{12}Z & \tilde{A}_{12} \\ Z\tilde{A}_{11} + \tilde{A}_{21} - \left(\tilde{A}_{22} + Z\tilde{A}_{12}\right)Z & \tilde{A}_{22} + Z\tilde{A}_{12} \end{pmatrix}\begin{pmatrix} Y_{11} & Y_{12} \\ Y_{12}^T & Y_{22} \end{pmatrix} + (*)^T\right)\mathcal{N}_{B^T} \prec 0. \qquad (13)$$

$$\mathcal{N}_C^T\left(\begin{pmatrix} X_{11} & X_{12} \\ X_{12}^T & X_{22} \end{pmatrix}\begin{pmatrix} \tilde{A}_{11} + \tilde{A}_{12}Z & \tilde{A}_{12} \\ Z\tilde{A}_{11} + \tilde{A}_{21} - \left(\tilde{A}_{22} + Z\tilde{A}_{12}\right)Z & \tilde{A}_{22} + Z\tilde{A}_{12} \end{pmatrix} + (*)^T\right)\mathcal{N}_C \prec 0.$$

где $\mathcal{N}_{B^T}$ и $\mathcal{N}_C$ – ядра матриц B^T и C соответственно; $XY = I$.

Обозначим сумму матриц в первом и втором неравенствах (13) в виде блочных симметрических матриц $\begin{pmatrix} \mathcal{U}_{11} & \mathcal{U}_{12} \\ \mathcal{U}_{12}^T & \mathcal{U}_{22} \end{pmatrix}$ и $\begin{pmatrix} \mathcal{L}_{11} & \mathcal{L}_{12} \\ \mathcal{L}_{12}^T & \mathcal{L}_{22} \end{pmatrix}$ соответственно. Исследуем сначала второе неравенство. В силу того, что $\mathcal{N}_C = \begin{pmatrix} 0 \\ I_{n-p} \end{pmatrix}$, то

$$\mathcal{L}_{22} = X_{22}\left(\tilde{A}_{22} - \tilde{B}_{21}\left(\tilde{B}_{11}^T\tilde{B}_{11}\right)^{-1}\tilde{B}_{11}^T\tilde{A}_{12}\right) + X_{12}^TA_{12} + (*)^T \prec 0.$$

Так как блок $\tilde{A}_{22} - \tilde{B}_{21}\left(\tilde{B}_{11}^T\tilde{B}_{11}\right)^{-1}\tilde{B}_{11}^T\tilde{A}_{12}$ гурвицев, то последнее неравенство разрешимо при $X_{12} = 0$. Рассмотрим решение системы (13) в классе блочно-диагональных матриц функции Ляпунова. Так как $X_{22} = Y_{22}^{-1}$, то неравенство $\mathcal{L}_{22} \prec 0$ можно переписать в виде

$$\mathcal{L}_{22} = \left(\tilde{A}_{22} - \tilde{B}_{21}\left(\tilde{B}_{11}^T\tilde{B}_{11}\right)^{-1}\tilde{B}_{11}^T\tilde{A}_{12}\right)Y_{22} + (*)^T \prec 0. \qquad (14)$$

Рассмотрим первое неравенство (13). Так как

$$\mathcal{N}_{B^T} = \begin{pmatrix} \mathcal{N}_{B_{11}^T} & 0 \\ 0 & I_{n-p} \end{pmatrix},$$

то

$$\mathcal{N}_{B^T}^T\begin{pmatrix} \mathcal{U}_{11} & \mathcal{U}_{12} \\ \mathcal{U}_{12}^T & \mathcal{U}_{22} \end{pmatrix}\mathcal{N}_{B^T} = \begin{pmatrix} \mathcal{N}_{B_{11}^T}^T\mathcal{U}_{11}\mathcal{N}_{B_{11}^T} & \mathcal{N}_{B_{11}^T}^T\mathcal{U}_{12} \\ \mathcal{U}_{12}^T\mathcal{N}_{B_{11}^T} & \mathcal{L}_{22} \end{pmatrix} \prec 0, \qquad (15)$$

где$\mathcal{N}_{B_{11}^T} \in R^{p\times(p-m)}$ – ядро матрицы B_{11}^T.

Видим, что (14) является необходимым условием разрешимости (15), поэтому, (14) можно отбросить. Сформулируем соответствующее утверждение:

Утверждение 2. *Достаточным условием существования статического регулятора, обеспечивающего стабилизацию системы с матрицей (12), является разрешимость линейного матричного неравенства (15) относительно матриц*$Y_{11} \succ 0$ *и* $Y_{22} \succ 0$.

Полученное решение не является единственным. Несколько иное решение задачи стабилизации при условии (9) получено в работе [5].

5. Заключение

В статье получено необходимое и достаточное условие приведения произвольной линейной системы к канонической управляемой форме, выражаемое в виде рангового равенства. В силу того что это условие накладывает довольно серьезные ограничения, далеко не каждая система может быть приведена к такому виду. На основании полученного результата рассмотрено решение задачи стабилизации, когда ранг произведения матриц выхода и входа образует полный ранг. Сформулированы достаточные условия разрешимости, выражаемые в виде линейного матричного неравенства.

Литература

1. Квакернаак Х, СиванР. Линейные оптимальные системы управления. – М.: Мир, 1977.
2. Параев Ю.И. Уравнения Ляпунова и Риккати. – Томск: изд. Том. ун-та, 1989.
3. Nemirovskii A.A. Several NP-hard problems arising in robust stability analysis // Math. Control, Signals, Systems. 1994. V. 6. P. 99-105.
4. SadabadiM. S., PeaucelleD. From static output feedback to structured robust static output feedback: A survey // Annual Reviews in Control. 2016. Vol. 42. P. 11-26.
5. Prempain E., Postlethwaite I. Static output stabilization with H∞ performance for a class plants // Sytem& Control Letters. 2001. V.43. P. 159-166.
6. Elias A. A Novel relaxation of the static output feedback problem for a class of plants // Automatica. 2023.V. 158.
7. Баландин Д.В., Коган М.М. Синтез законов управления на основе линейных матричных неравенств. М.: Физматлит, 2007. 280 с.

ЧИСЛЕННОЕ МОДЕЛИРОВАНИЕ ДИНАМИЧЕСКОЙ СИСТЕМЫ АСТРОЦИТА УЛЛАХА-ЮНГА

Л.П. Никифорова, С.Ю. Маковкин

Нижегородский государственный университет им. Н.И. Лобачевского

В данной статье проводится численное исследование модели астроцита Уллаха-Юнга, чтобы изучить механизмы, порождающие колебания в этих клетках. Модель описывается нелинейной системой дифференциальных уравнений третьего порядка, описывающей динамику концентрации ионов кальция внутри астроцита. Основное внимание уделяется бифуркации Андронова–Хопфа, приводящей астроцит к автоколебательному режиму и связанной с изменением устойчивости равновесного состояния. Исследуются внешние воздействия на астроцит посредством импульсов одинаковой амплитуды и длительности, а также поведение системы при изменении управляющего параметра IP3*. В ходе исследования проводится анализ устойчивых состояний, бифуркаций, и определение критических точек. Результаты исследования позволяют глубже понять сложные взаимодействия и динамические режимы в системе астроцитов.

Ключевые слова: динамика астроцита, бифуркация Андронова-Хопфа, астроцит Уллаха-Юнга, старший показатель Ляпунова, исследование биффуркации

Уравнения модели Уллаха-Юнга [1] представляют собой систему нелинейных дифференциальных уравнений третьего порядка, поскольку это система "обычных" нелинейных дифференциальных уравнений, для их интегрирования удобно использовать метод Рунге-Кутта 4-го порядка с постоянным шагом. В одиночном астроците ключевым механизмом, приводящим к автоколебательному режиму, выступает бифуркация Андронова–Хопфа. Эта бифуркация связана с изменением устойчивости стационарного состояния концентрации кальция [2, 3].

В данной работе исследуется динамика системы, в которой IP3* выступает в качестве управляющего параметра, а v4 остается фиксированным. Для анализа системы определяются максимальные и минимальные значения, позволяющие идентифицировать устойчивые состояния, изучить бифуркации и переходы между ними, а также выявить критические точки. На графике, отражающем максимальную и минимальную амплитуду концентрации кальция (рис. 1), виден диаметр предельного цикла. При изменении значений управляющего параметра IP3*, можно увидеть его влияние на динамику системы. Это описывает качественное поведение рассматриваемой динамической системы.

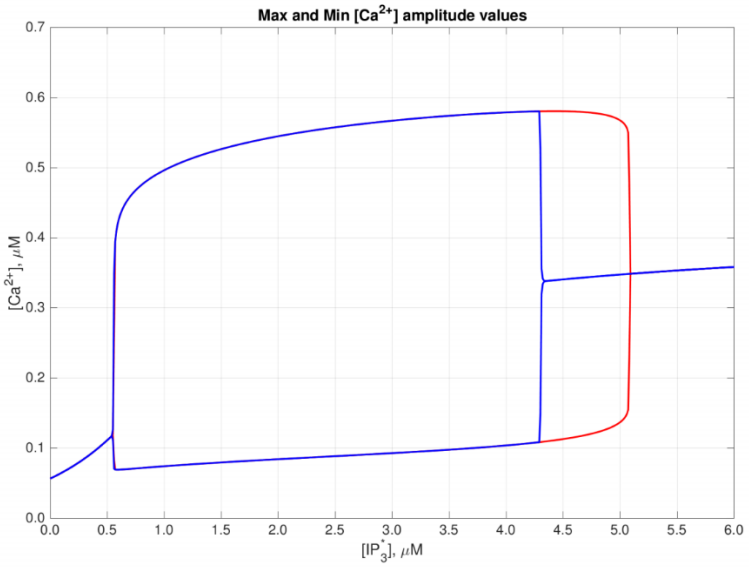


Рис. 1.

В статье рассматривается отклик астроцита на внешние импульсы с амплитудой A, длительностью τ и интервалом следования T, который меняется. С биологической точки зрения это означает, что если нейрон становится дефективным, он начинает чаще уменьшать свой период воздействия, данное поведение будет считаться патологическим режимом. При значениях интервала следования, близких к характерным временам релаксации системы, сигнал отклика становится более сложным и в некоторых случаях нерегулярным [4]. При дальнейшем уменьшении периода стимуляции концентрация кальция переходит в квазистационарное состояние. Для определения режима работы астроцита необходимо провести численный анализ, одним из методов которого является расчет старшего показателя Ляпунова. Этот показатель отражает степень расширения или сжатия эллипсоида в фазовом пространстве.

Были рассчитаны показатели Ляпунова, которые демонстрируют, что при значении $\lambda>0$ в системе присутствует странный аттрактор, указывающий на хаотический режим. Это свидетельствует о том, что динамическая система ведет себя непредсказуемо и нерегулярно, где траектории хаотически перемещаются, отражая сложные взаимодействия. При $\lambda=0$ система описывается предельным циклом, представляющим замкнутую траекторию в фазовом пространстве, по которой система движется периодически, такой цикл может быть кратным, то есть система повторяет один и тот же путь несколько раз. При $\lambda<0$ система находится в устойчивом фокусе, при этом она стремится к определенному равновесному состоянию, а малые отклонения от него вызывают возвращение к этому равновесному состоянию. Однако в данной модели астроцита значение параметра $\lambda<0$ не применимо, так как астроцит изначально находится в устойчивом фокусе. В данном исследовании ожидается, что значения λ будут равны или больше нуля. Эти два состояния представлены на рисунке 2, где верхний график иллюстрирует максимальную амплитуду кальция, а нижний – показатели Ляпунова.

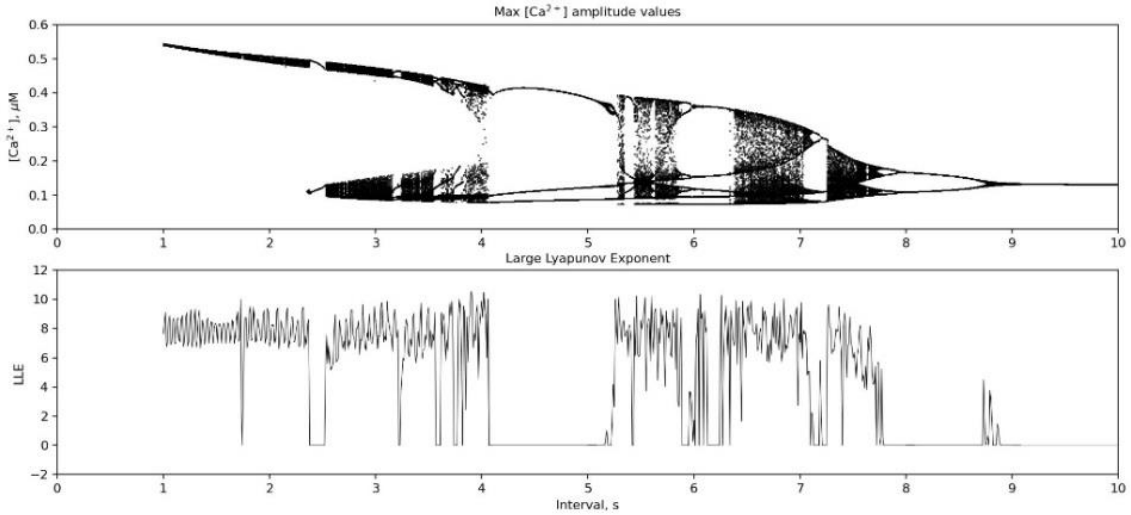


Рис. 2.

В результате исследования динамики модели астроцита по Уллаху-Юнгу было выявлено, что внешние импульсные воздействия и изменения управляющего параметра могут приводить к различным режимам поведения системы, включая автоколебания и хаотические состояния. Анализ показателей Ляпунова показал, что астроцит способен переходить от устойчивого поведения к хаотическому в зависимости от внешних условий, что указывает на сложность и гибкость его динамики. Это подчеркивает важность дальнейшего изучения механизмов регулирования активности астроцитов и их роли в работе нервной системы.

Литература

1. Ullah G., Jung P., Cornell-Bell A.H. // Cell Calcium. 2006. Vol. 39. P. 197.
2. Makovkin, S.Y., Shkerin, I.V., Gordleeva, S.Y., Ivanchenko, M.V. // Chaos, Solitons Fractals. 2020. Vol.138, 109951.

3. Matrosov V.V., Kazantsev V.B. Bifurcation mechanisms of regular and chaotic network signaling in brain astrocytes // Chaos. 2011. Vol. 21, № 2. 023103
4. Гордлеева С.Ю., Матросов В.В., Казанцев В.Б. // Известия высших учебных заведений. Прикладная нелинейная динамика. 2012. Т. 20. №. 3. С. 29.

АНАЛИЗ ГРАНИЧНОГО УЗЛА СЕТИ IAB В ПОЛУДУПЛЕКСНОМ РЕЖИМЕ: ИССЛЕДОВАНИЕ ЗАДЕРЖЕК С ПОМОЩЬЮ МОДЕЛИ ПОЛЛИНГОВОГО ОБСЛУЖИВАНИЯ С ДВУМЯ ОЧЕРЕДЯМИ[1*]

Д.И. Николаев, Е.А. Гайдамака

Российский университет дружбы народов имени Патриса Лумумбы

В работе рассматривается модель граничного узла сети интегрированного доступа и транзита как поллинговой системы массового обслуживания, состоящей из двух очередей и одного обслуживающего прибора. Цикл работы системы состоит из трех фаз. Получены аналитические выражения основных показателей производительности системы, таких как сквозная задержка, длительности фаз и цикла. Для проверки аналитических результатов построен симулятор на языке Python. Имитационное моделирование с помощью метода Монте-Карло показало относительную погрешность, не превышающую 3%, что позволяет рекомендовать построенную аналитическую модель для инженерных расчетов.

Ключевые слова: 5G, интегрированный доступ и транзит, поллинг, полудуплекс, теория массового обслуживания, миллиметровые волны, имитационное моделирование.

1. Введение

Сети пятого (5G) и шестого (6G) поколений требуют высокой плотности точек доступа из-за использования радиоволн миллиметрового диапазона, которые имеют небольшой радиус покрытия [1]. В таких условиях технология интегрированного доступа и транзита (IAB) [2] позволяет развернуть плотные сети на основе ретрансляционных узлов, которые являются более экономичными по сравнению с полноценными базовыми станциями [3]. Одной из ключевых проблем в этих сетях остаются задержки при передаче данных, особенно при многошаговой ретрансляции в полудуплексном режиме.

Работа направлена на анализ производительности граничного узла сети IAB, функционирующего в полудуплексном режиме. Основное внимание уделяется исследованию задержек при передаче пакетов с помощью поллинговой модели обслуживания для оценки соответствия системы требованиям стандартов 5G [4].

В статье для моделирования работы граничного узла сети интегрированного доступа и транзита использована поллинговая модель с двумя очередями и ненулевым временем переключения. В одну очередь поступают пакеты нисходящего потока от родительского узла, в другую – пакеты восходящего потока от дочерних узлов (пользовательских устройств). Циклический порядок опроса очередей моделирует полудуплексный режим передачи данных. Для анализа применены методы теории массового обслуживания, а распределения искомых характеристик получены в виде производящих функций, а также преобразований Лапласа и Лапласа-Стилтьеса. Проведено численное сравнение аналитической модели и симулятора.

Разработанная поллинговая модель с двумя очередями позволяет вычислить ключевые показатели системы, включая время ожидания и общее время пребывания пакетов в узле. Численные расчеты показали, что результаты аналитической модели согласуются с данными симулятора, что подтверждает адекватность предложенной модели.

Несмотря на упрощённый характер модели, она позволяет достаточно точно предсказывать временные характеристики системы. Сравнение с имитационной моделью подтверждает корректность разработанных решений, что позволяет рекомендовать использование модели для оптимизации работы сетей IAB и разработки более эффективных алгоритмов управления трафиком в сетях 5G, а также для их дальнейшего развития в сетях 6G.

[1*] Исследование выполнено за счет гранта Российского научного фонда № 24-19-00804, https://rscf.ru/project/24-19-00804/.

2. Модель граничного узла сети IAB

2.1. Системная модель

Ретрансляционные сети IAB состоят из множества базовых станций (БС), соединенных друг с другом беспроводными каналами. БС образуют древовидный граф с одной БС в корне, которую соединяют с ядром сети проводным каналом. Эту БС называют IAB-донор, остальные БС – IAB-узлами. IAB-узлы в листьевых вершинах имеют прямую связь с абонентскими устройствами (АУ).

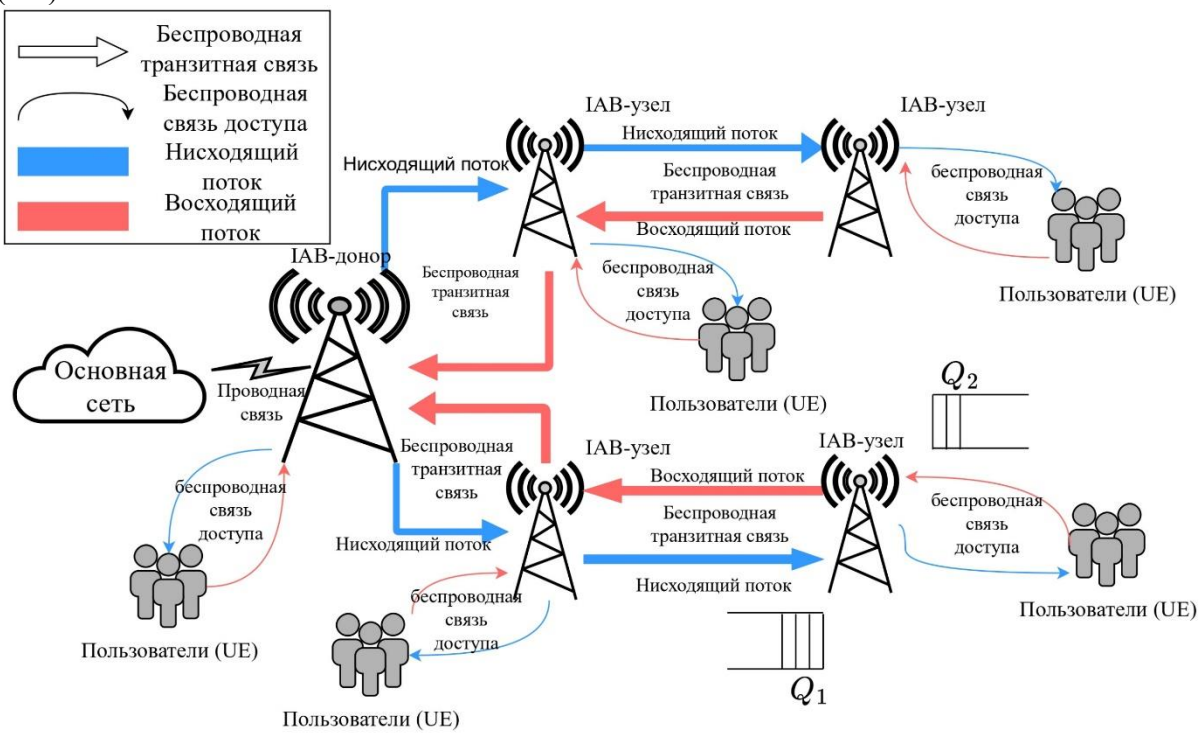


Рис. 1. Пример сети IAB

Рассмотрим далее пример сети с применением технологии IAB. На рис. 1 схематически представлена топология сети в виде связующего дерева (SP, Spanning Tree). В данной структуре IAB-узлы имеют единственный родительский узел.

Основное внимание уделяется потокам пакетов данных, проходящим граничный узел IAB, который представлен листьевой вершиной дерева. В качестве предмета исследования выступают характеристики этих потоков.

Пакеты данных, поступающие в узел IAB, можно разделить на две основные категории:

- пакеты, передаваемые по нисходящему каналу (downlink) от родительского узла, т.е по транзитному соединению;
- пакеты, передаваемые по восходящему каналу (uplink) от дочерних узлов или от АУ, ассоциированных с рассматриваемым узлом IAB, т.е. по каналу доступа.

Важно упомянуть, что технология IAB поддерживает как полнодуплексный, так и полудуплексный режимы передачи данных, но в данной работе мы ограничиваемся лишь полуплексным режимом, при этом учитываем характерные для него ограничения на передачу данных на узле:

- запрещено одновременно осуществлять прием и передачу данных по каналу доступа;
- запрещено одновременно осуществлять прием и передачу данных в транзитном соединении;
- запрещено одновременно осуществлять прием данных по каналу доступа и передачу данных в транзитном соединении и наоборот.

2.2. Математическая модель

Ограничения, вызванные полудуплексным режимом передачи данных, а также разделение каналов на восходящий и нисходящий, приводят к необходимости введения двух отдельных очередей. Это, в свою очередь, обусловливает применение модели поллингового обслуживания [5-7], учитывающей описанные выше ограничения.

Будем моделировать граничный узел сети IAB как поллинговую СМО с ненулевым временем переключения прибора, состоящую из двух очередей и одного обслуживающего прибора. Пакеты, поступающие от родительского узла по нисходящему каналу и от АУ по восходящему каналу, соответствуют заявкам, поступающим в СМО с интенсивностью λ_1 и λ_2 соответственно. Обозначим бесконечную очередь, куда поступают заявки от родительского узла – Q_1, от АУ – Q_2. Передача от узла IAB до АУ или до родительского узла соответствует обслуживанию заявок первого или второго типа с интенсивностью μ_1 или μ_2. Время между поступлениями заявок и длительность обслуживания одной заявки распределены экспоненциально с указанными интенсивностями.

С учетом накладываемых ограничений полудуплексного режима передачи данных, при моделировании граничного узла сети IAB с помощью модели поллинга с ненулевым временем переключения прибора внутри цикла поллинга можно выделить четыре ключевые фазы работы – две фазы переключения прибора и две фазы обслуживания прибором очередей. Эти фазы характеризуются изменениями числа заявок в очередях Q_1 и Q_2. Для упрощения модели объединим две фазы переключения прибора в одну фазу, которая будет ассоциироваться с периодом поступления заявок до начала т.н. «цикла обслуживания», который начинается с обслуживания первой очереди и завершается обслуживанием последней заявки из последней очереди. Время переключения между очередями внутри цикла обслуживания примем равным нулю. Таким образом, цикл поллинга будет состоять из трех фаз:

1. наполнение Q_1 и Q_2 – поступление пакетов downlink от родительского узла и uplink от АУ;
2. опустошение Q_1 – обслуживание downlink к АУ;
3. опустошение Q_2 – обслуживание uplink к родительскому узлу.

Заявки из очереди выбираются на обслуживание в соответствии с дисциплиной FCFS (First Come First Served), то есть в порядке поступления. В результате мы получаем несимметричную систему поллинга с различающимися для двух очередей интенсивностями переключения, поступления и обслуживания заявок. Модель функционирует в непрерывном времени, с циклическим порядком обслуживания и пуассоновскими входящими потоками, зависящими от состояния системы. Последняя особенность исходит из ограничений, накладываемых полудуплексным режимом передачи данных, то есть разделения во времени периодов приёма и передачи данных в нисходящем и восходящем каналах. Вытекающей отсюда особенностью построенной модели поллинга является разделение фаз наполнения и опустошения очередей. Из-за такого разделения заявки поступают в каждую очередь только в соответствующую фазу наполнения, после чего система не принимает новые заявки до окончания текущего цикла обслуживания. В таких предположениях входящие потоки активны только во время фазы наполнения очередей, а во время фаз опустошения очередей входящие потоки не активны. То есть по окончании фазы опустошения последней очереди все присутствующие в системе на момент начала цикла обслуживания заявки оказываются обслужены и покинут систему.

Основные характеристики построенной модели кратко представлены в таблице 3. Ряд вероятностно-временных характеристик для данной модели был получен в [8]. В частности, были получены функции распределения времён пребывания заявок для каждой из фаз обслуживания (1), квантили времён пребывания заявок на фазах обслуживания (2), средняя длительность цикла поллинга (3) и средние времена пребывания заявок (4).

$$V_1(t) = 1 - \Lambda_1 e^{-c_1 t}; \qquad V_2(t) = 1 - \gamma_1 e^{-c_1 t} - \gamma_2 e^{-c_2 t},\ t \geq 0, \tag{1}$$

$$Q^1_{\alpha_1} = \frac{1}{c_1} \ln \frac{\Lambda_1}{1-\alpha_1}; \qquad \gamma_1 e^{-c_1 Q^2_{\alpha_2}} + \gamma_2 e^{-c_2 Q^2_{\alpha_2}} = 1 - \alpha_2, \quad \alpha_1, \alpha_2 \in [0,1], \tag{2}$$

$$\mathcal{C} = \frac{\rho_1 + \rho_2 + 1}{s}, \tag{3}$$

$$\overline{v_1} = \frac{\rho_1}{s}; \qquad \overline{v_2} = \frac{\rho_1 + \rho_2}{s}, \tag{4}$$

где

$$\rho_i = \frac{\lambda_i}{\mu_i}, \quad \Lambda_i = \frac{\lambda_i}{\lambda_i + s}, \quad c_i = \frac{\mu_i s}{\lambda_i + s}, \quad i = 1, 2,$$

$$\gamma_1 = \Lambda_1 \frac{\mu_2(\lambda_1 + s) - \mu_1 s}{\mu_2(\lambda_1 + s) - \mu_1(\lambda_2 + s)}, \qquad \gamma_2 = \Lambda_2 \frac{\mu_1(\lambda_2 + s) - \mu_2 s}{\mu_1(\lambda_2 + s) - \mu_2(\lambda_1 + s)}.$$

3. Численный анализ

График на рис. 2 демонстрирует зависимость средней длины цикла поллинга от интенсивности переключения s, рассчитанную с использованием аналитической модели и имитационного моделирования. Наблюдается высокая степень точности аналитической модели, с относительной погрешностью не превышающей 3% (см. таблица 1), что подтверждает корректность предложенных расчетов. На графике также представлена относительная погрешность аналитической модели по сравнению с имитационной для различных значений s.

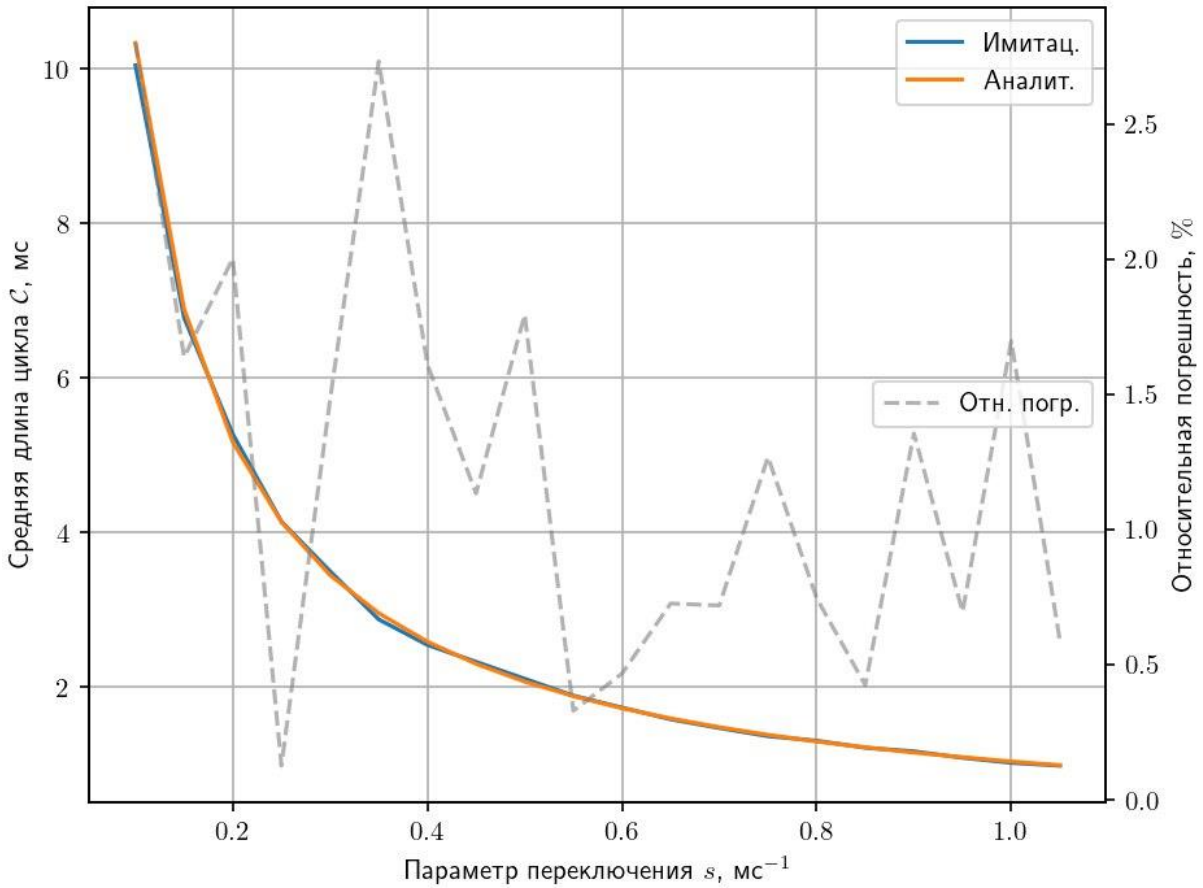


Рис. 2. Средняя длина цикла поллинга в зависимости от интенсивности переключения

В Таблице 1 приведены значения длины цикла поллинга C для различных значений интенсивности переключения s, полученные средствами имитационного моделирования и расчитанные по аналитическим формулам. Значения, полученные аналитически и в ходе моделирования, демонстрируют согласованность: абсолютная погрешность не превышает 0.09 мс для всех тестируемых значений s, а относительная погрешность имеет максимум в точке $s = 0.3$, равный 1.8%.

Таблица 1. Результаты имитационного моделирования

Зависимость длины цикла C от параметра переключения s				
s	**C аналит.**	**C имитац.**	**Абсолютн. погрешность**	**Относит. погрешность**
0,1	10,3333	10,4222	0,0889	0,0086
0,2	5,1667	5,2266	0,0599	0,0116
0,3	3,4444	3,3803	0,0641	0,0186
0,4	2,5833	2,5625	0,0208	0,0081
0,5	2,0667	2,0605	0,0062	0,0030
0,6	1,7222	1,7126	0,0097	0,0056
0,7	1,4762	1,4608	0,0154	0,0104
0,8	1,2917	1,2791	0,0126	0,0097
0,9	1,1481	1,1597	0,0115	0,0100
1	1,0333	1,0262	0,0071	0,0069

График на рис. 3 иллюстрирует 95-процентный квантиль времени пребывания заявок на первой фазе обслуживания, то есть время, пока прибор обслуживает первую очередь, рассчитанный как для экспоненциального, так и для детерминированного обслуживания. В случае экспоненциального обслуживания результаты аналитической модели совпадают с данными имитационного моделирования с расхождением менее 2%, что подтверждает устойчивость результатов аналитического подхода. В свою очередь, имитационная модель позволяет рассмотреть также и более сложные распределения времён обслуживания, в частности, на данном графике построен 95% квантиль времени пребывания заявок на первой фазе обслуживания с детерминированным обслуживанием, длительность которого совпадает со средней длительностью обслуживания в экспоненциальном случае.

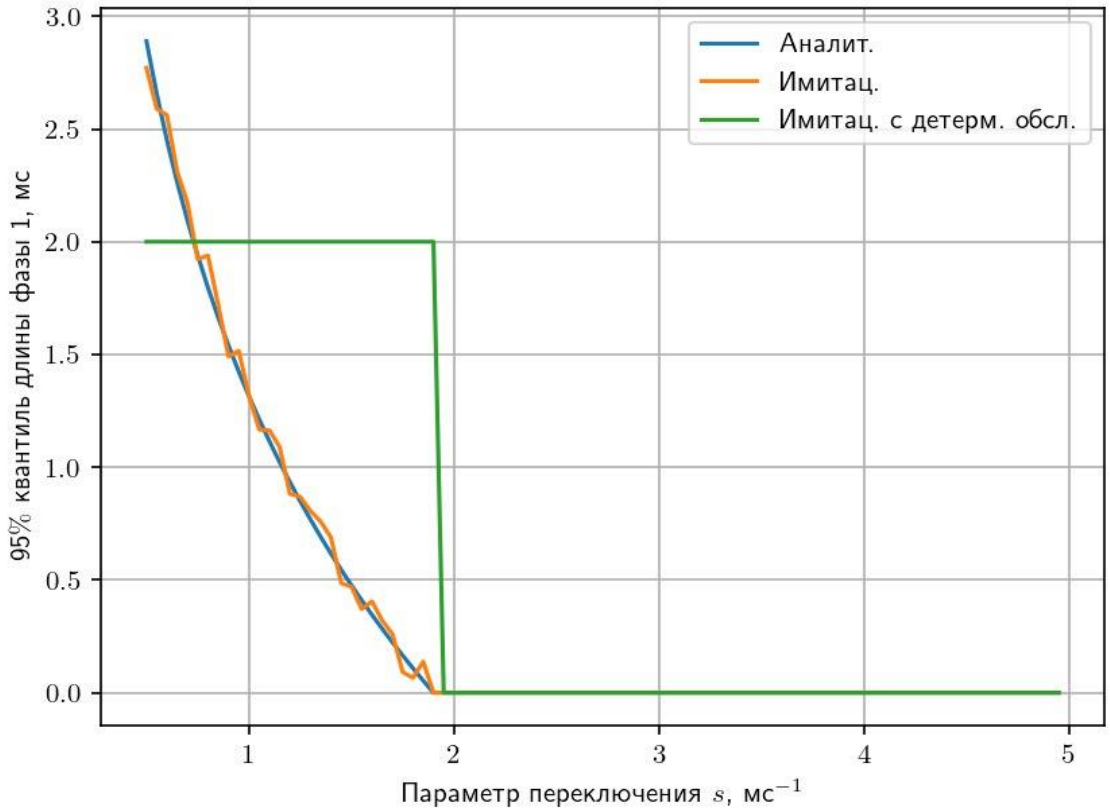


Рис. 3. 95% квантиль времени пребывания заявки на первой фазе обслуживания

4. Заключение

В данной работе представлена модель граничного узла сети IAB, функционирующего в полудуплексном режиме с двумя отдельными очередями для обработки восходящего и нисходящего потоков, позволяющая провести анализ временных характеристик системы. Разработанная модель позволяет оценить ключевые метрики системы, такие как средняя длина цикла поллинга, средние задержки и квантили времени пребывания заявок на фазах обслуживания.

Высокая точность аналитической модели подтверждена результатами имитационного моделирования. Это указывает на корректность использованных допущений и возможность применения предложенной в статье системы поллинга для оптимизации задержек в сетях IAB. Модель показала свою эффективность при различных уровнях интенсивности переключения, что позволяет рекомендовать её использование в задачах планирования сетей 5G и 6G, а также для разработки алгоритмов управления трафиком.

Использование поллинговой модели для описания работы узла IAB позволяет детально учитывать специфику полудуплексного режима, важного для сетей пятого и шестого поколений. Основные ограничения полудуплекса – необходимость поочередного использования каналов для восходящего и нисходящего потоков – были интегрированы в модель, что позволило разработать эффективный аналитический подход к оценке временных характеристик.

Для дальнейших исследований может быть полезным расширить модель, добавив динамическое распределение приоритетов обслуживания заявок, что позволит учитывать изменяющуюся нагрузку на каналы в условиях реальной сети.

Литература

1. Сети 5G/6G: архитектура, технологии, методы анализа и расчета / Д. А. Молчанов [и др.]. – Москва: РУДН, 2022. – С. 516.
2. 3GPP, “Study on Integrated Access and Backhaul,” Technical Report (TR) 38.874, 3GPP, 12 2018. Version 16.0.0.
3. ETSI Technical Specification TS 138 474V16.0.0 (2020-07) 5G NG-RAN.F1 data transport (3GPP TS 38.874 version 16.0.0 Release 16).
4. 3GPP TS 38.101–1 version 15.9.0 Release 15.
5. Вишневский В. М., Семёнова О. В. Системы поллинга: Теория и применение в широкополосных беспроводных сетях. – Москва: Техносфера, 2007. – С. 312.
6. Рыков В. В. К анализу поллинг-систем // Автоматика и телемеханика. – 2009. –№ 6. – С. 90-114.
7. Takagi H. Analysis of polling systems. – MIT Press, 1986. – P. 175.
8. Nikolaev, D., Gorshenin A., Gaidamaka, Yu.: Polling model for analysis of round-trip time in the IAB network. In: Vishnevskiy, V.M., Samouylov, K.E., Kozyrev, D.V. (eds.) Distributed Computer and Communication Networks: Control, Computation, Communications. Springer Nature Switzerland, Cham (2025). Accepted paper.

ЧИСЛЕННОЕ ИССЛЕДОВАНИЕ ВЛИЯНИЯ НАЧАЛЬНОГО НЕСОВЕРШЕНСТВА НА ФОРМУ ЗАКРИТИЧЕСКОГО ДЕФОРМИРОВАНИЯ ЦИЛИНДРИЧЕСКОЙ ОБОЛОЧКИ ПРИ ОСЕВОМ ДИНАМИЧЕСКОМ СЖАТИИ

М.В. Оболенский, О.В. Воронков, С.С. Андропенков, Н.Н. Куриков

ООО «Системы инженерного анализа», ул. Сахарова, д. 4, 603145 Н. Новгород, Россия,
АО «Атомэнергопроект», ул. Савушкина, д. 82, лит. А, 197183 Санкт-Петербург, Россия

Численно исследуется упругопластическая динамическая потеря устойчивости и закритическая деформация прямолинейной круглой алюминиевой трубы при осевом ударном сжатии. Постановка задачи основана на описании эксперимента из литературы. Рассматривается влияние малых начальных несовершенств геометрической формы трубы на ее деформацию и силу сопротивления сжатию. Результаты моделирования сравниваются с соответствующими экспериментальными результатами. Делается вывод, что деформированная форма трубы существенно зависит от ее малых начальных несовершенств, поэтому носит случайный характер. Пиковое усилие сопротивления трубы при образовании двух первых складок варьируется в пределах 4-8%, третьей складки – 33%; перемещение ударника – 8%.

Ключевые слова: МКЭ, ударное нагружение, контакт, динамическая потеря устойчивости, закритическое упругопластическое деформирование, начальное несовершенство формы, Abaqus.

1. Введение

Проблеме потери устойчивости оболочек вращения посвящено значительное количество работ, например [1-3], т.к. она играет важную роль при решении широкого круга технических задач. Применение метода конечных элементов к данной проблеме позволяет значительно расширить круг рассматриваемых аспектов, включая дополнительные источники нелинейности механического поведения конструкции. В частности, при статической и динамической закритической деформации тонкостенных оболочек от осевого сжатия значительную роль играет неидеальность реальной конструкции [4, 5]. При этом возникают вопросы: 1) о способах задания начальных несовершенств; 2) о достоверности получаемых при этом численных результатов. По первому вопросу в работах [6, 7] предлагается задание несовершенств, повторяющих собственные формы конструкции, как наиболее сильно снижающие критическую силу. Альтернативный подход [8] – задание начальных несовершенств на основе типовых технологических отклонений формы. Второй вопрос может быть разрешен путем валидации численных результатов на экспериментальных данных [9] и накоплением опыта моделирования. При этом, на данный момент нет общепринятого подхода к проблеме достоверного моделирования закритической деформации цилиндрических оболочек, поэтому она является актуальной.

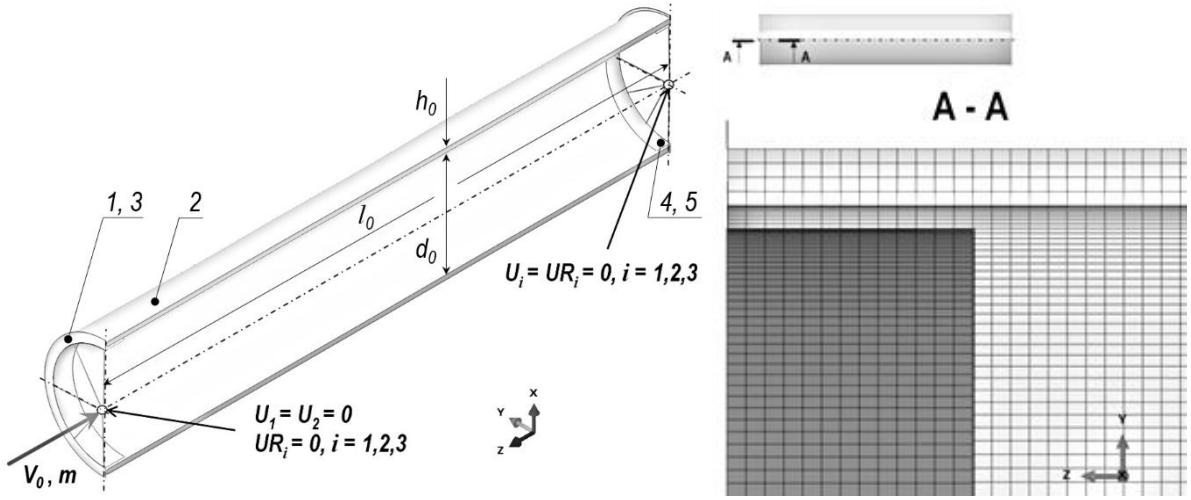
2. Постановка задачи

Рассматриваемый пример основан на описании экспериментальной установки и результатах натурных испытаний, приведенных в [9]. Решается задача динамического осевого сжатия прямой тонкостенной трубы круглого поперечного сечения, в ходе которого она претерпевает упругопластическую деформацию, связанную с локальной потерей устойчивости ее стенками и дальнейшим закритическим деформированием.

2.1. Расчетная схема и основные характеристики

Расчетная схема экспериментальной установки показана на рис. 1(а) в осевом разрезе. Сборка состоит из двух втулок: верхней (поз. 1) и нижней (поз. 4), между которыми размещена тонкостенная алюминиевая труба (поз. 2, $h_0 \times d_0 \times l_0$ = 1,5×35,5×190 мм). К верхней втулке приварен стальной сердечник (поз. 3, не показан), который выполняет функцию центрирования установки и служит в качестве направляющей. Нижняя втулка жестко закреплена на массивном основании (поз. 5, не показано). Стальные втулки и сердечник по отношению к тонкостенной трубе можно считать массивными твердыми телами. На расчетной схеме втулки отражены своими контактными поверхностями. Сердечник учтен граничными условиями и присоединенной массой верхней втулки. Нагружение сборки осуществляется падением ударника на втулку верхнюю против оси *Z* системы координат. Взаимодействие ударника с верхней втулкой не рассматривается, ударник учтен присоединенной массой втулки верхней и величиной ее начальной скорости, определяемой из условия сохранения энергии при их взаимодействии.

Характеристики материала трубы: плотность ρ = 2710 кг/м^3, модуль Юнга E = 71 ГПа, коэффициент Пуассона μ = 0,3. Применяется билинейная кривая деформирования σ(ε), проходящая через характерные точки: ($\sigma_т$, $\sigma_т/E$), ($\sigma_в$, $1 + \sigma_т/E - \sigma_т/\sigma_в$) в предположении, что до момента бифуркации истинная кривая деформирования достаточно близка к технической, и бифуркация происходит при $\sigma = \sigma_в$. Обозначено: σ – истинное напряжение; ε – логарифмическая деформация; $\sigma_т$ – предел текучести; $\sigma_в$ – предел прочности. Свойства материала аппроксимируются моделью изотропной линейной упругости и изотропной пластичности J_2 с законом течения по нормали к поверхности текучести (пластическая деформация является инвариантом НДС).



а) б)

Рис. 1. Расчетная схема (а) и фрагмент КЭМ (б)

На рис. 1(а) обозначено: V_0 = 4,11 м/с – начальная скорость главного узла верхнего твердого тела, соответствующего втулке верхней, сердечнику и ударнику; m = 86,83 кг – масса верхнего твердого тела; U_i – перемещение; UR_i – вращение; i = 1, 2, 3 – обозначает ось системы координат *X*, *Y* и *Z* соответственно. Показаны U_i и UR_i – закрепленные узловые степени свободы модели.

Между контактными поверхностями втулок и трубой, а также между наружным поверхностями трубы учитывается условие скользящего контактного взаимодействия с коэффициентом трения Амонтона-Кулона f = 0,18 (сталь-алюминий) [10].

2.2. Метод решения

Решение выполняется методом конечных элементов в Лагранжевых переменных при явной схеме интегрирования по времени [11], реализованных в коммерческом программном пакете *Abaqus* [12]. Уравнение движения системы записывается:

$$\ddot{\mathbf{u}}\big|_t = \mathbf{M}^{-1}\cdot(\mathbf{P}-\mathbf{I})\big|_t\ ,\ \mathbf{I}\big|_t = (\mathbf{K}\cdot\mathbf{u}+\mathbf{C}\cdot\dot{\mathbf{u}})\big|_t\ , \tag{1}$$

где **u** – вектор узловых перемещений; **M** – диагональная матрица массы; **P** – вектор внешних сил; **I** – вектор внутренних сил; **K** – матрица жесткости; **C** – матрица демпфирования; $|_t$ – символ означает: “на момент *t*”. **P** может зависеть от $\dot{\mathbf{u}}$ или **u**, но не от более высоких производных перемещения по времени.

Для интегрирования по времени применяется центрально-разностный оператор:

$$\dot{\mathbf{u}}\big|_{t+\Delta t/2} = \dot{\mathbf{u}}\big|_{t-\Delta t/2} + (\Delta t\big|_{t+\Delta t} + \Delta t\big|_t)\,\ddot{\mathbf{u}}\big|_t / 2\ ,\ \mathbf{u}\big|_{t+\Delta t} = \mathbf{u}\big|_t + \Delta t\big|_{t+\Delta t}\,\dot{\mathbf{u}}\big|_{t+\Delta t/2}\ , \tag{2}$$

где $\Delta t\big|_{t+\Delta t}$ – текущее временное приращение; $\Delta t\big|_t$ – предыдущее временное приращение. Решение выполняется с максимальными стабильными временными приращениями.

Контактное взаимодействие между деталями разрешается методом штрафной жесткости с использованием численного механизма «поверхность – поверхность». Применяется линейная штрафная функция, а жесткость определяется в процессе расчета на основании жесткостей контактирующих конечных элементов.

2.3. КЭ-сетка

Деформированная форма трубы после воздействия ударника на сборку известна из эксперимента [9], показана на рис. 2(а, б). В силу этого параметры КЭ-сетки можно назначить без исследования сеточной сходимости на основании правил, выработанных в ходе практического опыта моделирования.



а) б) в)

Рис. 2. Деформированная форма трубы [9]: вид спереди (а) и слева (б); НГФ4а (в) в масштабе деформации 50

При назначении параметров КЭ-сетки трубы в области образования складок учитывались следующие положения.

1. Из-за несимметричности деформированной формы трубы нельзя пользоваться симметрией сборки в исходном состоянии для уменьшения размерности КЭМ.
2. Вдоль поверхности трубы (в осевом и окружном направлениях) желательно, чтобы КЭ имели AR ~ 1 (отношение длин сторон КЭ).
3. Поперек одной полуволны деформированной формы стенки детали желательно иметь ~20 КЭ, что соответствует линейному размеру КЭ ls = 0,5 мм.
4. С учетом положения 0, h0 / ls = 3, следовательно, необходимо применять КЭ сплошной среды. Для задач быстрой динамики, подразумевающих большие деформации материала, для решения которых применяется Abaqus/Explicit [12], эффективными являются КЭ типа C3D8R – 8-узловые одноточечного интегрирования с контролем химерных форм искажения типа «песочные часы» [13].
5. С учетом положения 0, по толщине стенки трубы необходимо иметь минимум 4 КЭ для предотвращения возникновения химерных форм деформации КЭМ при локальном упругопластическом изгибе стенки.

Таким образом, необходимые размеры КЭ в области образования складок на трубе составляют ($z \times t \times h$): 0,5 × 0,5 × 0,375 мм, где z, t и h означают осевое, окружное направление и направление по толщине стенки трубы соответственно. В прочих областях для уменьшения размерности КЭМ допускается увеличение размеров КЭ в осевом и окружном направлениях до значения $AR \sim 4$, т.е. до 1,5 мм.

Фрагмент КЭМ, разработанной с учетом вышеупомянутых положений и правил, показан на рис. 1(б). КЭМ содержит 304 тыс. узлов и 244 тыс. КЭ.

2.4. Несовершенства геометрической формы (НГФ)

На основании конструкторской экспертизы рассматриваемого объекта – прямолинейной тонкостенной круглой трубы, предполагается возможность следующих вариантов основных отличий реальной конструкции от идеализированной геометрической модели:

НГФ1: идеализированная геометрическая форма трубы: круглое кольцо в поперечном сечении, постоянная толщина стенок, торцы трубы плоские и перпендикулярные ее оси, образующие прямолинейные и параллельные оси трубы.

НГФ2: скругление стенки на верхнем торце трубы радиусом равным ее толщине, что приближенно учитывает возможное наличие скруглений или фасок на кромках реальной трубы.

НГФ3: отклонение перпендикулярности верхнего торца трубы. Верхний торец повернут относительно оси, сонаправленной оси Y мировой системы координат, см. рис. 1(а), проходящей через ось в плоскости торца, на 0,5°. Такое отклонение соответствует 14-й степени точности [14].

НГФ4: линейная суперпозиция собственных форм потери устойчивости трубы.

НГФ5: овальность поперечного сечения [15] эллиптической формы с отклонением наибольшего и наименьшего диаметров на 0,5% от номинального, что укладывается в допуск круглости 13-й степени точности [14]. Толщина стенок трубы при этом постоянна, соответствует номинальному значению.

НГФ6: линейная суперпозиция НГФ4а и НГФ5.

Для задания НГФ4 моделирование осевого сжатия трубы выполняется в два шага: 1) линейный расчет собственных форм потери устойчивости решателем *Abaqus/Standard* [12], в результате которого получаются нормализованные по своему наибольшему значению поля перемещений конструкции, соответствующие ее учитываемым собственным формам; 2) полученные поля перемещений применяются перед расчетным шагом сжатия к КЭ-сетке трубы в виде линейной суперпозиции с заданными масштабными коэффициентами. Расчет на шаге 1 выполняется при тех же граничных условиях, что и на шаге 2; начальные условия: исходная недеформированная конфигурация сборки в состоянии покоя; нагрузки: к главному узлу верхнего твердого тела приложена единичная сила в направлении $-Z$ мировой системы координат, см. рис. 1(а). Учитываются первые три отличные собственные формы, величины применяемых масштабных коэффициентов k_i обратно пропорциональны величине критической нагрузки F_i. Для НГФ4 значение k_1 принимается из условия, что суммарное отклонение формы трубы от идеальной укладывается в допуск круглости и цилиндричности 13-й степени точности [14]; для НГФ6 – суммарное отклонение формы трубы от идеальной укладывается в допуск радиального биения 14-й степени точности [14]. В данном случае для трубы рассматриваемого диаметра наибольшее отклонение составляет 0,614 мм при допуске 0,8 мм. Итоговая форма трубы, соответствующая НГФ4а, в масштабе деформации 50 показана на рис. 2(в) в наложении на идеализированную форму трубы, которая приведена в полупрозрачном отображении.

3. Результаты

Расчетные деформированные формы верхней части трубы на момент $t = t_{max}$ показаны на рис 3. Нумерация изображений: цифра соответствует НГФ; а) вид спереди; б) вид слева. На рис. 4 приведены графики перемещения ударника по времени и соответствующие им зависимости перегрузки ударника a/g от перемещения. Обозначено: a – ускорение ударника, g = 9,81 м/с^2 – ускорение свободного падения.

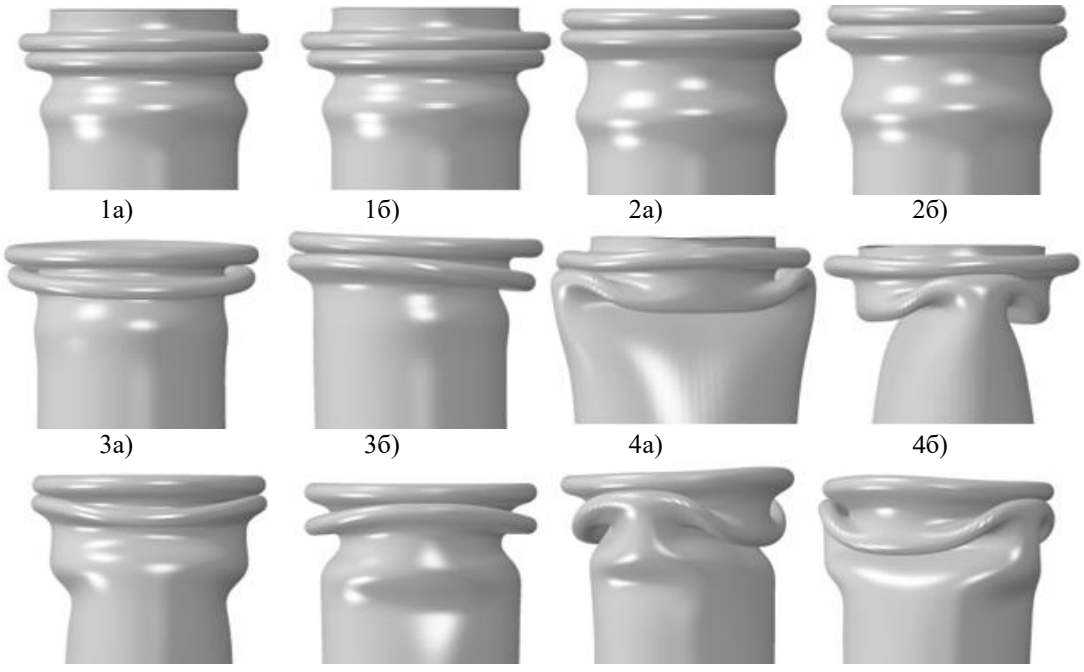
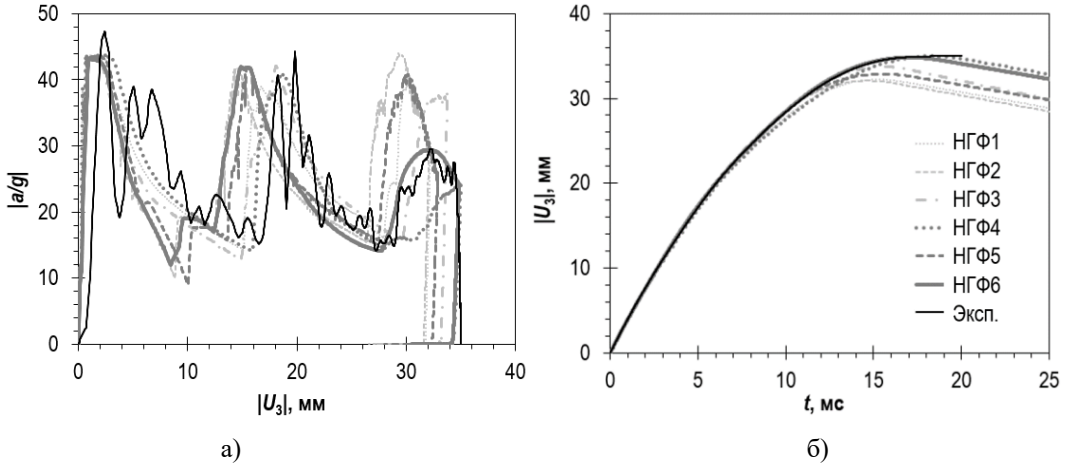


Рис. 3. Деформированные формы верхней части трубы



а) б)

Рис. 4. Нормализованное замедление ударника $|a/g|$ (а) и перемещение ударника вдоль оси Z $|U_3|$ (б)

Из приведенных данных можно отметить следующее: 1) для НГФ 1 и 2 деформированная форма трубы осесимметрична, что ожидаемо в силу осевой симметрии КЭМ и нагрузки; для прочих НГФ складки не осесимметричны, что тоже ожидаемо из-за привнесенной несимметричности исходной формы трубы; 2) форма первой складки для всех НГФ близка к осесимметричной, что выражается в примерно одинаковом первом пиковом значении силы сопротивления сжатию для всех рассмотренных случаев; 3) результирующая деформированная форма трубы существенно зависит от применяемого НГФ и в общем случае носит вероятностный характер; 4) деформированная форма трубы с НГФ6 близко совпадает с полученной в эксперименте, см. рис. 2(а, б); 5) пиковое усилие сопротивления трубы при образовании двух первых складок варьируется в зависимости от НГФ в пределах 4-8%, третьей складки – до 33%; перемещение ударника – до 8%.

4. Заключение

Приведены некоторые результаты численного исследования влияния начального несовершенства на деформированную форму тонкостенной прямолинейной круглой трубы при осевом динамическом сжатии с потерей устойчивости ее стенками и закритическом деформировании, которые показывают, что результирующая деформированная форма трубы существенно зависит от начального несовершенства и в общем случае носит вероятностный характер. Пиковое усилие сопротивления трубы при образовании двух первых складок варьируется в зависимости от начального несовершенства в пределах 4-8%, третьей складки – до 33%; перемещение ударника – до 8%.

Литература

1. Вольмир А.С. Нелинейная динамика пластинок и оболочек. М.: Наука, Гл. ред. физ.-мат. лит. 1972. 432 с.
2. Narayana Y.V., Gunda J.B., Reddy P.R., Markandeya R. Nonlinear buckling and post-buckling analysis of imperfect cylindrical shells subjected to axial compressive load. J of Structural Engineering, Vol. 42, No. 2, 2015, p. 78-85.
3. Баженов В.Г., Нагорных Е.В., Самсонова Д.А. Исследование влияния толщины и начальных несовершенств геометрии цилиндрической упругопластической оболочки с упругим заполнителем на формы потери устойчивости при внешнем давлении. Мат. XIII международной конф. АММАИ'2020. М.: Изд-во МАИ, 2020. С. 257-259.
4. Баженов В.Г., Игоничева Е.В. Динамическая потеря устойчивости и закритическое поведение тонкой цилиндрической оболочки с начальными несовершенствами под действием осевой ударной нагрузки. Прикладные проблемы прочности и пластичности: журнал. №6, 1977. С. 98-106.
5. Li Z., Cao Y., Pan G. Influence of geometric imperfections on the axially loaded composite conical shells with and without cutout. J AIP Advances, No. 10, Article ID 095106, 2020, 10 p. DOI 10.1063/5.0021103.
6. Speicher G., Saal H. Numerical calculation of limit loads for shells of revolution with particular regard to the applying equivalent initial imperfection. Buckling of Shell Structure, on Land, in the Sea, and in the Air. Elsevier Applied Science, London, 1991, p. 466-475.
7. Papadopoulos V., Papadrakakis M. Finite element analysis of cylindrical panels with random initial imperfections. J. Eng. Mech. No. 130, 2004, p. 867-876.
8. Hilburger M.W., Nemeth M.P., Starnes J.H. Shell buckling design criteria based on manufacturing imperfection signatures. AIAA J. No. 44(3), 2006, p. 654-663.
9. Баженов В.Г., Рябов А.А., Птицын С.О. Динамическая потеря устойчивости и закритическая деформация цилиндрических оболочек. Проблемы прочности и пластичности: журнал. Т. 80, №2. Н.Новгород.: ННГУ им. Н.И. Лобачевского, 2018. С. 209 – 218. DOI: https://doi.org/10.32326/1814-9146-2018-80-2-209-218.
10. Физические величины: справочник под ред. Григорьева И.С., Мейлихова Е.З. / Бабичев А.П., Бабушкина Н.А., Братковский А.М. и др. / М.: Энергоиздат, 1991. 1232 с.
11. Zienkiewicz O.C., Taylor R.L., Fox D.D. The finite element method for solid and structural mechanics. 7th edn. Butterworth-Heinemann Publ. GB, Oxford, 2014. ISBN: 978-1-85617-634-7
12. 3DS SIMULIA User Assistance 2023 [Электронный ресурс]: Abaqus Online Documentation. URL: https:// help.3ds.com/ (дата обращения 23.09.2024).
13. Belytschko T., Liu W.K., Moran B., Elkhodary K.I. Nonlinear Finite Elements for Continua and Structures. 2nd edn. Wiley, UK, Chichester, 2014. ISBN 978-1-118-63270-3.
14. ГОСТ 24643-81. Основные нормы взаимозаменяемости. Допуски формы и расположения поверхностей. Числовые значения. М.: ИПК Издательство стандартов, 1981. 10 с.
15. ГОСТ 24642-81. Основные нормы взаимозаменяемости. Допуски формы и расположения поверхностей. Основные термины и определения. М., 1981. 45 с.

ПРИМЕНЕНИЕ НЕЙРОННЫХ СЕТЕЙ ДЛЯ АНАЛИЗА ЭКГ СИГНАЛОВ НА НАБОРЕ ДАННЫХ PTBXL[1*]

В.В. Разин, А.А. Краснов

Нижегородский государственный университет имени Н.И. Лобачевского

В статье описывается применение нейронных сетей для классификации заболеваний сердца посредством анализа ЭКГ сигналов на наборе данных PTBXL. Рассматриваются рекуррентные и сверточные модели глубокого обучения, анализирующие 12-канальные 10-секундные ЭКГ записи с частотой дискретизации 500 Гц. Описывается применение альтернативного подхода обучения с рассмотрением наборов классов как отдельных сущностей с целью повышения качества диагностики. Также приводится использование замен с порогами, которая заключается в анализе ошибок нейросетей на обучающей и валидационной выборках и проведении замен в случаях, когда вероятность предсказания конкретного класса или набора ниже определенного числа (порога), поиск которого происходит дихотомией.

Ключевые слова: нейронные сети, замена с порогами, рекуррентные сети, сверточные сети, multilabel classification, глубокое обучение, диагностика ЭКГ, искусственный интеллект, анализ ЭКГ.

1. Введение

Электрокардиография является одним из основных методов диагностики сердечных заболеваний. Данный метод позволяет записать электрическую активность сердца и выявить нарушения в его функционировании. Диагностика ЭКГ позволяет выявить наличие или отсутствие аритмий, ишемии, инфаркта миокарда и других сердечных патологий. По результатам анализа ЭКГ врач может определить дальнейшие методы обследования и лечения пациента. Кроме того, регулярное проведение электрокардиографии позволяет контролировать эффективность проводимой терапии и динамику заболевания. Важность диагностики ЭКГ заключается также в том, что данное исследование позволяет выявить скрытые формы сердечных заболеваний, которые могут не проявляться в явной симптоматике, но представлять угрозу для здоровья пациента. Раннее выявление различных патологий позволяет своевременно начать лечение и предотвратить развитие серьезных осложнений. Регулярное проведение электрокардиографии позволяет выявить риск развития сердечно-сосудистых заболеваний в самом раннем возможном сроке и предпринять необходимые меры по профилактике и лечению.

Развитие методов искусственного интеллекта [1-2], рост вычислительных мощностей компьютеров дают уверенную надежду на совершенствование качества диагностики ЭКГ. Использование методов глубокого и машинного обучений, а также их различных модификаций, уже успешно зарекомендовало себя в диагностики ЭКГ записей [3-15]. В рамках работы рассматривается использование программного обеспечения и компьютерных мощностей для проведения анализа ЭКГ нейронными сетями с целью упрощения работы кардиологов и других профильных врачей этой области, а также использования как автономного решения.

2. Постановка задачи

2.1. Данные

В качестве входных данных выступает набор записей ЭКГ сигналов из датасета PTBXL [16], размещенным PhysioNet. Рассматриваются все записи с частотой дискретизации 500 Гц. Ко всем

[1*] Исследование выполнено при поддержке Министерства образования и науки Российской Федерации, проект FSWR-2024-0005.

данным применяется предобработка, которая состоит из следующей последовательности действий:

- прохода фильтром Notch;
- прохода фильтром нижних частот;
- прохода фильтром верхних частот;
- транспонирование исходных данных.

Данные разбиваются на обучающую, валидационную и тестовую выборки таким образом, что записи одних и тех же пациентов присутствуют лишь в одной из описанных выше выборок.

2.2. Задача multilabel классификации

Задача заключается в определении 5 независимых ключевых классов PTB-XL: NORM (нормальная ЭКГ), CD (нарушение проводимости), MI (инфаркт миокарда), HYP (гипертрофия), STTC (нарушение ST/T сегмента).

2.3. Метрики

В качестве метрики рассматривается макроусредненная F1 мера. Данная метрика выбрана в силу того, что она позволяет достигать баланса между Precision и Recall, что берется за основу в рамках этой работы.

2.4. Модели

В рамках данной работы рассматривается 10 нейронных сетей, обладающих наибольшей классифицирующей способностью:

- сверточная сеть [17] (далее CNN);
- глубоко разделяемая сверточная сеть [18] (далее SCNN);
- сеть, основанная на AlexNet [19] (далее AlexNet);
- комбинация сверточной сети с двунаправленными [20] LSTM и GRU (далее CBB);
- сеть, основанная на Inception [21] (далее Inception);
- рекуррентная сеть LSTM [22] (далее LSTM);
- рекуррентная сеть GRU [23] (далее GRU);
- комбинация LSTM и GRU (далее LG);
- сеть, основанная на VGG16 [24] (далее VGG16);
- сеть, основанная на Xception [25] (далее Xception).

3. Результаты

По результатам многочисленных экспериментов было зафиксировано соотношение обучающей и валидационной выборок как $\frac{8}{9}$ и $\frac{1}{9}$ соответственно (относительно их объединения). Данные разбиты в указанном выше соотношении на выборки по принадлежности к конкретному набору классов. Производилось многократное обучение (с ограничением по времени в 100 часов на каждую нейросеть) с сохранением лучшего результата посредством анализа F1 меры на валидационной выборке. Также применялись функции обратного вызова, которые понижают коэффициент обучения, если некоторое количество эпох нет увеличения метрики на валидационной выборке или вовсе останов обучения в таком случае, а еще сохранение лучшей конфигурации нейросети после каждой эпохи обучения, если повышается F1 мера на валидационном множестве. На вход нейросети получают обработанные данные, а на выходе выдают вектор вероятностей принадлежности к каждом классу (5 элементов), где каждая из них может быть от 0 до 1. В Таблице 1 приводятся результаты такого обучения (метрика на тестовой выборке приводится в процентах, с округлением до сотых).

Таблица 1. Результаты обучения нейронных сетей предсказания каждого класса

Нейронная сеть	F1-мера
CNN	68,93
SCNN	71,89
AlexNet	71,67
CBB	70,15
Inception	72,75
LSTM	66,63
GRU	66,44
LG	73,23
VGG16	71,42
Xception	**74,57**

Далее модели обучались не с предсказанием каждого класса, а предсказанием целого набора классов, то есть на выходе нейросетей получается вектор вероятностей из 32 элементов, где сумма вероятностей всегда дает 1. Для различия в названиях к прошлым добавляется приставка «class», означающую принадлежность к данному типу обучения. В Таблице 2 приводятся результаты такого обучения (метрика на тестовой выборке приводится в процентах, с округлением до сотых).

Таблица 2. Результаты обучения нейронных сетей предсказания целого набора классов

Нейронная сеть	F1-мера
classCNN	71,70
classSCNN	71,79
classAlexNet	72,15
classCBB	72,08
classInception	72,75
classLSTM	58,06
classGRU	70,97
classLG	12,19
classVGG16	71,67
classXception	**74,85**

Далее проводится эксперимент с анализом ошибок моделей. Анализируются все ошибки моделей на конкретном множестве (обучающая или валидационная выборки) и дихотомией ищется такой порог, что когда вероятность предсказания конкретного набора ниже, чем это число, то тогда всегда происходит замена на другой диагноз. В худшем случае предсказания модели поменяются, в лучшем можно достигать исправления тех ошибок, что встречаются чаще всего (одного набор на конкретный один другой). Замена с порогами при анализе ошибок на обучающей выборке приводится в Таблице 3, а при анализе ошибок на валидационном множестве в Таблице 4 (метрика на тестовой выборке приводится в процентах, с округлением до сотых).

Таблица 3. Проведение замен с порогами у моделей при анализе ошибок на обучающей выборке

Нейронная сеть	F1-мера	
	До замен	После замен
classXception	74,85	**74,86**
Xception	74,57	74,28
LG	73,23	72,99
Inception	72,75	72,70
classCBB	72,08	71,98
SCNN	71,89	71,80
classVGG16	71,67	71,76
classAlexNet	72,15	71,69
VGG16	71,42	71,46
classCNN	71,70	71,19
classSCNN	71,79	71,14

AlexNet	71,67	71,06
classGRU	70,97	70,61
CBB	70,15	70,18
classInception	69,51	69,61
LSTM	66,63	66,95
GRU	66,44	66,49
classLSTM	58,06	60,26
classLG	12,19	31,82

Таблица 4. Проведение замен с порогами у моделей при анализе ошибок на валидационной выборке

Нейронная сеть	**F1-мера**	
	До замен	После замен
classXception	74,85	**74,88**
Xception	74,57	74,28
LG	73,23	72,99
Inception	72,75	72,73
classCBB	72,08	71,98
SCNN	71,89	71,69
classVGG16	71,67	71,76
classAlexNet	72,15	71,63
VGG16	71,42	71,37
classCNN	71,70	71,17
classSCNN	71,79	71,14
AlexNet	71,67	71,06
classGRU	70,97	70,63
CBB	70,15	70,16
classInception	69,51	69,61
LSTM	66,63	66,74
GRU	66,44	66,44
classLSTM	58,06	60,26
classLG	12,19	27,30

Матрица ошибок модели classXception после замен с порогами при анализе ошибок на валидационной выборке, демонстрирующая наибольшие показатели на тестовом множестве, приводится на рис. 1.



Рис. 1. Матрица ошибок нейросети classXception после замен с порогами при анализе ошибок на валидационной выборке

4. Заключение и выводы

По результатам экспериментов можно наблюдать эффективность обучения рекуррентных и сверточных нейросетей для проведения диагностики по ЭКГ программными средствами. Рассмотрение альтернативного подхода к обучению позволяет повышать показатели целевых метрик некоторых моделей глубокого обучения. Проведение замен с порогами на обучающей и валидационной выборках также позволяет повышать классифицирующую способность некоторых моделей.

В дальнейшем планируется рассмотрение новых архитектур нейросетей, методов машинного обучения, а также различных модификаций с целью повышения качества анализа ЭКГ в

рамках предсказания 5 ключевых классов PTBXL, а также более сложных задач из данного набора и других.

Литература

1. Разин, В.В. Применение искусственного интеллекта и цифровых технологий для реше-ния проблем управления и социально-экономических задач // Актуальные проблемы управления : Сборник научных статей по итогам X Юбилейной Всероссийской научно-практической конференции, Нижний Новгород, 24 октября 2023 года. Нижний Новго-род: Национальный исследовательский Нижегородский государственный университет им. Н.И. Лобачевского, 2024. С. 650-653.
2. Разин, В.В. Аллергодианостика средствами искусственного интеллекта // Биосистемы: организация, поведение, управление : Тезисы докладов 77-й Всероссийской с между-народным участием школы-конференции молодых ученых, Нижний Новгород, 15–19 апреля 2024 года. Нижний Новгород: Нижегородский госуниверситет им. Н.И. Лоба-чевского, 2024. С. 302.
3. Разин В.В., Краснов А.А., Золотых Н.Ю. Глубокое обучение в задачах диагностики за-болевания по ЭКГ на наборе данных PTB-XL // Математика и ИТ – вместе в цифровое будущее. Сборник трудов Молодежной школы (Н. Новгород, 25–29 апреля 2022 г.). Нижний Новгород: Изд-во Нижегородского госуниверситета, 2022. С.78-86.
4. Краснов А.А., Разин В.В., Золотых Н.Ю. Машинное обучение в задачах диагностики заболевания по ЭКГ на наборе данных PTB-XL // Математика и ИТ – вместе в цифро-вое будущее. Сборник трудов Молодежной школы (Н. Новгород, 25–29 апреля 2022 г.). Нижний Новгород: Изд-во Нижегородского госуниверситета, 2022. С. 44-50.
5. Разин В.В., Краснов А.А. Использование искусственного интеллекта в задачах диагно-стики заболевания по ЭКГ на наборе данных PTB-XL // Путь в науку: прикладная ма-тематика, информатика и информационные технологии. Тезисы докладов Всероссий-ской молодежной научно-практической конференции (Ярославль, 17–21 апреля 2023 г.). Ярославль: Ярославский государственный университет им. П.Г. Демидова, 2023. С. 93-96.
6. Разин В.В., Краснов А.А. Применение машинного обучения в задачах диагностики за-болевания по ЭКГ на наборе данных PTB-XL // Математика и ИТ – вместе в цифровое будущее. Сборник трудов II Молодежной школы (Нижний Новгород, 24–28 апреля 2023 года). Нижний Новгород: Национальный исследовательский Нижегородский гос-ударственный университет им. Н.И. Лобачевского, 2023. С. 73-79.
7. Разин В.В., Краснов А.А. Применение глубокого обучения в задачах диагностики за-болевания по ЭКГ на наборе данных PTB-XL // Математика и ИТ – вместе в цифровое будущее. Сборник трудов II Молодежной школы (Нижний Новгород, 24–28 апреля 2023 года). Нижний Новгород: Национальный исследовательский Нижегородский гос-ударственный университет им. Н.И. Лобачевского, 2023. С. 66-72.
8. Разин В.В., Краснов А.А. Использование глубокого обучения в задачах диагностики заболевания по ЭКГ на наборе данных PTB-XL // Биосистемы: организация, поведение, управление : Тезисы докладов 76-й Всероссийской с международным участием шко-лы-конференции молодых ученых, Нижний Новгород, 11–14 апреля 2023 года. Ниж-ний Новгород: Нижегородский госуниверситет им. Н.И. Лобачевского, 2023. С. 267.
9. Разин В.В., Краснов А.А., Карчков Д.А. Классификация болезней сердца на наборе данных PTB-XL при помощи машинного обучения // Академический форум молодых ученых стран Большой Евразии «Континент науки»: Сборник тезисов докладов, Москва, 01–04 ноября 2023 года. Москва: Центр научно-технических решений, 2023. С. 570-573.
10. Краснов А.А., Разин В.В., Карчков Д.А. Классификация болезней сердца на наборе данных PTB-XL при помощи глубокого обучения // Академический форум молодых ученых стран Большой Евразии «Континент науки»: Сборник тезисов докладов, Москва, 01–04 ноября 2023 года. Москва: Центр научно-технических решений, 2023. С. 548-551.
11. Razin V. et al. Determining the presence of heart disease by ecg in the PTB-XL dataset using deep learning // Математическое моделирование и суперкомпьютерные технологии: Труды XXIII

Международной конференции, Нижний Новгород, 13–16 ноября 2023 го-да. – Нижний Новгород: Национальный исследовательский Нижегородский государ-ственный университет им. Н.И. Лобачевского, 2023. С. 173-176.
12. Краснов А.А., Разин В.В., Карчков Д.А. Использование искусственного интеллекта для классификации ЭКГ // Биосистемы: организация, поведение, управление : Тезисы до-кладов 77-й Всероссийской с международным участием школы-конференции молодых ученых, Нижний Новгород, 15–19 апреля 2024 года. Нижний Новгород: Нижегород-ский госуниверситет им. Н.И. Лобачевского, 2024. С. 185.
13. Razin V.V., Krasnov A.A. Determining the presence of heart disease by ECG in the PTB-XL dataset using machine learning // Математика и ИТ – вместе в цифровое будущее. Сбор-ник трудов III Молодежной школы (Нижний Новгород, 22–26 апреля 2024 года). Ниж-ний Новгород: Национальный исследовательский Нижегородский государственный университет им. Н.И. Лобачевского, 2024. С. 38-42.
14. Razin V.V., Krasnov A.A. Solving the problem of diagnosing a disease by ECG on the PTB-XL dataset using machine learning // Математика и ИТ – вместе в цифровое будущее. Сборник трудов III Молодежной школы (Нижний Новгород, 22–26 апреля 2024 года). Нижний Новгород: Национальный исследовательский Нижегородский государствен-ный университет им. Н.И. Лобачевского, 2024. С. 43-48.
15. Razin V. et al. Solving the Problem of Diagnosing a Disease by ECG on the PTB-XL Dataset Using Deep Learning // International Conference on Neuroinformatics. – Cham: Springer Nature Switzerland, 2023. С. 13-21.
16. Wagner P. et al. PTB-XL, a large publicly available electrocardiography dataset //Scientific data. – 2020. – Т. 7. – №. 1. – С. 1-15.
17. Albawi S., Mohammed T. A., Al-Zawi S. Understanding of a convolutional neural network //2017 international conference on engineering and technology (ICET). – Ieee, 2017. – С. 1-6.
18. Hu Z. et al. 3D separable convolutional neural network for dynamic hand gesture recogni-tion //Neurocomputing. – 2018. – Т. 318. – С. 151-161.
19. Ismail Fawaz H. et al. Inceptiontime: Finding alexnet for time series classification //Data Mining and Knowledge Discovery. – 2020. – Т. 34. – №. 6. – С. 1936-1962.
20. Huang Z., Xu W., Yu K. Bidirectional LSTM-CRF models for sequence tagging //arXiv pre-print arXiv:1508.01991. – 2015.
21. Wang C. et al. Pulmonary image classification based on inception-v3 transfer learning model //IEEE Access. – 2019. – Т. 7. – С. 146533-146541.
22. Hochreiter S., Schmidhuber J. Long short-term memory //Neural computation. – 1997. – Т. 9. – №. 8. – С. 1735-1780.
23. Cho K. et al. Learning phrase representations using RNN encoder-decoder for statistical ma-chine translation //arXiv preprint arXiv:1406.1078. – 2014.
24. Simonyan K., Zisserman A. Very deep convolutional networks for large-scale image recog-nition //arXiv preprint arXiv:1409.1556. – 2014.
25. Jinsakul N. et al. Enhancement of deep learning in image classification performance using xception with the swish activation function for colorectal polyp preliminary screening //Mathematics. – 2019. – Т. 7. – №. 12. – С. 1170.

МОДЕЛИРОВАНИЕ ПРОЦЕССА ПРОПИТКИ ПОЛИДИСПЕРСНОЙ СРЕДЫ ПОЛИМЕРНЫМ СВЯЗУЮЩИМ

С.В. Русаков, В.Г. Гилев

Пермский государственный национальный исследовательский университет

В работе рассматривается задача о создании композитного материала на основе полидисперсной смеси (песка) и полимерного связующего. Предлагается математическая модель двухфазного (жидкость, газ) течения в пористой среде. Обсуждаются результаты серии вычислительных и натурных экспериментов. Поскольку движение фронта насыщения жидкой фазы ограничено временем отверждения полимерного связующего основным результатом процесса пропитки является глубина проникновения связующего в слой песчаной смеси. В результате исследования проведены оценки влияния различных механизмов на рассматриваемый процесс, в том числе капиллярного давления, возникающего в пористой среде.

Ключевые слова: полидисперсная среда, пористость, проницаемость, капиллярное давление.

1. Введение

Эпоксидные полимеры по целому ряду свойств выделяются среди прочих полимерных материалов, играя важную роль в аэрокосмической, автомобильной, судостроительной и других отраслях промышленности. Широкое их применение в технике связано, во-первых, с высокой технологичностью эпоксидных смол, во-вторых, с уникальным сочетанием эксплуатационных характеристик продуктов их отверждения. Особое место занимают композиционные материалы с микроскопическими наполнителями, такими, например, как кварцевый песок, лузга, опилки, полимерные порошки, шламы, стружки – прикладное, но крайне востребованное направление материаловедения. В частности, они используются в изделиях строительного и декоративного назначения [1], для устранения дефектов в виде износов и сквозных протираний поверхности деталей со сложнопрофильной поверхностью [2]. Влияние наполнителя на механические характеристики эпоксидополимерных композиционных материалов рассмотрены в ряде работ (см. например [3]). Для изготовления монолитных конструкций широко используется метод предварительного смешения эпоксидной составляющей с тем или иным наполнителем и последующим формированием готового изделия. Для реставрации трещин, сквозных отверстий такой метод неэффективен. Здесь предпочтительно предварительное заполнение пор наполнителем с последующей пропиткой клеевым материалом. Такой процесс в научной литературе практически не затрагивается. В настоящей работе некоторые аспекты этой проблемы исследуются с помощью математического моделирования.

2. Математическая постановка задачи

На слой полидисперсной среды толщиной L размещается полимеризующаяся эпоксидная композиция (далее просто «композиция») толщиной l_0 в жидкой фазе, которая начинает проникать в пористую среду (см. Рис. 1). При этом, в качестве полидисперсной среды используется смесь «мелкого» и «крупного» песка, характеризующаяся значением C – доли «мелкого» песка в смеси в процентах.

В процессе перехода композиции от жидкого состояния в гель-фракцию её вязкость меняется от конкретного конечного значения до бесконечного. В этом случае процесс «просачивания» завершается, в следствии чего, эпоксидная композиция проникает на конечную глубину.

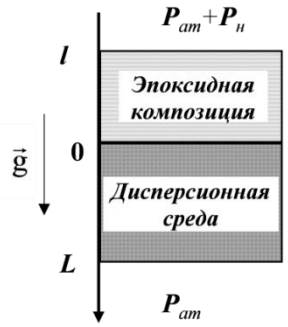


Рис. 1 Геометрия задачи

Расчёт глубины проникновения будем производить с помощью следующей математической модели, являющейся обобщением закона Дарси [4]:

$$\frac{\partial}{\partial x}\left(\frac{kF(S_э)}{\mu_э}\left(\frac{\partial P}{\partial x}-\rho_э g\right)\right)+\frac{\partial}{\partial x}\left(\frac{kF(S_в)}{\mu_в}\left(\frac{\partial P}{\partial x}-\rho_в g\right)\right)=0, \tag{1}$$

$$m\frac{\partial S_э}{\partial t}=\frac{\partial}{\partial x}\left(\frac{kF(S_э)}{\mu_э}\left(\frac{\partial P}{\partial x}-\rho_э g\right)\right), \tag{2}$$

$$S_э+S_в=1. \tag{3}$$

В уравнениях (1)-(3) используются обозначения.

Независимые переменные: t – время в секундах, x – пространство в метрах;

Искомые функции: $P(t, x)$ – давление, $S_э(t, x)$ и $S_в(t, x)$ – насыщенность эпоксидной композиции и воздуха соответственно; параметры модели: $k=k(C)$ и $m=m(C)$ – проницаемость и пористость песчаной смеси; $\rho_э=\rho_э(T)$ и $\rho_в=\rho_в(T)$ – плотность композиции и воздуха в зависимости от температуры T; g – значение ускорения свободного падения; $\mu_э=\mu_э(t;T)$ и $\mu_в=\mu_в(T)$ – динамическая вязкость композиции и воздуха; $F(S)$ – функция Бакли-Леверетта (использовался вариант $F(S)=S^2$) .

В этих обозначениях начальные и граничные условия имеют следующий вид:

$$S_э(0,0)=1,\quad S_в(0,x)=1,\ x\in(0,L], \tag{4}$$

$$P(0,0)=P_H+P_{am}+\rho_э g l_0,\quad P(0,x)=P_{am},\ x\in(0,L], \tag{5}$$

$$S_э(t,0)=1,\quad P(t,0)=P_H+P_{am}+\rho_э g l(t),\quad P(t,L)=P_{am}, \tag{6}$$

где P_{am} – атмосферное давление, P_H – давление приложенное к поверхности композиции (см. Рис. 1). Кроме того, этой величиной можно смоделировать работу капиллярных сил на границе фронта насыщенности.

Проведем обезразмеривание в изотермической постановке $T=const$, выбирая в качестве характерного значения длины начальную величину слоя эпоксидной композиции l_0 , времени – t_*, давления – $\rho_э g l_0$, вязкости – $\mu_э(0)$, плотности – $\rho_э$.

$$\bar{x}=\frac{x}{l_0},\quad \bar{t}=\frac{t}{t_*},\quad \bar{P}=\frac{P-P_{am}}{\rho_э(T)gl_0},\quad \bar{\mu}_э=\frac{\mu_э(\bar{t})}{\mu_э(0)},\quad \bar{\mu}_в=\frac{\mu_в}{\mu_э(0)},\quad \bar{\rho}_в=\frac{\rho_в}{\rho_э},\quad t_*=\frac{m(C)\mu_э(0)}{k(C)\rho_э(T)g}.$$

Обозначим $\Phi_э=\dfrac{F(S_э)}{\bar{\mu}_э},\quad \Phi_э=\dfrac{F(S_в)}{\bar{\mu}_в}$, тогда систему уравнений (1)-(3) можно преобразовать к виду:

$$\frac{\partial}{\partial\bar{x}}\left((\Phi_э+\Phi_в)\frac{\partial\bar{P}}{\partial\bar{x}}-(\Phi_э+\bar{\rho}_в\Phi_э)\right)=0, \tag{7}$$

$$\frac{\partial S_э}{\partial\bar{t}}=\frac{\partial}{\partial\bar{x}}\left(\Phi_э\left(\frac{\partial\bar{P}}{\partial\bar{x}}-1\right)\right). \tag{8}$$

Безразмерная запись начальных и граничных условий (4)-(6) для давления имеет вид:

$$\bar{P}(0,0)=1+\bar{P}_H,\quad \bar{P}(0,\bar{x})=0,\quad \bar{P}(\bar{t},0)=\bar{l}(\bar{t})+\delta,\quad P(\bar{t},\bar{L})=0,\quad x\in(0,\bar{L}], \tag{9}$$

где $\bar{P}_H=\dfrac{P_H}{\rho_э g l_0}$. Задача (7)-(9) решалась численно с учетом особенностей рассматриваемых уравнений.

3. Численные метод решения

Использование балансового соотношения (3) позволяет свести решение задачи к нахождению только одного значения насыщенности, поэтому в дальнейшем ограничимся обозначением $S=S_э$, тогда $1\text{-}S=S_в$. Также будем опускать символ '-' над безразмерными величинами.

Рассмотрим решения уравнения (7), которое напрямую не зависит от времени и для любого значения t может быть записано в виде:

$$\frac{\partial}{\partial x}\left(q(S(x))\frac{\partial P}{\partial x}-f(S(x))\right)=0,\quad q(S(x))=(\Phi_э+\Phi_в),\quad f(S(x))(\Phi_э+\bar{\rho}_в\Phi_э). \tag{10}$$

Проинтегрировав уравнение (10) по переменной x и проведя элементарные преобразования, получим в узлах разностной сетки $x_i,\ i=0,\ldots,N$:

$$\frac{dP(x_i)}{dx}=q^{-1}(x_i)(A+f(x_i)),\quad i=0,\ldots,N, \tag{11}$$

где A – константа интегрирования. Тогда, проинтегрировав разностное уравнение (11) по формуле трапеций, можно найти

$$P(x_0)=l(t)+P_H,\quad P(x_{i+1})=P(x_i)+\frac{h}{2}\left(\frac{dP(x_i)}{dx}+\frac{dP(x_{i+1})}{dx}\right),\quad i=0,\ldots,N-1. \tag{12}$$

Для определения значения константы A воспользуемся следующей процедурой: задав произвольные значения A_1 и A_2, найдем численные решения по формуле (12). Введем обозначения

$$P_1=P(x_N;A_1),\quad P_2=P(x_N;A_2), \tag{13}$$

и построим искомое решение в виде линейной комбинации:

$$P(x_i;A)=\alpha P(x_i;A_1)+(1-\alpha)P(x_i;A_2),\quad i=0,\ldots,N. \tag{14}$$

С учётом краевого условия при $x=x_N$ получаем соотношение

$$\alpha P(x_N;C_1)+(1-\alpha)P(x_N;C_2)=\alpha P_1+(1-\alpha)P_2=0. \tag{15}$$

Отсюда имеем

$$\alpha=\frac{P_2}{P_2-P_1},\quad A=\alpha A_1+(1-\alpha)A_2. \tag{16}$$

Таким образом, из соотношения (11) можно получить значения градиента давления в узлах разностной сетки. Для решения уравнения (8) воспользуемся явной конечно-разностной схемой первого порядка аппроксимации по пространству и времени:

$$\begin{aligned}&\frac{S_i^{(n+1)}-S_i^{(n)}}{\tau}=\frac{Q_i^{(n)}-Q_{i-1}^{(n)}}{h},\quad i=1,\ldots,N,\\ &Q_i^{(n)}=\Phi_э(S_i^{(n)})\left(\frac{dP(x_i)}{dx}-1\right),\quad i=0,\ldots,N,\end{aligned} \tag{17}$$

где n – номер шага по времени, а $\tau=\dfrac{h}{c_\tau}$ – величина шага по времени. В силу нелинейности уравнения величина параметра c_τ определялась эмпирически в ходе его численного решения.

4. Результаты моделирования

В качестве дисперсного материала использовались два сорта кварцевого песка с двумя характерными размерами частиц, условно названные «мелкими» и «крупными». Для исследования статистических характеристик порошков применялся оптический микроскоп. Серии изображений обрабатывались на компьютере при помощи программы ImageJ. После обработки изображений были построены гистограммы распределения по диаметру частиц. Исследуемые образцы имели следующие характеристики: «мелкие» – средний диаметр 0.15 мм, медиана – 0.146; «крупные» – средний диаметр 0.53 мм, медиана – 0.526.

В процессе измерений дисперсная смесь засыпалась в стеклянную трубку, сверху наливался эпоксидный состав (смола L, отвердитель EPH 621) и система помещалась в воздушный термостат при температуре T=45 °C до полной полимеризации. Измеряемыми параметрами являлись: высота эпоксидного состава до и после полимеризации и глубина проникновения в пористую среду. Для каждой концентрации «мелкого» песка в «крупном» проводилась серия измерений, результаты которых усреднялись.

В ходе натурного эксперимента во всех случаях бралось $l_0 = 0.1$ *м* и варьировалась концентрация «мелкого» песка. Было получено, что плотное пропитывание (отсутствие пустот в композитном материале) в данном случае реализуется только для $C \le 40\%$. Поэтому обсудим только эту часть расчётов. Во всех случаях измерялась глубина пропитки l_z, которая очевидно является следствием совместного действия гидростатического и капиллярного давления. На рис. 2 показана толщина полученного слоя композитного материала (глубина пропитки l_z), которая убывает практически линейно с ростом концентрации C. Далее была проведена серия расчётов, позволившая оценить зависимость $P_H(C)$, которую будем считать эквивалентной капиллярному давлению.

В ходе расчетов по рассматриваемой модели методом перебора подбиралась величина $P_H(C)$, обеспечивающая глубину пропитки, соответствующую экспериментальной. Полученные результаты представлены на рис.3.

Для дальнейшей обработки воспользуемся следующим предположением для капиллярного давления:

$$P_{кап} = \alpha \frac{m(C)\cos\left(\varphi(C)\right)}{R_{эф}(C)} \approx P_H(C), \quad R_{эф}(C) = \sqrt{\frac{8\,k(C)\;f(C)}{m(C)}}, \quad f(C) = \frac{0.5035}{m(C)^{1.1}}.$$

Здесь $R_{эф}(C)$ – эффективный радиус пор, определяемый через пористость $m(C)$, проницаемость $k(C)$, структурный коэффициент $f(C)$ [5] и угол смачивания $\varphi(C)$. Пористость и проницаемость определялись в ходе предварительного экспериментального исследования свойств полидисперсной смеси, а константа α и угол смачивания при обработке экспериментальных данных настоящей работы. В результате получено $\alpha = 0.000552$, $\varphi(C) = 0.0162 \cdot C$.

Используя оценку величины капиллярного давления, удалось численно решить задачу определения величины $l_0^*(C)$, при которой весь слой эпоксидной композиции полностью поглощается полидисперсной средой.

В результате была получена зависимость $l_0^*(C) = 0.046 - 5.6 \cdot 10^{-4} C$ ($C \le 40\%$). Если экстраполировать эту зависимость на увеличения концентрации мелкого песка, можно получить, что такое возможно только при значениях $C \le 80\%$. Отметим, что в натурном эксперименте получено $l_0^*(0) \approx 0.05$ *м*, что удовлетворительно согласуется с результатом расчёта.

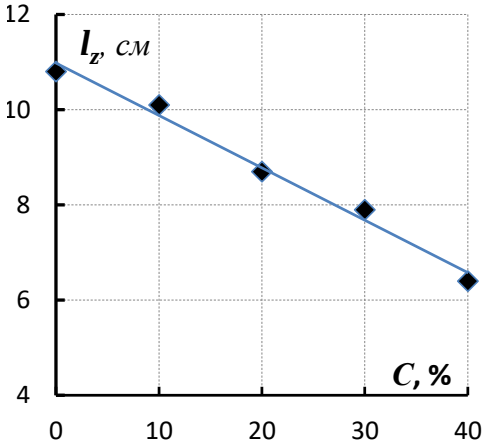


Рис. 2. Глубины пропитки $l_z(C)$

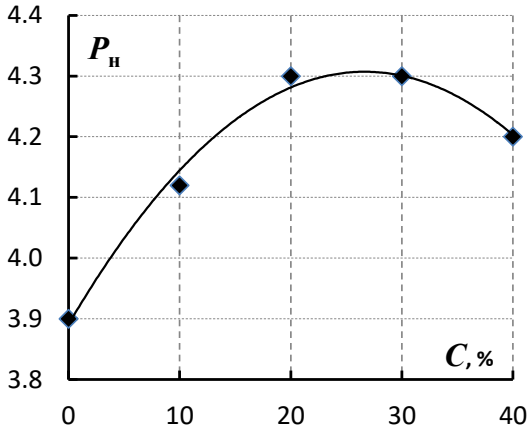


Рис. 3. Безразмерная величина давления $P_H(C)$

Таким образом, получены результаты, показывают, что с помощью предложенной математической модели имеется возможность оценивать капиллярные силы, возникающие в пористой среде при её пропитке полимерной композицией и тем самым прогнозировать максимальную глубину проникновения эпоксидной композиции в полидисперсную среду. С технологической точки зрения, дисперсная среда с преобладанием крупнозернистого песка является более предпочтительней, поскольку обеспечивает большую глубину пропитки.

Литература

1. Старокадомский Д.Л., Решетник М. Н. Влияние мезо- и микрочастиц кварцевого песка различного происхождения на физико-механические свойства эпоксидного композита для изделий строительного и декоративного назначения // Композитный Мир. – 2021. – № 2(95). – С. 64-70.
2. Михальченков А.М., Комогорцев В.Ф., Кононенко А.С. Абразивостойкий композит на основе эпоксидной смолы и кварцевого дисперсного наполнителя // Клеи. Герметики. Технологии. – 2017. – №11. – С.33-38.
3. Хапёрских С.А., Ананьева Е.С. Влияние природы наполнителя на твердость и износ дисперсно-наполненных композиционных материалов на основе эпоксидно-дианового связующего // Ползуновский Вестник. – 2021. – № 4. – С. 163-172.
4. Баренблатт Г.И., Ентов В.М., Рыжик В.М. Теория нестационарной фильтрации жидко-сти и газа. Москва: Недра, 1972. – 288 с.
5. Котяхов Ф.И. Физика нефтяных и газовых коллекторов. Москва: Недра, 1977. – 287 с.

ПРОБЛЕМА ФРОБЕНИУСА ДЛЯ ТРЕХ НАТУРАЛЬНЫХ ЧИСЕЛ СПЕЦИАЛЬНОГО ВИДА

В.П. Савельев

Нижегородский государственный инженерно-экономический университет

В настоящей работе решается задача нахождения числа Фробениуса для трех натуральных чисел вида $a, a+m, a+m+1$. Если ввести натуральное число $k = \left\lfloor \frac{a}{m} \right\rfloor$, где $\lfloor x \rfloor$ означает целую часть числа $\boldsymbol{x}$, то любое натуральное число можно представить в виде $a = km + l, l \in \{0,1,2, \dots, m-1\}$. Для указанных трех чисел получена формула числа Фробениуса при условии, что $k \geq l$.

Ключевые слова: множество Фробениуса, число Фробениуса, аддитивная полугруппа чисел.

1. Введение

Проблеме нахождения числа Фробениуса для конечного множества натуральных чисел посвящено много работ [1]. Эта проблема напрямую связана с решением ряда известных задач дискретной математики, таких, как задача о рюкзаке, задача о размене монет [2=4], некоторые задачи теории графов [5]. Формула для вычисления числа Фробениуса в общем случае известна лишь для двух натуральных чисел: если наибольший общий делитель натуральных чисел a и b равен единице, то

$$frob(a,b) = ab - a - b. \quad (1)$$

В общем случае, такой формулы нет даже для трех натуральных чисел. Однако, в некоторых частных случаях формулы найдены, или предложены алгоритмы вычисления числа Фробениуса [6-8].

2. Вывод формулы

Для исходного множества натуральных чисел $a, a+m, a+m+1$ определим аддитивную полугруппу $G(a, a+m, a+m+1)$, порожденную этими числами. Отметим, что каждый элемент полугруппы можно представить в виде линейной комбинации

$$L = x_1 a + x_2(a+m) + x_3(a+m+1), x_i \in Z_+, i = \overrightarrow{1,3}. \quad (2)$$

Множество $N \backslash G(a, a+m, a+m+1) = \overline{G(a, a+m, a+m+1)}$ называется множеством Фробениуса. Числом Фробениуса $frob(a, a+m, a+m+1)$ называется наибольшее натуральное число, принадлежащее множеству Фробениуса $\overline{G(a, a+m, a+m+1)}$.

Для доказательства, что некоторое число является числом Фробениуса, полезно использовать следующее утверждение [6,стр. 61].

Утверждение 1. Пусть $a_i \in N, i = \overrightarrow{1,n}$, и $d = min\{a_1, a_2, \dots, a_n\}$. Если каждое из d последовательных натуральных чисел, начиная с некоторого числа b, принадлежит аддитивной полугруппе $G(a_1, a_2, \dots, a_n)$, то ей принадлежат все натуральные числа, начиная с числа b.

Утверждение 2. Пусть $m \in N+1, a = km + l, l \epsilon \{0, \dots m-1\}, k \epsilon N, k \geq l$, тогда

$$frob(a, a+m, a+m+1) = \left(k + m - 2 - \left\lfloor \frac{k-l-1}{m+1} \right\rfloor\right) a + m(m-1) - 1. \quad (3)$$

Во-первых, покажем, что уравнение

$$L = x_1 a + x_2(a+m) + x_3(a+m+1) = \left(k + m - 2 - \left\lfloor \frac{k-l-1}{m+1} \right\rfloor\right) a + m(m-1) - 1 \quad (4)$$

не имеет решения в целых неотрицательных числах.

С учетом того, что $a = km + l$, уравнение (4) запишется в виде

$$x_1 km + x_2(k+1)m + x_3(k+1)m + lx_1 + lx_2 + (l+1)x_3 + 1 =$$

$$\left(k+m-2-\left\lfloor\frac{k-l-1}{m+1}\right\rfloor\right)km+l\left(k+m-2-\left\lfloor\frac{k-l-1}{m+1}\right\rfloor\right)+m(m-1). \quad (5)$$

С учетом целочисленности слагаемых, это значит, что имеет место соотношение

$$lx_1+lx_2+(l+1)x_3+1-l\left(k+m-2-\left\lfloor\frac{k-l-1}{m+1}\right\rfloor\right)=im, i\in Z. \quad (6)$$

С учетом (6) соотношение (5) примет вид:

$$(x_1+x_2+x_3)k+x_2+x_3+i=\left(k+m-2-\left\lfloor\frac{k-l-1}{m+1}\right\rfloor\right)k+m-1. \quad (7)$$

Из соотношения (7) следует, что $x_2+x_3+i-m+1$ должно быть кратным числу k, то есть,

$$x_2+x_3+i-m+1=jk, j\in Z. \quad (8)$$

С учетом соотношения (8), соотношение (7) примет вид:

$$x_1+x_2+x_3+j-\left(k+m-2-\left\lfloor\frac{k-l-1}{m+1}\right\rfloor\right)=0. \quad (9)$$

Умножив соотношение (9) на и вычитая его из соотношения (5), получим:

$$x_3+1-lj=im, i\in Z. \quad (10)$$

Таким образом, найдено выражение для множителя x_3

$$x_3=im+lj-1, \quad (11)$$

А из соотношения (8) находим выражение для множителя x_2

$$x_2=j(k-l)+m-i(1+m), j\in Z. \quad (12)$$

Если предположить, что $x_3=im+lj-1\geq 0$, то это будет означать, что хотя бы один из параметров i и j является натуральным числом. Предположим, что $i\epsilon N$. Если предположить, что $x_2\geq 0$, то из неравенства

$$j(k-l)\geq i(1+m)-m \quad (13)$$

следует, что и $j\epsilon N$. Таким образом, будем считать, что $j\epsilon N$.

Из неравенства (13) следует ограничение сверху на параметр i:

$$i\leq\frac{j(k-l)+m}{1+m}. \quad (14)$$

С учетом значений коэффициентов (11) и (12) найдем значение коэффициента x_1 из соотношения (9):

$$x_1=k+m-2-\left\lfloor\frac{k-l-1}{m+1}\right\rfloor-j-im-lj+1+j(l-k)-m+i(1+m)=k-1-j+i-$$
$$jk-\left\lfloor\frac{k-l-1}{m+1}\right\rfloor=k+1-2-j-jk+i-\left\lfloor\frac{k-l-1}{m+1}\right\rfloor=(k+1)(1-j)-2+i-\left\lfloor\frac{k-l-1}{m+1}\right\rfloor.$$

Если $\frac{k-j-1}{m+1}\epsilon Z$, то

$$x_1\leq(k+1)(1-j)-2+\frac{j(k-l)+m}{1+m}-\frac{k-l-1}{1+m}=(k+1)(1-j)-2-\frac{k-l}{1+m}(1-j)+1=$$
$$\left(k+1-\frac{k-l}{1+m}\right)(1-j)-1<0.$$

Если $\frac{k-l-1}{m+1}-$ не является целым числом, то $-\left\lfloor\frac{k-l-1}{m+1}\right\rfloor<-\frac{k-l-1}{m+1}+1$

$$x_1<(k+1)(1-j)-2+\frac{j(k-l)+m}{1+m}-\frac{k-l-1}{1+m}+1=(k+1)(1-j)-1-\frac{k-l}{1+m}(1-j)+1=$$
$$\left(k+1-\frac{k-l}{1+m}\right)(1-j)\leq 0.$$

Итак, у любого решения $X=(x_1,x_2,x_3,)$ уравнения (2), определяемого параметрами $i\in N, j\in N$, одна из компонент будет отрицательна, то есть, число

$f=\left(k+m-2-\left\lfloor\frac{k-l-1}{m+1}\right\rfloor\right)a+m(m-1)-1$ не может быть представлено в виде линейной целочисленной неотрицательной комбинации чисел $a, a+m, a+m+1$.

Далее, число f удобнее представить в виде

$$f=\left(k+m-2-\left\lfloor\frac{k-l-1}{m+1}\right\rfloor\right)a+m(m-1)-1. \quad (15)$$

Во-вторых, покажем, что каждое из натуральных чисел, начиная с числа

$$\left(k+m-2-\left\lfloor\frac{k-l-1}{m+1}\right\rfloor\right)a+m(m-1), \quad (16)$$

и заканчивая числом

$$\left(k+m-2-\left\lfloor\frac{k-l-1}{m+1}\right\rfloor\right)a+m(m-1)+km+l-1, \quad (17)$$

может быть представлено в виде линейной целочисленной неотрицательной комбинации чисел $a, a+m, a+m+1$.

Сначала рассмотрим частный случай, когда $k = l$, то есть, $\left\lfloor \frac{k-l-1}{m+1} \right\rfloor = -1$. В этом случае надо показать, что каждое из натуральных чисел, начиная с числа

$$(k + m - 1)a + m(m - 1), \tag{18}$$

и заканчивая числом

$$(k + m - 1)a + m(m - 1) + k(m + 1) - 1, \tag{19}$$

может быть представлено в виде линейной целочисленной неотрицательной комбинации чисел $a, a + m, a + m + 1$.

Действительно, положим

$$x_1 = k + m - 1 - p, x_2 = p - q, x_3 = q.$$

Тогда линейная комбинация (2) примет вид:

$$L = (k + m - 1)a + pm + q.$$

При последовательном задании натуральных значений параметров p и q

$$p = \overrightarrow{m - 1, k + m - 1},\ q = \overrightarrow{0, m - 1} \tag{20}$$

линейная комбинация (2) примет все натуральные значения от числа (18) до числа (19).

$$(k + m - 1)a + m(k + m - 1) + m - 1 = (k + m - 1)a + m(m - 1) + m(k + 1) - 1.$$

Заметим, что при изменении параметров p и q коэффициенты x_1, x_2, x_3 линейной комбинации (2) остаются неотрицательными.

Теперь рассмотрим общий случай, когда $k > l$.

В этом случае положим

$$x_1 = k + m - 2 - \left\lfloor \frac{k-l-1}{m+1} \right\rfloor - p, x_2 = p - q - t, x_3 = q + t. \tag{21}$$

Тогда линейная комбинация (3) примет вид:

$$L = \left(k + m - 2 - \left\lfloor \frac{k-l-1}{m+1} \right\rfloor\right) a + pm + q + t. \tag{22}$$

На первом этапе при фиксированном значении $t = 0$ и последовательном задании натуральных значений параметров p и q

$$p = \overrightarrow{m - 1, k + m - 2 - \left\lfloor \frac{k-l-1}{m+1} \right\rfloor},\ q = \overrightarrow{0, m - 1} \tag{23}$$

линейная комбинация (2) примет все натуральные значения от числа (16) до числа

$$\left(k + m - 2 - \left\lfloor \frac{k-l-1}{m+1} \right\rfloor\right) a + m\left(k + m - 2 - \left\lfloor \frac{k-l-1}{m+1} \right\rfloor\right) + m - 1 =$$
$$(k + m - 2 - \left\lfloor \frac{k-l-1}{m+1} \right\rfloor)a + m\left(m - 1 - \left\lfloor \frac{k-l-1}{m+1} \right\rfloor\right) + km - 1. \tag{24}$$

На втором этапе при фиксированных значениях

$$p = k + m - 2 - \left\lfloor \frac{k-l-1}{m+1} \right\rfloor, q = m - 1, \tag{25}$$

и последовательном задании натуральных значений параметра

$$t = \overrightarrow{0, m\left\lfloor \frac{k-l-1}{m+1} \right\rfloor + l} \tag{26}$$

линейная комбинация (2) примет все натуральные значения от числа

$$\left(k + m - 2 - \left\lfloor \frac{k-l-1}{m+1} \right\rfloor\right) a + m\left(k + m - 2 - \left\lfloor \frac{k-l-1}{m+1} \right\rfloor\right) + m \tag{27}$$

до числа (17).

Пример. $frob(12{,}15{,}16) = (4 + 3 - 2 - 0)12 + 3 * 2 - 1 = 65.$

Заметим, что для такого рода троек чисел имеется формула Брауэра [9], которая упрощает вычисление числа Фробениуса:

$$frob(a, b, c) = dfrob\left(\frac{a}{d}, \frac{b}{d}, c\right) + (d - 1)c. \tag{28}$$

В соответствии с этой формулой

$$frob(12{,}15{,}16) = 3$$
$$frob(4{,}5{,}16) + 2 * 16 = 3 * 11 + 32 = 65.$$

Литература

1. Alfonsin J. R. The Diophantine Frobenius Problem. Oxford University Press, 2005. 243 P.

2. Шевченко В.Н. Задача о размене, задача Фробениуса и задача групповой минимизации // Комбинаторно-алгебраические методы в прикладной математике. – Горький: Издательство Горьковского университета, 1982. – С. 166-179.
3. E.S. Selmer, Two popular problems in number theory, I. Changing coins, Normat 29(3) (1981), p. 81-87.
4. A. Tripathi, The coin exchange problem for arithmetic progressions, Am. Math. Monthly 101 (1994), p. 779-781.
5. Фомичев В.М. Оценка экспонента некоторых графов с помощью чисел Фробениуса для трех аргументов // Прикладная дискретная математика. 2014. №2(24). С. 88 – 96.
6. Савельев В.П. Задачи дискретной и непрерывной оптимизации. Число Фробениуса. Изд –во LAP LAMBERT Academic Publishing, 2020. – 134 с.
7. F. Curtis. On formulas for the Frobenius number of a numerical semigroup. Math. Scand. 67, 1990. p. 190-192.
8. Савельев В.П., Шевченко В.Н. О двух свойствах числа Фробениуса // Международный научно-исследовательский журнал, выпуск №9 (123), 2022, С. 1-5.
9. A. Brauer and J.E. Shockley, On a problem of Frobenius, J. Reine Angewandte Math. 211 (1962), p. 215-220.

МАТЕМАТИЧЕСКИЕ МОДЕЛИ ДИНАМИКИ МАЛОГО БИЗНЕСА В УСЛОВИЯХ ВЫДЕЛЕНИЯ ГРАНТОВ

В.П. Савельев, Н.И. Сутягина

Нижегородский государственный инженерно-экономический университет

Стабильное развитие малого бизнеса в регионе – необходимое условие жизнеспособности сельских территорий. Поэтому применение разнообразных методов исследования деятельности предприятий малого бизнеса является актуальным и значимым как в теории, так и на практике. Построенные ранее математические модели [1, 2, 3, 4] развития предприятий, оказывающих услуги жителям сельских территорий, позволяют определить условия для стабильного существования малого бизнеса. В настоящее время актуальным представляется рассмотрение влияния грантовой поддержки социально-значимых проектов на устойчивое развитие региональной системы малого бизнеса. Проводимый анализ доказывает, что в соответствующей математической модели область устойчивости в пространстве параметров расширяется.

Ключевые слова: динамическая система, линеаризация, малый бизнес, состояние равновесия, фазовый портрет.

Аналогично построенным моделям [1-4] предполагаем, что трудоспособное население сельской территории состоит из двух частей: первая часть работает в сфере производства товаров и услуг для населения рассматриваемой сельской территории (малый бизнес), вторая часть населения работает, в основном, на внешний сервис. Соответственно, для построения математической модели, описывающей динамику финансовых потоков, мы вводим понятие состояния системы, которое в данном случае является двумерной функцией времени (x(t),y(t)), где x(t) – объем предоставляемых услуг населению сельской территории предприятиями малого бизнеса, y(t) – объем произведенных товаров, выполненных работ и оказанных услуг предприятиями, работающими на внешний сервис.

1. Линейная модель

Предполагается, что товары и услуги предприятий внешнего сервиса пользуются спросом также и внутри сельской территории. Динамика изменения состояния $\big(x(t), y(t)\big)$ задается линейной системой дифференциальных уравнений:

$$\begin{cases} \dot{x} = -ax + by + p, \\ \dot{y} = R + dx - cy. \end{cases} \tag{1}$$

где (в единицу времени) $ax -$ текущие расходы предприятий малого бизнеса (зарплата, закупка оборудования и т.д.), $by -$ текущий доход малого бизнеса от домашних хозяйств работников внешнего сервиса, $p -$ грантовая поддержка предприятий малого бизнеса, R – доход предприятий внешнего сервиса от продажи товаров и услуг вне данной сельской территории, dx – доход предприятий внешнего сервиса от домашних хозяйств работников малого бизнеса, cy – текущие расходы предприятий внешнего сервиса (зарплата, закупка оборудования и т.д.).

При естественном предположении, что $a > d, c > b,$ находим состояние равновесия системы (1):

$$x^* = \frac{bR + pc}{ac - bd},\ y^* = \frac{aR + dp}{ac - bd}. \tag{2}$$

После введения новых переменных $x = X + \frac{bR + pc}{ac - bd},\ y = Y + \frac{aR + dp}{ac - bd}$ система (1) примет вид однородной системы с состоянием равновесия $(0,0)$:

$$\begin{cases} \dot{X} = -aX + bY, \\ \dot{Y} = dX - cY. \end{cases} \tag{3}$$

Из уравнения $\lambda^2+(a+c)\lambda+ac-bd=0$ находятся собственные значения матрицы:

$$\lambda_{1,2}=\frac{-(a+c)\pm\sqrt{(a+c)^2-4(ac-bd)}}{2}=\frac{-(a+c)\pm\sqrt{(a-c)^2+4bd}}{2} \quad (4)$$

Поскольку собственные значения матрицы действительные и отрицательные, состояние равновесия $\left(\frac{bR+pc}{ac-bd},\frac{aR+dp}{ac-bd}\right)$ является устойчивым узлом (см. рис. 1).

То есть, грантовая поддержка предприятий малого бизнеса сельских территорий способствует увеличению как объема предоставляемых услуг населению предприятиями малого бизнеса, так и объема произведенных товаров, выполненных работ и оказанных услуг предприятиями, работающими на внешний сервис.

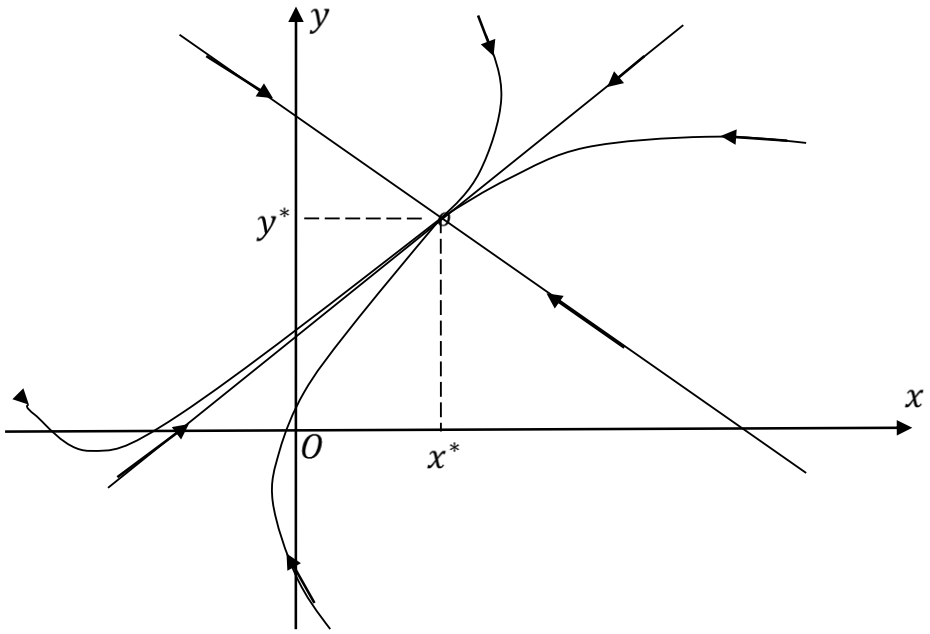


Рис. 1. Фазовый портрет системы (1)

Поведение фазовых траекторий на оси Ox определяется вектором $(\dot{x},\dot{y})=(-ax+p,R+dx)$ Это значит, что траектории пересекают ось Ox снизу вверх при $x>-\frac{R}{d}$, и сверху вниз при $x<-\frac{R}{d}$. Поведение фазовых траекторий на оси Oy определяется вектором $(\dot{x},\dot{y})=(p+by,R-cy)$, то есть, траектории пересекают ось Oy слева направо при $y>-\frac{p}{b}$, и справа налево при $y<-\frac{p}{b}$.

2. Нелинейная модель

В линейной модели (1) предполагалось, что доход by системы предприятий типа X зависит лишь от объема y произведенных товаров и услуг предприятиями типа Y. Однако, такое предположение справедливо лишь при постоянном объеме x услуг малого бизнеса. Очевидно, чем больше развит рынок внутренних услуг, тем больший доход могут получить предприятия малого бизнеса. В соответствии с этим составим систему нелинейных дифференциальных уравнений

$$\begin{cases}\dot{x}=-ax+bxy+p,\\ \dot{y}=R+dx-cy-bxy,\end{cases} \quad (5)$$

где (в единицу времени) $bxy-$ доход системы предприятий малого бизнеса от работников предприятий внешнего сервиса, $cy-$ оплата труда работников предприятий типа Y за вычетом их затрат на внутренний сервис. Динамическая система (5) имеет два состояния равновесия $M_1(x_1,y_1),M_2(x_2,y_2)$ с координатами:

$$x_1 = \frac{(R+p)b - ac + \sqrt{\left[(R+p)b - ac\right]^2 + 4bcp(a-d)}}{2b(a-d)},\ y_1 = \frac{(R+p)b + ac - \sqrt{\left[(R+p)b - ac\right]^2 + 4bcp(a-d)}}{2bc} \quad (6)$$

$$x_2 = \frac{(R+p)b - ac - \sqrt{\left[(R-p)b - ac\right]^2 + 4bcp(a-d)}}{2b(a-d)},\ y_2 = \frac{(R+p)b + ac + \sqrt{\left[(R+p)b - ac\right]^2 + 4bcp(a-d)}}{2bc} \quad (7)$$

В окрестности первого состояния равновесия линеаризованная система (6) имеет вид:

$$\begin{cases} \dot{X} = \frac{(R+p)b - ac - \sqrt{\left[(R+p)b - ac\right]^2 + 4bcp(a-d)}}{2c} X + \frac{(R+p)b - ac + \sqrt{\left[(R+p)b - ac\right]^2 + 4bcp(a-d)}}{2(a-d)} Y, \\ \dot{Y} = -\frac{(R+p)b + ac - 2cd - \sqrt{\left[(R+p)b - ac\right]^2 + 4bcp(a-d)}}{2c} X - \frac{(R+p)b + ac - 2cd + \sqrt{\left[(R+p)b - ac\right]^2 + 4bcp(a-d)}}{2(a-d)} Y \end{cases} \quad (8)$$

где $X = x - x_1, Y = y - y_1$.

Собственные значения матрицы линейной системы (8) находятся из уравнения

$$\lambda^2 + A\lambda + B = 0,$$

где

$$A = \frac{\sqrt{\left[(R+p)b - ac\right]^2 + 4bcp(a-d)} - \left[(R+p)b - ac\right]}{2c} + \frac{(R+p)b + ac - 2cd + \sqrt{\left[(R+p)b - ac\right]^2 + 4bcp(a-d)}}{2(a-d)}$$

$$B = \frac{\sqrt{\left[(R+p)b - ac\right]^2 + 4bcp(a-d)} - \left[(R+p)b - ac\right]}{2c} \cdot \frac{(R+p)b + ac - 2cd + \sqrt{\left[(R+p)b - ac\right]^2 + 4bcp(a-d)}}{2(a-d)} +$$

$$+ \frac{(R+p)b + ac - 2cd - \sqrt{\left[(R+p)b - ac\right]^2 + 4bcp(a-d)}}{2c} \cdot \frac{(R+p)b - ac + \sqrt{\left[(R+p)b - ac\right]^2 + 4bcp(a-d)}}{2(a-d)} =$$

$$= \sqrt{\left[(R+p)b - ac\right]^2 + 4bcp(a-d)} > 0$$

При выполнении простого условия

$$(R+p)b - cd > 0 \quad (9)$$

коэффициент A также положителен, так как выполняется неравенство

$$(R+p)b + ac - 2cd = (R+p)b - cd + c(a-d) > 0.$$

В этом случае собственные числа матрицы либо оба действительные и отрицательные, либо комплексно-сопряженные с отрицательной действительной частью. Соответствующее состояние равновесия $M_1(x_1, y_1)$ является либо устойчивым узлом, либо устойчивым фокусом.

В окрестности второго состояния равновесия линеаризованная система имеет вид:

$$\begin{cases} \dot{X} = \frac{(R-p)b - ac + \sqrt{\left[(R-p)b + ac\right]^2 - 4bcp(a-d)}}{2c} X + \frac{(R-p)b - ac - \sqrt{\left[(R-p)b + ac\right]^2 - 4bcp(a-d)}}{2a} Y, \\ \dot{Y} = -\frac{(R-p)b + ac + \sqrt{\left[(R-p)b + ac\right]^2 - 4bcp(a-d)}}{2c} X - \frac{(R-p)b + ac - \sqrt{\left[(R-p)b + ac\right]^2 - 4bcp(a-d)}}{2a} Y. \end{cases} \quad (10)$$

Уравнение для собственных значений матрицы системы (11) имеет вид:

$$\lambda^2 + C\lambda + D = 0, \text{ где } D = -\sqrt{\left[(R-p)b - ac\right]^2 + 4bcp(a-d)} < 0$$

Поскольку $D < 0$, собственные значения действительные и разных знаков, то есть, второе состояние равновесия является седлом. Фазовый портрет системы (5) представлен на рисунке 2. Поведение фазовых траекторий системы (5) на оси Ox определяется вектором $(\dot{x}, \dot{y}) = (-ax + p, R + dx)$. Это значит, что фазовые траектории пересекают ось Ox снизу вверх при условии $x > -\frac{R}{d}$, и сверху вниз при условии $x < -\frac{R}{d}$.

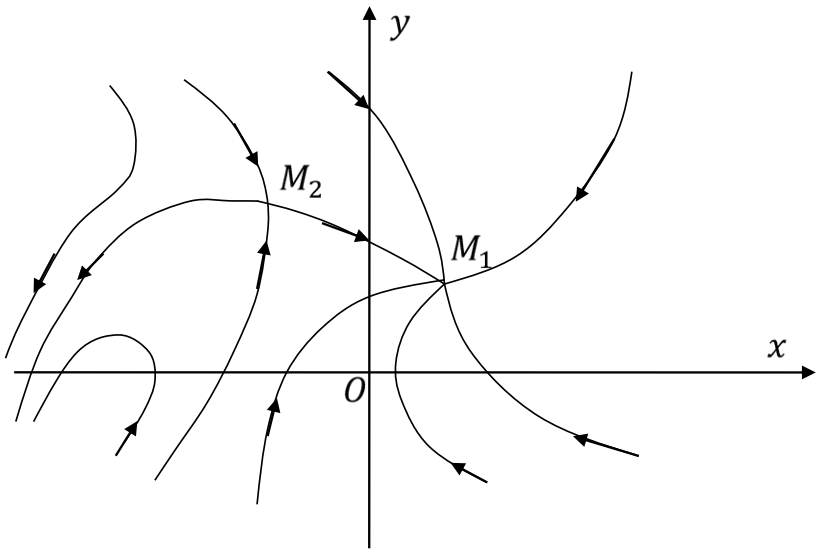


Рис. 2. Фазовый портрет системы (5) с устойчивым узлом

Заметим, что и нелинейная модель демонстрирует положительный эффект грантовой поддержки малого бизнеса, заключающийся в расширении области притяжения устойчивого состояния равновесия, расположенной справа от ω- сепаратрис седловой точки M_2. В случае $p = 0$ седловая точка расположена на оси Ox.

Литература

1. Савельев В.П., Сутягина Н.И. Простые модели динамики малого бизнеса, Труды XX Международной конференции «Математическое моделирование и суперкомпью-терные технологии», г. Н. Новгород, ноябрь 2020 г., С. 329-332.
2. Савельев В.П., Сутягина Н.И. Математические модели динамики регионального бизнеса // Проблемы информатики, №2, 2021, С. 49-58.
3. Савельев В.П., Сутягина Н.И. Модель регионального бизнеса в условиях конкуренции, Труды XXI Международной конференции «Математическое моделирование и суперкомпьютерные технологии», г. Н. Новгород, ноябрь 2021 г., С. 298-300.
4. Савельев В.П., Сутягина Н.И. Модель регионального бизнеса в условиях выплат по кредиту, Труды XXII Международной конференции «Математическое моделирование и суперкомпьютерные технологии», г. Н. Новгород, ноябрь 2022 г., С. 119-122.
5. Андронов А.А., Леонтович Е.А., Гордон И.И., Майер А.Г. Качественная теория динамических систем второго порядка. М.: Наука, 1966.

ИССЛЕДОВАНИЕ ДЛИН ПЕРИОДОВ ЦЕПНЫХ ДРОБЕЙ ДЛЯ КВАДРАТНОГО КОРНЯ ИЗ СТЕПЕНЕЙ ПРОСТЫХ ЧИСЕЛ

С.В. Сидоров, Д.В. Усов

Нижегородский государственный университет им. Н.И. Лобачевского

В докладе представлено экспериментальное исследование длин $l(p^{2n+1})$ периодов цепных дробей для квадратных корней из p^{2n+1}, где p – простое число. Выдвинута гипотеза о связи минимального решения (x_0, y_0) уравнения Пелля $x^2 - py^2 = \pm 1$ с величиной $l(p^{2n+1})$. А именно, предположено существование предела $\lim\limits_{n\to\infty} \frac{p^n \cdot \ln(x_0 + y_0\sqrt{p})}{l(p^{2n+1})} = L$, где $L = \frac{\pi^2}{12 \cdot \ln(2)} = 1.18656911\ldots$ – константа Леви. Из этой гипотезы выведено существование предела $\lim\limits_{n\to\infty} \frac{l(p^{2n+1})}{l(p^{2n-1})} = p$ при фиксированном p.

Ключевые слова: цепная дробь, простое число, константа Леви, период цепной дроби, уравнение Пелля.

1. Введение

Одним из важных разделов теории чисел являются цепные дроби (см. [1-7]). Они возникают, например, в связи с диофантовыми приближениями действительных чисел и при решении диофантовых уравнений второй степени. В частности, существует классический метод (см. [8, 9]) нахождения решений уравнения Пелля $x^2 - Dy^2 = \pm 1$ и обобщённого уравнения Пелля $x^2 - Dy^2 = N$, где $D \in \mathbb{N}$ – не является полным квадратом, а $N \in \mathbb{Z}$. Всякое положительное действительное число x можно представить в виде простой цепной дроби

$$x = q_0 + \cfrac{1}{q_1 + \cfrac{1}{q_2 + \cdots}},$$

где $q_0, q_1, q_2, \cdots$ – натуральные числа, называемые неполными частными. Часто цепную дробь записывают следующим образом: $[q_0, q_1, q_2, \cdots]$.

Лагранжем и Лежандром было доказано (см. [2, 5]), что если $D \in \mathbb{N}$ не является полным квадратом, то цепная дробь для $\sqrt{D}$ является периодической и имеет вид

$$\sqrt{D} = [q_0, \overline{q_1, \cdots, q_{l-1}, q_l}],$$

где l – длина периода, а черта сверху означает периодическую часть. Более того, блок $q_1, q_2, \cdots, q_{l-1}$ является палиндромом, т.е. $q_i = q_{l-i}$ для $l = 1,2, \cdots, l-1$, а $q_l = 2q_0$. Длину периода цепной дроби для $\sqrt{D}$ будем обозначать $l(D)$.

Например, $\sqrt{19} = [4, \overline{2,1,3,1,2,8}], l(19) = 6$.

2. Описание экспериментов и основные результаты

В неопубликованной заметке K. Matthews [10] приведена таблица для значений $l(p^{2n+1})$ при $p = 2$ и $0 \le n \le 35$ и выдвинута гипотеза о том, что существует предел $\lim\limits_{n\to\infty} \frac{l(2^{2n+1})}{2^n} = 0.7427\ldots$ Из этого, в частности, следует, что существует предел $\lim\limits_{n\to\infty} \frac{l(2^{2n+1})}{l(2^{2n-1})} = 2$.

Целью нашей работы является продолжение этого исследования, а именно, численное изучение закономерностей поведения длин периодов цепных дробей для $\sqrt{p^{2n+1}}$, где p – простое число, $n = 0,1,2, \ldots$ Были проведены эксперименты для всех простых $p \le 59$. Результаты приведены ниже в таблицах.

Таблица 1. Результаты экспериментов для $p = 2$

n	$l(2^{2n+1})$	$\frac{l(2^{2n+1})}{l(2^{2n-1})}$	$\frac{l(2^{2n+1})}{2^n}$
0	1	-	1
1	2	2	1
2	4	2	1
3	4	2	0.5
4	12	3	0.75
5	24	2	0.75
6	48	2	0.75
7	96	2	0.75
8	196	2.0416	0.765625
9	368	1.8775	0.71875
10	760	2.0652	0.742188
11	1524	2.0052	0.744141
12	3064	2.0104	0.748047
13	6068	1.9804	0.740723
14	12168	2.0052	0.742676
15	24360	2.0019	0.743408
16	48668	1.9978	0.742615
17	97160	1.9963	0.741272
18	194952	2.0065	0.743683
19	389416	1.9974	0.742752

n	$l(2^{2n+1})$	$\frac{l(2^{2n+1})}{l(2^{2n-1})}$	$\frac{l(2^{2n+1})}{2^n}$
20	778832	2	0.742752
21	1557780	2.000148941	0.742807
22	3116216	1.99916	0.742963
23	6229836	1.99916	0.742654
24	12462296	2.000421199	0.74281072
25	24923320	1.999897932	0.7427728
26	49849604	2.000118925	0.74281698
27	99694536	1.999906278	0.74278217
28	199394616	2.00005561	0.74280282
29	398783628	1.999971895	0.74279239
30	797556364	1.999972687	0.742782246
31	1595117676	2.000006204	0.7427845504
32	3190297400	2.000038899	0.742798997
33	6380517544	1.999975784	0.74279
34	12761088588	2.000008385	0.742793117
35	25522110948	1.99999481	0.7427911899
36	51044281208	2.000002324	0.742792053
37	102088450460	1.999997807	0.7427912384
38	204177067944	2.000001636	0.742791846
39	408353857832	1.999998638	0.74279134029

Результаты для других p сведены в одну таблицу.

Таблица 2. Результаты экспериментов для $p \leq 59$

p	n	$\frac{l(p^{2n+1})}{p^n}$	$\frac{l(p^{2n+1})}{l(p^{2n-1})}$	$\frac{p^n \ln(x_0 + y_0\sqrt{p})}{l(p^{2n+1})}$	p	n	$\frac{l(p^{2n+1})}{p^n}$	$\frac{l(p^{2n+1})}{l(p^{2n-1})}$	$\frac{p^n \ln(x_0 + y_0\sqrt{p})}{l(p^{2n+1})}$
2	39	0.74279	1.9999986	1.1865695	29	6	4.16474	29.0010284	1.1865542
3	19	1.10988	2.9999420	1.1865767	31	6	6.7587	31.0016083	1.1865615
5	15	1.21664	5.0000045	1.1865720	37	6	2.09997	37.0023149	1.1865776
7	9	2.33313	6.9988100	1.1866703	41	6	3.50519	41.0014473	1.1865629
11	9	2.5226	11.000035	1.1865621	43	6	7.45724	43.0008423	1.1865654
13	9	3.02066	12.999823	1.1865743	47	6	3.84658	46.9996326	1.1865708
17	8	1.76535	17.000009	1.1865705	53	6	4.96992	53.0001364	1.1865697
19	8	4.91244	19.000098	1.1865669	59	5	5.87071	59.0002668	1.1865729
23	6	3.26222	23.001095	1.1865445					

Результаты экспериментов позволяют выдвинуть следующие гипотезы.

Гипотеза 1. *Пусть $l(p^{2n+1})$ – длина периода цепной дроби для $\sqrt{p^{2n+1}}$, где p – простое число. Тогда существует*

$$\lim_{n\to\infty} \frac{l(p^{2n+1})}{l(p^{2n-1})} = p.$$

Гипотеза 2. *Пусть (x_0, y_0) – минимальное положительное решение уравнения Пелля $x^2 - py^2 = \pm 1$, где p – простое число. Тогда существует*

$$\lim_{n\to\infty} \frac{p^n \cdot \ln(x_0 + y_0\sqrt{p})}{l(p^{2n+1})} = L,$$

где $L = \frac{\pi^2}{12\ln(2)} = 1.1865691104156254528217\ldots$ – константа Леви.

В работе [11] была выдвинута похожая гипотеза с некоторыми дополнительными условиями.

Заметим, что Гипотеза 1 является следствием Гипотезы 2. Действительно, рассмотрим величину $\alpha(n) = \frac{p^n \cdot \ln(x_0 + y_0\sqrt{p})}{l(p^{2n+1})}$. Тогда согласно Гипотезе 2 имеем $\lim_{n\to\infty} \alpha(n) = L$, поэтому $\lim_{n\to\infty} \frac{\alpha(n)}{\alpha(n-1)} = 1$. Поскольку $\frac{\alpha(n)}{\alpha(n-1)} = \frac{p \cdot l(p^{2n-1})}{l(p^{2n+1})}$, то $\lim_{n\to\infty} \frac{p \cdot l(p^{2n-1})}{l(p^{2n+1})} = 1$, следовательно, $\lim_{n\to\infty} \frac{l(p^{2n+1})}{l(p^{2n-1})} = p$.

Литература

1. Mollin R. A. Fundamental number theory with applications. – Crc Press, 1997.
2. Davenport, H. The higher arithmetic: An introduction to the theory of numbers. Cambridge University Press, 1999. 241 p.
3. Niven, I., Zuckerman, H.S., Montgomery, H.L. An introduction to the theory of numbers. John Wiley & Sons, 1991. 529 p.
4. Olds, C. Continued Fractions. Random House, 1963. 162 p.

5. Perron, O. Die Lehre von den Kettenbrüchen: Band I: Elementare Kettenbrüche. Springer-Verlag, 2013. 194 p. DOI:10.1007/978-3-663-12289-0.
6. Rockett, A., Szusz, P. Continued fractions. World Scientific, 1992. 200 p.
7. Rosen, K.H.: Elementary number theory and its applications. Pearson Education London, 2011. 752 p.
8. Matthews K. R., Robertson J. P. On solving a binary quadratic Diophantine equation //Rocky Mountain Journal of Mathematics. – 2021. – Т. 51. – №. 4. – p. 1369-1385.
9. Mollin R. A. Simple continued fraction solutions for Diophantine equations //Expositiones Mathematicae. – 2001. – Т. 19. – №. 1. – p. 55-73.
10. Matthews K. R. On the continued fraction expansion of sqrt(2^2n+1) period.pdf (numbertheory.org).
11. Тахтаджян Л. А., Асимптотическая формула для суммы длин периодов квадратичных иррациональностей дискриминанта D // Зап. научн. сем. ЛОМИ, 1979, том 91, С. 134-144.

ТРЕХМЕРНАЯ РЕКОНСТРУКЦИЯ ОБЪЕКТА ПО ДВУМЕРНЫМ УЛЬТРАЗВУКОВЫМИ СНИМКАМ С ИСПОЛЬЗОВАНИЕМ РОБОТА-МАНИПУЛЯТОРА[1*]

В.С. Соколов

Научно-технологический университет «Сириус»

Ультразвуковая диагностика является одним из наиболее распространённых методов обследования в медицине. Автоматизация данной процедуры позволит уменьшить нагрузку на медицинский персонал, а также даст доступ к медицинской помощи в отдалённых регионах и при чрезвычайных ситуациях. В данной работе предлагается прототип робототехнической системы для проведения трёхмерной реконструкции объекта по набору двумерных ультразвуковых снимков.

Ключевые слова: робототехника, автоматизация, ультразвук, промышленный манипулятор, датчик по силе, управление, реконструкция

1. Введение

С момента появления первых манипуляторов и зарождения науки робототехники идут попытки их внедрения не только в промышленность, но также и в медицину [1-2]. Медицинские приложения имеют своей спецификой необходимость более точного контроля за усилиями в контактных операциях [3-5], это может быть, как контакт хирургического инструмента с мягкими тканями, так и контакт ультразвукового датчика с телом. На данный момент известно некоторые количество коммерческих разработок хирургических робототехнических систем. Данные системы ввиду дороговизны и безопасности [6] так ещё и не получили широко распространения.

Автору данной статьи видеться, что более перспективной областью применения робототехнических систем в медицине являются различные малоинвазивные методы диагностики, такие как: ультразвук, физиотерапия и реабилитационная медицина. Для таких приложений вопросы безопасности менее критичны, а самое главное ожидаемая стоимость таких систем значительно ниже, чем для хирургических приложений. Уже существует ряд разработок, где манипулятор передвигает ультразвуковой датчик вместо руки врача, но в подавляющем большинстве случаев это системы для телемедицины , т.е. врач контролирует движение и усилие прикладываемое к ультразвуковому датчику с помощью какой-либо хаптик системы (haptic feedback).

В данной работе предлагается прототип системы, которая помимо автоматического сбора двухмерных ультразвуковых снимков в B-режиме, позволяет осуществлять трёхмерную реконструкцию объекта при линейном паттерне сканирования с помощью робота манипулятора (рис. 1).

2. Описание эксперимента

2.1. Сканирование фантома

Для эксперимента был изготовлен искусственный фантом на основе желатина, в объём которого был погружен объект в форме пирамиды для последующей трёхмерной реконструкции Сбор ультразвуковых снимков по линейному паттерну сканирования, как представлено на рис. 1. Для управления манипулятором IRB1600 фирмы ABB используется программа, написанная на языке RAPID и осуществляющая линейное сканирование с записью положения ультразвукового датчика на основе показаний энкодеров в файл, для последующей оффлайн обработки и трёхмерной

[1*] Исследование выполнено при поддержке гранта Российского научного фонда No 24-11-20009.

реконструкции. В ходе эксперимента количество полученных снимков порядка сотни, снимки отражающие характерную геометрию сечений пирамиды представлены на рис. 2.

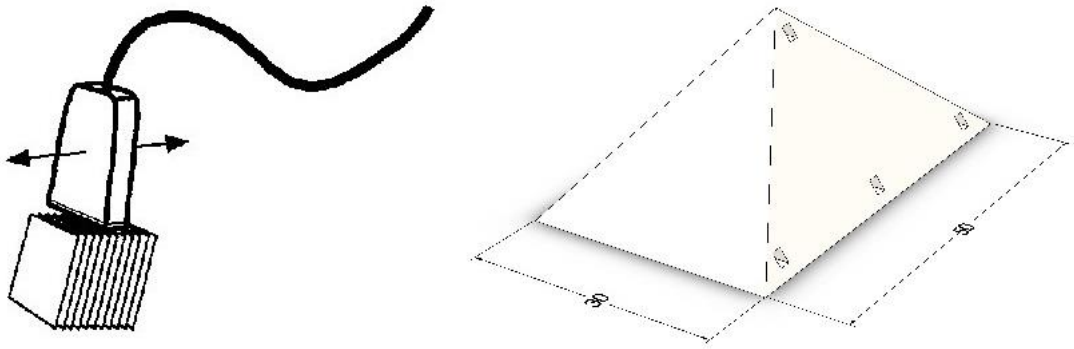


Рис. 1. Линейный паттерн сканирования (слева) и объект для реконструкции (справа)

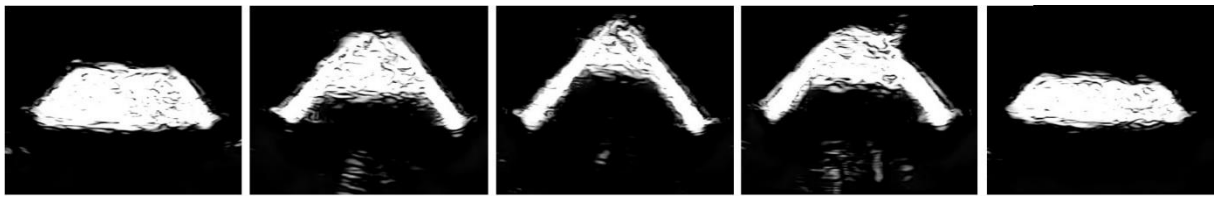


Рис. 2. Ультразвуковые снимки пирамиды погруженной в желатин при линейном сканировании

2.2. Трёхмерная реконструкция

Трёхмерная реконструкция на основе полученных снимков и их положений в пространстве осуществляется с помощью алгоритма шагающих кубов (The Marching Cubes Algorithm), который реализован в используемой программе 3D Slicer [7]. Результат реконструкции представлен на рис. 4. В качестве валидации адекватности реконструкции проведено сравнение оцененных габаритных размеров объекта с его реальными размерами, как видно из сравнения габаритных размеров на рис. 1 и рис. 3, достигнута удовлетворительная точность при условии использования некалиброванного ультразвукового датчика, что является само по себе отдельной сложной задачей.

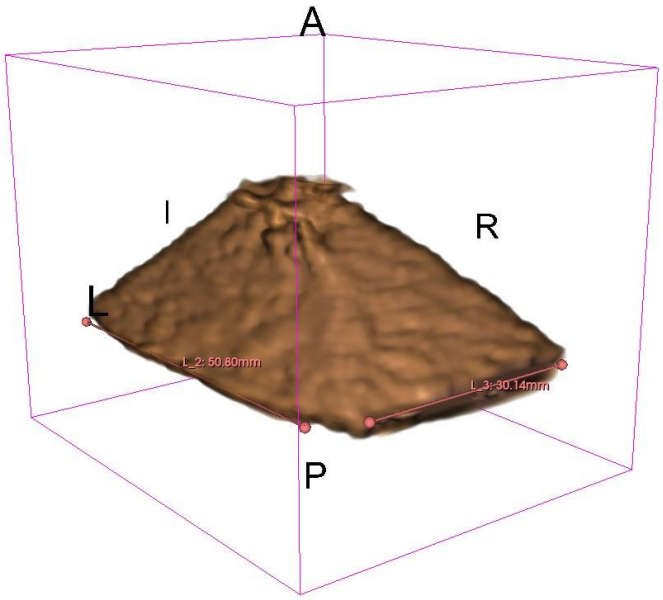


Рис. 3. Результат реконструкции с оценками габаритных размеров

3. Заключение

В данной работе предложен подход использования робота-манипулятора для автоматизации процесса ультразвукового сканирования, а также показана возможность осуществлять трёхмерную реконструкции на основе двумерных снимков. В медицинской практике для получения трёхмерной реконструкции используются специальные механические приспособления для слежения за положением датчика, либо используются более дорогие датчики с двухмерной решёткой пьезоэлементов, тогда как в предложенном прототипе, положения считываются непосредственно с энкодеров манипулятора. Дальнейшее улучшение прототипа предполагает разработку собственного регулятора для поддержания контакта ультразвукового датчика с поверхностью, чтобы сделать решение независимым от конкретного производителя манипулятора и их проприетарных библиотек.

Литература

1. Peters BS, Armijo PR, Krause C, Choudhury SA, Oleynikov D. Review of emerging surgical robotic technology, Surg Endosc, 2018, vol. 32, no. 4, p. 1636-1655.
2. Han J, Davids J, Ashrafian H, Darzi A, Elson DS, Sodergren M. A systematic review of robot-ic surgery: From supervised paradigms to fully autonomous robotic approaches, Int J Med Robot, 2022, vol. 18, no. 2.
3. Hogan N. Contact and Physical Interaction, Annual Review of Control, Robotics, & Autono-mous Systems, 2022, vol. 5, p. 179-203.
4. M. Vukobratovic and A. Tuneski. Contact control concepts in manipulation robotics/spl mi-nus/an overview, in IEEE Transactions on Industrial Electronics, vol. 41, no. 1, p. 12-24.
5. A. Bicchi and V. Kumar. Robotic grasping and contact: a review, Proceedings 2000 ICRA, USA, 2000, vol.1, p. 348-353.
6. Baowei Fei, Wan Sing Ng, Sunita Chauhan, Chee Keong Kwoh. The safety issues of medical robotics, Reliability Engineering & System Safety, 2001, vol. 1, no. 2, p. 183-192.
7. 3D Slicer, URL: https://www.slicer.org/.

ПОСТРОЕНИЕ ОБОБЩЕННОГО H_∞-ОПТИМАЛЬНОГО УПРАВЛЕНИЯ ДИСКРЕТНОЙ СИСТЕМОЙ ПО АПРИОРНЫМ И ЭКСПЕРИМЕНТАЛЬНЫМ ДАННЫМ[1*]

А.В. Степанов

Нижегородский государственный архитектурно-строительный университет
Научно-технологический университет «Сириус»

В статье рассматривается метод синтеза управления гашения начальных и внешних возмущений для дискретных систем с полностью неизвестными параметрами. Представленный подход основан на использовании априорных и экспериментальных данных, что позволяет реализовать обобщенное $\boldsymbol{H}_\infty$-оптимальное управление без построения математической модели объекта. Применение линейных матричных неравенств и техники двойственных систем позволяет находить робастные законы управления. В работе приводятся результаты численных экспериментов, демонстрирующие эффективность метода.

Ключевые слова: обобщенная $\boldsymbol{H}_\infty$-норма, неопределенность, робастное управление, экспериментальные данные, двойственные системы, линейные матричные неравенства.

1. Введение

В последнее время активные разработки в теории управления направлены на построение законов управления для динамических объектов в условиях неопределенности их математических моделей, действующих внешних возмущений и неизвестных начальных условий. При таком подходе предполагается, что с объектом можно провести серию экспериментов, задавая входные воздействия и измеряя выходные переменные. Задача сводится к тому, чтобы на основе полученных измерений и априорных данных, не идентифицируя неизвестные параметры объекта, определить параметры обратных связей, при которых гарантируется определенное качество управления.

В [1] было установлено, что для полной характеризации линейной стационарной динамической системы может быть использована единственная траектория, полученная при выполнении так называемого условия неисчезающего возбуждения. С учетом этого результата в [3] для объектов с неизвестными матрицами уравнения динамики состояния и заданными матрицами уравнения целевого выхода были предложены варианты прямого синтеза законов управления на основе экспериментальных данных, в которых выполняется указанное условие. В [4] было показано, что для построения законов управления по экспериментальным данным достаточно выполнения условия информативности данных по отношению к синтезируемому закону управления, которое менее ограничительно по сравнению с условием неисчезающего возбуждения. В [5] построение H_2- и H_∞-оптимальных законов управления полностью неопределенным дискретным объектом по измерениям входа и выхода осуществляется на основе матричного варианта S-леммы [6], а в [5] – на основе леммы Питерсена [2]. В [8,9,10] параметры обратной связи по состоянию находятся на основе априорных данных и измерений входа и выхода дискретного неопределенного объекта, на который действует неизмеряемое возмущение из определенного класса.

В настоящей работе синтез обобщенных H_∞-оптимальных, обеспечивающих гашение начального и/или внешнего возмущения с полностью неизвестными матрицами уравнений динамики состояния и целевого выхода осуществляется на основе априорных и экспериментальных данных для дискретных объектов.

[1*] Работа выполнена при финансовой поддержке РНФ (проект № 24-11-20023).

2. Постановка задачи

Рассмотрим дискретную систему следующего вида:

$$\begin{aligned} x(t+1) &= Ax(t) + Bu(t) + w(t) \qquad x(0) = x_0 \\ z(t) &= C\,x(t) + D\,u(t), \end{aligned} \tag{2.1}$$

где $x(t) \in R^{n_x}$ – состояние системы, $u(t) \in R^{n_u}$ – управляющее воздействие, $w(t) \in R^{n_w}$ – внешние возмущения, а $z(t) \in R^{n_z}$ – целевой выход. Матрицы A, B, C и D неизвестны.

Информация о неизвестных параметрах системы (2.1) извлекается из конечного набора измерений ее траектории. Измерения состояния дискретной системы $x_0, x_1, \dots, x_N$ и целевого выхода $z_0, \dots, z_{N-1}$ получены при выбранных управлениях $u_0, \dots, u_{N-1}$ и некотором неизвестном возмущении $w_0, \dots, w_{N-1}$. Составим матрицы

$$\Phi = (x_0 \cdots x_{N-1}), \quad \Phi_+ = (x_1 \cdots x_N), U = (u_0 \cdots u_{N-1}), \quad W = (w_0 \cdots w_{N-1}), \quad Z = (z_0 \cdots z_{N-1})$$

Для матриц с экспериментальными данными имеет место соотношение

$$\begin{aligned} \Phi_+ &= A_{real}\Phi + B_{real}U + W \\ Z &= C_{real}\Phi + D_{real}U, \end{aligned} \tag{2.2}$$

в которых A_{real}, B_{real}, C_{real} и D_{real} - реальные неизвестные матрицы уравнений системы.

Обозначим:

$$\Delta_{real} = \begin{pmatrix} A_{real} & B_{real} \\ C_{real} & D_{real} \end{pmatrix}, \quad \widehat{\Phi} = \begin{pmatrix} \Phi \\ U \end{pmatrix}, \quad \widetilde{\Phi} = \begin{pmatrix} \Phi_+ \\ Z \end{pmatrix}, \quad \widehat{W} = \begin{pmatrix} W \\ 0 \end{pmatrix}$$

и запишем уравнения (2.2) в виде линейной матричной регрессии

$$\widetilde{\Phi} = \Delta_{real}\widehat{\Phi} + \widehat{W} \tag{2.3}$$

Допустим, что возмущение в эксперименте удовлетворяет условию

$$\sum_{i=0}^{N-1} w(t_i)w^{\mathrm{T}}(t_i) = WW^{\mathrm{T}} \le \Omega \tag{2.4}$$

В частности, если возмущение при всех t удовлетворяет ограничению $||w(t)||_\infty \le d_w$ для некоторого заданного d_w, которое будем называть уровнем возмущения, то $\Omega = d_w^2 n_w N I_{n_x}$.

В случае, когда суммарная энергия возмущения во время эксперимента ограничена $\sum_{i=0}^{N-1}|w(t_i)|^2 \le \alpha^2$, то $\Omega = \alpha^2 I$. Если в (2.1) $w(t) = B_v v(t), v(t) \in \mathrm{R}^{n_v}$ для некоторой матрицы B_v и $||v(t)||_\infty \le d_v$, то $\Omega = d_v^2 n_v N B_v B_v^{\mathrm{T}}$.

Из (2.4) следует, что

$$\widehat{W}\widehat{W}^T \le \begin{pmatrix} \Omega & \star \\ 0 & 0 \end{pmatrix} = \widehat{\Omega} \tag{2.5}$$

Определим множество Δ_p матриц Δ порядка $(n_x + n_z) \times (n_x + n_u)$, которые могли бы генерировать полученные в эксперименте матрицы Φ, Φ_+ и Z при выбранных управлениях U и некоторых допустимых возмущениях W, удовлетворяющих ограничению (2.4). Для этих матриц равенство $\widetilde{\Phi} = \Delta\widehat{\Phi} + \widehat{W}$ должно выполняться при некоторой $\widehat{W}$, удовлетворяющей (2.5). Следовательно,

$$\Delta_p = \left\{\Delta\colon \widetilde{\Phi} = \Delta\widehat{\Phi} + \widehat{W}, \quad \widehat{W}\widehat{W}^{\mathrm{T}} \le \widehat{\Omega}\right\}$$

и $\Delta \in \Delta_p$ тогда и только тогда, когда

$$\left(\widetilde{\Phi} - \Delta\widehat{\Phi}\right)\left(\widetilde{\Phi} - \Delta\widehat{\Phi}\right)^{\mathrm{T}} \le \widehat{\Omega} \tag{2.6}$$

Очевидно, что $\Delta_{real} \in \Delta_p$.

Для дальнейшего применения представим последнее неравенство в виде

$$\begin{pmatrix} \Delta & I_{n_x+n_z} \end{pmatrix}\Psi_1\begin{pmatrix} \Delta & I_{n_x+n_z} \end{pmatrix}^{\mathrm{T}} \le 0, \tag{2.7}$$

где симметрическая матрица Ψ_1 порядка $2n_x + n_u + n_z$ разбивается на блоки соответствующего порядка и вычисляется следующим образом:

$$\Psi_1 = \left(\begin{array}{cc|cc} \Phi\Phi^{\mathrm{T}} & * & * & * \\ U\Phi^{\mathrm{T}} & UU^{\mathrm{T}} & * & * \\ \hline -\Phi_+\Phi^{\mathrm{T}} & -\Phi_+U^{\mathrm{T}} & \Phi_+\Phi_+^{\mathrm{T}} - \Omega & * \\ -Z\Phi^{\mathrm{T}} & -ZU^{\mathrm{T}} & Z\Phi_+^{\mathrm{T}} & ZZ^{\mathrm{T}} \end{array}\right). \tag{2.8}$$

Таким образом, множество матриц Δ, согласованных с полученными экспериментальными данными, удовлетворяет неравенству (2.7).

Стоит отметить, что при невырожденности информационной матрицы "размер"' множества Δ_p определяется матрицами регрессоров $\widehat{\Phi}$ и в конечном итоге зависит от реального объекта, выбираемых в эксперименте управлений U и возмущений W.

Далее, пусть имеется дополнительная информация о том, что неизвестная матрица Δ_{real} удовлетворяет ограничению

$$(\Delta - \Delta_*)(\Delta - \Delta_*)^{\mathrm{T}} \le \rho^2 I,\ \Delta_* = \begin{pmatrix} \mathrm{A}_* & \mathrm{B}_* \\ \mathrm{C}_* & \mathrm{D}_* \end{pmatrix} = \begin{pmatrix} \Delta_*^{(1)} \\ \Delta_*^{(2)} \end{pmatrix} \tag{2.9}$$

в котором Δ_* и ρ – заданные матрица и параметр, характеризующие центр и размер области неопределенности. Запишем это неравенство в виде

$$\begin{pmatrix} \Delta & I_{n_x+n_z} \end{pmatrix} \Psi_2 \begin{pmatrix} \Delta & I_{n_x+n_z} \end{pmatrix}^{\mathrm{T}} \le 0, \tag{2.10}$$

где

$$\Psi_2 = \left(\begin{array}{cc|cc} I_{n_x} & \star & \star & \star \\ 0_{n_u \times n_x} & I_{n_u} & \star & \star \\ \hline -A_* & -B_* & \Delta_*^{(1)}\Delta_*^{(1)\mathrm{T}} - \rho^2 I_{n_x} & \star \\ -C_* & -D_* & \Delta_*^{(2)}\Delta_*^{(1)\mathrm{T}} & \Delta_*^{(2)}\Delta_*^{(2)\mathrm{T}} - \rho^2 I_{n_z} \end{array} \right). \tag{2.11}$$

Обозначим: Δ_a – множество матриц, которые удовлетворяют неравенству (2.10), и $\Delta_{set} = \Delta_p \cap \Delta_a$ – множество матриц, которые удовлетворяют неравенствам (2.7) и (2.10).

Качество системы (2.1) при заданной матрице Δ, замкнутой управлением вида линейной стационарной обратной связи по состоянию $u(t) = \Theta x(t)$, будем оценивать уровнем гашения внешнего и начального возмущений, т.е. обобщенной H_∞-нормой

$$\gamma_{g\infty}(\Delta, \Theta) = \sup_{x_0,\, w} \frac{\|z\|}{\left(x_0^{\mathrm{T}} R^{-1} x_0 + \|w\|^2\right)^{1/2}},$$

где $R = R^{\mathrm{T}} > 0$ - весовая матрица, $||\xi||^2 = \sum_{t=0}^{\infty} |\xi(t)|^2$.

При отсутствии внешнего возмущения, т.е. при $w(t) \equiv 0$, обобщенная H_∞ - норма принимает вид так называемой γ_0-нормы

$$\gamma_0(\Delta, \Theta) = \sup_{x_0 \ne 0} \frac{\|z\|}{\left(x_0^{\mathrm{T}} R^{-1} x_0\right)^{1/2}},$$

которая характеризует "наихудшее" значение квадратичного функционала на траекториях системы, когда начальное состояние принадлежит эллипсоиду $x^{\mathrm{T}} R^{-1} x \le 1$.

Если начальное состояние нулевое, то обобщенная H_∞- норма (при $R \to 0$) превращается в обычную H_∞ -норму

$$\gamma_\infty(\Delta, \Theta) = \sup_{w \ne 0} \frac{\|z\|}{\|w\|}.$$

Качество неопределенной системы (2.1), замкнутой управлением вида $u(t) = \Theta x(t)$, будем оценивать минимальной верхней границей уровня гашения внешнего и начального возмущений, т.е. минимальной верхней границей обобщенной H_∞ - нормы для всех матриц объекта, согласованных с экспериментальными и априорными данными:

$$\gamma_*(\Theta) = \sup_{\Delta \in \mathbf{\Delta_{set}}} \gamma_{g\infty}(\Delta, \Theta). \tag{2.12}$$

Робастное обобщенное H_∞-оптимальное управление определяется как управление с матрицей параметров Θ_*, при которой достигается минимальное значение этой границы, которая является решением минимаксной задачи

$$\inf_{\Theta} \sup_{\Delta \in \mathbf{\Delta_{set}}} \gamma_{g\infty}(\Delta, \Theta) = \inf_{\Theta} \gamma_*(\Theta) = \gamma_*(\Theta_*). \tag{2.13}$$

Задача заключается в том, чтобы, не имея и не строя математическую модель системы, синтезировать робастное обобщенное H_∞-оптимальное управление с матрицей параметров Θ, при которой обобщенная H_∞-норма замкнутой системы для всех объектов, матрицы которых согласованы с полученными экспериментальными данными, ограничена заданной константой, т.е. $\gamma_*(\Theta) < \gamma$.

3. Обобщенная H_∞-норма в терминах двойственной системы

Обобщенная H_∞-норма от входа v к выходу z

$$\gamma_{g\infty} = \sup_{x_0,\, v} \frac{\|z\|}{\left(x_0^{\mathrm{T}} R^{-1} x_0 + \|v\|^2\right)^{1/2}} \tag{3.1}$$

устойчивой дискретной системы

$$\begin{aligned} x(t+1) &= Ax(t) + Bv(t), \\ z(t) &= Cx(t) \end{aligned} \tag{3.2}$$

удовлетворяет условию $\gamma_{g\infty} < \gamma$ тогда и только тогда, когда разрешимы относительно матрицы $Y = Y^{\mathrm{T}} > 0$ следующие линейные матричные неравенства:

$$\begin{pmatrix} -Y & \star & \star & \star \\ Y\mathcal{A}^{\mathrm{T}} & -Y & \star & \star \\ \mathcal{B}^{\mathrm{T}} & 0 & -\gamma^2 I & \star \\ 0 & \mathcal{C}Y & 0 & -I \end{pmatrix} < 0, \quad \begin{pmatrix} Y & \star \\ I & \gamma^2 R^{-1} \end{pmatrix} > 0. \tag{3.3}$$

Неравенство (3.3) означает, что для положительно определенной функции $V(x) = x^{\mathrm{T}} Y^{-1} x$ с $Y > \gamma^{-2} R$ в силу системы (3.2) выполняется неравенство

$$\Delta V(x) + |z|^2 - \gamma^2 |v|^2 < 0 \quad \forall\, x, v \tag{3.4}$$

Обобщенную H_∞-норму системы (3.2) можно характеризовать в терминах двойственной системы, поскольку она удовлетворяет условию $\gamma_{g\infty} < \gamma$ тогда и только тогда, когда существует положительно определенная квадратичная форма $V_a(x_a) = x_a^{\mathrm{T}} P x_a$ с $P > R$, для которой по траектории двойственной системы

$$\begin{aligned} x_a(t+1) &= A^{\mathrm{T}} x_a(t) + C^{\mathrm{T}} v_a(t), \\ z_a(t) &= B^{\mathrm{T}} x_a(t) \end{aligned} \tag{3.5}$$

выполняется неравенство

$$\Delta V_a\big(x_a(t)\big) + |z_a(t)|^2 - \gamma^2 |v_a(t)|^2 < 0. \tag{3.6}$$

4. Синтез обобщенных H_∞-оптимальных законов управления

Опишем основные шаги получения верхней границы обобщенной H_∞-нормы и соответствующих матриц параметров Θ законов управления для неопределенной замкнутой системы

$$\begin{aligned} x(t+1) &= (A + B\Theta)x(t) + w(t), \\ z(t) &= (C + D\Theta)x(t) \end{aligned} \tag{4.1}$$

Предполагая, что параметры Θ выбраны так, что замкнутая система устойчива, и учитывая введенные обозначения, представим эти уравнения в виде

$$\begin{aligned} x(t+1) &= (I_{n_x}\ 0_{n_x \times n_z})\, \Delta \begin{pmatrix} I_{n_x} \\ \Theta \end{pmatrix} x(t) + w(t), \\ z(t) &= (0_{n_z \times n_x}\ I_{n_z})\, \Delta \begin{pmatrix} I_{n_x} \\ \Theta \end{pmatrix} x(t), \end{aligned} \tag{4.2}$$

где Δ – неизвестная $(n_x + n_z) \times (n_x + n_u)$-матрица, Θ - $(n_u \times n_x)$ – матрица параметров регулятора.

Уравнения двойственной дискретной системы имеют вид

$$x_a(t+1) = \begin{pmatrix} I \\ \Theta \end{pmatrix}^{\mathrm{T}} \Delta^{\mathrm{T}} \begin{pmatrix} I \\ 0 \end{pmatrix} x_a(t) + \begin{pmatrix} I \\ \Theta \end{pmatrix}^{\mathrm{T}} \Delta^{\mathrm{T}} \begin{pmatrix} 0 \\ I \end{pmatrix} w_a(t), \qquad (4.3)$$
$$z_a(t) = x_a(t).$$

Рассмотрим систему, которую назовем расширенной, с дополнительными искусственными входом $w_\Delta(t) \in L_2(l_2)$ и выходом $z_\Delta(t)$, определяемую уравнениями

$$\begin{aligned} x(t+1) &= \begin{pmatrix} I \\ \Theta \end{pmatrix}^{\mathrm{T}} w_\Delta(t), \\ \widehat{z}(t) = \widehat{x}(t), \quad z_\Delta(t) &= \begin{pmatrix} I \\ 0 \end{pmatrix} \widehat{x}(t) + \begin{pmatrix} 0 \\ I \end{pmatrix} \widehat{w}(t), \end{aligned} \qquad (4.4)$$

в которых $\hat{x}(t)$ – состояние, $\widehat{w}(t)$ – возмущение, $\hat{z}(t)$ – целевой выход. Заметим, что при $w_\Delta(t) = \Delta^{\mathrm{T}} z_\Delta(t)$ уравнения (4.4) совпадают с уравнениями системы (4.3). Допустим, что дополнительные входной и выходной сигналы в системе (4.4) при всех $t \geq 0$ удовлетворяют двум неравенствам

$$\begin{pmatrix} w_\Delta(t) \\ z_\Delta(t) \end{pmatrix}^{\mathrm{T}} \Psi_1 \begin{pmatrix} w_\Delta(t) \\ z_\Delta(t) \end{pmatrix} \leq 0, \quad \begin{pmatrix} w_\Delta(t) \\ z_\Delta(t) \end{pmatrix}^{\mathrm{T}} \Psi_2 \begin{pmatrix} w_\Delta(t) \\ z_\Delta(t) \end{pmatrix} \leq 0, \qquad (4.5)$$

где матрицы Ψ_1 и Ψ_2 заданы в (2.8) и (2.11).

Множество всех таких сигналов $w_\Delta(t)$ обозначим через W_Δ. При $w_\Delta(t) = \Delta^{\mathrm{T}} z_\Delta(t)$ для всех $\Delta \in \Delta_{set}$, как следует из (2.7) и (2.10), выполняются неравенства

$$\begin{pmatrix} w_\Delta(t) \\ z_\Delta(t) \end{pmatrix}^{\mathrm{T}} \Psi_1 \begin{pmatrix} w_\Delta(t) \\ z_\Delta(t) \end{pmatrix} = z_\Delta^{\mathrm{T}}(t) \begin{pmatrix} \Delta^{\mathrm{T}} \\ I \end{pmatrix}^{\mathrm{T}} \Psi_1 \begin{pmatrix} \Delta^{\mathrm{T}} \\ I \end{pmatrix} z_\Delta(t) \leq 0,$$
$$\begin{pmatrix} w_\Delta(t) \\ z_\Delta(t) \end{pmatrix}^{\mathrm{T}} \Psi_2 \begin{pmatrix} w_\Delta(t) \\ z_\Delta(t) \end{pmatrix} = z_\Delta^{\mathrm{T}}(t) \begin{pmatrix} \Delta^{\mathrm{T}} \\ I \end{pmatrix}^{\mathrm{T}} \Psi_2 \begin{pmatrix} \Delta^{\mathrm{T}} \\ I \end{pmatrix} z_\Delta(t) \leq 0.$$

Таким образом, $w_\Delta(t) = \Delta^{\mathrm{T}} z_\Delta(t) \in W_\Delta$ и, следовательно, система (4.3) с $\Delta \in \Delta_{set}$, двойственная исходной неопределенной системе, "погружена" в расширенную систему (4.4), (4.5). С учетом леммы 3.1. [9] это позволяет получить верхнюю границу обобщенной H_∞-нормы неопределенной системы, используя соответствующее свойство расширенной системы.

Как показано в Теореме 4.1 [9] верхняя граница обобщенной H_∞-нормы неопределенной дискретной системы (2.1) при законе управления $u(t) = \Theta x(t)$, где $\Theta = QP^{-1}$, меньше γ, если следующие линейные матричные неравенства разрешимы относительно $P = P^{\mathrm{T}} > 0$, Q, $\mu_1 \geq 0$ и $\mu_2 \geq 0$

$$\begin{pmatrix} -P & \star & \star & \star & \star \\ 0 & -P + I - \sum_{i=1}^{2} \mu_i \Xi_{11}^{(i)} & \star & \star & \star \\ 0 & -\sum_{i=1}^{2} \mu_i \Xi_{21}^{(i)} & -\sum_{i=1}^{2} \mu_i \Xi_{22}^{(i)} - \gamma^2 I & \star & \star \\ P & -\sum_{i=1}^{2} \mu_i \Xi_{31}^{(i)} & -\sum_{i=1}^{2} \mu_i \Xi_{32}^{(i)} & -\sum_{i=1}^{2} \mu_i \Xi_{33}^{(i)} & \star \\ Q & -\sum_{i=1}^{2} \mu_i \Xi_{41}^{(i)} & -\sum_{i=1}^{2} \mu_i \Xi_{42}^{(i)} & -\sum_{i=1}^{2} \mu_i \Xi_{43}^{(i)} & -\sum_{i=1}^{2} \mu_i \Xi_{44}^{(i)} \end{pmatrix} < 0, \qquad (4.6)$$

где $P > R$,

$\Xi_{11}^{(1)} = \Phi_+ \Phi_+^{\mathrm{T}} - \Omega$, $\Xi_{21}^{(1)} = Z\Phi_+^{\mathrm{T}}$, $\Xi_{22}^{(1)} = ZZ^{\mathrm{T}}$,
$\Xi_{31}^{(1)} = -\Phi\Phi_+^{\mathrm{T}}$, $\Xi_{32}^{(1)} = -\Phi Z^{\mathrm{T}}$, $\Xi_{33}^{(1)} = \Phi\Phi^{\mathrm{T}}$,
$\Xi_{41}^{(1)} = -U\Phi_+^{\mathrm{T}}$, $\Xi_{42}^{(1)} = -UZ^{\mathrm{T}}$, $\Xi_{43}^{(1)} = U\Phi^{\mathrm{T}}$, $\Xi_{44}^{(1)} = UU^{\mathrm{T}}$,

$$\Xi_{11}^{(2)} = \Delta_*^{(1)}\Delta_*^{(1)\mathrm{T}} - \rho^2 I_{n_x}, \quad \Xi_{21}^{(2)} = \Delta_*^{(2)}\Delta_*^{(1)\mathrm{T}}, \quad \Xi_{22}^{(2)} = \Delta_*^{(2)}\Delta_*^{(2)\mathrm{T}} - \rho^2 I_{n_z},$$
$$\Xi_{31}^{(2)} = -A_*^{\mathrm{T}}, \quad \Xi_{32}^{(2)} = -C_*^{\mathrm{T}}, \quad \Xi_{33}^{(2)} = I_{n_x},$$
$$\Xi_{41}^{(2)} = -B_*^{\mathrm{T}}, \quad \Xi_{42}^{(2)} = -D_*^{\mathrm{T}}, \quad \Xi_{43}^{(2)} = 0_{n_u \times n_x}, \quad \Xi_{44}^{(2)} = I_{n_u}.$$

5. Иллюстративный пример

Для иллюстрации метода был проведен численный эксперимент с дискретной системой пятого порядка, $n_x = 5$ с двумя управлениями $n_u = 2$, возмущением при $n_w = 5$ и двумя целевыми выходами $n_z = 2$ с матрицами, каждый элемент которых был выбран случайно на интервале $[-1,1]$. Таким образом, система содержит 49 неизвестных параметров. В эксперименте были измерены состояния системы при различных входных воздействиях и уровнях возмущений. В эксперименте начальные условия и компоненты вектора управления выбирались случайно на интервале $[-1,1]$, а возмущение - также случайно на интервале $[-d, d]$. Было сделано $N = 50$ измерений. Весовая матрица начального возмущения имеет вид $R = 0{,}01I_5$.

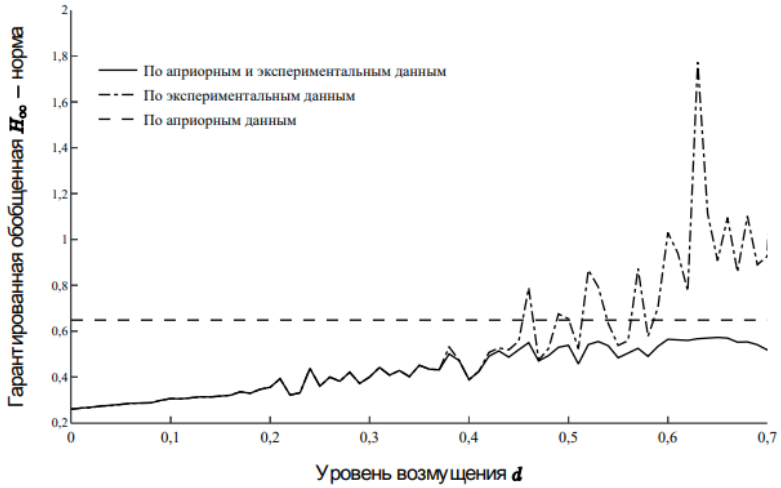


Рис. 1. Оценки обобщенной H_∞-нормы

На рис. 1 приведены три гарантированные оценки обобщенной H_∞-нормы, полученные только по априорной информации, только по экспериментальным данным и совместно по априорной информации и экспериментальным данным в зависимости от уровня возмущения d в эксперименте. Априорная информация состояла в том, что неизвестные матрицы системы удовлетворяют условию (2.9), в котором $\rho = 0{,}1$ и $A_* = A_{real} + (\rho/2)I$, $B_* = B_{real}$, $C_* = C_{real}$, $D_* = D_{real}$.

Полученные результаты говорят о том, что гарантированные оценки норм замкнутой неопределенной системы, синтезированной при использовании как априорных, так и экспериментальных данных, начиная с некоторого уровня возмущений в эксперименте значительно меньше соответствующих оценок норм замкнутой системы при управлениях, синтезированных только по априорным или только по экспериментальным данным.

Таким образом, результаты эксперимента показали, что использование как априорных, так и экспериментальных данных позволяет синтезировать управление, которое существенно снижает уровень возмущений по сравнению с подходами, основанными только на одном из этих источников данных.

6. Заключение

Представлен метод синтеза обобщенного H_∞-оптимального управления для дискретных систем в условиях неопределенности. Он позволяет реализовать робастное управление на основе априорных и экспериментальных данных без необходимости точного знания модели системы.

Данный подход с использованием линейных матричных неравенств может быть успешно применен в различных областях, где системы подвержены внешним возмущениям и имеют неопределенные параметры. Иллюстративный пример со случайно сгенерированным объектом пятого порядка демонстрирует, что при совместном применении априорных и экспериментальных данных качество системы управления значительно улучшается.

Литература

1. Willems J.C., Rapisarda P., Markovsky I., De Moor B. A note on persistency of excitation // Syst. Control Lett. 2005. V. 54. P. 325-329.
2. Petersen I. R. A stabilization algorithm for a class of uncertain linear systems // Syst. Control Lett. 1987. V. 8. P. 351-357.
3. De Persis C., Tesi P. Formulas for Data-Driven Control: Stabilization, Optimality and Robustness // IEEE Trans. Automat. Control. 2020. V. 65. No. 3. P. 909-924.
4. Waarde H. J., Eising J., Trentelman H.L., Camlibel M.K. Data Informativity: a New Perspective on Data-Driven Analysis and Control // IEEE Trans. Automat. Control. 2020. V. 65. No. 11. P. 4753-4768.
5. Waarde H.J., Camlibel M.K., Mesbahi M.} From Noisy Data to Feedback Controllers: Nonconservative Design via a Matrix S-Lemma // IEEE Trans. Automat. Control. 2022. V. 67. No. 1. P. 162-175.
6. Якубович В.А. S-процедура в нелинейной теории управления // Вестн. Ленинград. ун.-та. Математика. 1977. Т. 4. С. 73-93.
7. Bisoffi A., De Persis C., Tesi P. Data-driven Control via Petersen's Lemma // Automatica. 2022. V. 145. Article 110537.
8. Berberich J., C.W. Scherer C.W., Allgower F. Combining Prior Knowledge and Data for Robust Controller Design // IEEE Trans. Automat. Control. 2023. V. 68. No. 8. P. 4618-4633.
9. Коган М.М., Степанов А.В. Синтез субоптимальных робастных регуляторов на основе априорных и экспериментальных данных // АиТ. 2023. № 8. С. 24-42.
10. Коган М.М., Степанов А.В. Как улучшить робастное управление линейной нестационарной системой с помощью экспериментальных данных // АиТ. 2024. № 6. С. 115-139.

ВЫЧИСЛИТЕЛЬНЫЙ ЭКСПЕРИМЕНТ ПО АНАЛИЗУ ХАРАКТЕРИСТИК ХАОСА ДЛЯ КОНКРЕТНЫХ ДИНАМИЧЕСКИХ СИСТЕМ

С.Н. Стребуляев, Сингх Дхананджай, А.Д.Черепанов

Нижегородский государственный университет им. Н.И. Лобачевского

Проводится компьютерное исследование устойчивости электромеханической системы электропривода и расчет робастной устойчивости колебаний переднего колеса трехколесного шасси самолета. Рассматривается типовая структурная схема системы электропривода, построены границы областей устойчивости в плоскостях различных ее параметров. Рассматриваются также математические модели динамических систем, описывающих явление хаоса, приведенные в работах Лоренца и Дуффинга Приведены результаты компьютерного моделирования по схемам качественной теории динамических систем. Определяются состояния равновесия, строится глобальная структура фазового пространства и анализируются результаты многофакторного вычислительного эксперимента с целью определения областей возникновения хаоса.

Ключевые слова: математическая модель, адекватность, компьютерные технологии, хаос, электропривод.

1. Введение

В настоящее время при исследовании, в том числе и устойчивости, конкретных динамических систем получило широкое распространения методология математического моделирования и вычислительного эксперимента [1]. Эта методология охватывает многие сферы, от разработки комплексов технических систем и управления ими до анализа сложных технических, экономических, биологических и ряда других систем и процессов. Математическое моделирование развивается в недрах фундаментальных наук: механики и физики, которые отличают наивысший уровень теоретических исследований. При исследовании ряда новых процессов и систем необходимо построение адекватных математических моделей и проведение расчетов на ЭВМ, то есть проведение поискового вычислительного эксперимента. Вопрос об адекватности математической модели является очень важным, требующем высокой квалификации исследователей. Анализ адекватности математической модели реальному изучаемому объекту или процессу должен включаться в алгоритм и разрабатываемое программное обеспечение Динамика роста вычислительных возможностей современных компьютеров достаточно высока и позволяет надеяться на успешное решение все более сложных задач.

2. Робастная устойчивость системы электропривода

Из анализа структурных схем электроприводов была получена обобщенная схема исследуемой динамической системы. Ввиду сложности и громоздкости символьных преобразований, связанных с получением уравнений движения, характеристических уравнений и получением областей устойчивости было принято решение об использовании системы аналитических вычислений (САВ) Maple. Были получены аналитические выражения главной передаточной функции и характеристического уравнение вида:

$$Q_6(p) = q_0p^6 + q_1p^5 + q_2p^4 + q_3p^3 + q_4p^2 + q_5p^1 + q_6 = 0$$

где

$$q_0 = T_5b_2d_3T_2T_4;$$

$$q_1 = T_5b_2d_3T_4 + T_5b_1d_3T_2T_4 + T_5b_2d_2T_2T_4 + T_5b_2d_3T_2;$$

$$q_2 = C_e C_m T_3 b_2 d_3 T_2 + T_5 b_1 d_3 T_2 + T_5 b_1 d_2 T_2 T_4 + T_5 b_2 d_1 T_2 T_4 + T_5 b_2 d_3 + T_5 b_2 d_2 T_2 + T_5 b_1 d_3 T_4 + T_5 b_2 d_2 T_4;$$

$$q_3 = C_e C_m T_3 b_1 d_3 T_2 + T_5 b_1 d_2 T_4 + C_e C_m T_3 b_2 d_2 T_2 + K_1 T_1 T_3 T_5 c_2 b_2 + T_5 b_2 d_1 T_4 + T_5 b_1 d_2 T_2 + T_5 b_2 d_1 T_2 + T_5 b_2 d_2 + T_5 b_1 d_3 + C_e C_m T_3 b_2 d_3 + T_5 b_1 d_1 T_2 T_4;$$

$$q_4 = C_e C_m T_3 b_1 d_2 T_2 + C_e C_m T_3 b_1 d_3 + T_5 b_1 d_2 + C_e C_m T_3 b_2 d_2 + K_2 C_m T_1 T_3 a_2 c_2 + T_5 b_1 d_1 T_4 + T_5 b_1 d_1 T_2 +$$

$$C_e C_m T_3 b_2 d_1 T_2 + K_1 T_1 T_3 T_5 c_1 b_2 + K_1 T_1 T_3 T_5 c_2 b_1 + T_5 b_2 d_1;$$

$$q_5 = C_e C_m T_3 b_2 d_1 + K_2 C_m T_1 T_3 a_1 c_2 + T_5 b_1 d_1 + C_e C_m T_3 b_1 d_1 T_2 + K_1 T_1 T_3 T_5 c_1 b_1 + C_e C_m T_3 b_1 d_2 + K_2 C_m T_1 T_3 a_2 c_1;$$

$$q_6 = C_m K_2 T_1 T_3 a_1 c_1 + C_e C_m T_3 b_1\, d_1;$$

В дальнейшем, с помощью разработанного программного обеспечения, был проведен многофакторный вычислительный эксперимент по анализу границ областей устойчивости (с использованием режима анимации) в плоскостях различных параметров рассматриваемой динамической системы (рис. 1).

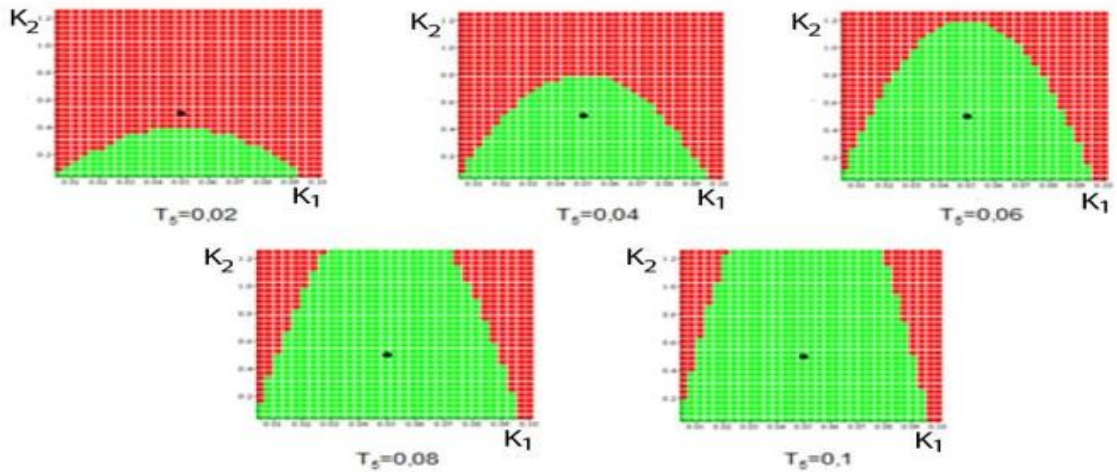


Рис. 1. Границы областей устойчивости в плоскости параметров (*K*1, *K*2) для разных значений *T*5

3. Колебания передней стойки шасси самолета (шимми)

Для расчета устойчивости переднего колеса трехколесного шасси самолета были рассмотрены результаты теоретических исследований ряда Российских ученых: Келдыша М.В., Гоздека В.С., Неймарка Ю.И.,Фуфаева Н.А., Журавлева В.Ф., Климова Д.М. и других. В настоящей работе используется математическая модель, приведенная в работе [2].

Подставляя выражения для потенциальной, кинетической энергий и функции рассеивания, согласно [2], в уравнение Лагранжа получаем следующие уравнения движения, с учетом неголономных связей

$$\begin{cases} A\ddot{\varphi} + [k + \rho N + lN(\sigma - 1)]\psi + D\ddot{\theta} + C\omega\dot{\theta} - cN\theta - (la + \sigma N)\xi = 0 \\ D\ddot{\psi} - C\omega\dot{\psi} - cN(\sigma - 1)\psi + B\ddot{\theta} + h\dot{\theta} - ca\xi - \beta\varphi = 0 \end{cases}$$

$$\begin{cases} \dot{\psi} + c\theta + \xi + V\theta + V\varphi = 0 \\ \dot{\theta} + \dot{\varphi} - \alpha V\xi + \beta V\varphi - \gamma V\psi = 0 \end{cases}$$

В дальнейшем в полученном характеристическом уравнении проводился переход от размерных параметров к безразмерным r_i и построение границ областей устойчивости в каждой паре плоскостей этих параметров. В каждой плоскости параметров определялся минимальный радиус окружности с центром в некотором номинальном значении параметров, таким образом, чтобы все точки внутри этой окружности лежалив области устойчивости. Наименьший из полученных радиусов и объявлялся радиусом робастной устойчивости (рис. 2).

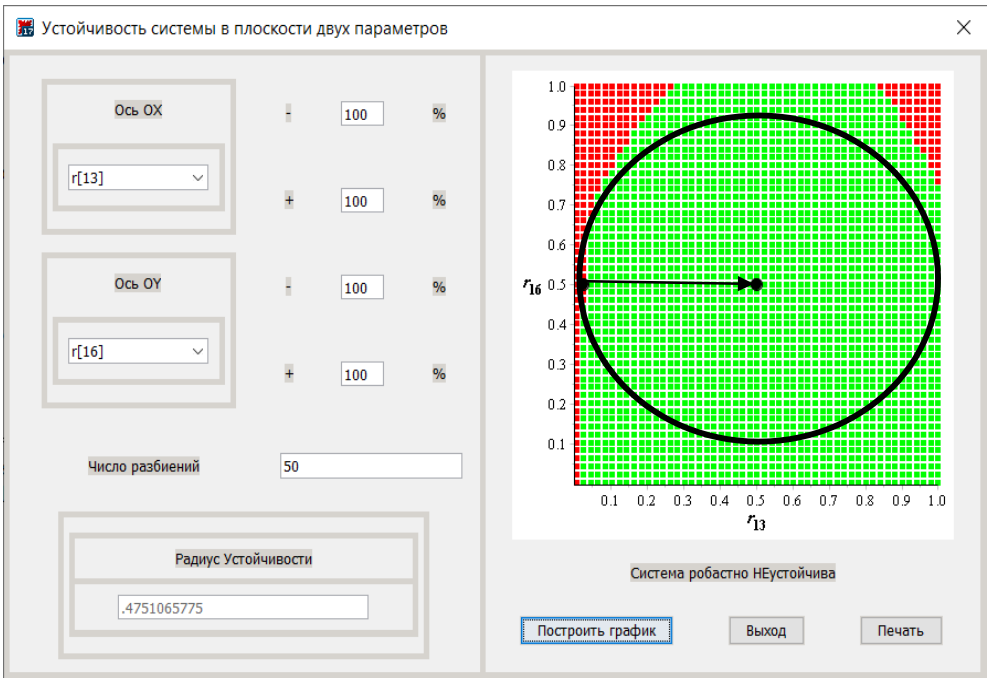


Рис. 2. Интерфейс разработанного программного обеспечения

4. Хаотические режимы в модели Лоренца

На следующем этапе рассматривается общеизвестная модель Лоренца [3], в виде системы трех нелинейных дифференциальных уравнений:

$$\begin{cases} \dot{x} = \sigma(y - x) \\ \dot{y} = rx - y - xz \\ \dot{z} = xy - bz \end{cases}$$

В дальнейшем определяются состояния равновесия в количестве трех (при r>1) и одном (при r<1). На базе этих зависимостей создано универсальное программное обеспечение. Основными этапами расчетов являются: вычисление состояний равновесия, построение интегральных кривых и фазового портрета относительно каждого из состояний равновесия и построение глобальной структуры фазового пространства (рис. 3). Также реализован алгоритм, позволяющий при изменении трех параметров системы, определять наличие или отсутствие хаоса (рис. 4). В процессе компьютерного моделирования используются возможности CAB Maple в области численного решения систем нелинейных дифференциальных с визуализацией результатов расчетов и виде графиков в режиме анимации.

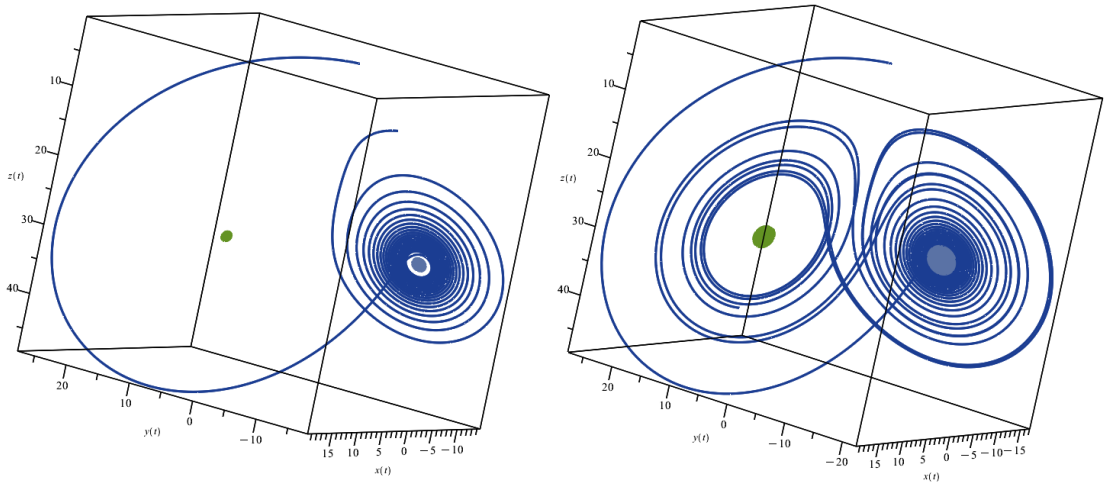


Рис. 3. Глобальная структура фазового пространства

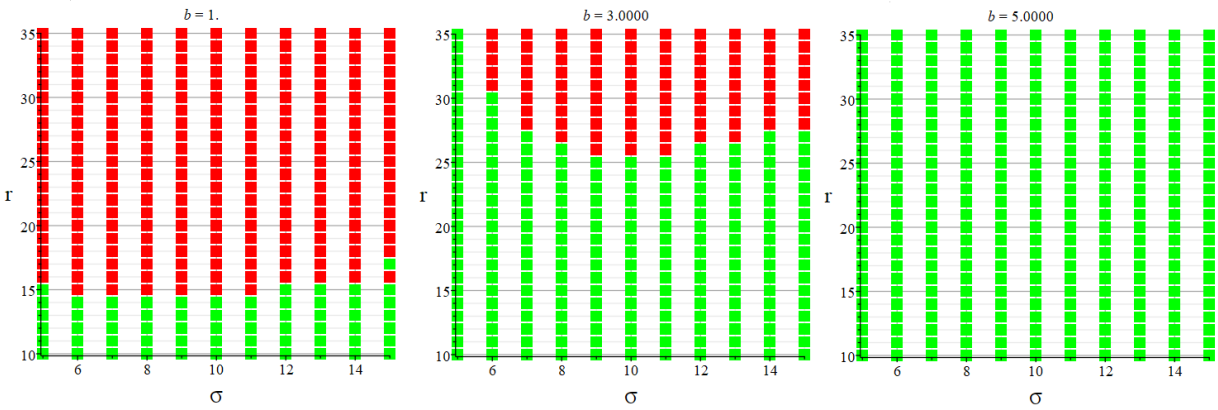


Рис. 4. Области наличия и отсутствия хаоса при изменении параметров r, σ, b

5. Хаотические режимы в модели Дуффинга

Нелинейный осциллятор, описываемый уравнением Дуффинга и находящийся под действием гармонической внешней силы, содержащий, в общем случае, постоянную составляющую:

$$\ddot{x} + \alpha\dot{x} + kx + \gamma x^3 = B_0 + Bcos\omega t.$$

Для уравнения рассматриваются различные виды упругих характеристик: мягкая $(k > 0, \gamma < 0)$, жесткая $(k > 0, \gamma > 0)$, с нулевой $(k = 0, \gamma > 0)$ и отрицательной $(k < 0, \gamma > 0)$ линейной жесткостью.

Как при мягкой, так и при жесткой упругой характеристике хаотические колебания существуют при достаточно большой амплитуде внешней силы в интервале частот ω, где соответствующая амплитудно-частотная характеристика неоднозначна.

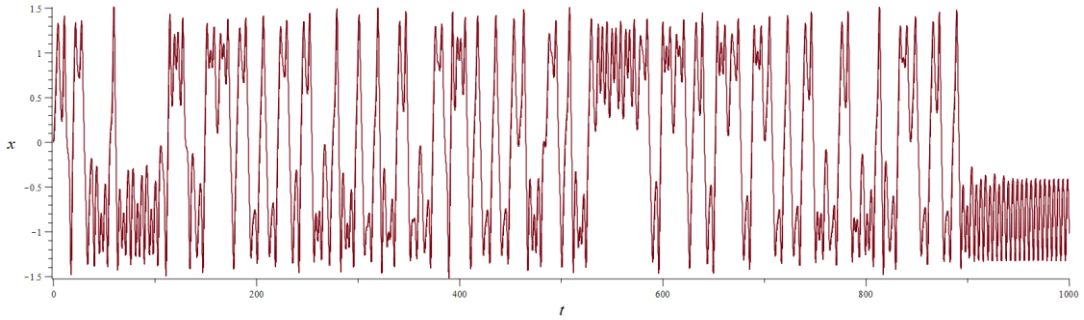


Рис. 5. График $x(t)$ при параметрах: $\alpha = 0.1; k = -1;\ \gamma = 1;\ B_0 = 0; B = 0.3;\ \omega = 1.4$

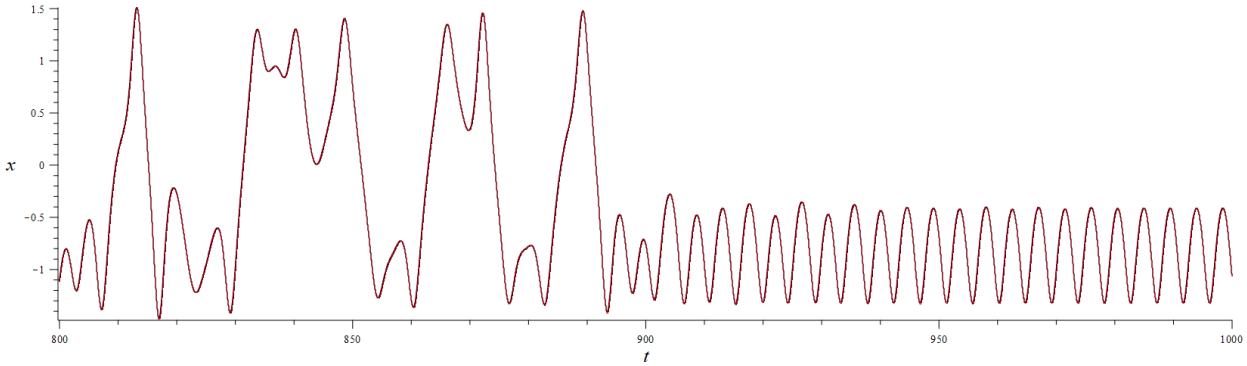


Рис. 6. График $x(t)$ при параметрах: $\alpha = 0.1; k = -1;\ \gamma = 1;\ B_0 = 0; B = 0.3;\ \omega = 1.4$

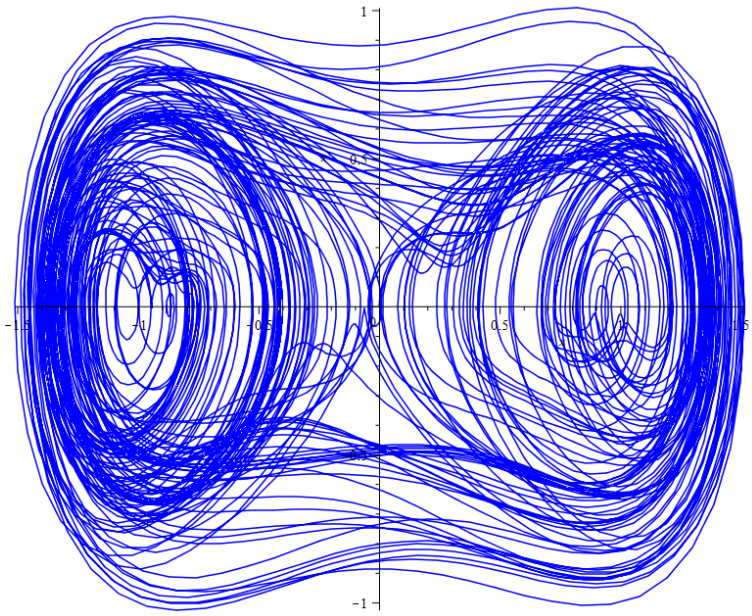


Рис. 7. Фазовый портрет при параметрах: $\alpha = 0.1; k = -1;\ \gamma = 1;\ B_0 = 0; B = 0.3;\ \omega = 1.4$

6. Заключение

В рассмотренной работе приведены результаты математического моделирования и расчета характеристик устойчивости колебаний в технических системах- системе электропривода и шимми переднего колеса трехколесного шасси самолета. Рассмотрение указанных систем проводилось в полной мере с учетом всехих характеристик без каких-либо упрощений. Использование CAB Maple для достижения результатов исследования показало эффективность этого программного обеспечения. С использованием теоретического материала и адекватных математических моделей, приведенных в работах известных Российских ученых, в сочетании с появившимися возможностями вычислительной техники и программного обеспечения удалось получить новые научные результаты

Литература

1. Стребуляев С.Н., Сироткина Д.А. // Учебно-исследовательские проекты по анализу динамического качества технических систем и процессов с использованием их математических моделей, Современные информационные технологии и ИТ-образование. 2021. Т.17. №2. С.392-403.
2. Неймарк Ю.И. // Динамика неголономных систем / Ю.И. Неймарк, Н.А. Фуфаев. – Москва, Наука, 1967.
3. Неймарк Ю.И., Ланда П.С. Стохастические и хаотические колебания / Неймарк Ю. И., Ланда П.С. – Москва, Изд. Наука, Гл. ред. физ.-мат. лит., 1987.
4. С.П Кузнецов // Динамический хаос – Москва, 2001.

РОСТ МИНОРОВ РАНГОВОГО ПОРЯДКА МАТРИЦЫ ОГРАНИЧЕНИЙ ТРЕХИНДЕКСНОЙ ПЛАНАРНОЙ ТРАНСПОРТНОЙ ЗАДАЧИ

Е.Б. Титова

Нижегородский государственный университет им. Н.И. Лобачевского

Для матрицы T, где T – матрица ограничений трехиндексной планарной транспортной задачи, проведено сравнение роста среднего значения суммы квадратов миноров рангового порядка базисной системы строк матрицы T и базисной системы столбцов матрицы T.

Ключевые слова: трехиндексная планарная транспортная задача, асимптотика среднего значения квадрата минора, миноры рангового порядка.

1. Введение

Пусть $\mathbb{Q}-$ поле рациональных чисел, $\mathbb{Q}^n-$ линейное пространство столбцов с компонентами из $\mathbb{Q}$, а $\mathbb{Z}^n-$ столбцы из $\mathbb{Q}^n$ с целочисленными компонентами. Для $(m\times n)$-матрицы A с элементами из $\mathbb{Q}$ обозначим через $L_A=\{u\in\mathbb{Q}^m|uA=0\}$ левое нуль-пространство матрицы A, а через $L-$ какой-нибудь его базис. Можно считать, что $L-$ целочисленная $(m-r\ \times m)$-матрица, где $r-$ ранг матрицы A. Множество $L_A^Z=L_A\cap\mathbb{Z}^m$ назовем левым модулем матрицы A. Множество целочисленных линейных комбинаций строк матрицы L содержится в L_A^Z, но, вообще говоря, с ним не совпадает. Это выполняется тогда и только тогда, когда матрица L унимодулярна (т.е. НОД базисных миноров равен 1). В этом случае L будем называть базисом левого модуля матрицы A. Аналогично определим правое нуль-пространство $R_A=\{x\in\mathbb{Q}^n|Ax=0\}$ и правый модуль $R_A^Z=R_A\cap\mathbb{Z}^n$ матрицы A.

Известно, что если матрица A содержит базисную систему столбцов $A^{\mathcal{B}}$ $(|\mathcal{B}|=r)$, содержащую базисный минор, равный 1, то (см., например, [1]) $|\det A^{\mathcal{B}}(\mathcal{J})|=|\det L_A^Z(\bar{\mathcal{J}})|$, где $\bar{\mathcal{J}}-$ множество, дополнительное к $\mathcal{J}$. Аналогично, для базисной системы строк $A^{\mathcal{A}}$ $(|\mathcal{A}|=r)$: $|\det A^{\mathcal{A}}(\mathcal{J})|=|\det R_A^Z(\bar{\mathcal{J}})|$. Таким образом, количественный анализ базисных миноров матриц $A^{\mathcal{B}}$ и L_A^Z, $A^{\mathcal{A}}$ и R_A^Z равносилен.

В данной работе для матрицы ограничений трехиндексной планарной транспортной задачи T проведено сравнение роста среднего значения суммы квадратов миноров рангового порядка базисной системы её строк матрицы $\mathrm{T}^{\mathcal{A}}$ и базисной системы столбцов матрицы $\mathrm{T}^{\mathcal{B}}$.

2. Постановка задачи

Рассмотрим ограничения-равенства планарной трехиндексной транспортной задачи (см., например, [2]) с неизвестными $x_{j_1j_2j_3}\geq 0$ ($j_i=1,\dots,n_i$ при $i=1,2,3$):

$$\sum_{j_1=0}^{n_1}x_{j_1j_2j_3}=b_{0j_2j_3},$$

$$\sum_{j_2=0}^{n_2}x_{j_1j_2j_3}=b_{j_10j_3},$$

$$\sum_{j_3=0}^{n_3}x_{j_1j_2j_3}=b_{j_1j_20},$$

и обозначим через $T(n_1,n_2,n_3)$ ее матрицу.

Из [3] следует, что матрицу $T(n_1,n_2,n_3)$ можно представить в виде строчечных блоков:

$$T(n_1,n_2,n_3)=\begin{pmatrix} I_{n_1}\times E_{n_2}\times E_{n_3} \\ E_{n_1}\times I_{n_2}\times E_{n_3} \\ E_{n_1}\times E_{n_2}\times I_{n_3} \end{pmatrix},$$

где $A\times B-$ кронекерово произведение матриц A и B (определение и свойства см., например, в [4]), E_n- единичная матрица n-го порядка, $I_n=(1,\dots,1)-n$-мерная строка, все компоненты которой равны 1.

Матрица $T(n_1,n_2,n_3)$ имеет число столбцов $N_T=n_1n_2n_3$, число строк $M_T=n_1n_2+n_1n_3+n_2n_3$ и ранг $r_T=n_1n_2+n_1n_3+n_2n_3-n_1-n_2-n_3+1$. Матрица $T(n_1,n_2,n_3)$ содержит базисную систему строк $T^{\mathcal{A}}$ ($|\mathcal{A}|=r$), имеющую минор порядка r, равный 1. Аналогично можно выделить базисную систему столбцов $T^{\mathcal{B}}$ ($|\mathcal{B}|=r$), имеющую минор порядка r, равный 1.

3. Основные результаты

Рассмотрим левое нуль-пространства матрицы $T(n_1,n_2,n_3)$.

В [1] дано описание построения левого нуль-пространства. Следуя ему, матрицу $L(n_1,n_2,n_3)$ можно представить в блочном виде:

$$L(n_1,n_2,n_3)=\begin{pmatrix} I_1\times I_{n_2}\times E_{n_3} & -I_{n_1}\times I_1\times E_{n_3} & 0_{n_3\times n_1n_2} \\ I_1\times E_{n_2}\times I_{n_3} & 0_{n_2\times n_1n_3} & -I_{n_1}\times E_{n_2}\times I_1 \\ 0_{n_1\times n_2n_3} & E_{n_1}\times I_1\times I_{n_3} & -E_{n_1}\times I_{n_2}\times I_1 \end{pmatrix},$$

где $0_{m\times n}-$ нулевая $(m\times n)$-матрица.

Непосредственно перемножая матрицы, получаем, что $L(n_1,n_2,n_3)\cdot T(n_1,n_2,n_3)=0$.

Матрица $L(n_1,n_2,n_3)$ имеет число строк $M_L=n_1+n_2+n_3$, число столбцов $N_L=n_1n_2+n_1n_3+n_2n_3$ и имеет размерность $d_L(n_1,n_2,n_3)=M_T-r_T=n_1+n_2+n_3-1$. Она представляет собой матрицу инцидентности полного ориентированного графа, следовательно, вполне унимодулярна.

Пусть $Q_n^*=\begin{pmatrix} 1 & \dots & 1 \\ -1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & -1 \end{pmatrix}=\begin{pmatrix} I_{n-1} \\ -E_{n-1} \end{pmatrix}-$ матрица $n\times(n-1)$ порядка.

Теорема 1. *Базис собственных векторов матрицы* $L(n_1,n_2,n_3)\cdot L^T(n_1,n_2,n_3)$ *составляют следующие столбцовые блоки матрицы* V*:*

$$v_1=\left(-I_{n_3\times 1}\,\middle|\, I_{n_2\times 1}\,\middle|-I_{n_1\times 1}\right)^T,$$
$$v_2=\left(Q^*_{n_3}\middle|0_{n_2\times n_3-1}\middle|\, 0_{n_1\times n_3-1}\right)^T,$$
$$v_3=\left(0_{n_3\times n_2-1}\middle|Q^*_{n_2}\,\middle|\, 0_{n_3\times n_1-1}\right)^T,$$
$$v_4=\left(0_{n_3\times n_1-1}\middle|\, 0_{n_2\times n_1-1}\,\middle|Q^*_{n_1}\right)^T,$$
$$v_5=\left\{\left(n_1I_{n_3\times 1}\,\middle|\, 0_{n_2\times 1}\,\middle|-n_3I_{n_1\times 1}\right)^T,\left(0_{n_3\times 1}\middle|\, n_1I_{n_2\times 1}\,\middle|\, n_2I_{n_1\times 1}\right)^T\right\}.$$

Им соответствуют собственные числа $\lambda_1=0$, $\lambda_2=n_1+n_2$ *кратности* (n_3-1), $\lambda_3=n_1+n_3$ *кратности* (n_2-1), $\lambda_4=n_2+n_3$ *кратности* (n_1-1) *и* $\lambda_5=n_1+n_2+n_3$ *кратности* 2.

Д о к а з а т е л ь с т в о. Выпишем матрицу $L(n_1,n_2,n_3)\cdot L^T(n_1,n_2,n_3)$:

$$L(n_1,n_2,n_3)\cdot L^T(n_1,n_2,n_3)=\begin{pmatrix} (n_1+n_2)E_{n_3} & I_{n_3\times n_2} & -I_{n_3\times n_1} \\ I_{n_2\times n_3} & (n_1+n_3)E_{n_2} & I_{n_2\times n_1} \\ -I_{n_1\times n_3} & I_{n_1\times n_2} & (n_2+n_3)E_{n_1} \end{pmatrix},$$

где $I_{n\times m}-(n\times m)$-матрица, все компоненты которой равны 1.

Непосредственной проверкой получим, что $\left(L(n_1,n_2,n_3)\cdot L^T(n_1,n_2,n_3)\right)V=V\Lambda$, где $\Lambda=diag(\lambda_1,\dots,\lambda_5)$.

Выделим в матрице $T(n_1,n_2,n_3)$ базисную систему столбцов $T^{\mathcal{B}}$ ($|\mathcal{B}|=r$), а в матрице $L(n_1,n_2,n_3)$ базисную систему строк $L_{\overline{\mathcal{B}}}$.

Для построения базиса левого модуля L_T^Z достаточно удалить любую из строк матрицы $L(n_1,n_2,n_3)$, тогда получим матрицу $L_{\overline{\mathcal{B}}}$.

Следствие 1. *Определитель Грама матрицы* $L_{\overline{\mathcal{B}}}$ *равен*

$$det(L_{\bar{B}} \cdot (L_{\bar{B}})^T) = (n_1+n_2)^{n_3-1} \cdot (n_1+n_3)^{n_2-1} \cdot (n_2+n_3)^{n_1-1} \cdot (n_1+n_2+n_3)^2. \quad (1)$$

Обозначим среднее значение квадрата минора рангового порядка матрицы $T^{\mathcal{B}}$ через

$$\delta_{\mathcal{B}}(n) = \det\left(\left(T^{\mathcal{B}}\right)^T \cdot T^{\mathcal{B}}\right) / \binom{M}{r}. \quad (2)$$

Очевидно, что $L_{\bar{B}} \cdot T = 0 \Leftrightarrow L_{\bar{B}} \cdot T^{\mathcal{B}} = 0$. Известно [1], что

$$|\det\left(\left(T^{\mathcal{B}}\right)^T \cdot T^{\mathcal{B}}\right)| = |\det(L_{\bar{B}} \cdot (L_{\bar{B}})^T)|.$$

Поэтому для исследования поведения $\delta_{\mathcal{B}}$ используем значение определителя Грама матрицы $L_{\bar{B}}$.

Теорема 2. *При $n \to \infty$ среднее значение квадрата минора рангового порядка базисной системы столбцов $T^{\mathcal{B}}$ матрицы ограничений трехиндексной планарной транспортной задачи*

$$\delta_{\mathcal{B}}(n) \sim \sqrt{2\pi(3n-1)}\left(\frac{2}{e}\right)^{3n-1}. \quad (3)$$

Доказательство. Величину $\binom{M}{r} = \binom{M}{M-r} = \frac{3n^2 \cdot (3n^2-1) \cdot \ldots \cdot (3n^2-3n+2)}{(3n-1)!}$ оценим следующим образом:

$$\frac{(3n^2-3n+2)^{3n-1}}{(3n-1)!} < \binom{M}{M-r} < \frac{(3n^2)^{3n-1}}{(3n-1)!}.$$

Оценив по формуле Стирлинга $(3n-1)! \sim \sqrt{2\pi(3n-1)}\left(\frac{3n-1}{e}\right)^{3n-1}$, получаем при $n \to \infty$

$$\binom{M}{M-r} \sim \frac{1}{\sqrt{2\pi(3n-1)}} \cdot \left(\frac{3n^2}{3n-1}\, e\right)^{3n-1} \sim \frac{1}{\sqrt{2\pi(3n-1)}} \cdot (n \cdot e)^{3n-1}.$$

При $n_1 = n_2 = n_3 = n$ из формулы (1) следует

$$\det(L_{\bar{B}} \cdot (L_{\bar{B}})^T) = 9 \cdot 2^{3(n-1)} \cdot n^{3n-1}.$$

Подставляя в (2) оценку величины $\binom{M}{r}$ и значение определителя, получаем утверждение теоремы.

Следствие 2. *Среднее значение квадрата минора базисной системы столбцов матрицы ограничений трехиндексной планарной транспортной задачи при $n \to \infty$ стремится к 0 с экспоненциальной скоростью.*

Рассмотрим правое нуль-пространство матрицы $T(n_1, n_2, n_3)$. Оно имеет размерность $d_R(n_1, n_2, n_3) = N_T - r_T = n_1n_2n_3 - n_1n_2 - n_1n_3 - n_2n_3 + n_1 + n_2 + n_3 - 1$.

Выделим в матрице $T(n_1, n_2, n_3)$ базисную систему строк $T^{\mathcal{A}}$ ($|\mathcal{A}| = r$), а в матрице $R(n_1, n_2, n_3)$ базисную систему столбцов $R_{\bar{\mathcal{A}}}$. В [5] автором был получен определитель Грама матрицы $R_{s,k}$ – правого модуля k-индексной транспортной задачи с s суммированиями $(1 \le s \le k)$ и получена асимптотика среднего значения $\alpha_{\mathcal{A}}(n)$ квадрата минора рангового порядка базисной системы строк матрицы ограничений k-индексной транспортной задачи с s суммированиями.

Следствие 3 [5]. *Среднее значение $\alpha_{\mathcal{A}}(n)$ квадрата минора рангового порядка базисной системы строк $T^{\mathcal{A}}$ матрицы ограничений трехиндексной планарной транспортной задачи*

$$\alpha_{\mathcal{A}}(n) \sim (n-1)\left(\frac{3}{e}\right)^{3(n-1)^2}.$$

Следствие 4. *Среднее значение квадрата минора базисной системы строк матрицы ограничений трехиндексной планарной транспортной задачи при $n \to \infty$ стремится к бесконечности с экспоненциальной скоростью.*

Полученные асимптотики роста среднего значения квадратов миноров базисной системы строк и базисной системы столбцов показывают, что при решении трехиндексной планарной транспортной задачи выбор базисного минора в базисной системе столбцов может привести к получению целочисленного решения, что в общем случае неверно, т.к. матрица ограничений трехиндексной планарной транспортной задачи неунимодулярна при $n_1 \ge n_2 \ge n_3 \ge 3$.

Литература

1. Шевченко В.Н. Многогранники многоиндексных транспортных задач: алгебраический подход. // Материалы конференции «Дискретный анализ и исследование операций». Новосибирск: Изд-во ин-та математики, 2004, С. 64-70.
2. Шевченко В.Н. Характеристические многочлены многоиндексных транспортных задач. //Дискретная математика Т.15. Вып. 2. 2003. С. 83-88.
3. Воеводин В.В., Кузнецов Ю.А. СМБ. Матрицы и вычисления. М.: Наука, 1984.
4. Мишина А.П., Проскуряков И.В. Высшая алгебра. Линейная алгебра, многочлены, общая алгебра. М.: Наука, 1965.
5. Титова Е.Б., Шевченко В.Н. О минорах матрицы ограничений многоиндексных транспортных задач. // Дискретная математика. 2012. Т. 24. Вып. 4. С. 147-157.

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ ДИНАМИКИ СФЕРИЧЕСКОГО РОБОТА С ЭЛЕКТРОМАГНИТНЫМ ПРИВОДОМ[1*]

***А.М. Тузиков*[1, 2]**

[1]Нижегородский государственный университет им. Н.И. Лобачевского,
[2]Научно-технологический университет «Сириус»

Рассматривается задача движения по прямой с заданной скоростью сферического робота с электромагнитным приводом, состоящим из электромагнитов и постоянных магнитов. Построена математическая модель динамики сферического робота с электромагнитным приводом, включающая в себя описание процесса взаимодействия электромагнитов и постоянных магнитов. Исследован вопрос об оптимальной, с точки зрения энергетических затрат, конфигурации магнитов. Получены оценки для величин токов в электромагнитах, обеспечивающих необходимые силовые моменты для реализации движения робота по прямой с заданной скоростью. Приводятся результаты численных экспериментов.

Ключевые слова: сферический робот, электромагнетизм, математическая модель.

1. Введение

В течение последних десятилетий активно разрабатывались мобильные роботы, в том числе сферические, которые могут быть использованы в различных областях человеческой деятельности. Сферический мобильный робот имеет форму шара, в который встроены все необходимые механизмы, устройства управления и источники энергии. Благодаря этому он обладает рядом преимуществ перед роботами на колёсах и ножках. Во-первых, сферическая форма не позволяет роботу перевернуться. Во-вторых, он может быстро маневрировать и передвигаться благодаря всенаправленному движению. В-третьих, он легко преодолевает препятствия. В-четвёртых, он защищён от воздействия внешней среды, такой как жидкость, пыль и газ. В-пятых, он энергоэффективен, так как имеет только одну точку контакта с поверхностью. Самое важное преимущество сферического мобильного робота – его способность быстро восстанавливать равновесие после столкновения.

Система привода сферического робота может быть реализована различными способами, например, с использованием вращающихся маховиков или набором движущихся масс, расположенных на ортогонально связанных осях, а также при помощи маятника, соединенным со сферической оболочкой шарниром и приводимый в движение системой двигателей [1–6].

В данной работе рассматривается сферический робот, состоящий из двух сферических оболочек: внешней с радиусом R и внутренней с радиусом r (R>r). На внутренней сфере расположены 12 электромагнитов, а на внешней – 18 постоянных неодимовых магнитов. Движение робота осуществляется за счет взаимодействия магнитов на разных оболочках. Основной целью работы является получение оценки величин токов в электромагнитах необходимых для реализации заданного движения. Для этого была построена математическая модель квазистационарного электромагнитного взаимодействия размещенных на разных движущихся оболочках магнитов.

2. Математическая модель электромагнитного привода

Схематичное изображение рассматриваемого робота с одним постоянным и одним электромагнитом приведено на рис. 1. Введем систему координат, *Oxyz* центр которой совпадает с геометрическим центром сферического робота, ось *Oz* направлена вертикально вверх, ось *Ox*

[1*] Работа выполнена при финансовой поддержке РНФ (проект № 24-11-20023).

направлена горизонтально слева направо. В данной работе рассматривается движение по прямой вдоль оси *Ox*, поэтому расположим все магниты в плоскости *Oxz*, *y=0*.

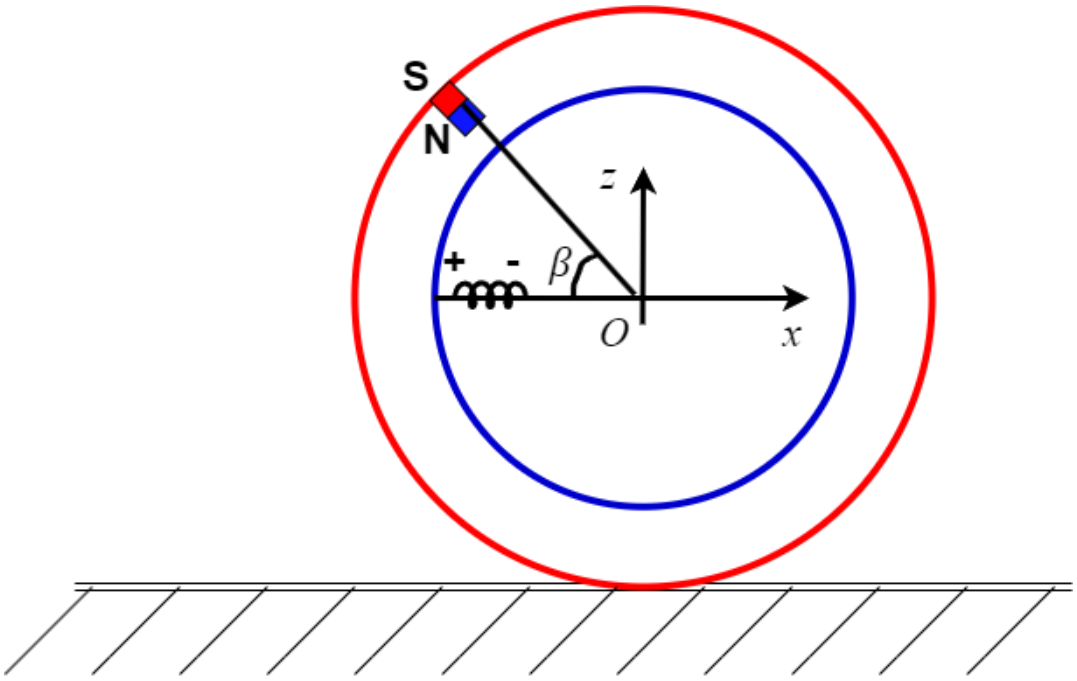


Рис. 1. Схематичное изображение сферического робота

Для формирования математической модели электромагнитного привода в целом рассмотрим сначала взаимодействие двух магнитов, один из которых является электромагнитом, а другой постоянным магнитом. Пусть электромагнит представляет собой соленоид, а в качестве постоянного магнита возьмем неодимовый магнит цилиндрической формы. В первую очередь, используя закон Био-Савара-Лапласа, записывается уравнение для поля конечномерного соленоида в точке расположения постоянного магнита. В свою очередь, сила с которой действует поле соленоида $\boldsymbol{B}$ на постоянный магнит можно найти с помощью закона Ампера. Определив силу, не составляет труда найти вращающий момент, приложенный к внешней оболочке, который создается вследствие взаимодействия постоянного и электромагнита.

Результатом моделирования взаимодействия двух магнитов является зависимость вращающего момента Q, приложенного к внешней сфере, от угла β между постоянным и электромагнитом, при этом так как движение происходит по прямой вдоль *Ox*, то вращающий момент имеет только компоненту *y*. Для следующих параметров системы [7]: радиус внешней сферы $R = 0.066$ м, радиус внешней сферы $r = 0.065$ м, длина соленоида $h_1 = 0.025$ м, внутренний радиус соленоида $a_1 = 0.004$ м, внешний радиус соленоида $a_2 = 0.014$ м, количество витков в соленоиде $N = 1200$, ток в соленоиде $I = 2$ А, длина постоянного магнита $h_2 = 0.024$ м, радиус постоянного магнита $a_3 = 0.011$ м была получена зависимость *y*-составляющей вращающего момента Q от угла β между постоянным магнитом и электромагнитом (рис. 2). Из представленной зависимости видно, что при $\beta = \pm 10°$ значение вращающего момента достигает своего максимального значения. На основании этого факта можно сделать вывод, что для достижения максимальной эффективности постоянные магниты нужно располагать эквидистантно с шагом 20°, кроме того, полюса соседних постоянных магнитов, направленные внутрь сферического робота, должна быть разной полярности. При этом шаг расположения электромагнитов нужно выбирать так, чтобы избежать “мертвых” зон, когда все постоянные магниты и электромагниты находятся друг напротив друга, в данной работе расположение электромагнитов эквидистантно с шагом 30°.

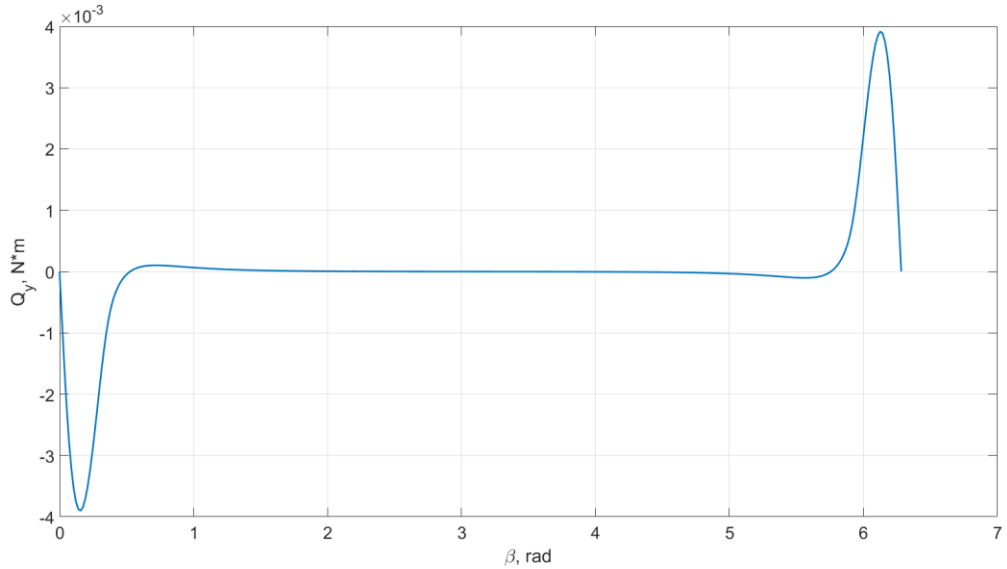


Рис. 2. Зависимость y-составляющей вращающего момента Q от угла β между постоянным и электромагнитом

3. Оценка величин тока

Рассмотрим движение робота вдоль оси Ox с растущей скоростью от $v = 0$ м/с до $v = 3$ км/ч $= 0.83$ м/с. Для расчета необходимого для этого вращающего момента воспользуемся подходом, описанным в [8]. В итоге получаем зависимость y-составляющей вращающего момента от времени (рис. 3).

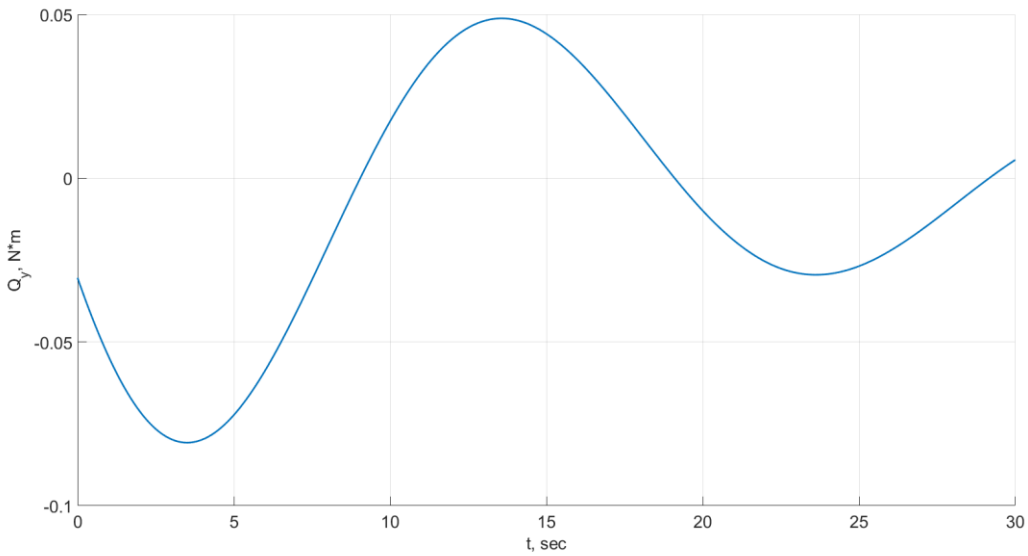


Рис. 3. Зависимость y-составляющей вращающего момента Q от времени

Определив вращающий момент, мы можем получить оценку токов необходимых для реализации заданного движения используя метод наименьших квадратов. Зависимости токов на двух соседних соленоидах от времени приведены на рис. 4. Анализируя эти зависимости, можно сделать вывод, что для реализации заданного вращающего момента (рис. 3) максимально необходимый ток равен примерно 4 А, что является вполне реализуемым, кроме этого огибающая этих токовых сигналов качественно похожа на зависимость y-составляющей вращающего момента от

времени. Также стоит отметить, что формы сигналов для всех четных и нечетных электромагнитов одинаковы с точностью до знака, поэтому достаточно найти токовые зависимости для двух соседних электромагнитов.

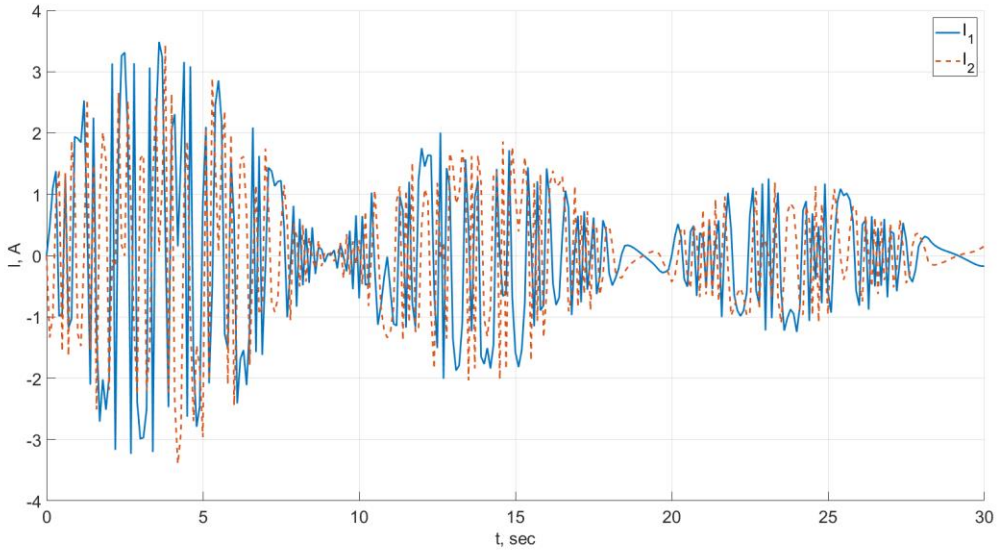


Рис. 4. Зависимость токов на двух соседних соленоидах от времени

4. Заключение

Построена математическая модель сферического робота с электромагнитным приводом, включающая в себя описание взаимодействия между постоянными магнитами и электромагнитами. Представлено одно из решений вопроса об оптимальном расположении магнитов на разных оболочках. Получена оценка токовых сигналов, необходимых для реализации заданного движения сферического робота. Из этих оценок можно сделать вывод, что для осуществления заданного вращающего момента необходимые токи имеют вполне реализуемые значения, кроме этого, огибающая этих токовых сигналов качественно похожа на зависимость *y*-составляющей вращающего момента от времени. Также стоит отметить, что формы сигналов для всех четных и нечетных электромагнитов одинаковы с точностью до знака, поэтому достаточно найти токовые зависимости для двух соседних электромагнитов.

Литература

1. Balandin D.V., Komarov M.A., Osipov G.V. A motion control for a spherical robot with pendulum drive // J. Comput. Syst. Sci. Int. 2013. Vol. 52, No. 4. P. 650–663. DOI: https://doi.org/10.1134/S1064230713040047.
2. Hogan F.R., Forbes J.R., Barfoot T.D. Rolling Stability of a Power-Generating Tumbleweed Rover // J. Spacecr. Rocket. 2014. Vol. 51, No. 6. P. 1895–1905. DOI: https://doi.org/10.2514/1.A32883.
3. Hogan F., Forbes J. Modeling of spherical robots rolling on generic surfaces. Modeling of spherical robots rolling on generic surfaces // Multibody Syst. Dyn. 2015. Vol. 35, No. 1. P. 91–109. DOI: https://doi.org/10.1007/s11044-014-9438-3.
4. Ho C.-H., Wu C.-K., Tu J.-Y., Hsu S.-K. Conceptual design of spherical vehicle system for future transportation // J. Autom. Control Eng. 2015. Vol. 3, No. 3. P. 191–195. DOI: https://doi.org/10.12720/joace.3.3.191-195.
5. Bai Y., Svinin M., Yamamoto M. Adaptive trajectory tracking control for the ball-pendulum system with time-varying uncertainties // Proceeding of International Conference on Intelligent Robots and Systems, IROS, 2017, Vancouver, BC, Canada, 2017, P. 2083–2090. DOI: https://doi.org/10.1109/IROS.2017.8206026.

6. Ivanova T.B., Kilin A.A., Pivovarova E.N. Controlled motion of a spherical robot with feedback // J. Dyn. Control Syst. 2018. Vol. 24, No. 3. P. 497–510. DOI: https://doi.org/10.1007/s10883-017-9387-2.
7. He J., Li G., Zhou R. Torque Modelling and Validation for a Spherical Motor with Stepped Permanent Magnets // J. Electr. Eng. Technol. 2020. Vol. 15. P. 2661–2673. DOI: https://doi.org/10.1007/s42835-020-00538-y.
8. Balandin D.V., Biryukov R.S., Tuzikov A.M. Path tracking control of a spherical robot with pendulum-like driver // CCIS. 2024. Vol. 1914. P. 176–189. DOI: https://doi.org/10.1007/978-3-031-52470-7_15.

МОДИФИКАЦИЯ ГРАДИЕНТНОГО СПУСКА В ЗАДАЧЕ БИЛИНЕЙНОЙ ОПТИМИЗАЦИИ

Г.В. Уткин, С.В. Сидоров

Нижегородский государственный университет им. Н.И. Лобачевского

В работе исследуется модификация метода градиентного спуска, предложенная в статье [1]. Модификация метода основывается на учете асинхронной сходимости переменных. Градиентный спуск с рассматриваемой модификацией проверяется на примере задачи билинейной оптимизации.

Ключевые слова: методы оптимизации, градиентный спуск, машинное обучение

1. Введение

Метод градиентного спуска является одним из основных численных методов современной оптимизации. Приложения данного метода возникают в задачах машинного обучения (включая линейную регрессию, логистическую регрессию и нейронные сети), компьютерного зрения и в анализе данных. Например, градиентный спуск используется для минимизации ошибок в модели машинного обучения (минимизации функции потерь).

Существует достаточное количество альтернатив и модификаций данного метода (SGD, Adam, Momentum, Mini-Batch GD и др.). В данной работе мы исследуем предложенную идею в [1] для модификации градиентного спуска, важным аспектом которой является учет особенности асинхронной сходимости переменных. Заметим, что описание метода в [1] содержит опечатки в формулах, затрудняющих понимание самого метода. Здесь мы постарались устранить эти недостатки. Наше изложение будет следовать работе [1] с уточнением применяемых формул.

2. Обзор применения. Постановка задачи

К рассмотрению предлагается следующая оптимизационная модель (задача билинейной оптимизации):

$$\underset{A,x}{\operatorname{argmin}}\, G(A,x), \text{где } G(A,x) = \frac{1}{2}\|b - Ax\|_2^2 + \lambda\|x\|_1 + R(A) \quad (1)$$

где $b \in \mathbb{R}^{M\times 1}$ – наблюдение, описываемое с помощью переменных $A \in \mathbb{R}^{M\times N}$ и $x \in \mathbb{R}^{N\times 1}$. На переменную x накладывается ограничение на разреженность в виде l_1 нормы, чтобы избежать переобучения. Слагаемое $R(A)$ отвечает за регуляризацию.

Подойти к решению данной оптимизационной задачи можно с разных сторон. Например, в текущей работе мы рассмотрим стандартный подход к решению – использование метода градиентного спуска, но с некоторой модификацией, предложенной в [1].

3. Модификация градиентного спуска

Классический метод градиентного спуска для задачи (1) заключается в следующем:

$$\begin{cases} A^{t+1} = A^t - \eta_1 \dfrac{\partial G}{\partial A} \\ x^{t+1} = x^t - \eta_2 \dfrac{\partial G}{\partial x} \end{cases}$$

где $\eta_1, \eta_2 \geq 0$ небольшие шаги, называемые скоростью обучения.

Заметим, что в статье [1] в соответствующих формулах стоит знак «+». Поскольку исходная задача на минимум, то необходимо двигаться в направлении антиградиента (направление наискорейшего локального убывания), т.е. со знаком «−».

Для решения задачи (1) можно использовать стохастический метод градиентного спуска с итеративным проецированием (см. [2, 3]). Несмотря на это, во всех модификациях градиентных спусков база остается прежней – оптимизация одной переменной при фиксировании значений другой, происходящая итеративно. В рассматриваемой задаче может наблюдаться асинхронная сходимость переменных (см, например, [4]).

В выражении частной производной по A для задачи (1) можно видеть, что градиент по A будет «затухать» при стремлении x к 0.

$$\left(\frac{\partial G}{\partial A}\right)^{\top} = x(Ax - b)^{\top}$$

Поэтому ранняя сходимость x может вызывать преждевременную сходимость A, что может помешать качественно проводить обучение исследуемой модели.

В таком случае желательно синхронизировать сходимость переменных (можно поправлять x, когда его быстрая сходимость влияет на поиск решения). Предполагается зависимость $A = A(x)$, тогда имеем целевую функцию оптимизации $G(A(x), x)$, а именно:

$$G(A(x), x) = \frac{1}{2}\|b - Ax\|_2^2 + \lambda\|x\|_1 + R(A) = G\big(A(x)\big) + G(x) + R(A),$$

где $G(x) = \lambda\|x\|_1$ и $G\big(A(x)\big) = \frac{1}{2}\|b - Ax\|_2^2$

Данная идея из оригинальной статьи [1] проиллюстрирована на рисунке ниже (взят из [1]).

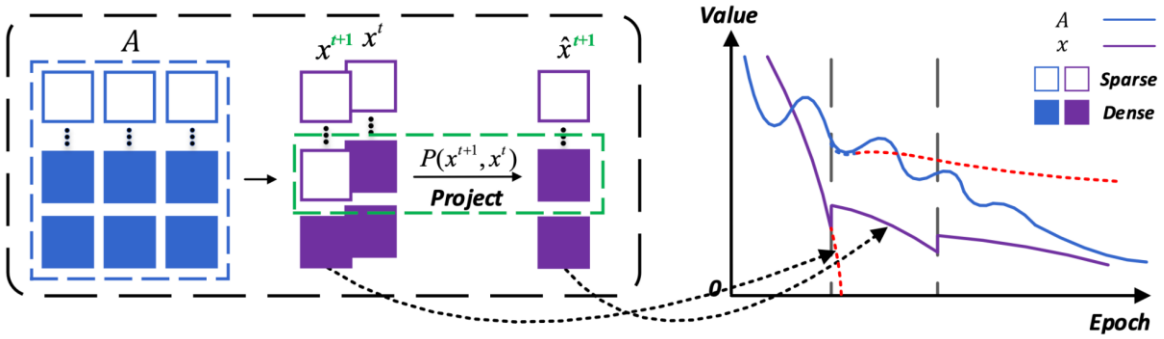


Рис. 1. Иллюстрация принципа работы предложенного алгоритма

Проделывая все математические выкладки с учетом такой зависимости, можно получить формулу, по которой следует корректировать переменную x:

$$\tilde{x}_j^{t+1} = x_j^t + \eta_2 x_j \sum_{i=1}^{M} \tilde{g}_i \frac{\partial A_{ij}}{\partial x_j}$$

Или в матричной форме:

$$\tilde{x}^{t+1} = x^t + \eta_2 \begin{pmatrix} \sum_{i=1}^{M} \tilde{g}_i \frac{\partial A_{i1}}{\partial x_1} \\ \vdots \\ \sum_{i=1}^{M} \tilde{g}_i \frac{\partial A_{iN}}{\partial x_N} \end{pmatrix} \odot \begin{pmatrix} x_1 \\ \vdots \\ x_N \end{pmatrix} = x^t + \eta_2 c^t \odot x^t,$$

где $\tilde{G} = (Ax^t - b)^{\top}$, а бинарная операция $\odot$ – это произведение Шура-Адамара (покомпонентное умножение матриц одинаковой размерности).

Теперь следует определить, когда требуется корректировать x. Для этого предлагается характеристика для проверки, сошлась ли переменная с некоторым пороговым значением α. В оригинальной статье с помощью двух пороговых параметров α_x и α_A определяется асинхронность, а именно отслеживаются нормы A и x. Принцип выявления асинхронности логичен, а именно

пока норма A превышает пороговое значение α_A, корректируется x если он меньше порогового α_x.

В окончательном варианте правило корректировки x следующее:

$$\tilde{x}^{t+1} = \begin{cases} x^t + \eta_2 c^t \odot x^t, \text{если } R(x) < \alpha_x \text{ и } R(A) > \alpha_A, \\ x^{t+1}, \quad \text{иначе.} \end{cases} \tag{2}$$

Конечно, на практике явно узнать зависимость $A(x)$, чтобы посчитать частную производную, не получится. Поэтому выражение приближается отношениями переменных между итерациями, то есть:

$$\frac{\partial A_j}{\partial x_j} \approx \frac{\Delta A_j}{\Delta x_j}$$

при этом, если $\Delta x_j \to 0$, то значение производной принимают за единицу.

Заметим, что данную модификацию можно применять и к другим градиентным методам. Возврат в любом случае происходит по правилу обновления градиентного спуска.

4. Реализация метода

Модификация обновления вектора x методом градиентного спуска с рассмотренной модификацией описана на псевдокоде на Рис. 2. Сама реализация была написана на языке Python 3.10.12.

Вход: $A^t, A^{t+1}, x^t, x^{t+1}, b, \eta_2, \alpha_A, \alpha_x$
Выход: $\tilde{x}^{t+1}$

1. $\tilde{x}^{t+1} = x^{t+1}$
2. **Если** $\|x^{t+1}\|_1 \leq \alpha_x$ и $\|A^{t+1}\|_1 > \alpha_A$ **То**
3. **Цикл по** i **от** 0 **до** n **Выполнять**
4. $c^t \coloneqq 2(A^t x^t - b)\frac{\Delta A}{\Delta x}$
5. **Конец.**
6. $\tilde{x}^{t+1} = x^t + \eta_2 c^t \odot x^t$
7. **Конец.**
8. **Возвратить** $\tilde{x}^{t+1}$

Рис. 2. Алгоритм рассмотренной модификаици градиентного спуска

Под операцией $\odot$, как и ранее, понимается покомпонентное умножение векторов, а под обозначением ΔA понимается отношение между итерациями, то есть $\Delta A = A^{t+1} - A^t$. Под выражением $\frac{\Delta A}{\Delta x}$ понимается деление каждого столбца ΔA на соответствующий элемент вектора Δx.

5. Эксперименты

В работе [1] метод проверялся на известной тестовой оптимизационной функции Била:

$beale(x_1, x_2) = (1.5 - x_1 + x_1 x_2)^2 + (2.25 - x_1 + x_1 x_2^2)^2 + (2.625 - x_1 + x_1 x_2^3)^2$

с дополнительными ограничениями:

$$F(x_1, x_2) = beale(x_1, x_2) + \big||x_1|\big| + x_2^2,$$

которая сводится к задаче (1). Действительно,

$$A = \begin{pmatrix} 1 - x_2 \\ 1 - x_2^2 \\ 1 - x_2^3 \end{pmatrix}, b = \begin{pmatrix} 1.5 \\ 2.25 \\ 2.625 \end{pmatrix}, x = x_1$$

Рассмотренная модификация встраивалась в известные методы градиентного спуска SGD, Momentum и Adam (см, соответственно, [5-7]), то есть в случае ранней сходимости будем обновлять x^{t+1} по правилу (2), а если ранней сходимости нет, то x^{t+1} будет оставаться таким, как и в методе SGD, Momentum или Adam.

На следующих рисунках были изображены траектории метода до точки минимума (рисунок слева отображает траектории, а рисунок справа – значение функции на определенной итерации). Выбранные параметры алгоритмов следующие:

1. SGD: начальная точка с координатами $(1.5, 1.8), \alpha_{x_1} = 1, \alpha_{x_2} = 0.5, \eta_2 = 0.001$;
2. Momentum: начальная точка с координатами $(1.5, 1.8), \alpha_{x_1} = 1, \alpha_{x_2} = 0.5, \eta_2 = 0.005$; Внутренний параметр Momentum $\beta_1 = 0.55$;
3. Adam: начальная точка с координатами $(1.5, 1.8)$, $\alpha_{x_1} = 1, \alpha_{x_2} = 0.5, \eta_2 = 0.1$; Внутренние параметры метода Adam стандартные $\beta_1 = 0.9, \beta_2 = 0.99$;.

Оказалось, что метод сошелся в каждом случае и траектории с внедрением рассмотренной модификации короче, чем в исходном методе.

Мы проверим работу метода на следующей задаче:

$$\underset{A,x}{\operatorname{argmin}} G(A, x) = \frac{1}{2}\|b - Ax\|_2^2 + \|x_1 + x_2\|_1 + (x_1^2 + x_2^2), \tag{3}$$

$$A = \begin{pmatrix} x_1 - x_2 \\ x_1 - x_2^2 \\ x_1 - x_2^3 \end{pmatrix}, b = \begin{pmatrix} 1 \\ 2 \\ 4 \end{pmatrix}, x = (x_1 + x_2)$$

Параметры данного метода были выбраны такими же, как выше, а в качестве начальной точки взята $(0.1, -2.5)$. На рис. 3 приведены результаты эксперимента, которые подтверждают ожидания: метод сходится быстрее и с более короткой траекторией.

В дальнейшем предполагается продолжение исследований для изучения сходимости метода с предложенной модификацией и применения для конкретной практической задачи (например, для CNN training, а также для Network Pruning).

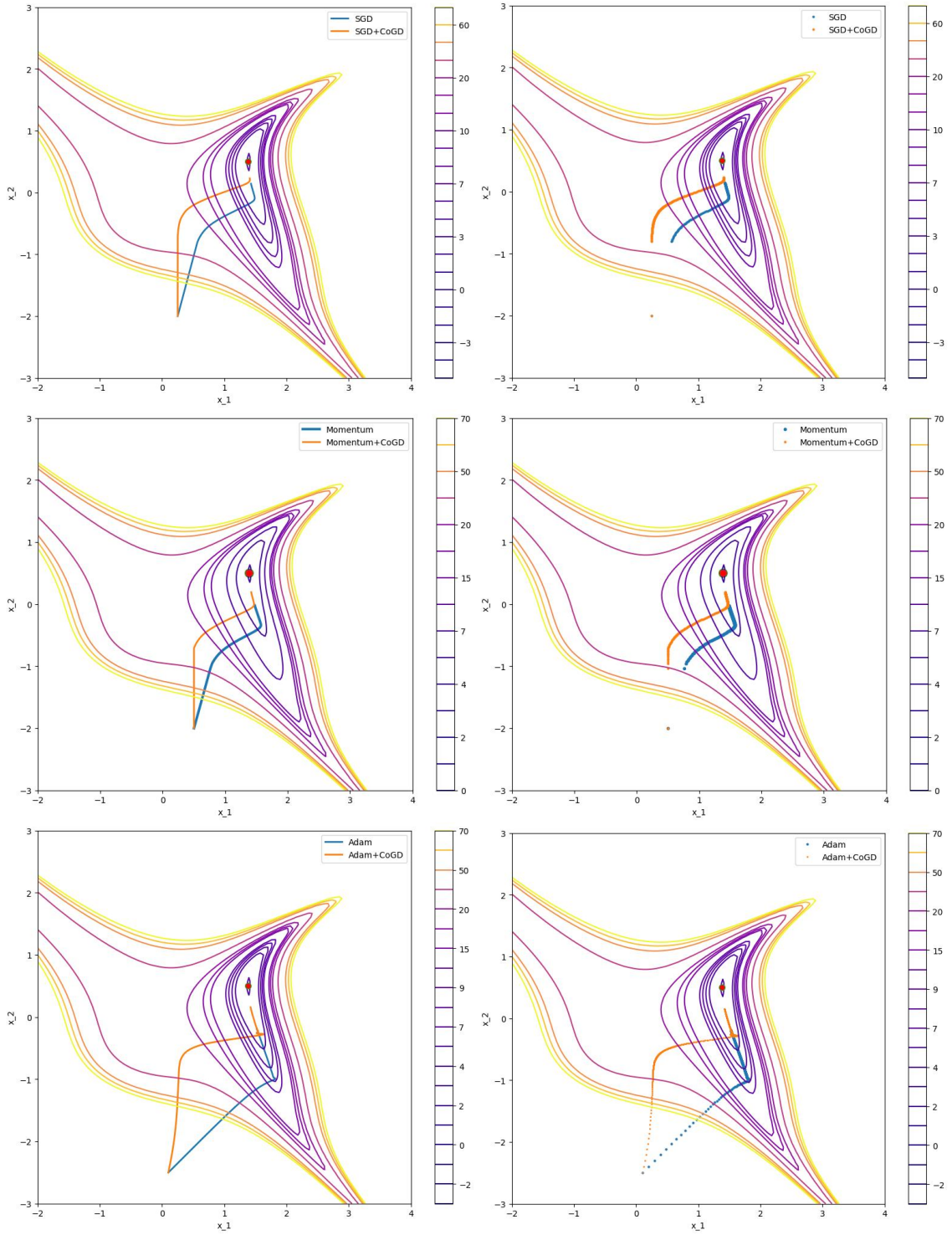


Рис. 3. Поиск минимума методами SGD, Momentum, Adam и с внедрением модификации (для функции (3))

Литература

1. Wang R., Zhang B., Zhuo L., Ye Q., Doermann D. Cogradient Descent for Dependable Learning // arXiv preprint arXiv:2106.10617. – 2021. DOI: 10.48550/arXiv.2106.10617.
2. Aharon M., Elad M. “Sparse and Redundant Modeling of Image Content Using an Image-Signature-Dictionary” // SIAM Journal on Imaging Sciences. 1 (3), 2008. P. 228–247. %DOI:10.1137/07070156x.

3. Pinter, J.D. Yair Censor and Stavros A. Zenios, Parallel Optimization – Theory, Alghorithms, and Applications. -- Oxford University Press. -- New York, 1997. xxviii+539 P.
4. J. Wu, L. Cong, Y. Wang, Q. Hu, C. Jian Quantized convolutional neural networks for mobile devices // In CVPR. – 2016. – 1–11 P. DOI: 10.48550/arXiv.1512.06473.
5. Taddy, Matt Stochastic Gradient Descent. Chap. 10.4 in Business Data Science: Combining Machine Learning and Economics to Optimize, Automate, and Accelerate Business Decisions. 1st ed. New York, – 2019.
6. Rumelhart, David E., Hinton Geoffrey E., Williams Ronald J. Learning representations by back-propagating errors. – 1986. – 533–536 P.
7. Diederik P. Kingma, Jimmy Ba. Adam: A method for stochastic optimization // arXiv preprint arXiv:1412.6980, – 2017. DOI: 10.48550/arXiv.1412.6980.

ИДЕНТИФИКАЦИЯ КЛАСТЕРОВ КОЛЛОИДНЫХ СБОРОК МЕТОДОМ МАШИННОГО ОБУЧЕНИЯ[1*]

Л.Т. Хусаинова, К.С. Колегов

Астраханский государственный университет им. В.Н. Татищева

Данная работа посвящена проблеме идентификации кластеров коллоидных сборок методом машинного обучения на примере двухмерных покрытий (однослойные сборки). В работе обсуждается актуальность применения коллоидных систем в различных областях науки и техники, а также важность эффективной идентификации кластеров частиц. Здесь описана классификация сборок частиц: изолированные частицы, димеры, цепочки и кластеры. В качестве инструмента для идентификации таких объектов на различных снимках (экспериментальных и компьютерных симуляциях) была использована модель YOLOv8. Описаны особенности и текущие результаты обучения модели на специально подготовленном датасете. Точность распознавания составила примерно 93%. Этот инструмент полезен, к примеру, для приложений в области фотоники и функциональных покрытий, так как их физические свойства зависят от морфологии структуры.

Ключевые слова: идентификация кластеров, коллоидные сборки, морфология структур, нейронная сеть, алгоритм YOLO.

1. Введение

Самосборка коллоидных структур стала важным методом производства для наноструктурированных материалов с широким спектром применений, начиная от оптоэлектроники и заканчивая определением химических и биологических аналитов или созданием биомиметических поверхностей со специфическими свойствами или свойствами смачиваемости [1]. Коллоидные сборки, в силу своей уникальной природы, представляют собой объекты интенсивного внимания в научном и инженерном сообществе. Эта актуальность обусловлена не только их широким спектром применения в различных областях, но и потенциалом этих систем в решении важных проблем современного мира.

Одним из вызовов, стоящих перед исследователями, является эффективная и точная идентификация кластеров в коллоидных сборках. Точное выделение и классификация этих кластеров является ключевым этапом для понимания их свойств и воздействия на окружающую среду. Коллоидные сборки, представляющие собой дисперсные системы с частицами размером от нанометров до микрометров, привлекают все большее внимание исследователей благодаря своим уникальным свойствам и широкому спектру применений. Однако вопросы, связанные с идентификацией и классификацией кластеров в коллоидных сборках, остаются актуальными и требуют внимания. Целью данной работы является анализ возможностей модели YOLOv8 в идентификации сборок коллоидных частиц по экспериментальным и искусственно сгенерированным снимкам.

2. Постановка задачи

В данной работе рассматриваются однослойные сборки, содержащие как бинарные смеси частиц, так и частицы одного размера. Эти сборки, будучи малыми строительными блоками материи, обладают способностью объединяться и образовывать кластеры. В частности, каждый из связанных компонентов в кластере, принадлежит к одной из следующих категорий, которые определяют конфигурацию, принимаемую частицами [2]:

[1*] Исследование выполнено за счет гранта Российского научного фонда № 22-79-10216.

– изолированные частицы (частица *j* принадлежит к этой группе, если у нее нет соседей на определенном расстоянии от ее центра; это расстояние выбирается индивидуально в зависимости от самого изображения, то есть носит интуитивный характер в предположении, что расстояние между двумя соседними частицами должно быть на порядок меньше суммы радиусов этих частиц; это связано с тем, что точное критическое расстояние для всех возможных случаев определить проблематично);

– димеры (две частицы *j* и *k* принадлежат к этой группе, если у *j* есть только сосед *k* на определенном расстоянии от ее центра и если у *k* есть только сосед *j* на определенном расстоянии от ее центра массы);

– цепи (в цепочке все частицы имеют двух соседей, за исключением частиц на двух концах цепи, которые имеют только одного соседа);

– кластеры (частицы составляют кластер, если они принадлежат связной компоненте, которая не попадает ни в одну из предыдущих категорий; в этом наборе, к примеру, мы можем различать петли; частицы принадлежат циклу, если все они имеют двух соседей на определенном расстоянии от их центроида; петлю можно интерпретировать как своего рода «замкнутую» цепь или кольцо).

3. Метод решения

В качестве основного метода решения поставленной задачи было выбрано использование модели YOLOv8 [3]. Данная модель (обобщенная архитектура представлена на рис. 1) обладает не только высокой скоростью, но и высокой точностью определения и распознавания объектов, а также поддерживает полный спектр задач компьютерного зрения, включая используемую в данном исследовании сегментацию.

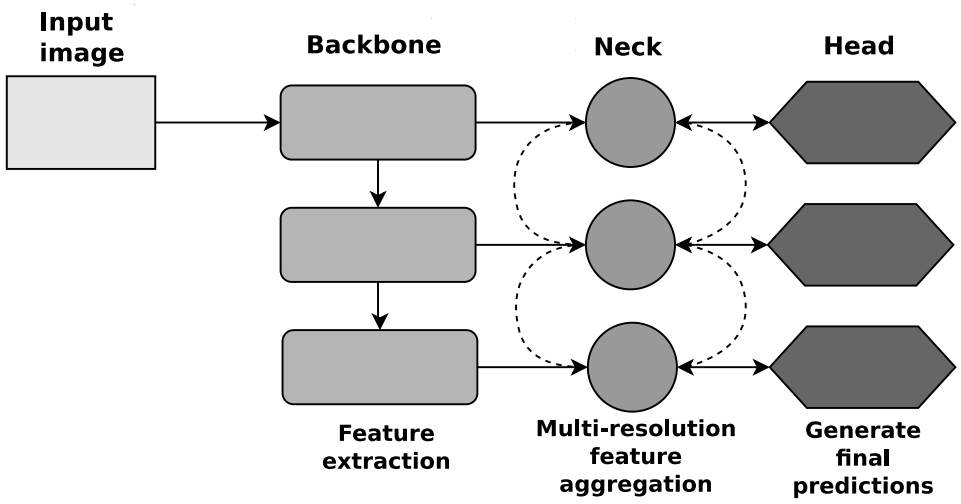


Рис. 1. Архитектура сети YOLOv8

Архитектура модели состоит из трех частей: Backbone, Neck и Head. Backbone служит для извлечения признаков из входного изображения. Обычно это глубокая сверточная нейронная сеть. Neck обрабатывает признаки, извлеченные из Backbone. Может включать дополнительные слои и механизмы для уточнения извлеченных признаков. Head делает окончательные прогнозы на основе предыдущих частей архитектуры.

В отличие от техники обнаружения объектов на изображении сегментация позволяет классифицировать каждый пиксель входного изображения, в том числе определить, к какому именно классу данный пиксель принадлежит. Использование техники сегментации в задаче идентификации кластеров коллоидных сборок является лучшим вариантом, так как при анализе входных изображений было выявлено, что различные кластеры могут быть расположены близко друг к другу, что влияет на точность их определения, а сегментация, в свою очередь, позволяет их точнее разделить между собой. В отличие от архитектуры YOLOv8, модель YOLOv8-seg, которая непосредственно отвечает за сегментацию изображения, имеет две части Head: Detection Head и

Segmentation Head. Первая часть предсказывает ограничивающие рамки и вероятности классов для обнаружения объектов, аналогично стандартной модели YOLOv8. Вторая часть предсказывает маски сегментации для каждого обнаруженного объекта. Обычно эта часть включает в себя масштабирование признаков до исходного размера и применение попиксельной классификации для присваивания каждого пикселя объекту или фону.

В ходе работы над поставленной задачей был создан и размечен датасет, состоящий из различных изображений коллоидных сборок (обработано около 700 изображений). Работа над датасетом была осуществлена с использованием платформы Roboflow [4].

В ходе эксперимента моделью автоматически был выбрал оптимизатор AdamW, который отлично подходит для пользовательских датасетов с небольшим количеством классов. Также автоматически были выбраны гиперпараметры “Learning rate”, равный $1x10^{-2}$, и “momentum”, равный 0,937.

4. Полученные результаты

Модели YOLO уже предобучены на большом наборе данных для одновременного обнаружения на изображении нескольких классов, в нашем случае: изолированные частицы, димеры, цепочки и кластеры. Эксперимент проводился на GPU NVIDIA Quadro RTX A6000 с автоматически рассчитанным batch size, равным 58. Модель тренировалась с использованием Python 3.8 и PyTorch 1.12. Изображения в датасете были предобработаны и аугментированы с помощью инструментов Roboflow. В ходе обучения нейронной сети на подготовленном нами датасете удалось добиться высокого уровня распознавания кластеров – средняя точность (mAP) около 93%. На рис. 2 представлены результаты валидации модели.

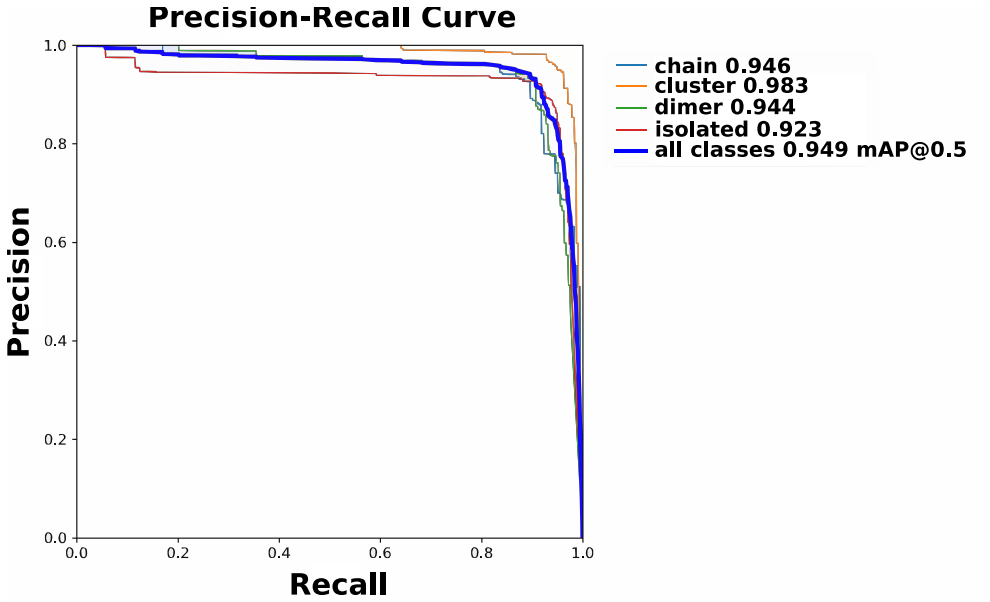


Рис. 2. Результаты валидации модели

Для более наглядного пояснения взглянем на несколько показательных примеров. На рис. 3 представлены изображения коллоидных сборок, с одной стороны (рис. 3a) размеченных человеком вручную, а с другой – предсказанных моделью после обучения (рис. 3b). Эти результаты демонстрируют, что наша модель обладает хорошей способностью обнаруживать кластеры коллоидных сборок. Однако было замечено, что при больших скоплениях частиц падает точность идентификации кластеров. Ознакомиться с датасетом и обученной моделью можно здесь [5].

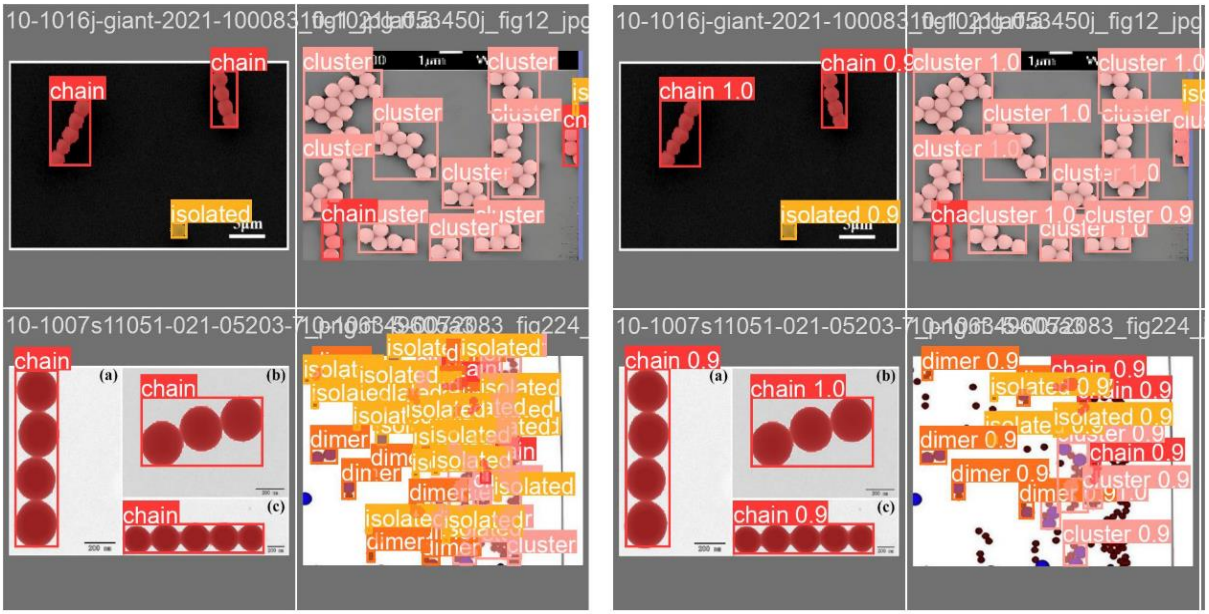


Рис. 3. Примеры кластеров коллоидных сборок: (a) размеченные вручную, (b) предсказанные моделью

Также в целях уменьшения времени обучения модели был проведен эксперимент, в котором модель обучалась на разном количестве графических процессорах NVIDIA Quadro RTX A6000. В табл.1 можно увидеть, что при обучении модели на 8 GPUs скорость обучения увеличилась примерно в 1.7 раз в сравнении с 1 GPU. Эта разница не слишком заметна в рамках данного датасета. Увеличение числа GPU в данном случае несильно влияет на скорость обучения модели. Связано это с тем, что обучающий датасет довольно мал, чтобы увидеть значительную разницу во времени расчета.

Таблица 1. Зависимость времени обучения от количества видеокарт

Количество GPU	**Время, с**
1	389
3	263
6	247
8	232

5. Выводы

В ходе работы были рассмотрены однослойные сборки, содержащие как бинарные смеси частиц, так и частицы одного размера, для возможности идентификации образованных кластеров по снимкам. На основе изображений этих сборок был создан и размечен соответствующий датасет. В качестве метода идентификации выбрана модель YOLOv8. Модель обучена на созданном нами датасете. Анализ полученных результатов показал, что при большом скоплении частиц на изображении точность идентификации кластеров падает. Установлено, что увеличение количества GPU, на котором обучается модель, сокращает время обучения примерно в 1.7 раза. Такое незначительное ускорение расчета, скорее всего, связано с относительно малым размером обучающего датасета.

Литература

1. Lotito V, Zambelli T. Approaches to self-assembly of colloidal monolayers: A guide for nanotechnologists // Adv. Colloid Interface Sci. 2017. No. 246. P. 217-274. DOI: 10.1016/j.cis.2017.04.003.
2. Lotito V, Zambelli T. Pattern detection in colloidal assembly: A mosaic of analysis tech-niques. // ScienceDirect. 2020. No. 9. P. 108. DOI: 10.1016/j.cis.2020.102252.

3. YOLOv8 – Ultralytics YOLO Docs. URL: https://docs.ultralytics.com/.
4. Roboflow: Computer vision tools for developers and enterprises. URL: https://roboflow.com/.
5. Roboflow dataset. URL: https://universe.roboflow.com/masters-dissertation-98uy9/segmentation-test-r5wms/model/80.

СОДЕРЖАНИЕ

Н.Ф. Алексиадис
Задача о базисах в функциональной системе рациональных функций с рациональными коэффициентами (алгоритмический подход) ... 3

Д.В. Баландин, Р.С. Бирюков, М.М. Коган
Многокритериальные минимаксные задачи оптимального управления ... 8

Борисов Н.А., Егоров К.С.
Некоторые задачи и алгоритмы навигации автономных транспортных средств ... 14

А.А. Бусалов, А.В. Калинин, А.А. Тюхтина
Исследование эффективности распараллеливания итерационного алгоритма решения нелинейной системы переноса излучения и статистического равновесия ... 18

М.А. Быкова, А.А. Урмаев, А.И. Яушева, Д.В. Пачурина, Д.И. Попугаев
Гибридный алгоритм решения задачи размещения компонент интегральных схем ... 21

А.А. Воденеева, К. И. Ковалев, Е.А. Козинов, А.Ю. Пирова, А.В. Устинов, В.Д. Волокитин, И.Б. Мееров
Оптимизация функций умножения плотной ленточной матрицы на вектор ... 27

М.Д. Вульф, Д.С. Жариков, Д.С. Коломенский
Обработка экспериментов соударения капли с подложкой и бинарный классификатор разрушения капли ... 31

И.Е.З. Гебриал, К.А. Леонтьева, И.А. Кочеткова
Ресурсная система массового обслуживания с управлением перераспределением ресурса для анализа нарезки беспроводной сети ... 35

И.В. Дель, А.В. Старченко
Краткосрочный прогноз скорости ветра на основе искусственных нейронных сетей с применением метода декомпозиции на вариационные моды ... 39

А.Г. Дёгтев, В.В. Разин
Перспективы классификации нормативных требований к предпринимателям посредством применения классических методов машинного обучения ... 42

Н.С. Дюкина, А.В. Домрачева
Численно-аналитический метод расчета осадки системы «сооружение-грунт» ... 48

А.А. Живцова, К.Е. Самуйлов
Оценка качества политики активации каналов, полученной с помощью обучения с подкреплением ... 51

Ю.В. Захарова, М.Ю. Сахно
Генетический алгоритм для построения расписания на многоядерном процессоре с учетом взаимного влияния работ ... 56

Д.А. Збруев, Н.А. Борисов
Моделирование алгоритма движения робота в условиях складских помещений ... 59

В.И. Звонилов
Графы тригональных кривых и жесткие изотопии особых вещественных алгебраических кривых бистепени (4,3) на гиперболоиде ... 65

А.Д. Ивлев
Реализация алгоритмов подготовки начального состояния системы кубитов с учетом ограничений современных квантовых компьютеров ... 68

Е.А. Козинов, А.Ю. Пирова, В.Д. Волокитин, Е.А. Панова, А.В. Линев, И.Б. Мееров
Анализ производительности OpenBLAS на процессорах с архитектурой RISC-V и x86 75

А.В. Кондратьева, М.И. Кузнецов
Вложение фильтрованных деформаций градуированных неальтернирующих гамильтоновых алгебр Ли в общую алгебру Ли картановского типа 79

Р.А. Кононов, Н.А. Поспелов, О.В. Масленников, В.И. Некоркин
Формирование специализированных популяций в рекуррентных нейронных сетях, обученных с подкреплением задаче контекстно-зависимого принятия решения..... 81

Н.А. Кудряшкин, Р.А. Кононов, В.С. Тиселько, О.В. Масленников
Структура и динамика спайковых нейронных сетей в обработке стохастических входных сигналов 84

С.А. Лавренков, А.В. Васюков
Определение упругих свойств композитных материалов методом решения обратной задачи колебаний пластин на вибрационном стенде 87

Р.Е. Ломовский, М.И. Кумсков, В.М. Буданов
Решение задачи семантической сегментации для работы видео модуля беспилотного робота 90

О.В. Любимцев, А.А. Туганбаев
Лиево-нильпотентные центрально существенные кольца 96

В.В. Маносин, М.И. Кумсков, В.М. Буданов
Распознавание препятствий с использованием семантической сегментации при движении автономного мобильного робота 99

А.В. Мухин
Условия представления произвольной линейной системы в канонической управляемой форме..... 105

Л.П. Никифорова, С.Ю. Маковкин
Численное моделирование динамической системы астроцита Уллаха-Юнга 111

Д.И. Николаев, Е.А. Гайдамака
Анализ граничного узла сети IAB в полудуплексном режиме: исследование задержек с помощью модели поллингового обслуживания с двумя очередями..... 114

М.В. Оболенский, О.В. Воронков, С.С. Андропенков, Н.Н. Куриков
Численное исследование влияния начального несовершенства на форму закритического деформирования цилиндрической оболочки при осевом динамическом сжатии 120

В.В. Разин, А.А. Краснов
Применение нейронных сетей для анализа ЭКГ сигналов на наборе данных PTBXL..... 126

С.В. Русаков, В.Г. Гилев
Моделирование процесса пропитки полидисперсной среды полимерным связующим 132

В.П. Савельев
Проблема Фробениуса для трех натуральных чисел специального вида 137

В.П. Савельев, Н.И. Сутягина
Математические модели динамики малого бизнеса в условиях выделения грантов 141

С.В. Сидоров, Д.В. Усов
Исследование длин периодов цепных дробей для квадратного корня из степеней простых чисел 145

В.С. Соколов
Трехмерная реконструкция объекта по двумерным ультразвуковыми снимкам с использованием робота-манипулятора 149

А.В. Степанов
Построение обобщенного H_∞-оптимального управления дискретной системой по априорным и экспериментальным данным 152

С.Н. Стребуляев, Сингх Дхананджай, А.Д.Черепанов
Вычислительный эксперимент по анализу характеристик хаоса для конкретных динамических систем..... 159

Е.Б. Титова
Рост миноров рангового порядка матрицы ограничений трехиндексной планарной транспортной задачи..... 164

А.М. Тузиков
Математическое моделирование динамики сферического робота с электромагнитным приводом..... 168

Г.В. Уткин, С.В. Сидоров
Модификация градиентного спуска в задаче билинейной оптимизации..... 173

Л.Т. Хусаинова, К.С. Колегов
Идентификация кластеров коллоидных сборок методом машинного обучения 179

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ И СУПЕРКОМПЬЮТЕРНЫЕ ТЕХНОЛОГИИ

Труды XXIV Международной конференции

Нижний Новгород, 18–21 ноября 2024 г.

Под ред. проф. Д.В. Баландина

Отв. за выпуск К.А. Баркалов

Публикуется в авторской редакции

Издательство Нижегородского госуниверситета им. Н.И. Лобачевского.
603950, Н. Новгород, пр. Гагарина, 23.