

Министерство науки и высшего образования РФ

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



МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ И СУПЕРКОМПЬЮТЕРНЫЕ ТЕХНОЛОГИИ

Труды XXII Международной конференции

Нижний Новгород, 14–17 ноября 2022 г.

Нижний Новгород
Издательство Нижегородского госуниверситета
2022

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

- М34 **Математическое моделирование и суперкомпьютерные технологии.** Труды XXII Международной конференции (Н. Новгород, 14–17 ноября 2022 г.) / Под ред. проф. Д.В. Баландина. – Нижний Новгород: Изд-во Нижегородского госуниверситета, 2022. – 170 с.

Отв. за выпуск К.А. Баркалов

ISBN 978-5-91326-771-9

Сборник материалов Двадцать второй Международной конференции «Математическое моделирование и суперкомпьютерные технологии», состоявшейся 14–17 ноября 2022 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/hpc2022>.

Поддержка конференции

Научно-образовательный математический центр
«Математика технологий будущего»

ISBN 978-5-91326-771-9

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

© Авторы статей, 2022
© ННГУ им. Н.И. Лобачевского, 2022

ПРЕДИСЛОВИЕ

14–17 ноября 2022 г. в рамках международного конгресса «Суперкомпьютерные дни в России» на базе Нижегородского государственного университета им. Н.И. Лобачевского при поддержке научно-образовательного математического центра «Математика технологий будущего» проведена XXII 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 моделирование 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).

НЕЛИНЕЙНАЯ НЕСТАЦИОНАРНАЯ СИСТЕМА ИНТЕГРО-ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ ПЕРЕНОСА ИЗЛУЧЕНИЯ И СТАТИСТИЧЕСКОГО РАВНОВЕСИЯ В ДИФФУЗИОННОМ ПРИБЛИЖЕНИИ

А.А. Бусалов¹, А. В. Калинин^{1,2}, А. А. Тюхтина^{1,2}

¹Нижегородский государственный университет им Н.И. Лобачевского

²Институт прикладной физики РАН

В работе рассматривается начально-краевая задача для нелинейной системы интегро-дифференциальных уравнений переноса излучения и статистического равновесия в диффузионном приближении. Предлагается и численно исследуется итерационный линеаризующий алгоритм решения рассматриваемой задачи. Анализируются результаты численного решения, и обсуждается эффективность предложенного итерационного алгоритма.

Ключевые слова: нелинейное уравнение переноса, диффузионное приближение, уравнения статистического равновесия, итерационные методы.

1. Введение

Вопросы физического, математического и численного моделирования процессов переноса излучения обсуждаются в работах [1-5]. Более полная информация о возможных приложениях и современных подходах к теоретическому и численному анализу задач теории переноса излучения содержится в [5,6].

В настоящей работе исследуется возможность применимости линеаризующего итерационного алгоритма для нелинейной нестационарной системы уравнений переноса излучения и статистического равновесия для модели двухуровневого атома в диффузионном приближении. Обсуждаются особенности программной реализации. Соответствующая стационарная задача исследовалась в работе [7]. Для рассматриваемой в работе нестационарной задачи приводится конечно-разностная схема для аппроксимации диффузионного уравнения переноса, обсуждаются ее основные свойства. Демонстрируются результаты численного анализа предложенного итерационного алгоритма, подтверждающие его сходимость.

2. Постановка задачи

Нелинейная система интегро-дифференциальных уравнений переноса излучения и статистического равновесия в плоско-параллельном слое, соответствующая модели двухуровневого атома в рамках предположения о полном перераспределении излучения по частоте имеет вид [1,2]:

$$\frac{1}{c} \frac{\partial \varphi(t, x, \mu, \nu)}{\partial t} + \mu \frac{\partial \varphi(t, x, \mu, \nu)}{\partial x} + h\nu_{12} \frac{\kappa(\nu)}{2} [B_{12}C_1(t, x) - B_{21}C_2(t, x)] \varphi(t, x, \mu, \nu) = h\nu_{12} \frac{\kappa(\nu)}{2} A_{21}C_2(x). \quad (1)$$

$$\left[C_{12}n_e(x) + B_{12} \int_I \int_{-1}^1 \frac{\kappa(\nu)}{2} \varphi(t, x, \mu, \nu) d\mu d\nu \right] C_1(t, x) = \left[A_{21} + C_{21}n_e(x) + B_{21} \int_I \int_{-1}^1 \frac{\kappa(\nu)}{2} \varphi(t, x, \mu, \nu) d\mu d\nu \right] C_2(t, x), \quad (2)$$

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

Рассматриваемая система дополняется граничными и начальными условиями

$$\varphi(t, x_1, \mu, \nu) = \psi_1(t, x_1, \mu, \nu), \quad \mu > 0, \quad (4)$$

$$\varphi(t, x_2, \mu, \nu) = \psi_2(t, x_2, \mu, \nu), \quad \mu < 0, \quad (5)$$

$$\varphi(0, x, \mu, \nu) = \varphi^0(x, \mu, \nu). \quad (6)$$

Здесь $\varphi(t, x, \mu, \nu)$ – интенсивность излучения; $C_1(t, x), C_2(t, x)$ – населенности уровней в основном и возбужденном состояниях; $x \in [x_1; x_2]$, $x_2 - x_1 = d > 0$, $\mu \in [-1; 1]$, $t \in [0; T]$, $T > 0$, $h, \nu_{12}, \nu_0, A_{21}, B_{12}, B_{21}, C_{12}, C_{21}$ – заданные положительные числа, удовлетворяющие условию $B_{12}C_{21} - B_{21}C_{12} \geq 0$; $n_e(x), f(x), k(\nu)$ – положительные и ограниченные в своих областях определения функции. Физический смысл коэффициентов системы и приведенных условий обсуждается в [1, 2].

Для построения диффузионного приближения [7, 8], решение уравнения (1) будем искать в виде разложения по сферическим функциям

$$\varphi(t, x, \mu, \nu) = \sum_{k=0}^{\infty} (2k+1) \sum_{i=-k}^k \varphi_{k,i}(t, x, \nu) Y_{ki}(\mu),$$

ограничиваясь только первыми двумя членами ряда:

$$\varphi(t, x, \mu, \nu) \approx \frac{1}{2} [\varphi_0(t, x, \nu) + 3\mu\varphi_1(t, x, \nu)],$$

$$\text{где } \varphi_0(t, x, \nu) = \int_{-1}^1 \varphi(t, x, \mu, \nu) d\mu; \quad \varphi_1(t, x, \nu) = \int_{-1}^1 \mu \varphi(t, x, \mu, \nu) d\mu.$$

Для того чтобы получить уравнения относительно функций $\varphi_0(t, x, \nu)$ и $\varphi_1(t, x, \nu)$, проинтегрируем уравнение (1) с весами 1 и μ по интервалу $\mu \in [-1; 1]$:

$$\begin{cases} \frac{1}{c} \frac{\partial \varphi_0(t, x, \nu)}{\partial t} + \frac{\partial \varphi_1(t, x, \nu)}{\partial x} + h\nu_{12} \frac{\kappa(\nu)}{2} [B_{12}C_1(t, x) - B_{21}C_2(t, x)] \cdot \varphi_0(t, x, \nu) = \\ = h\nu_{12} \kappa(\nu) A_{21} C_2(t, x), \\ \frac{1}{3} \frac{\partial \varphi_0(t, x, \nu)}{\partial x} + h\nu_{12} \frac{\kappa(\nu)}{2} [B_{12}C_1(t, x) - B_{21}C_2(t, x)] \cdot \varphi_1(t, x, \nu) = 0. \end{cases} \quad (7)$$

В диффузионном приближении граничные условия, соответствующие (4), (5) имеют вид [8]:

$$\varphi_1 + \frac{1}{2} \varphi_0 \Big|_{x=x_1} = 0, \quad \varphi_1 - \frac{1}{2} \varphi_0 \Big|_{x=x_2} = 0. \quad (8)$$

Система (7) может быть сведена к уравнению второго порядка

$$\begin{aligned} \frac{1}{c} \frac{\partial \varphi_0(t, x, \nu)}{\partial t} - \frac{\partial}{\partial x} \frac{1}{3h\nu_{12} \frac{\kappa(\nu)}{2} [B_{12}C_1(t, x) - B_{21}C_2(t, x)]} \frac{\partial}{\partial x} \varphi_0(t, x, \nu) + \\ + h\nu_{12} \frac{\kappa(\nu)}{2} [B_{12}C_1(x) - B_{21}C_2(t, x)] \cdot \varphi_0(t, x, \nu) = h\nu_{12} \kappa(\nu) A_{21} C_2(t, x). \end{aligned} \quad (9)$$

Уравнения статистического равновесия (2), (3) принимают вид:

$$\begin{aligned} \left[C_{12} n_e(x) + B_{12} \int_I \kappa(\nu) \varphi_0(t, x, \nu) d\nu \right] C_1(t, x) = \left[A_{21} + C_{21} n_e(x) + B_{21} \int_I \kappa(\nu) \varphi_0(t, x, \nu) d\nu \right] C_2(t, x), \\ C_1(t, x) + C_2(t, x) = f(x). \end{aligned}$$

Для решения возникающей системы уравнений предлагается и численно исследуется линеаризующий итерационный алгоритм. Введем следующие обозначения:

$$a^k(t, x, \nu) = h\nu_{12} \frac{\kappa(\nu)}{2} [B_{12}C_1^k(t, x) - B_{21}C_2^k(t, x)], \quad b^k(t, x, \nu) = h\nu_{12} \kappa(\nu) A_{21} C_2^k(t, x), \quad D^k(t, x, \nu) = \frac{2}{3a^k} -$$

коэффициент диффузии.

С учетом введенных обозначений, уравнение (9) может быть записано в виде:

$$\frac{1}{c} \frac{\partial \varphi_0^k(t, x, \nu)}{\partial t} - \frac{\partial}{\partial x} D^k \frac{\partial}{\partial x} \varphi_0^k(t, x, \nu) + a^k(t, x, \nu) \cdot \varphi_0^k(t, x, \nu) = b^k(t, x, \nu), \quad (10)$$

где k - номер итерации.

3. Разностная аппроксимация диффузионного уравнения

Введем разностную сетку по переменным t (индекс n), x (индекс i), и аппроксимируем уравнение (10). Для решения уравнения диффузионного типа с переменными коэффициентами диффузии разностная схема строится следующим образом:

$$\frac{1}{\Delta t} (\varphi_i^{n+1} - \varphi_i^n) - \frac{1}{\Delta x^2} \left(\chi_{i+1/2}^{n+1} (\varphi_{i+1}^{n+1} - \varphi_i^{n+1}) - \chi_{i-1/2}^{n+1} (\varphi_i^{n+1} - \varphi_{i-1}^{n+1}) \right) + a_i^{n+1} \varphi_i^{n+1} = b_i^{n+1}.$$

Получившуюся систему сеточных уравнений решаем методом простой итерации, в соответствии с предложенным линеаризующим алгоритмом.

$$\frac{1}{\Delta t} (\varphi_i^{n+1,k+1} - \varphi_i^n) - \frac{1}{\Delta x^2} \left(\chi_{i+1/2}^{n+1,k} (\varphi_{i+1}^{n+1,k+1} - \varphi_i^{n+1,k+1}) - \chi_{i-1/2}^{n+1,k} (\varphi_i^{n+1,k+1} - \varphi_{i-1}^{n+1,k+1}) \right) + a_i^{n+1,k} \varphi_i^{n+1,k+1} = b_i^{n+1,k},$$

$$\chi_{i+1/2}^{n+1,k} = \frac{1}{2} [D(x_i, t, \varphi_i) + D(x_{i+1}, t, \varphi_{i+1})], \quad \chi_{i-1/2}^{n+1,k} = \frac{1}{2} [D(x_{i-1}, t, \varphi_{i-1}) + D(x_i, t, \varphi_i)],$$

$$\varphi_i^{n+1,k+1} + 2D^{n+1,k} \frac{\varphi_i^{n+1,k+1} - \varphi_{i-1}^{n+1,k+1}}{\Delta x} = 0, \quad \varphi_1^{n+1,k+1} + 2D^{n+1,k} \frac{\varphi_1^{n+1,k+1} - \varphi_0^{n+1,k+1}}{\Delta x} = 0.$$

Можно показать, что данная разностная схема является абсолютно устойчивой, консервативной, монотонной с порядком аппроксимации $O(\Delta t + \Delta x^2)$. На каждом шаге итерации с номером k строится трехдиагональная СЛАУ, которая решается методом прогонки.

4. Численные исследования

4.1. Оптически прозрачная среда

В области $\Pi = \{0 \leq x \leq 1\}$ (размеры приведены в сантиметрах) решается уравнение переноса с нулевыми граничными условиями, при условии отсутствия входящего потока. Начальное значение полного количества частиц в системе $\varphi_0 = 1 \left[Bm \cdot cm^{-2} \right]$. Коэффициент Эйнштейна для спонтанного испускания $A_{21} \sim 0.21 \times 10^8 \left[сек^{-1} \right]$. Энергия при приходе с одного уровня на другой задавалась следующим значением $h\nu_{12} = 3.2 \left[эВ \right]$. При таких значениях коэффициенты B_{12}, B_{21} пропорциональны $\sim 10^{13} \left[Дж^{-1} \cdot см^3 \cdot сек^{-2} \right]$. Значения коэффициентов C_{12}, C_{21} были взяты из [9,10]. В качестве результата рассматривался скалярный поток $\varphi_0 \left[Bm \cdot см^{-2} \right]$ частиц на характерные моменты времени.

Расчеты проводились на регулярной пространственной сетке с равномерным шагом по пространству $\Delta x = 0.01 \left[см \right]$. Итерационный процесс повторяется до условия достижения заданной точности ($\varepsilon_0 = 10^{-4}, \varepsilon_1 = 10^{-16}$) по критерию: $\left| \varphi^{k+1} - \varphi^k \right| \leq \varepsilon_0 \varphi^{k+1} + \varepsilon_1$, $\varepsilon_0, \varepsilon_1$ - заданные константы.

На рисунке 1 приведены профили плотности излучения в разные моменты времени.

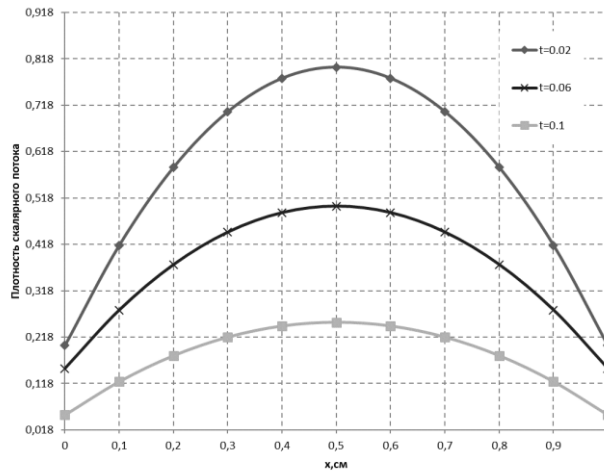


Рис.1. Распределение плотности скалярного потока

Для достижения относительной погрешности $\varepsilon_0 = 10^{-4}$ в момент времени $t=0.06$ потребовалось ~ 270 итераций, в момент времени $t=1.0$ потребовалось ~ 380 итераций.

4.2. Трехслойная среда

В области $\Pi = \{0 \leq x \leq 0.3\}$ (размеры приведены в сантиметрах) решается уравнение переноса с нулевыми граничными условиями третьего рода, при условии отсутствия входящего потока. Начальное значение полного количества частиц в системе $\varphi_0 = 0.01 \left[Bm \cdot cm^{-2} \right]$.

В задаче задавалось две области, различные по физическим свойствам. В таблице 1 приведены константы для каждой из областей.

Таблица 1. Параметры областей в задаче

Область	Коэффициент спонтанного испускания A_{21}	Коэффициенты B_{21}, B_{12}
1 $\Pi_1 = \{0 \leq x \leq 0.1\}$	$0.48 \times 10^7 \left[сек^{-1} \right]$	$10^{14} \left[Дж^{-1} \cdot см^3 \cdot сек^{-2} \right]$
2 $\Pi_2 = \{0.1 \leq x \leq 0.2\}$	$0.21 \times 10^8 \left[сек^{-1} \right]$	$10^{13} \left[Дж^{-1} \cdot см^3 \cdot сек^{-2} \right]$
1 $\Pi_3 = \{0.2 \leq x \leq 0.3\}$	$0.48 \times 10^7 \left[сек^{-1} \right]$	$10^{14} \left[Дж^{-1} \cdot см^3 \cdot сек^{-2} \right]$

Энергия при приходе с одного уровня на другой задавалась следующим значением $h\nu_{12} = 3.2 \left[эВ \right]$. При таких значениях коэффициенты. Значения коэффициентов C_{12}, C_{21} были взяты из [9,10]. В качестве результата рассматривался скалярный поток $\varphi_0 \left[Bm \cdot см^{-2} \right]$ частиц.

Расчеты проводились на регулярной пространственной сетке с равномерным шагом по пространству $\Delta x = 0.01 \left[см \right]$. Итерационный процесс повторяется до условия достижения заданной точности ($\varepsilon_0 = 10^{-4}, \varepsilon_1 = 10^{-16}$) по критерию: $\left| \varphi^{k+1} - \varphi^k \right| \leq \varepsilon_0 \varphi^{k+1} + \varepsilon_1$, $\varepsilon_0, \varepsilon_1$ – заданные константы.

На рисунке 2 приведен профиль плотности излучения в момент времени $t=0.01$.

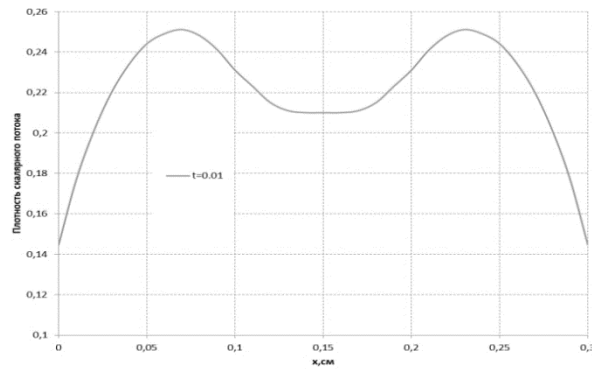


Рис.2. Распределение плотности скалярного потока

Для достижения относительной погрешности $\varepsilon_0 = 10^{-4}$ в момент времени $t=0.01$ потребовалось ~ 580 итераций

Таким образом, численные исследования позволяют сделать следующие вывод: итерационный процесс является сходящимся.

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. Владимиров В.С. Математические задачи односкоростной теории переноса частиц // Тр. Мат. Ин-та им. В.А. Стеклова АН СССР. 1961, вып 61. С. 2-158.
5. Сушкевич Т.А. Математические модели переноса излучения. М.: БИНОМ, 2006.
6. Гольдин В.Я. Квазидиффузионный метод решения кинетического уравнения. Журнал вычисл. Математики и матем.физики. 1964. №4. С.1078-1087.
7. Бусалов А.А. Нелинейная стационарная задача теории переноса в диффузионном приближении // Проблемы информатики. 2021, №2. С 6-14
8. Белл Д., Глесстон С. Теория ядерных реакторов. М., Атомиздат, 1974, 494с.
9. Ельяшевич М.А. Атомная и молекулярная спектроскопия. Эдиториал УРСС. 2001.
10. Аллен К.У. Астрофизические величины М.: Мир, 1977г.

ЗАДАЧА О РЮКЗАКЕ С ДВУМЯ НЕИЗВЕСТНЫМИ^{1*}

С.И. Веселов, А.С. Поднебеснов

Нижегородский государственный университет имени Н.И. Лобачевского

Описан полиномиальный алгоритм нахождения точного решения задачи о рюкзаке с двумя неизвестными, позволяющий прекратить вычисления, как только получено приближенное решение с наперед заданной точностью.

Ключевые слова: задача о рюкзаке, приближенный алгоритм, точность решения.

Здесь рассматривается задача о рюкзаке с двумя неизвестными. Задача о рюкзаке является известной NP-полной задачей дискретной оптимизации [1, 2, 3, 4, 5, 7, 11, 12, 13, 14]. Сведения о структуре множества задачи о рюкзаке приведены в работах [6, 8, 9, 10]. Она сводится к максимизации линейной функции на множестве M целочисленных решений системы неравенств

$$\begin{cases} \alpha x_1 + \beta x_2 \leq \gamma, \\ -x_2 \leq 0. \end{cases}$$

Основная идея решения данной задачи в том, что необходимо выбрать вершину выпуклой оболочки множества M , а затем перемещаться по ребрам выпуклой оболочки до тех пор, пока не найдется оптимальная точка или точка, являющаяся приближенным решением с заданной точностью.

Полагаем $l^T = (-\beta, \alpha)$, $a^T = (\alpha, \beta)$. Далее, указанным ниже способом, построим состоящую из фрагментов последовательность векторов $H = \{h_1, h_2, \dots\}$. Величину $u_i = a^T h_i$ назовем *весом* вектора h_i .

Первый фрагмент содержит только $h_1^T = (0, 1)$.

Полагаем $h_2^T = (-1, \lceil \alpha/\beta \rceil)$, $r = \lceil u_1/u_2 \rceil$, $h_3 = rh_2 - h_1$, $g = h_3 - h_2$, $s = \lfloor u_2/(u_2 - u_3) \rfloor$.

Векторы второго фрагмента вычисляются по формуле $h_i = h_2 + (i - 2)g$ ($i = 2, \dots, s + 1$).

Если $l \neq h_2 + sg$, то следующий фрагмент, начинающийся с вектора $h_{s+2} = h_2 + sg$, находим аналогично, выбрав вместо h_1 , h_2 соответственно h_{s+1} , h_{s+2} .

Любая пара соседних векторов является базисом решетки $\mathbb{Z}^2$ и нетрудно показать, что направляющие векторы граней для P_l содержатся в H . Первой вершиной является точка $T_1 = (\lfloor \gamma/\alpha \rfloor, 0)$. Вычисляем в T_1 величину невязки $\gamma_1 = \gamma - \alpha \lfloor \gamma/\alpha \rfloor$. Пусть T_k – последняя найденная вершина и γ_k величина невязки. Если найдется наименьший номер s и наибольшее целое число $d \geq 1$ такие, что

$$du_s \leq \gamma_k, \quad (1)$$

то полагаем $T_{k+1} = T_k + dh_s$, $\gamma_{k+1} = \gamma_k - du_s$, удаляем из H векторы h_i ($i \leq s$) и продолжаем вычисления до тех пор, пока не произойдет одно из событий:

1. Значение целевой функции в точке T_k больше значения в точке T_{k+1} . В этом T_k – точное решение.
2. $k = \lceil -\log \epsilon \rceil$ В этом случае T_{k+1} – приближенное решение.

Литература

1. Веселов С.И., Чирков А.Ю., “Оценки числа вершин целых полиэдров”, Дискретн. анализ и исслед. опер., сер. 2, 14:2 (2007), С. 14–31; J. Appl. Industr. Math., 2:4 (2008), P. 591–604. DOI: 10.1134/S1990478908040157.
2. Схрейвер А. Теория линейного и целочисленного программирования. Т. 2. М.: Мир, 1991. 360 с.
3. Чирков А.Ю., Грибанов Д.В. О близости оптимальных значений целочисленной задачи о многомерном ранце и целочисленной задачи о многомерном ранце с ограничением на мощность. Математическое моделирование и суперкомпьютерные технологии

^{1*} Работа выполнена при поддержке гранта Российского Научного Фонда (РНФ) №21-11-00194

- (ММиСТ'2021): Труды XXI международной конференции (Нижний Новгород, 22–26 ноября 2021 г.). Нижний Новгород: Издательство Нижегородского госуниверситета, 2021. С. 401-402.
4. Чирков А.Ю., Шевченко В.Н. О приближении оптимального решения целочисленной задачи о ранце оптимальными решениями целочисленной задачи о ранце с ограничениями на мощность // Дискретный анализ и исследование операций. 2006. Серия 2. Т13 №2. Р. 56-73.
 5. Шевченко В. Н. Качественные вопросы целочисленного программирования. М.: Физматлит, 1995. 190 с.
 6. Barany I., Howe R., Lovasz L. On integer points in polyhedra: a lower bound // Combinatorica. 1992. Vol. 12, № 2. P. 135-142.
 7. Chirkov A.Y., Griбанov D.V., Zolotykh N.Y. On the Proximity of the Optimal Values of the Multi-dimensional Knapsack Problem with and Without the Cardinality Constraint. In: Kochetov Y., Bykadorov I., Gruzdeva T. (eds) Mathematical Optimization Theory and Operations Research. MOTOR 2020. Communications in Computer and Information Science. 2020. Vol. 1275, P. 16-22, Springer, Cham. DOI: 10.1007/978-3-030-58657-7_2.
 8. Cook W., Hartmann M., Kannan R., and McDiarmid C. On integer points in polyhedra // Combinatorica. 1992. Vol. 12, No. 1. P. 27-37.
 9. Gomory R. E. Faces of integer polyhedron // Proc. Nat. Acad. Sci. USA. 1967. Vol. 57, No. 1. P. 16-18.
 10. Gomory R. E. Some polyhedra related to combinatorial problems // Linear Algebra and Its Applications. 1969. Vol. 2, No. 4. P. 451-558.
 11. Kohli R., Krishnamurti R. A total-value greedy heuristic for the integer knapsack problem // Operations Research Letters. 1992. Vol. 12, P. 65-71.
 12. Kohli R., Krishnamurti R. Joint performance of greedy heuristics for the integer knapsack problem // Discrete Applied Mathematics. 1995. Vol. 56, P. 37-48.
 13. Morgan D. A. Upper and lower bound results on the convex hull of integer points in polyhedra // Mathematika. 1991. Vol. 38. P. 321-328.
 14. Kannan R. A Polynomial Algorithm for the Two-Variable Integer Programming Problem // Journal of the ACM. 1980. Vol. 27, Issue 1. P. 118-122. DOI: 10.1145/322169.322179.

ПРИМЕНЕНИЕ АЛГОРИТМОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ПРОГНОЗИРОВАНИЯ ТИПА СОЦИАЛЬНОГО ЛИФТА^{1*}

А.В. Голубева, А.Л. Сочков

Нижегородский государственный университет им. Н.И. Лобачевского

Целью работы является создание программного обеспечения для прогнозирования типа социального лифта. Эта задача решается на базе алгоритмов машинного обучения. В качестве массива исходных данных используются оцифрованные факты биографий известных исторических личностей. Наиболее высокую точность при обучении продемонстрировала модель LightGBM (градиентный бустинг), которая и была положена в основу приложения. Был разработан также графический интерфейс пользователя. Созданное программное обеспечение на языке Python может быть использовано кадровыми службами для профориентации молодежи и студенчества.

Ключевые слова: машинное обучение, задача прогнозирования, социальный лифт.

1. Введение

Принятая в 2019 году Национальная стратегия развития искусственного интеллекта в РФ способствовала интенсификации работ по разработке и внедрению такого рода систем в различные сферы деятельности. Если примеры использования технологий искусственных нейронных сетей (ИНС) в инженерных задачах известны достаточно давно [1], то их применение для решения социально-экономических задач значительно активизируется лишь в последние годы. В качестве примера использования нейронных сетей для анализа развития человеческого капитала субъектов РФ можно привести статью [2]. В работе [3] разработана нейросетевая модель для прогнозирования валового регионального продукта Ставропольского края, а в статье [4] синтезирована ИНС для предсказания показателей социально-экономического развития Воронежской области.

Технологии искусственного интеллекта и машинного обучения позволяют также автоматизировать решение некоторых задач в сфере управления персоналом. Так, в статье [5] предпринята попытка создания ИНС для профориентации молодых сотрудников. В работе [6] разработана нейронная сеть, моделирующая скорость успеха одаренной личности. Возможность применения карт Кохонена в рассматриваемой области изучена в [7], а вопросы автоматизации тестирования исследованы в [8].

Целью данной работы является создание программного обеспечения на базе технологий машинного обучения, которое прогнозировало бы тип социального лифта для молодежи и студенчества, позволяющего им эффективно самореализовываться в процессе профессионального становления. Для достижения поставленной цели необходимо решить следующие задачи: подготовить массив исходных данных; осуществить машинное обучение модели; разработать пользовательский интерфейс приложения. Программирование велось на языке Python.

2. Разработка приложения для прогнозирования социального лифта

Создание программного обеспечения проводилось в три этапа. На первом этапе был подготовлен массив исходных данных. На втором этапе были обучены несколько моделей, и, в итоге, по определенным критериям была выбрана лучшая. На третьем этапе был разработан графический интерфейс пользователя данного приложения, позволяющий вести комфортный диалог с программой. Далее каждый этап рассмотрен подробнее.

^{1*} Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 19-29-07462.

2.1. Подготовка данных

Исходный датасет для обучения модели представляет собой набор признаков для каждой исторической личности, которых было 247 человек (Год рождения, Наличие диагноза при рождении, Ранняя одаренность, Место рождения, Уровень образования, Материальное состояние родителей, Наличие разнородных интересов, Вовлечение в общественную работу, Наличие контекста, Род деятельности, Лифт). В качестве вариантов социального лифта рассматривались «предпринимательство», «госслужба», «академическое сообщество» и вариант «без лифта».

В начале были импортированы несколько модулей библиотеки `sklearn`. Далее, была закодирована целевая переменная «Лифт» при использовании значений от 0 до 3, в зависимости от определенного класса по каждой строке.

Для работы с данными необходимо все категориальные признаки привести в числовой формат, и все числовые признаки нормализовать от 0 до 1. С использованием `one-hot-encoding` библиотеки `pandas` необходимо привести все категориальные признаки в формат 0 и 1.

С помощью кода, приведенного на рис.1, определим принадлежность каждого признака к категориальному типу данных (`object`) и к числовому типу (`int`, `float`).

```
df_chislo = pd.DataFrame()
df_catego = pd.DataFrame()
for col in df:
    type_col = df[col].dtype
    if type_col == 'object':
        print('категориальные', col)
        df_catego = pd.concat([df_catego, df[col]], axis = 1)
    else:
        print('числовые', col)
        df_chislo = pd.concat([df_chislo, df[col]], axis = 1)
```

Рис. 1. Отбор категориальных и номинальных признаков

С помощью кода, приведенного на рис.2, произведем нормализацию и быстрое кодирование признаков.

```
i_data = df_chislo.copy()
df_id = pd.DataFrame(df_catego[0])
del df_catego[0]
object= StandardScaler()
scale = object.fit_transform(i_data)
scalar_chislo = pd.DataFrame(scale, columns=df_chislo.columns[:])
scalar_chislo.head()

df_encoding_all = pd.DataFrame()
for col in df_catego:
    print(col)
    #type_col = df_catego[col].dtype
    df_encod = pd.get_dummies(df[col])
    df_encoding_all = pd.concat([df_encoding_all, df_encod], axis=1)
```

Рис. 2. Нормализация номинальных и кодирование категориальных признаков

После того, как были получены 3 датафрейма (ФИО, числовые переменные, категориальные переменные), их необходимо объединить в одну таблицу.

Следующим шагом необходимо разделить датафрейм на признаки (X) и целевую переменную (y).

Далее, с помощью `train_test_split` разделим датафрейм на тренировочную и тестовую выборки в соотношении 80 к 20 с параметром `random_state` равным 21. Если не указывать `random_state` в коде, то каждый раз, когда он запускается, генерируется новое случайное значение, а наборы данных обучения и тестов будут иметь разные значения. Число `random_state` случайным образом разделяет наборы тестовых и обучающих данных.

Подготовка данных и разбиение датасета завершены, теперь можно приступить к выбору модели машинного обучения.

2.2. Выбор модели машинного обучения

Модели обучения, которые были протестированы на данных: LogisticRegression – логистическая регрессия; LinearDiscriminantAnalysis – линейный дискриминантный анализ; KNeighborsClassifier – алгоритм ближайшего соседа; DecisionTreeClassifier – деревья решений; GaussianNB – наивный байес; SVC – метод опорных векторов; lgb.LGBMClassifier – градиентный бустинг; XGBClassifier – ансамбль методов деревьев; RandomForestClassifier – случайный лес.

Для того, чтобы оценить точность модели используем самый простой способ - перекрестную проверку (cross_val_score). На диаграмме размаха, приведенной на рис.3, можно увидеть, какая модель на первом проходе отработала лучше всего – это RFC, точность модели на тестовой выборке составила 70%. Такая точность модели недостаточна для того, чтобы сделать прогноз социального лифта, поэтому необходимо протестировать все модели с различными значениями гиперпараметров.

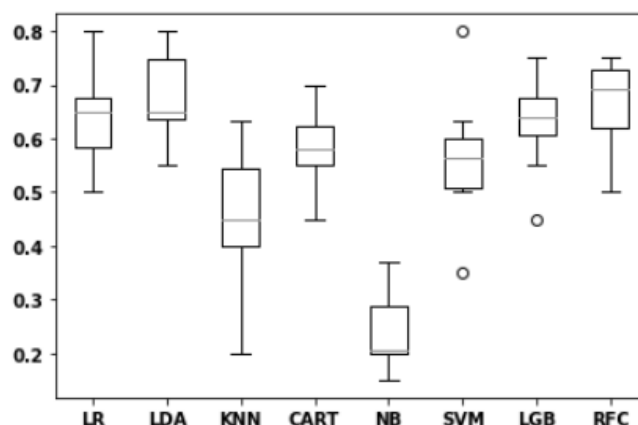


Рис. 3. Диаграмма размаха по каждой из модели обучения

Каждую модель проверим на трех различных наборах гиперпараметров, чтобы выбрать лучшую. В результате лучшие показатели продемонстрировала модель – LightGBM (градиентный бустинг), которая на обучающей выборке показала точность 94%, а на тестовой выборке - 80%.

После того, как модель была обучена, она была сохранена в файл формата pkl с использованием библиотеки pickle. Это необходимо для того, чтобы не производить все вышеизложенные манипуляции с данными, а сохранить созданную модель и применять для новых данных. Далее переходим к разработке графического пользовательского интерфейса приложения.

2.3. Графический интерфейс пользователя

Для удобства работы с моделью создадим графический пользовательский интерфейс (GUI) десктопного приложения с использованием библиотеки PyQt5. Она написана в объектно-ориентированном стиле и содержит несколько сотен классов. Иерархия наследования всех классов имеет очень большой размер. Тем не менее, чтобы показать зависимости, при описании компонентов иерархия наследования конкретного класса будет показываться.

Для создания пользовательского интерфейса с помощью программы Qt был разработан прототип визуального представления программы, а именно – расположение блоков, количество вкладок, размещение кнопок на вкладках, виджет индикатора выполнения модели.

Далее, прототип визуальной части программы был сохранен в файл формата .ui, и с помощью команды `ruic5 prototip.ui > main.py` был создан файл формата .py, позволяющий конвертировать визуальную часть из программы Qt в написанный скрипт на языке Python, чтобы до-

полнить визуальную часть классами и функциями обработки действий каждой из кнопок/вкладок и т.д.

Первая вкладка предусматривает загрузку данных из «Проводника». Кнопка «Запустить модель» подключает скрипт, обрабатывающий модель, описанную в п.2.2, и сохраняет результат в файл формата *xlsx* с предсказанным значением лифта для каждой строки входного файла. Вторая вкладка, изображенная на рис.4, представляет собой интерфейс ввода данных по одному человеку, для которого хотим предсказать социальный лифт.

Загрузить из excel файла Выбрать из списка

ФИО
Введите Фамилия[пробел]Имя[пробел]Отчество

Год рождения
Введите последовательность 4-х цифр

Наличие в детском возрасте огр-ых возм-ей здоровья 0/1

Раннее проявление одаренности 0/1

Место рождения (город/деревня)
город|деревня

Уровень образования
высшее|школа|поствысшее|нвысшее|среднее

Финансовое благополучие родительской семьи 0/1

Наличие разносторонних интересов 0/1

Вовлеченность в общественную работу 0/1

Соотв-ие сферы деят-ти актуальным вызовам времени 0/1

Род деятельности
|Менеджер|Военный|Искусство|Гуманитарий|Математик

0%

Запустить модель Закрыть

Рис. 4. Графический интерфейс для ввода данных одного человека

Результат работы приложения появляется в всплывающем окне на фоне описанного интерфейса.

Скрипт для работы с GUI включает в себя 3 класса: `class Thread (QThread)` – первый класс, который отвечает за асинхронное параллельное программирование; `class Ui_MainWindow (object)` – основной класс, в котором происходит отработка скрипта; `class MainWindow (QtWidgets.QMainWindow, Ui_MainWindow)` – главный класс, отвечающий за воспроизведение класса `Ui_MainWindow`.

Разработанное программное обеспечение позволяет прогнозировать и рекомендовать тип социального лифта, подходящий для конкретного человека. Приложение может быть использовано кадровыми службами для профессиональной ориентации молодежи и студенчества.

Литература

1. Сочков А.Л., Калин С.А. Использование технологии нейронных сетей для решения электротехнических задач. В кн.: Актуальные проблемы электроэнергетики: Тезисы докладов, Нижний Новгород, 01–31 декабря 1999 года. – Нижний Новгород: Нижегородский госу-

- дарственный технический университет им. Р.Е. Алексеева, 1999. С. 59-62. – EDN NNEDED.
2. Трифонов Ю.В., Сочков А.Л., Миронов Е.А. Типология российских регионов с точки зрения развития человеческого капитала на базе нейросетевого кластерного анализа // Вестник Нижегородского университета им. Н.И. Лобачевского. Серия: Социальные науки. 2021. № 4(64). С. 23-34. DOI 10.52452/18115942_2021_4_23. – EDN PRFXZA.
 3. Сочков А.Л., Соловьев А.Е. Прогнозирование экономического развития Ставропольского края на базе нейросетевой модели. В кн.: Актуальные проблемы управления: сборник научных статей по итогам VIII Всероссийской научно-практической конференции, Нижний Новгород, 16 ноября 2021 года / Национальный исследовательский Нижегородский государственный университет им. Н.И. Лобачевского. – Нижний Новгород: Национальный исследовательский Нижегородский государственный университет им. Н.И. Лобачевского, 2022. С. 386-391. – EDN TJZNTC.
 4. Азарнова Т.В., Трещевский Ю.И., Папин С.Н. Прогнозирование параметров социально-экономического развития региона с использованием аппарата нейронных сетей (на примере ВРП Воронежской области) // Современная экономика: проблемы и решения. 2020. № 3(123). С. 8-25. DOI 10.17308/merp.2020.3/2321. – EDN PKXNNJ.
 5. Азарнова Т.В., Степин В.В., Щепина И.Н. Повышение эффективности методов управления развитием персонала на основе нейросетевых моделей и нечетких экспертных технологий // Вестник Воронежского государственного университета. Серия: Экономика и управление. 2014. № 3. С. 121-130. – EDN TBPWSJ.
 6. Климов В.А., Сочков А.Л. Нейросетевое моделирование скорости успеха одаренной личности. В кн.: Современные исследования проблем управления кадровыми ресурсами: сборник научных статей VII Международной научно-практической конференции, Москва, 29–31 марта 2022 года. – Москва: Общество с ограниченной ответственностью «Эдельвейс», 2022. С. 68-73. – EDN BREWZS.
 7. Кузнецова М.А., Сочков А.Л. Использование карт Кохонена для определения сферы деятельности талантливого сотрудника. В кн.: Современные исследования проблем управления кадровыми ресурсами: Сборник научных статей VI Международной научно-практической конференции, Москва, 06–08 апреля 2021 года. – Москва: Общество с ограниченной ответственностью «Эдельвейс», 2021. С. 157-165. – EDN JBULTY.
 8. Семенова Е.А., Сочков А.Л. Применение многослойного персептрона для оценки профессионально-психологических качеств одаренной личности. В кн.: Современные исследования проблем управления кадровыми ресурсами: Сборник научных статей VI Международной научно-практической конференции, Москва, 06–08 апреля 2021 года. – Москва: Общество с ограниченной ответственностью «Эдельвейс», 2021. С. 150-157. – EDN IKNHDV.

О БИФУРКАЦИЯХ ДВУМЕРНЫХ ДИФФЕОМОРФИЗМОВ С ГОМОКЛИНИЧЕСКИМИ ТРАЕКТОРИЯМИ К НЕГРУБЫМ НЕПОДВИЖНЫМ ТОЧКАМ^{1*}

С.В. Гонченко, О.В. Гордеева

Нижегородский государственный университет им. Н.И. Лобачевского

В работе изучаются динамические свойства двумерных диффеоморфизмов с гомоклиническими траекториями к негрубым неподвижным точкам, а также их глобальные бифуркации в параметрических семействах общего положения.

Ключевые слова: трансверсальное пересечение, касание многообразий, вырожденная неподвижная точка, бифуркационная диаграмма, гиперболическое множество, топологическая схема Бернулли.

1. Введение

Работа посвящена исследованию динамических свойств и бифуркаций двумерных диффеоморфизмов с гомоклиническими траекториями к негрубым неподвижным точкам типа седло-узел и сложное седло. В случае грубого (гиперболического) седла окрестность его трансверсальной гомоклинической траектории была изучена в работе Л.П. Шильникова [1], а бифуркации гомоклинических касаний были исследованы в работе Н.К. Гаврилова и Л.П. Шильникова [2]. Структура траекторий в окрестности трансверсальной гомоклинической траектории к невырожденному седло-узлу, а также их бифуркации в однопараметрических семействах общего положения были исследованы в работе В.И. Лукьянова и Л.П. Шильникова [3]. Таким образом, настоящую работу можно рассматривать как логическое продолжение этих исследований на случаи большей коразмерности.

В параграфе 2 рассматривается случай вырожденной негрубой неподвижной точки с трансверсальной гомоклинической траекторией. В параграфе 3 изучаются бифуркации квадратичного гомоклинического касания к вырожденной негрубой неподвижной точке в однопараметрических семействах, в которых эта точка не бифурцирует. В параграфах 3 и 4 изучаются бифуркации в двухпараметрических семействах общего положения соответственно в следующих случаях, когда исходный диффеоморфизм имеет (а) квадратичное гомоклиническое касание к неподвижной точке типа невырожденный седло-узел; (б) трансверсальную гомоклиническую траекторию к неподвижной точке типа невырожденное сложное седло.

2. О структуре окрестности трансверсальной гомоклинической траектории к негрубой неподвижной точке

Рассмотрим двумерный диффеоморфизм f , удовлетворяющий условиям А) и В) ниже:.

А) f имеет $(n - 1)$ -кратно вырожденную, $n \geq 2$, неподвижную точку O с мультипликаторами λ_1, λ_2 , где $0 < \lambda_1 < 1$, и $\lambda_2 = 1$, которая является точкой либо типа седло-узел, если n - четное, либо типа сложное седло, если n - нечетное.

Как показано в [4], в достаточно малой окрестности $U_0(O)$ отображение f может быть записано в виде $\bar{x} = \lambda x + f(x, y)x, \bar{y} = y + y^n + g(y)$, где $f(0, 0) = 0, g(y) = O(y^{n+1})$. Тогда у f существуют локальные сильно устойчивое и центрально неустойчивое многообразия

^{1*} Работа выполнена при финансовой поддержке гранта 0729-2020-0036 министерства науки и высшего образования РФ.

вида $y = 0$ и $x = 0$ (в случае седло-узла последнее является многообразием с краем вида $\{x = 0, y > 0\}$). Эти многообразия продолжаются до глобальных, и мы предполагаем, что

В) Центральное неустойчивое $W^{cu}(O)$ и сильноустойчивое $W^{ss}(O)$ многообразия, точки O пересекаются трансверсально в точках некоторой гомоклинической траектории Γ_0 .

Обозначим через N множество траекторий, целиком лежащих в достаточно малой окрестности $U(O \cup \Gamma_0)$, см. рисунок 1. Пусть B_2 – топологическая схема Бернулли из двух символов «0» и «1». Рассмотрим ее подсистему $\Omega_{\bar{k}}$, выделяемую следующими условиями:

- 1) $\Omega_{\bar{k}}$ не содержит последовательностей, в которых есть два соседних символа «1»;
- 2) у последовательностей из $\Omega_{\bar{k}}$ длина любого отрезка из символов «0», не меньше $\bar{k}$

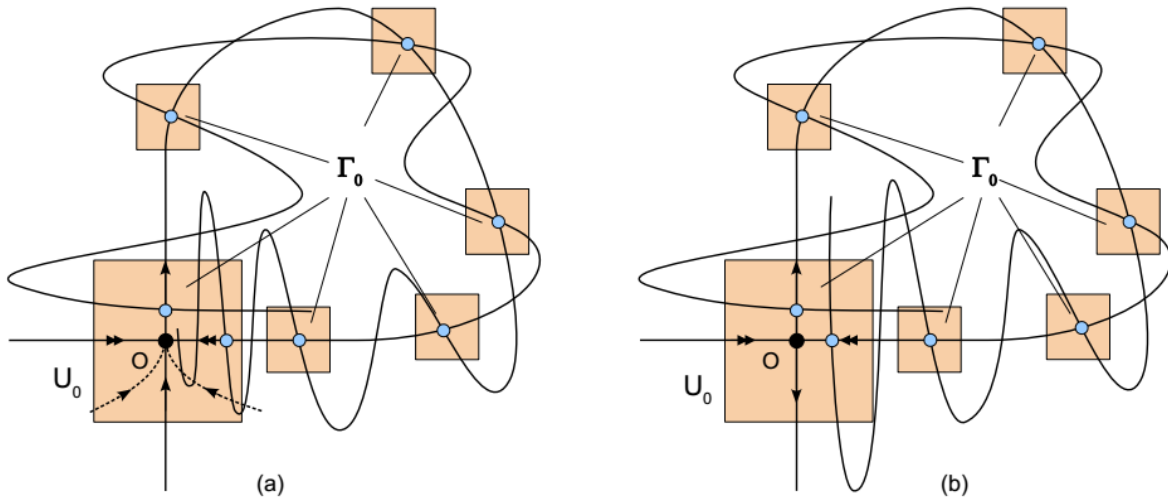


Рис. 1. Окрестность трансверсальной гомоклинической траектории Γ_0 к неподвижной точке типа (a) седло-узел; (b) сложное седло.

Теорема 1. Пусть двумерный диффеоморфизм f удовлетворяет условиям А) и В), тогда система $f|_N$ топологически сопряжена с $\Omega_{\bar{k}}$.

Доказательство теоремы проводится с помощью метода седловых отображений, предложенного Л.П. Шильниковым в [1], а также модификации этого метода, развитой в работах Лукьянова и Шильникова [3,4], основанной на построении инвариантных устойчивого и неустойчивого многообразий траекторий из множества N . Случай $n = 2$ теоремы 1 (точка O является невырожденным седло-узлом) был рассмотрен в работе [3]

3. О бифуркациях квадратичных гомоклинических касаний к негрубой неподвижной точке

Рассмотрим диффеоморфизм f_0 , имеющий квадратичное гомоклиническое касание многообразий $W^{cu}(O)$ и $W^{ss}(O)$ точки O , вырожденной в силу условий А). Вложим f_0 в однопараметрическое семейство f_μ , которое при изменении параметра μ расщепляет исходное гомоклиническое касание общим образом, но сохраняет тип вырождения неподвижной точки O .

Следующая теорема описывает основные бифуркации однообходных периодических траекторий в семействе f_μ , т.е. таких траекторий, которые имеют периоды k , где $k = \bar{k}, \bar{k} + 1, \dots$, и за период обходят окрестность $U(O \cup \Gamma_0)$, ровно один раз.

Теорема 2. В любой окрестности точки $\mu = 0$ существует счетное множество непересекающихся интервалов δ_k значений параметра μ таких, что при $\mu \in \delta_k$ диффеоморфизм f_μ имеет асимптотически устойчивую однообходную периодическую траекторию. Границами интервалов δ_k являются точки ℓ_k^+ и ℓ_k^- , отвечающие бифуркациям коразмерности один, седло-узловой и удвоения периода соответственно. При $k \rightarrow +\infty$ интервалы δ_k накапливаются к точке $\mu = 0$.

Замечание. Основным моментом в доказательстве теоремы 2 является рескейлинг метод, с помощью которого показывается, что отображение первого возвращения $T_k = f_\mu^k$ приводится к отображению, асимптотически близкому к одномерному отображению параболы $\bar{Y} = M - Y^2$, где $M \sim (\mu - \alpha_k)k^{\frac{2n}{n-1}}$ и $\alpha_k \sim k^{\frac{1}{n-1}}$. Подставляя границы $M = -\frac{1}{4}$ и $M = \frac{3}{4}$, отвечающие бифуркациям, седло-узловой и удвоения периода соответственно, в отображении параболы, получим, что интервалы δ_k имеют длины $l(\delta_k) \sim k^{-(2+2/n-1)}$, уменьшающиеся по степенному закону при $k \rightarrow +\infty$. Это дает такие их порядки: k^{-4} при $n = 2$ (невырожденный седло-узел), k^{-3} при $n = 3$ (невырожденное сложное седло), $k^{-8/3}$ при $n = 4$ и т.д. Заметим, что в случае квадратичного гомоклинического касания к гиперболическому седлу [2], длины соответствующих интервалов δ_k изменяются по экспоненциальному закону: $l(\delta_k) \sim \gamma^{-2k}$, где $\gamma > 1$ - неустойчивый мультипликатор седла.

4. О бифуркациях двумерных диффеоморфизмов с квадратичным гомоклиническим касанием к невырожденному седло-узлу

Рассмотрим C^r -гладкий диффеоморфизм f_0 ($r \geq 4$), удовлетворяющий условиям:

С) f_0 имеет неподвижную точку O типа невырожденный седло-узел: он имеет мультипликаторы λ_1, λ_2 , где $0 < \lambda_1 < 1$ и $\lambda_2 = 1$, и первую ляпуновскую величину l_1 не равную нулю;

Д) Инвариантные многообразия $W^{cu}(O)$ и $W^{ss}(O)$ имеют квадратичное касание в точках некоторой гомоклинической траектории Γ_0 .

Вложим f_0 в двухпараметрическое семейство f_μ , $\mu = (\mu_1, \mu_2)$, общего положения, где μ_1 - параметр расщепления гомоклинического касания, а μ_2 управляет бифуркациями неподвижной точки.

Теорема 3. В окрестности начала координат $\mu = 0$ на плоскости параметров (μ_1, μ_2) существуют две бифуркационные кривые $L_+ : \mu_2 = 0$ и $L_h : \mu_1 = \sqrt{-\mu_2} + O(|\mu_2|)$. При $\mu \in L_+$ диффеоморфизм f_μ имеет неподвижную точку типа седло-узел, близкую к O , которая при $\mu_2 < 0$ распадается на две неподвижные точки, седло O_1 и устойчивую O_2 , а при $\mu_2 > 0$ исчезает. При $\mu \in L_h$ у f_μ существует негрубая гомоклиническая траектория Γ_μ , близкая к Γ_0 , в точках которой квадратично касаются инвариантные многообразия $W^u(O_1)$ и $W^s(O_1)$. Кривые L_+ и L_h разбивают окрестность начала координат на три области:

В области **I** ($\mu_2 > 0$) диффеоморфизм f_μ не имеет неподвижных точек. В областях **II** и **III** у f_μ существуют две неподвижные точки, седло O_1 и узел O_2 . При $\mu \in \text{II}$ у O_1 нет гомоклинических траекторий, близких к Γ_μ , а при $\mu \in \text{III}$ у O_1 появляются две грубые гомоклинические траектории, близкие к Γ_μ .

Бифуркационная диаграмма к теореме 3 представлена на рисунке 2.

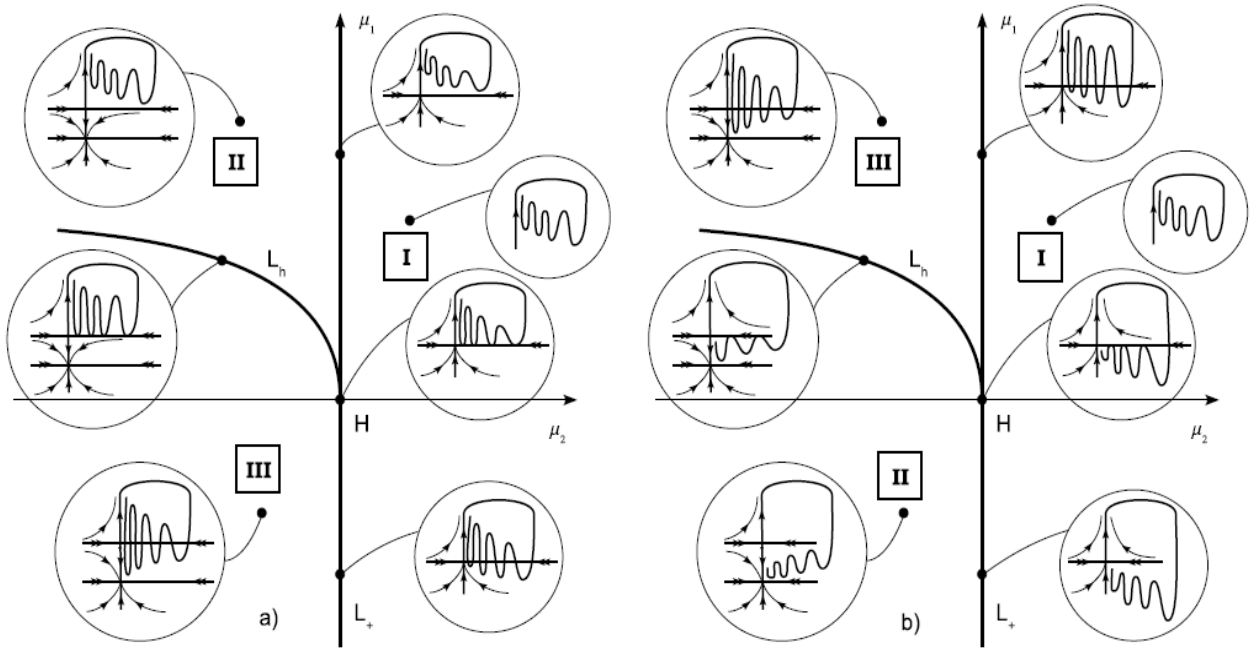


Рис. 2. Бифуркационная диаграмма к Теореме 3 в случаях касания «сверху»; б) касания «снизу»

Теорема 4. На плоскости параметров (μ_1, μ_2) в любой достаточно малой окрестности начала координат существует счетное множество непересекающихся областей Δ_k таких, что при $\mu \in \Delta_k$ диффеоморфизм f_μ имеет асимптотически устойчивую однообходную периодическую траекторию. Границами областей Δ_k являются бифуркационные кривые L_k^+ и L_k^- , отвечающие бифуркациям коразмерности один, седло-узловой и удвоения периода соответственно. При $k \rightarrow +\infty$ области Δ_k накапливаются к кривой $L_h \cup \{L_+ \cap \{\mu_1 < 0\}\}$.

Рисунок 3 представляет собой некоторую иллюстрацию к теореме 4. Здесь показано также изменение геометрии отображения первого возвращения T_k при варьировании параметров (μ_1, μ_2) внутри области Δ_k . Здесь также условно показаны две области: область RD, где доказательство теоремы 4 проводится с помощью рескелинг-метода (так же как и теоремы 2), и область Δ_{LS} , где этот метод не работает из-за специфики геометрии отображений первого отображения, отличной от геометрии отображений типа отображения Эно. В области Δ_{LS} проводятся доказательства, основанные на вычислениях отображений первого возвращения в исходных координатах с использованием методов работ [3,4].

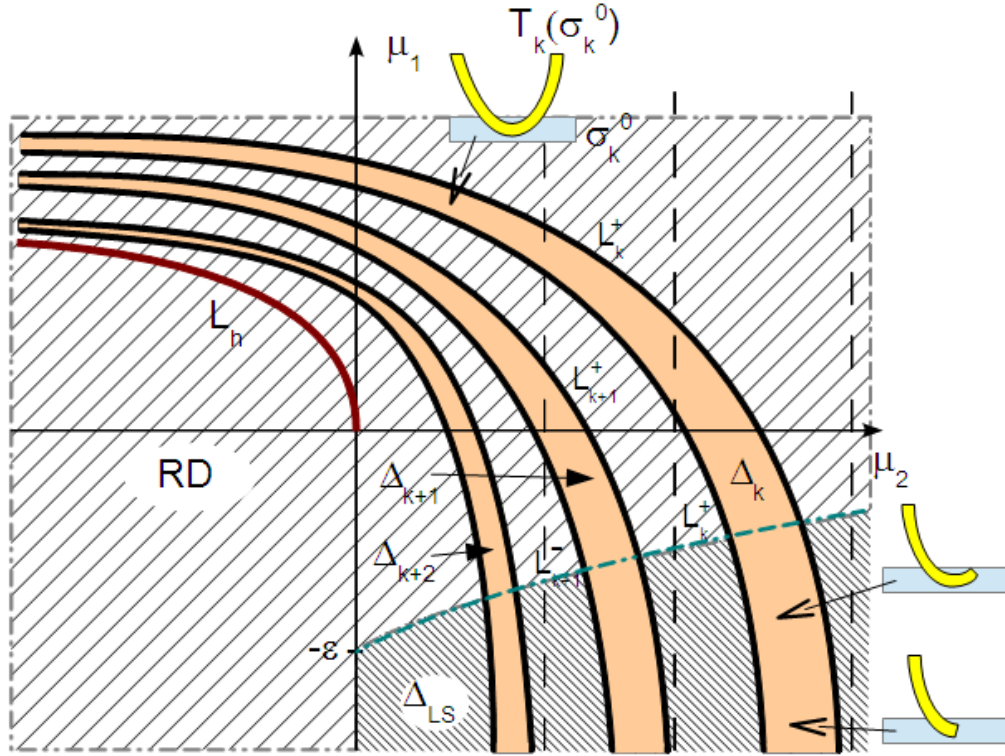


Рис. 3. Бифуркационная диаграмма в случае касания «сверху». В случае касания «снизу» она будет иметь такой же вид, с той лишь разницей, что теперь кривые L_k^+ и L_k^- поменяются местами (L_k^+ будет нижней, а L_k^- верхней границами области Δ_k)

5. О бифуркациях множества траекторий в окрестности трансверсальной гомоклинической орбиты к сложному седлу

Рассмотрим C^r – гладкий диффеоморфизм f_0 ($r \geq 4$), удовлетворяющий условиям:

Е) f_0 имеет неподвижную точку O типа невырожденное сложное седло с мультипликаторами λ_1, λ_2 , где $0 < \lambda_1 < 1$ и $\lambda_2 = 1$, и ляпуновскими величинами $l_1 = 0$ и $l_2 > 0$.

Ф) Инвариантные многообразия $W^{cu}(O)$ и $W^{ss}(O)$ пересекаются трансверсально в точках некоторой гомоклинической траектории Γ_0 .

Вложим f_0 в двухпараметрическое семейство f_μ общего положения, где $\mu = (\mu_1, \mu_2)$ – параметры, управляющие бифуркациями сложного седла.

Пусть U – достаточно малая фиксированная окрестность контура $O \cup \Gamma_0$. Обозначим через B_2 топологическую схему Бернулли из двух символов, а через N_μ множество траекторий диффеоморфизма f_μ , целиком лежащих в U .

Стандартно доказывается, что на плоскости параметров (μ_1, μ_2) существуют две бифуркационные кривые L_1 и L_2 , отвечающие седло-узловым бифуркациям неподвижной точки (см. рисунок 4), которые разбивают эту плоскость на две области E_1 и E_2 такие, что в E_1 существует только одна грубая седловая неподвижная точка O , а в области E_2 – три неподвижные точки: седло O_1 , устойчивая O_2 и седло O_3 ; на кривой L_1 точки O_1 и O_2 сливаются в седло-узел O_{12} , на

кривой L_2 точки O_2 и O_3 сливаются в седло-узел O_{23} . Основным результатом главы 4 составляет следующая

Теорема 5. Множество N_μ всегда содержит нетривиальное подмножество Ω_μ , траектории которого находятся во взаимно-однозначном соответствии с траекториями B_2 , и кроме того N_μ допускает полное описание следующего типа:

(i) $N_\mu = \Omega_\mu$ при $(\mu_1, \mu_2) \in E_1$; $N_\mu = \Omega_\mu + O_2 + O_3$ при $(\mu_1, \mu_2) \in E_2$, и здесь для любого такого μ множество Ω_μ является равномерно гиперболическим.

(ii) При $(\mu_1, \mu_2) \in L_1$ множество Ω_μ содержит седло-узел O_{12} , т.е. $N_\mu = \Omega_\mu + O_{12}$ («скачок гиперболичности»), и является неравномерно гиперболическим; при $(\mu_1, \mu_2) \in L_2$ множество Ω_μ равномерно гиперболично и $N_\mu = \Omega_\mu + O_{23}$.

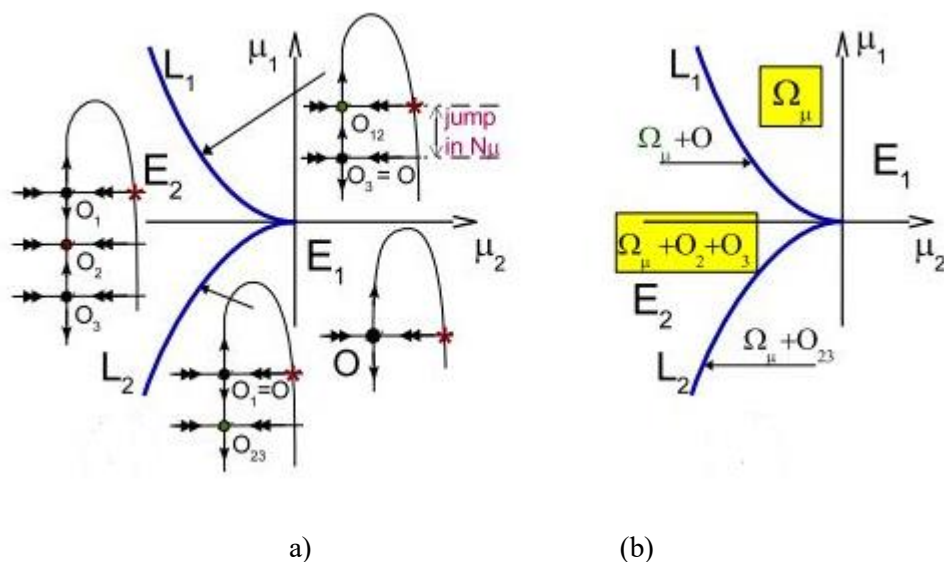


Рис. 4. Бифуркационная диаграмма к теореме 5

Литература

1. Шильников Л.П. Об одной задаче Пуанкаре –Биркгофа // Матем. сб. 1967. – т.74(116). – стр.378-397.
2. Гаврилов Н. К., Шильников Л. П. О трехмерных динамических системах, близких к системам с негрубой гомоклинической кривой //Часть 1. – Матем. Сборник, 1972, т.88, №. 4. 475-492; Часть 2. – Матем. сборник, 1973, т.90, №. 1, 139-156.
3. Лукьянов В.И., Шильников Л.П О некоторых бифуркациях динамических систем с гомоклиническими структурами.//ДАН СССР . – 1978. – т.243. – №1. – с. 26-29.
4. В.И.Лукьянов. О существовании гладких инвариантных слоений в окрестности некоторых негрубых неподвижных точек диффеоморфизма. Межвуз. сб. Горький. – 1979. – с.60-66.

ПРИМЕНЕНИЕ ЗАКОНОВ СОХРАНЕНИЯ В НЕЙРОСЕТЕВОМ РЕШЕНИИ ОДНОМЕРНОГО НЕЛИНЕЙНОГО УРАВНЕНИЯ ШРЁДИНГЕРА^{1*}

Ю.В. Гурьева, Е.П. Васильев, Л.А. Смирнов

Нижегородский государственный университет им. Н.И. Лобачевского

В работе рассматривается модификация Physics-Informed нейронных сетей, основанная на одном из наиболее важных свойств интегрируемых систем – законах сохранения. Этот метод подразумевает введение консервативных свойств нелинейной физической системы в нейронную сеть для контроля физических свойств решения в процессе обучения. Консервативное ограничение обеспечивается добавлением квадрата изменения сохраняющейся величины нейросетевого решения в функцию потерь. Данный подход применяется к одномерному нелинейному уравнению Шрёдингера (НУШ). Полученные результаты показывают увеличение точности предсказанного решения по сравнению со стандартным методом.

Ключевые слова: законы сохранения, Physics-Informed Neural Networks, нелинейное уравнение Шрёдингера.

1. Введение

Одномерное нелинейное уравнение Шрёдингера (НУШ) – это уравнение из классической теории поля, описывающее ряд явлений в различных областях физики, включая распространение электромагнитных волн в нелинейных оптических волокнах и волноводах, бозе-эйнштейновский конденсат, плазменные волны и др. Для моделирования распространения оптического импульса в нелинейном волокне на основе НУШ было разработано множество конечно-разностных численных схем. Одним из наиболее популярных и эффективных алгоритмов является псевдо-спектральный метод расщепления [1]. Тем не менее, данная задача остается достаточно сложной с точки зрения вычислений, особенно когда в системе появляются все более мелкие характерные масштабы вследствие самофокусировки волнового поля.

Основное ограничение расчета динамики нелинейных систем в том, что в какой-то момент такие масштабы таких неоднородностей становятся сопоставимы с шагом дискретизации пространственной координаты, который, в свою очередь, ограничивает шаг по времени в конечно-разностных схемах [2]. Поэтому для избежания численной неустойчивости необходимые объем обрабатываемых данных и число итераций сильно возрастают. Таким образом, существует потребность в альтернативных методах расчета нелинейной динамики в различных областях физики.

Машинное обучение представляет перспективной альтернативой благодаря своей эффективности для многомерных, обратных и некорректно-поставленных задач [3]. Методы глубокого обучения, в частности, позволяют выделять закономерности из большого объема неоднородных данных и делать точные предсказания динамики на основе физических законов (дифференциальных уравнений в частных производных) и ограничений (в том числе законов сохранения: массы, импульса, энергии и др.).

Такой подход позволяет снять ограничения классических алгоритмов решения дифференциальных уравнений в частных производных, которые подразумевают дискретизацию пространственно-временной области.

Метод глубокого обучения Physics-Informed Neural Network (PINN) [4] был разработан и применён для решения прямых [5] и обратных (восстановление параметров уравнений по данным) задач [6] нелинейной динамики. Physics-Informed нейронная сеть агрегирует наблюдаемые данные, начальные и граничные условия, а также описание физической системы (уравнения), включая соответствующие остатки в функцию потерь (loss-функцию). Таким образом, за-

^{1*} Работа выполнена при финансовой поддержке РФФИ (проект 0729-2021-013).

дача решения нелинейного дифференциального уравнения в частных производных сводится к задаче минимизации суммы квадратов невязок в точках пространственно-временной области [3, 4].

Метод PINN был применен ко многим физическим системам, в частности, к нелинейному уравнению Шрёдингера: рассматривались разные волновые решения НУШ [7], а также более сложные системы, взаимодействие светлого и тёмного солитонов [8], и волны-убийцы в случае дефокусировки [9].

Было предложено множество улучшений оригинальной версии PINN, основанных на законах сохранения, с целью улучшить качество аппроксимации решений нелинейных дифференциальных уравнений в частных производных. В [10] рассматривается включение законов сохранения в дифференциальной форме в функцию потерь наравне с основными уравнениями. В данной работе выбран способ учёта законов сохранения в интегральной форме [11, 12]. Одномерное НУШ является примером точнорешаемой задачи, то есть системы с неограниченным набором законов сохранения и соответствующих сохраняющихся физических величин [13]. Мы рассматриваем аналог закона сохранения массы (нормы волновой функции НУШ, число квантов) в нейросетевом решении НУШ методом PINN. Ожидается, модифицированная на основе закона сохранения Physics-Informed нейронная сеть улучшит качество выполнения закона сохранения (или уравнения непрерывности) и улучшит качество предсказанного решения.

2. Physics-Informed Neural Network для одномерного НУШ

В данной работе, закон сохранения нормы решения используется как ограничение в процессе обучения Physics-Informed нейронной сети, решающей одномерное нелинейное уравнение Шрёдингера с периодическими граничными условиями в заданной области:

$$\begin{aligned} i\psi_t + 0.5\psi_{xx} + |\psi|^2\psi &= 0, \quad x \in [-4, 4], \quad t \in [0, 2.5] \\ \psi(0, x) &= 2\text{sech}(x) \\ \psi(t, -4) &= \psi(t, 4) \\ \psi_x(t, -4) &= \psi_x(t, 4) \end{aligned} \quad (1)$$

где $\psi(t, x)$ – комплекснозначное решение (волновая функция НУШ). Соответственно, невязка уравнения определяется выражением

$$f := i\psi_t + 0.5\psi_{xx} + |\psi|^2\psi \quad (2)$$

Это выражение соответствует нейронной сети $f(t, x)$ с 2 выходами, аппроксимирующими действительную и мнимую часть решения $\psi(t, x) = u(t, x) + iv(t, x)$.

Для задачи (1), функция потерь в оригинальном методе PINN имеет следующий вид:

$$L_1 = L_0 + L_{bc} + L_f \quad (3)$$

где L_0 и L_{bc} обозначают остатки начального и граничного условий, соответственно:

$$L_0 = \frac{1}{N_0} \sum_{i=1}^{N_0} |\psi(0, x_0^i) - \psi_0^i|^2 \quad (3.1)$$

$$L_{bc} = \frac{1}{N_b} \sum_{i=1}^{N_b} |\psi(t_b^i, -4) - \psi(t_b^i, 4)|^2 + |\psi_x(t_b^i, -4) - \psi_x(t_b^i, 4)|^2 \quad (3.2)$$

Компонента L_f есть сумма квадратов остатков уравнения в случайно сгенерированных точках пространственно-временной области задачи:

$$L_f = \frac{1}{N_f} \sum_{i=1}^{N_f} |f(t_f^i, x_f^i)|^2 \quad (3.3)$$

Наборы точек (датасеты) для начальных и граничных условий обозначаются как $\{x_0^i, \psi_0^i\}_{i=1}^{N_0}$ и $\{t_b^i\}_{i=1}^{N_b}$. Датасет $\{x_f^i, t_f^i\}_{i=1}^{N_f}$ обозначает множество точек, случайно сгенерированных в пространственно-временной области, в которых вычисляются значения $f(t, x)$ и по ним считается невязка уравнения. Таким образом, L_0 и L_{bc} навязывают выполнение начальных и граничных условий задачи (1) для нейросетевого решения, а L_f – самого уравнения в случайных точках области. Параметры нейронной сети, аппроксимирующей решение $\psi(t, x)$, определяются алгоритмом минимизации суммы квадратов остатков (3). Общая схема PINN для одномерного НУШ в постановке (1) показана на рис. 1.

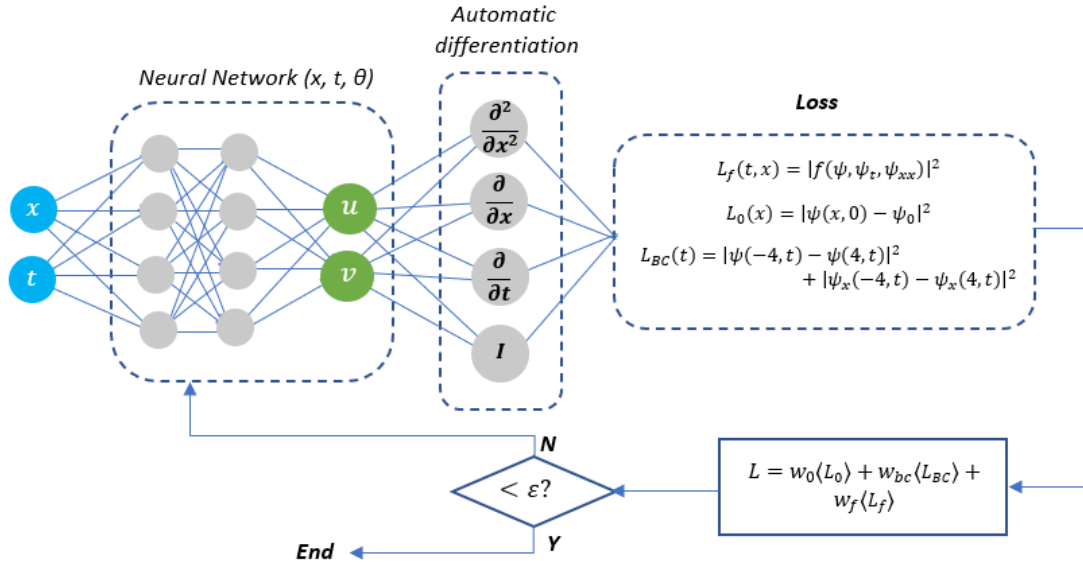


Рис. 1. Схема Physics-Informed нейронной сети для одномерного НУШ (1).

Для разработки модифицированного PINN, необходимо ввести ограничение на инвариантность сохраняющихся величин с помощью добавления соответствующих остатков в оригинальную функцию потерь сети.

3. Консервативная модификация нейронной сети

1D Нелинейное уравнение Шредингера – пример точно решаемой задачи, то есть динамической системы с неограниченным количеством первых интегралов, то есть сохраняющихся величин. Для первой серии экспериментов был рассмотрен простейший интеграл низшего порядка, соответствующий закону сохранения массы в механике сплошных сред:

$$P = \int |\psi(t, x)|^2 dx, \quad (4)$$

Начнем с рассмотрения консервативной модификации PINN, основанной на законе сохранения (4), подразумевающей наличие соответствующей невязки в функции потерь: $\propto (P(t) - P(0))^2$. Основываясь на начальном условии $\psi(0, x)$, вычисляется начальное значение P_0 .

3.1. Плоскости непрерывности (Continuity Planes)

Рассмотрим метод плоскостей непрерывности для включения законов сохранения в функцию потерь нейронной сети. В данном подходе, выбираются N_{pl} моментов времени и в них

устанавливаются «плоскости непрерывности» (в одномерном случае это линии) на которых вычисляются значения сохраняющейся величины на текущем нейросетевом решении: $\{P(t_p^i)\}_{i=1}^{N_{pl}}$. Консервативная модификация PINN устроена так, что значения в эти моменты времени $\{P(t_p^i)\}_{i=1}^{N_{pl}}$ приближаются к начальному P_0 за счет минимизации функции потерь. Следовательно, выражение для функции потерь оригинального PINN изменяется:

$$L_1 = L_0 + L_{bc} + L_f + L_p, \quad (5.1)$$

где компонента L_p обозначает агрегированную невязку сохраняющейся величины:

$$L_p \sim \frac{1}{N_{pl}} \sum_{i=1}^{N_{pl}} |P(t_p^i) - P_0|^2 \quad (5.2)$$

Для вычисления отклонения от начального значения на каждой плоскости, P_0 , интеграл в (4.1) заменяется суммой. Каждая плоскость непрерывности $[t_p^i, (x_0, x_1)]$ равномерно делится N_i точками, что дает выражение для сохраняющейся величины на каждой плоскости:

$$P(t_p^i) = \int_{x_0}^{x_1} |\psi(t_p^i, x)|^2 dx \approx \sum_{j=1}^{N_i} \frac{\Delta x}{N_i} |\psi(t_p^i, x_j)|^2 \quad (6)$$

Первая серия экспериментов была проведена с одной плоскостью непрерывности в качестве разных моментов времени с точки зрения решения $|\psi(t, x)|$ (такие, как интервалы фокусировки и дефокусировки). Оказалось, что точность результата сильно зависит от расположения плоскости (Таблица 1). Поэтому, для получения агрегированного результата требуется несколько плоскостей. Эксперимент был проведен также для пяти эквидистантных плоскостей.

Далее представлены результаты для плоскостей непрерывности. Рассматривается также процесс обучения нейронной сети – здесь приведены зависимости лосс-функции (функции потерь) и невязки сохраняющейся величины на плоскостях от итерации алгоритма минимизации (рис.2). Как и ожидалось, что дополнительные ограничения в функции потерь замедляют процесс обучения, в то время как закон сохранения выполняется лучше.

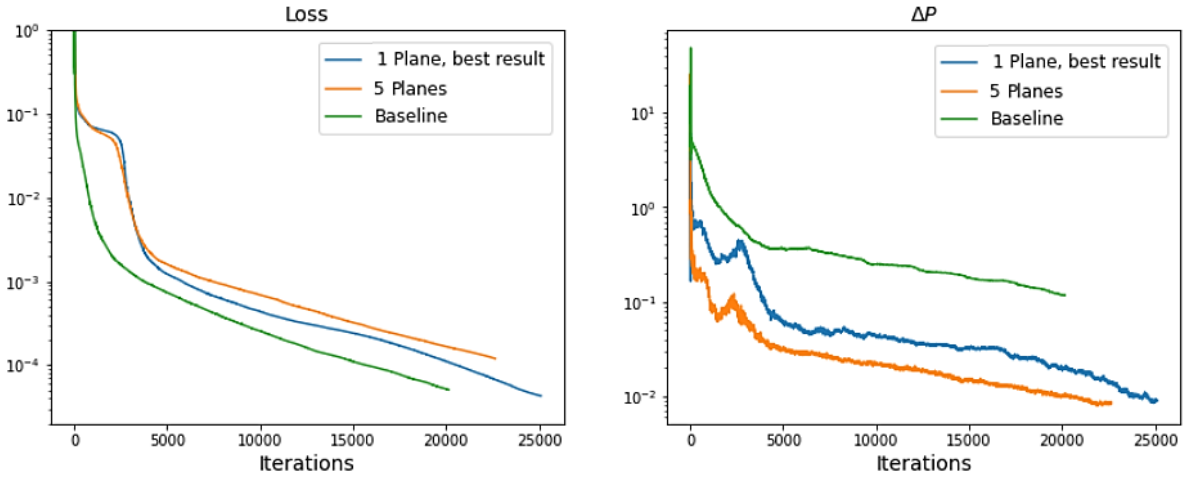


Рис. 2. Процесс обучения модифицированной PINN с применением плоскостей сохранения (лучшие результаты). (a): Зависимость лосс-функции от итерации, (b): Невязка сохраняющейся величины от итерации

В Таблице 1 приведены усредненные результаты вычисления точности для экспериментов с плоскостями непрерывности. Дополнительные ограничения в среднем улучшили точность предсказания действительной и мнимой частей решения, но физический смысл имеет его модуль. Точность его предсказания также улучшился в случае пяти плоскостей: была получена ошибка $\sim 7\%$, исходное значение 9% ; лучший результат ($\sim 4\%$) был получен одной удачно выбранной плоскостью на интервале дефокусировки. Таким образом, единственная удачно расположенная плоскость может уменьшить ошибку в 2.5 раза.

Таблица 1. Относительная L2-норма ошибки.

	Baseline	1 Plane, first focus	1 Plane, defocus	1 Plane, second focus	5 planes
$\langle \delta Re(\psi) \rangle$	0.33	0.32	0.12	0.32	0.22
$\langle \delta Im(\psi) \rangle$	0.32	0.29	0.10	0.29	0.19
$\langle \delta \psi \rangle$	0.09	0.09	0.04	0.10	0.07

3.2. Интегрирование закона сохранения методом Монте Карло

Для многомерных задач введение нескольких плоскостей непрерывности увеличивает вычислительные затраты ввиду соответствующей проблемы детерминированного численного интегрирования. В соответствии с этим, предлагается использовать интегрирование Монте Карло, суммируя по случайным точкам пространственно-временной области $\{x_f^i, t_f^i\}_{i=1}^{N_f}$:

$$\int_{t_0}^{t_1} \int_{x_0}^{x_1} |\psi(t, x)|^2 dx dt \approx \sum_{i=1}^{N_f} \frac{\Delta x \Delta t}{N_f} |\psi(t_f^i, x_f^i)|^2, \quad (7.1)$$

Следовательно, компонента лосс-функции, отвечающая за закон сохранения, имеет следующий вид:

$$L_p = \left(\sum_{i=1}^{N_f} \frac{\Delta x \Delta t}{N_f} |\psi(t_f^i, x_f^i)|^2 - P_0 \Delta t \right)^2 \quad (7.2)$$

Данный подход является проинтегрированным по времени законом сохранения и поэтому кажется менее детальным, чем предыдущий. Сравнительная динамика обучения и точность предсказания частично подтверждают это наблюдение: компоненты функции потерь убывают медленнее (рис. 3), а средняя относительная ошибка увеличилась с 9% до 10.7%. Следовательно, для одномерных задач такой метод консервативной модификации нейронной сети недостаточно эффективен. Тем не менее, предполагается, что он будет более вычислительно эффективен в более высокоразмерных задачах.

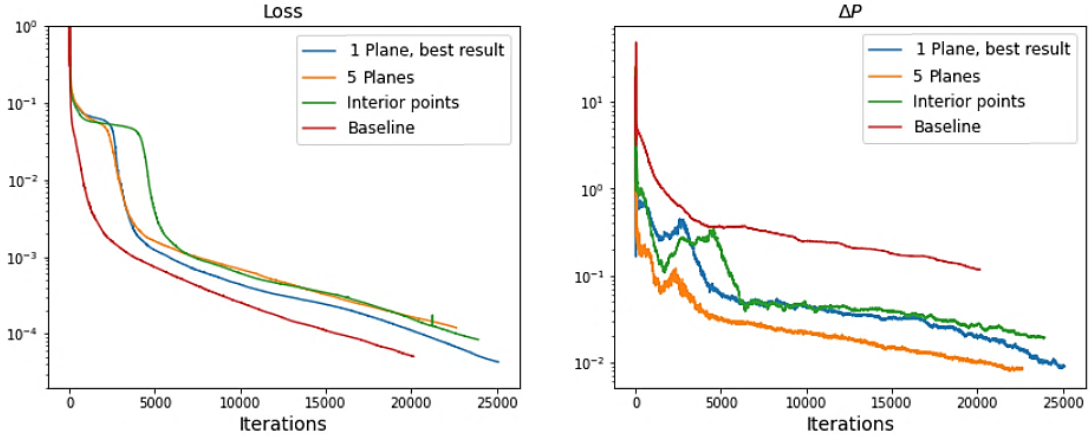


Рис.3. Сравнение динамики обучения модификаций PINN. (а): Функция потерь от итерации, (б): Невязка сохраняющейся величины от итерации.

4. Заключение

В данной работе законы сохранения были применены для обучения Physics-Informed нейронной сети в решении нелинейного одномерного уравнения Шрёдингера. Были рассмотрены два способа учёта изменения (невязки) сохраняющейся величины в лосс-функции нейронной сети. Для одномерной задачи метод плоскостей непрерывности был более эффективен чем интегрирование Монте Карло, который обычно даёт преимущество в более высокоразмерных задачах. Введение пяти эквидистантных плоскостей непрерывности уменьшило ошибку ориги-

нального метода примерно на 20%. Было также показано, что сходимость и точность модифицированной версии PINN чувствительны к расположению плоскостей непрерывности – одна удачно выбранная плоскость может уменьшить относительную ошибку нейросетевого решения на 60%. Поэтому имеет смысл рассмотреть другие стратегии расположения плоскостей сохранения по временному интервалу (например, случайный выбор плоскостей и др.). Также следует использовать другие сохраняющиеся величины нелинейного уравнения Шредингера (такие как энергия, импульс) или их комбинацию. Чтобы доказать обобщающую способность консервативной модификации PINN, нужно тестировать её для других систем (например, нелинейное уравнение Шредингера с дисперсией), а также двухмерных и трехмерных задач.

Благодарности

Проект 0729-2021-013, выполняемый в рамках Государственного задания на выполнение научно-исследовательских работ лабораториями, прошедшими конкурсный отбор в рамках национального проекта «Наука и университеты», в отношении которых принято решение Бюджетной комиссии Минобрнауки России (от 14.09.2021 № БК-П/23) о предоставлении из федерального бюджета субсидии на финансовое обеспечение государственного задания на выполнение научно-исследовательских работ.

Литература

1. Moxley F., Chuss D., Dai W.: A generalized finite-difference time-domain scheme for solving nonlinear Schrödinger equations // *Computer Physics Communications*, 184 (8), 1834–1841 (2013).
2. Васильев Е.П., Болотов Д.И., Болотов М.И., Смирнов Л.А.: Нейросетевой подход к решению задачи самовоздействия волновых полей в нелинейных средах // *Проблемы Информатики*. 2022. № 1. с. 5–16.
3. Karniadakis G.E., Kevrekidis I., Lu L., Perdikaris P., Wang S., Yang L.: Physics informed machine learning // *Nature Reviews Physics*. Vol.3. P.422–440.
4. Raissi M., Perdikaris P., Karniadakis G.E.: Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*. 2019. Vol. 378. P. 686–707.
5. Raissi M., Perdikaris P., Karniadakis G.E. Physics informed deep learning (part i): Data-driven solutions of nonlinear partial differential equations. arXiv:1711.10561 (2017).
6. Raissi M., Perdikaris P., Karniadakis G.E.: Physics informed deep learning (part ii): Data-driven discovery of nonlinear partial differential equations. arXiv.1711.10566. (2017).
7. Pu J.C., Li J., Chen Y.: Soliton, breather, and rogue wave solutions for solving the nonlinear Schrödinger equation using a deep learning method with physical constraints // *Chinese Physics B*. 2021. Vol.30, № 6. P. 060202.
8. Wu G.Z., Fang Y., Wang Y.Y., Wu G.C., Dai C.Q. Predicting the dynamic process and model parameters of the vector optical solitons in birefringent fibers via the modified PINN // *Chaos, Solitons & Fractals*. 2021. Vol. 152, P. 111393.
9. Wang L., Yan. Z.: Data-driven rogue waves and parameter discovery in the defocusing nonlinear Schrödinger equation with a potential using the PINN deep learning // *Physics Letters A*. 2021. Vol. 404, P.127408.
10. Wu G.Z., Fang Y., Wang Y.Y., Dai C.Q. Modified physics-informed neural network method based on the conservation law constraint and its prediction of optical solitons. arXiv.2108.13192. 2021.
11. Lin S., Chen Y. A two-stage physics-informed neural network method based on conserved quantities and applications in localized wave solutions // *Journal of Computational Physics*. 2022. Vol.457. P.111053.
12. Recommended practices in Nvidia Modulus. https://docs.nvidia.com/deeplearning/modulus/user_guide/theory/recommended_practices.html.
13. Zakharov V., Manakov S. On the complete integrability of a nonlinear Schrödinger equation // *Theoretical and Mathematical Physics*. 1974. Vol. 19 № 3. P. 551–559.

ПРИМЕНИМОСТЬ УПРОЩЕННОЙ МОДЕЛИ СООРУЖЕНИЯ ПРИ РАСЧЕТАХ СЕЙСМОСТОЙКОСТИ ПОДЗЕМНЫХ КОММУНИКАЦИЙ

Н.С. Дюкина

Нижегородский государственный университет им. Н.И. Лобачевского

Исследуется степень необходимой детализации расчетной модели сооружения при оценке прочности подземных коммуникаций, прилегающих к сооружению. В серии вычислительных экспериментов изучено влияние сооружения на грунтовое основание при статических и динамических воздействиях. Для анализа влияния детализации сооружения на сейсмическую реакцию и динамическую деформацию фундамента рассматриваются три расчетные модели сооружения. Инварианты вычислительных экспериментов: масса и центр массы сооружения, форма фундамента и габариты сооружения, динамическое воздействие волной сдвига, константы упругости для всех материалов. Показаны значительные различия во влиянии рассмотренных моделей сооружения на грунтовое основание при осадке и динамическом воздействии.

Ключевые слова: численное моделирование, расчетная модель, грунтовое основание, динамическое воздействие.

1. Введение

Расчет сооружений и примыкающих к ним трубопроводов на сейсмические воздействия обычно ведется либо с упрощенной моделью сооружения [1], либо с упрощенной моделью грунтового основания [2]. Упрощение модели сооружения допускается в [3] для атомных станций, отличающихся от гражданских сооружений значительной толщиной фундаментных плит, тогда как расчеты гражданских конструкций показывают максимальные напряжения именно в фундаментной плите [2]. Проходящие под фундаментом здания сейсмические волны изменяют параметры напряженно-деформированного состояния системы «здание-грунт». Поскольку грунтовое основание обладает податливостью и инерционностью, в нем распространяются волны исходного сейсмического воздействия и волны, связанные с присутствием и движением фундамента сооружения, которые часто оказывают большее влияние на поведение сооружения, чем исходное воздействие. Необходимо также принимать во внимание возможность локальных концентраций напряжений в грунтовом основании, приводящих к переходу грунтовой среды в сыпучее и упругопластическое состояние и, как следствие, изменению несущей способности основания. Описанные выше эффекты могут приводить к значительному усилению динамических и статических нагрузок на заглубленные в грунт трубопроводы. Поэтому требуется обособить допустимые границы идеализации сооружения и грунтового основания в задачах расчета сейсмостойкости. Представляется интересным исследование влияния детализации модели сооружения на напряженно-деформированное состояние грунтового основания и динамическую деформацию фундамента в расчетах сейсмостойкости.

2. Вычислительный эксперимент

Рассматривается прямоугольный массив грунта $2000 \times 2000 \times 200$ м. Грунт моделируется сплошной идеально-упругой средой с характеристиками $E = 2,4$ ГПа, $\nu = 0,4$, $\rho = 2000$ кг/м³. На дневной поверхности грунта в центре расположено сооружение – массив идеально-упругой среды $20 \times 20 \times 20$ м с механическими характеристиками $E = 20$ ГПа, $\nu = 0,25$, и массой $M=8000$ т. Численное моделирование выполнено в LS-Dyna. Длина ребра кубических конечных элементов для грунта принималась равной 2,5 м, для здания – 1 м. Для моделирования действия силы тяжести к расчетной области приложено постоянное сжимающее напряжение. На нижней границе грунтового массива задавались кинематические граничные условия, моделирующие сдвиг-

вую волну, распространяющуюся к дневной поверхности грунта. Определение кинематического воздействия на нижней границе грунта выполнено в соответствии с [4] по заданному на поверхности грунта тестовому импульсу вида $v_y = 0,6 \sin^2(2\pi(10t - 3))$.

Для анализа влияния детализации сооружения на сейсмический отклик и динамическую деформацию фундамента рассмотрено несколько расчетных моделей сооружения заданной массы. Масса сооружения, форма фундамента, упругие постоянные E, ν оставались неизменными во всех задачах.

Рассмотрено три модели сооружения:

1. Куб $20 \times 20 \times 20$ м (рис.1а), усредненная плотность $\rho = \frac{M}{V} = 1000 \text{ кг/м}^3$.
2. Две плиты $2 \times 20 \times 20$ м, соединенные 4 стойками $16 \times 5 \times 5$ м (рис.1б), плотность материала соответствует плотности железобетона $\rho = 2500 \text{ кг/м}^3$.
3. Две плиты $2 \times 20 \times 20$ м, соединенные стойкой $16 \times 10 \times 10$ м (рис.1в), плотность материала соответствует плотности железобетона $\rho = 2500 \text{ кг/м}^3$.

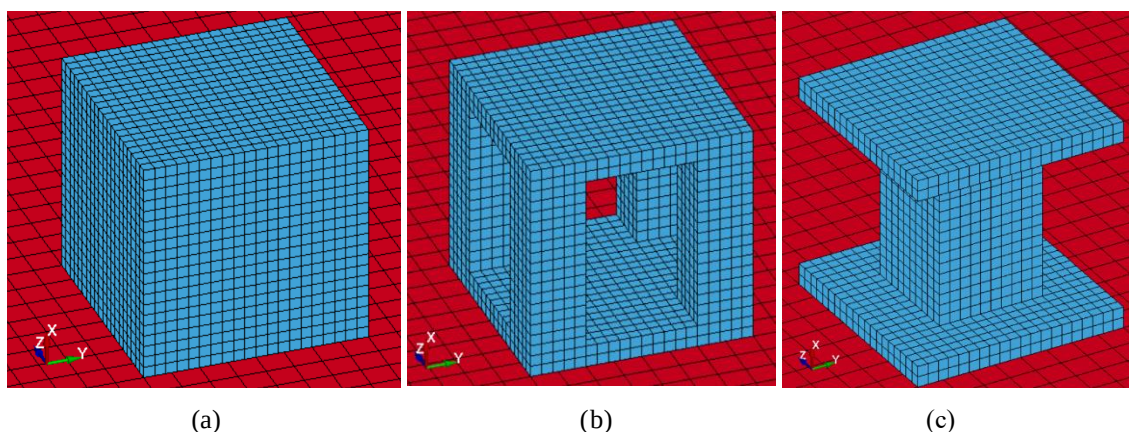


Рис. 1. Расчетные модели сооружения

Расчеты позволили обнаружить неоднородности распределения полей напряжений и скоростей в конструктивно сложных сооружениях, а также их различное влияние на грунтовое основание на этапе осадки сооружений от действия поля силы тяжести и на этапе вибраций от динамического воздействия. В соответствии с конструктивными особенностями зданий получены неоднородные поля напряжений на нижней плите сооружения и в приповерхностном слое грунта от осадки ($t = 0$ с) и при динамической реакции на сдвиговой импульс ($t = 0,33$ с) – напряжения σ_{xx} для этих случаев приведены на рис.2. Профили скоростей в моменты подхода ($t < 0,35$ с) и отражения ($t > 0,35$ с) сдвиговой волны от поверхности показывают различия в сдвиговых скоростях, не превышающее 17% для разных вариантов конструкции (рис.3,4). Более значительна разница в вертикальной компоненте скорости V_x , характеризующей раскачивание сооружения. Худший сценарий реализуется для однородного варианта конструкции.

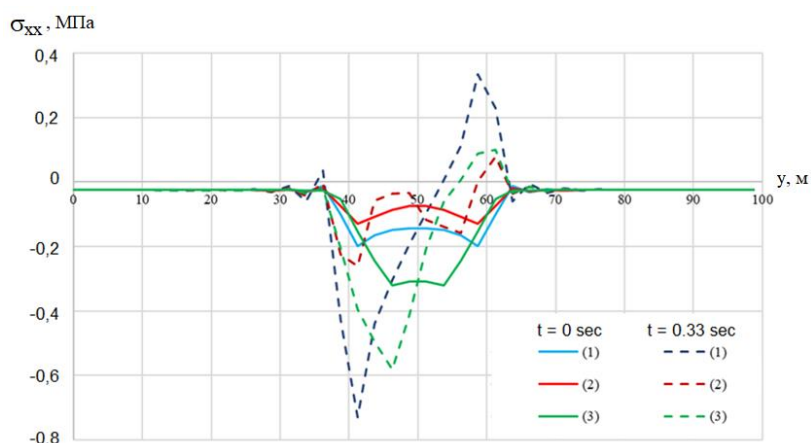


Рис.2. Профиль напряжений σ_{xx} в грунте на глубине 0,5м (интервал 40-60м профиля соответствует фундаменту сооружения)

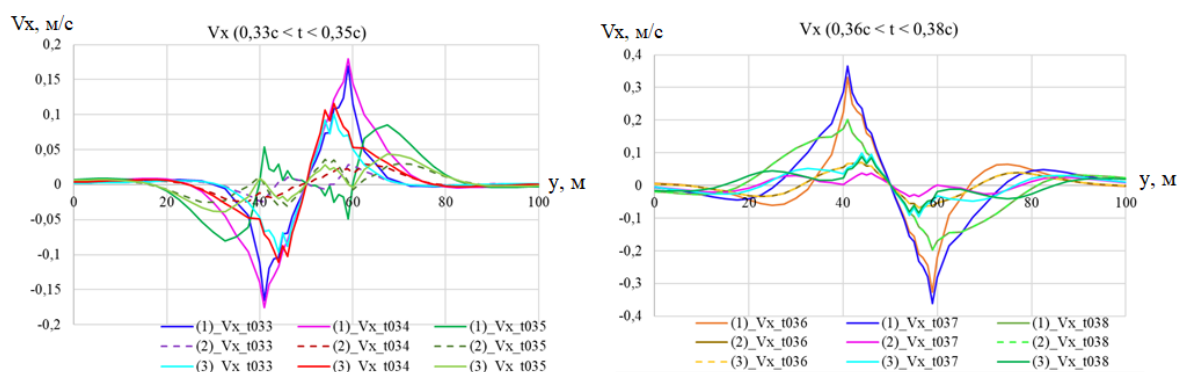


Рис.3. Профили скоростей V_x в грунте на глубине 0,5м под сооружением

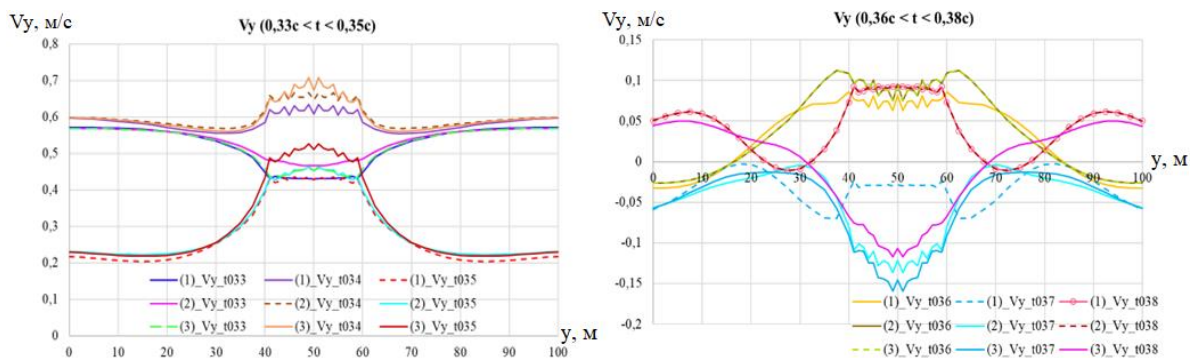


Рис.4. Профили скоростей V_y в грунте на глубине 0,5м под сооружением

3. Заключение

Выбор модели сооружения влияет на результаты численного моделирования поведения сооружения при динамических нагрузках, а также на изменение напряженно-деформированного состояния грунтового основания. В проведенном исследовании худший сценарий реализован для наиболее грубой модели сооружения, что, с одной стороны, позволяет гарантировать более благоприятные сценарии реакции сооружения и грунтового основания для реальных, более детальных моделей. В то же время это может привести к завышенным оценкам нагрузок при оценке сейсмостойкости конструкции и, как следствие, удорожанию строительства, поэтому при расчетах на сейсмические воздействия должна учитываться определенная степень детализации сооружения.

Литература

1. Тяпин А.Г. Особенности расчета жестких фундаментов на "подушках", устраиваемых под фундаментом в жестком основании, с помощью программы SASSI // Сейсмостойкое строительство. Безопасность сооружений. 2013. № 3. С. 17-23.
2. Ерофеев В.И., Сатанов А.А., Хазов П.А., Шишова М.А. Частоты и формы собственных колебаний высотного здания во Владивостоке при расчете на максимальное расчетное землетрясение // В сборнике: Живучесть и конструкционное материаловедение (ЖивКоМ - 2020). Сборник трудов V Международной научно-технической конференции в дистанционном формате. 2020. С. 237-241.
3. Seismic Analysis of Safety-Related Nuclear Structures and Commentary. ASCE4-98. Reston, Virginia, USA. 1999.
4. Баженов В.Г., Дюкина Н.С. Численное решение двумерных и трехмерных задач взаимодействия сооружений с грунтовым основанием при сейсмических воздействиях // Вестник ЦНИИСК им. В.А. Кучеренко "Исследования по теории сооружений". 2009. № 1. С. 172-178.

АНАЛИЗ ХАРАКТЕРИСТИК ПЕРЕМЕЖАЮЩЕЙСЯ ОБОБЩЕННОЙ СИНХРОНИЗАЦИИ ПРИ ПОМОЩИ РАСЧЕТА МЕРЫ МУЛЬТИСТАБИЛЬНОСТИ В ОДНОНАПРАВЛЕННО СВЯЗАННЫХ СИСТЕМАХ ЛОРЕНЦА

Е.В. Евстифеев^{1,2}, О.И. Москаленко^{1,2}

¹*Саратовский национальный исследовательский госуниверситет им. Н.Г. Чернышевского»,*

²*Региональный научно-образовательный математический центр
«Математика технологий будущего»*

На примере связанных систем Лоренца (систем со сложной, двулистной топологией аттрактора) при помощи метода расчета локальных ляпуновских показателей и метода вспомогательной системы, модифицированных для определения меры мультистабильности, были оценены основные характеристики перемежающейся обобщенной синхронизации при однонаправленной связи. В качестве характеристик рассматривались распределения длительностей ламинарных фаз при фиксированных значениях параметра связи, а также зависимость средней длительности ламинарных фаз от параметра надкритичности. Результаты показали, что данный подход позволяет получить более точную оценку характеристик перемежаемости по сравнению с традиционными методами и подходами.

Ключевые слова: перемежающаяся обобщенная синхронизация, мультистабильность, локальные показатели Ляпунова, метод вспомогательной системы.

1. Введение

Обобщенная хаотическая синхронизация [1–2] – это явление радиофизики, широко распространенное и обладающее отличным потенциалом к применению при разработке медицинского оборудования [3], скрытой передаче информации [4] и т.д. Перемежающаяся обобщенная синхронизация [5] – это режим, при котором наблюдается чередование фаз синхронной (ламинарных) и асинхронной (турбулентных) динамики связанных систем.

Мультистабильность [6] – универсальное явление систем различной природы, характеризующееся существованием нескольких аттракторов при фиксированных значениях управляющих параметров. Наличие мультистабильности вблизи границы обобщенной синхронизации означает, что в зависимости от выбора начальных условий систем в различные моменты времени может как наблюдаться режим обобщенной синхронизации, так и отсутствовать. Тогда, если ввести меру мультистабильности, характеризующую степень синхронизации в данный момент времени, то можно оценить временную динамику исследуемых систем и установить тип перемежаемости по статистическим характеристикам. Такой подход применялся в системах с простой топологией аттрактора [7]. Целью данной работы является применение такого метода в системах со сложной топологией аттрактора – системах Лоренца [8].

2. Методы и материалы

2.1. Методы

При анализе характеристик перемежающейся обобщенной синхронизации обычно используется метод вспомогательной системы [9]. Он заключается во введении дополнительной ведомой (вспомогательной) системы, идентичной исходной ведомой системе во всем за исключением начальных условий. Тогда, в случае синхронизации состояния ведомой и вспомогательной систем будут эквивалентны, что следует из определения обобщенной синхронизации. Тогда, можно рассчитать разность между состояниями ведомой и вспомогательной систем в каждый

момент времени и, ввести некоторую величину d , характеризующую отклонение между собой указанных состояний.

Проблема заключается в том, что при различных начальных условиях в определенный момент времени при наличии мультистабильности величина d может претерпевать значительные изменения. Более того, флуктуационный характер величины d также затрудняет оценку статистических характеристик. Следовательно, возникает необходимость в использовании более сглаженной временной зависимости, например меры мультистабильности.

В данной работе мера мультистабильности вводится как вероятность детектирования асинхронных участков временной динамики:

$$P_a \approx 1 - \sum_{i=1}^M \frac{n(\xi^i)}{M(M-1)}, \quad (1)$$

где $M \gg 1$ – число рассматриваемых осцилляторов, $n(\xi^i)$ – количество систем, находящихся в синхронизме с i -м осциллятором, ξ^i – исследуемая временная зависимость, характерная для i -го осциллятора. В случае вспомогательной системы $\xi(t) = d(t)$.

В работе также был рассмотрен другой, более универсальный подход, – метод локальных показателей Ляпунова [10]. Данный метод интересен тем, что он не ограничен случаем однонаправленной связи в отличие от метода вспомогательной системы [11] и может быть использован при взаимной связи. При расчете локальных ляпуновских показателей в качестве $\xi(t)$ выступает Λ_2^i – второй по старшинству условный (зависящий от величины параметра связи) показатель Ляпунова. В данном случае критерий синхронизации отличается: синхронному режиму соответствует момент $\Lambda_2^i < 0$. Однако, на практике из-за флуктуации данной величины более удобным оказывается другой, немного менее строгий, критерий: $\Lambda_2^i < \Delta$, Δ – малая по величине заданная наперед константа.

2.1. Исследуемая модель

Для демонстрации вышесказанного были взяты системы Лоренца, в которых была обнаружена перемежаемость типа перескоков [10]. Для получения меры мультистабильности было смоделировано множество ведомых систем Лоренца при одной ведущей. Уравнения, описывающие данную модель, выглядят следующим образом:

$$\begin{aligned} \dot{x}_{1,2}^i &= \sigma(y_{1,2}^i - x_{1,2}^i) + \varepsilon_{1,2}(x_{2,1}^i - x_{1,2}^i) \\ \dot{y}_{1,2}^i &= (r_{1,2} + \delta_{1,2}^i)x_{1,2}^i - y_{1,2}^i - x_{1,2}^i z_{1,2}^i \\ \dot{z}_{1,2}^i &= -bz_{1,2}^i + x_{1,2}^i y_{1,2}^i, \end{aligned} \quad (2)$$

где $i = 1, 2, \dots, N$ – индекс пары систем ($N \approx 100$), $\sigma = 10$, $b = 2$, $r_1 = 40$, $r_2 = 35$, $\delta_1^i = 0$, $\delta_2^i \in [0, 10^{-10}]$ – случайные числа, распределенные равномерно в указанном отрезке, характеризующие малую расстройку, введенную между вторыми системами каждой пары. Расстройка δ_2^i введена для того, чтобы избежать ловушку численного счета. Параметры связи $\varepsilon_1 = 0$, $\varepsilon_2 = \varepsilon$ в случае однонаправленной связи, и $\varepsilon_{1,2} = \varepsilon$ при взаимной связи. Значения этих параметров выбраны таким образом, чтобы после переходного процесса режим обобщенной синхронизации всегда возникал при некоторой критической величине параметра связи вне зависимости от начальных условий.

Стоит отметить, что при количестве рассматриваемых систем порядка 10000 по мере дальнейшего увеличения N и, соответственно, заполнения бассейна притяжения ведомой системы, повышение точности определяемых характеристик будет происходить со все более низкой интенсивностью.

Отметим, что при использовании метода вспомогательной системы под рассмотрение попадает $N(N-1)/2$ ведомых систем, если сравнивать их попарно между собой, в то время как при использовании метода локальных показателей Ляпунова будет рассматриваться N систем.

3. Результаты и выводы

При помощи модификаций методов расчета локального ляпуновского показателя и вспомогательной системы для определения меры мультистабильности были оценены основные характеристики перемежаемости, такие как распределения длительностей ламинарных фаз при фиксированном значении параметра связи, а также зависимость средней длительности ламинарных фаз от параметра надкритичности. Полученные характеристики полностью соответствуют как теоретическим зависимостям, так и результатам других работ [7, 11, 12].

Таким образом, установлено, что мера мультистабильности позволяет оценить характеристики перемежающейся обобщенной синхронизации. Можно сделать предположение, что данный вывод будет также справедлив при рассмотрении перемежаемости в случае других типов синхронизации при наличии мультистабильности в исследуемых системах.

Литература

1. Boccaletti S., Kurths J., Osipov G., et al. The synchronization of chaotic systems // *Physics Report*. 2002. Vol. 366, Issue 1-2. P. 1. DOI: 10.1016/S0370-1573(02)00137-0.
2. Rulkov N.F., Sushchik M.M., Tsimring L.S., et al. Generalized synchronization of chaos in directionally coupled chaotic systems // *Phys. Rev. E*. 1995. Vol. 51, No. 2. P. 980. DOI: 10.1103/PhysRevE.51.980
3. Hramov A.E., Koronovskii A.A., Ponomarenko V.I., et al. Detecting synchronization of self-sustained oscillators by external driving with varying frequency // *Phys. Rev. E*. 2006. Vol. 73. Art. No. 026208. DOI: 10.1103/PhysRevE.73.026208.
4. Moskalenko O.I., Koronovskii A.A., Hramov A.E. Generalized synchronization of chaos for secure communication: Remarkable stability to noise // *Physics Letters A*. 2010. Vol. 374, No. 29. P. 2925-2931. DOI: 10.1016/j.physleta.2010.05.024.
5. Hramov A.E., Koronovskii A.A. Intermittent generalized synchronization in unidirectionally coupled chaotic oscillators // *Europhysics Lett.* 2005. Vol. 70, No. 2. P. 169-175. DOI:10.1209/EPL/I2004-10488-6.
6. Pisarchik A., Feudel U. Control of multistability // *Physics Reports*. 2014. Vol. 540, No. 4. PP.167-218. DOI: 10.1016/J.PHYSREP.2014.02.007.
7. Москаленко О.И., Короновский А.А., Сельский А.О., Евстифеев Е.В. Метод определения характеристик перемежающейся обобщенной синхронизации, основанный на вычислении вероятности наблюдения синхронного режима // *Письма в ЖТФ*. 2022. Т. 48, № 2. С. 3-6. DOI: 10.21883/PJTF.2022.02.51910.18985.
8. Lorenz E.N. Deterministic Nonperiodic Flow // *J. Atmos. Sci.* 1963. V. 20. No. 2. P. 130. DOI: [https://doi.org/10.1175/1520-0469\(1963\)020<0130:DNF>2.0.CO;2](https://doi.org/10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2).
9. Abarbanel H.D.I., Rulkov N.F., Sushchik M.M. Generalized synchronization of chaos: The auxiliary system approach // *Phys Rev E*. 1996. Vol. 53, Iss. 5. P.4528. DOI: 10.1103/PhysRevE.53.4528.
10. Abarbanel H. D. I., Brown R., Kennel M.B. Variation of Lyapunov Exponents on a Strange Attractor // *Journal of Nonlinear Science*. 1991. Vol. 1. P. 175–199. DOI: 10.1007/BF01209065.
11. Moskalenko O.I., Koronovskii A.A., Khanadeev V.A. Intermittency at the boundary of generalized synchronization in mutually coupled systems with complex attractor topology // *Technical Physics*. 2019. Vol. 64, No. 3. P. 302-305. DOI: 10.1134/S1063784219030198.
12. Москаленко О.И., Евстифеев Е.В., Короновский А.А. Метод определения характеристик перемежающейся обобщенной синхронизации на основе расчета локальных показателей ляпунова // *Письма в ЖТФ*. 2022, том 46, вып. 16(163). С. 12-15. DOI: 10.21883/PJTF.2020.16.49846.18359.

ПОЧТИ КОМПЛЕКСНЫЕ МНОГООБРАЗИЯ С ГОМОЛОГИЧЕСКИ НУЛЕВОЙ ВЕЩЕСТВЕННОЙ ЧАСТЬЮ^{1*}

В.И. Звонилов

Нижегородский государственный университет им. Н.И. Лобачевского

Формула Рохлина [1] для комплексных ориентаций неособой плоской вещественной алгебраической кривой чётной степени, разбивающей свою комплексификацию, распространяется на случай пары $2n$ - и $(2n - 2)$ -мерных почти комплексных многообразий $B \supset A$, на которых действует гладкая инволюция **conj**, антиголоморфная относительно почти комплексной структуры, причём множество $\mathbb{R}A = \text{fix conj}|_A$ гомологично нулю как в A , так и в $\mathbb{R}B$. Приводятся примеры вещественных алгебраических многообразий любой размерности, гомологичных нулю в своей комплексификации.

Ключевые слова: почти комплексное многообразие, гомологически нулевая вещественная часть, комплексные ориентации вещественной части.

1. Определения

Пусть X – почти комплексное многообразие, размерности $2d$, на котором действует гладкая инволюция **conj** с непустым ориентируемым множеством вещественных точек $\mathbb{R}X = \text{fix conj}$. Пусть множество $\mathbb{R}X$, наделённое некоторой ориентацией, реализует нулевой класс в $H_d(X)$ (короче, $\mathbb{R}X \sim 0$ в X), и W является $(d + 1)$ -мерной ориентированной цепью с $\partial W = \mathbb{R}X$. Указанная ориентация многообразия $\mathbb{R}X$ называется *комплексной W -ориентацией*.

2. Примеры

2.1. Вращение

Пусть $C_1 \subset \mathbb{R}P^2$ – неособая вещественная алгебраическая кривая типа I (т.е. разбивающая свою комплексификацию) – задаётся уравнением $f(x_0, x_1, x_2^2) = 0$ чётной степени. Тогда она симметрична относительно прямой $L: x_2 = 0$. Пусть кривая C_1 не имеет вещественных точек пересечения с L и её комплексная ориентация симметрична относительно L . Ясно, что поверхность $C_2 \subset \mathbb{R}P^3$, задаваемая уравнением $f(x_0, x_1, x_2^2 + x_3^2) = 0$, получается вращением кривой C_1 в $\mathbb{R}P^3$ вокруг L и, очевидно, гомологична нулю в комплексификации. Продолжая этот процесс, можно для любого $n \geq 1$ получить неособую вещественную n -мерную гиперповерхность $C_n \subset \mathbb{R}P^{n+1}$ чётной степени гомологичную нулю в комплексификации.

2.2. Край трубчатой окрестности

Пусть $A \subset P^n: \begin{cases} f(x_0, \dots, x_n) = 0 \\ g(x_0, \dots, x_n) = 0 \end{cases}$ – неособое полное пересечение вещественных гиперповерхностей одинаковой степени m , $B \subset P^n: f^2 + g^2 = \varepsilon^2 x_0^{2m}$ и $\tilde{B}$ – неособое многообразие, полученное раздутием особенности $B \cap \{x_0 = 0\}$.

Теорема 1. Если $\mathbb{R}A \sim 0$ в $\mathbb{C}A$, то $\mathbb{R}\tilde{B} \sim 0$ в $\mathbb{C}\tilde{B}$ при малом ε .

2.3. Расслоения

Ясно, что если $\mathbb{R}X \sim 0$ в $\mathbb{C}X$ или $\mathbb{R}Y \sim 0$ в $\mathbb{C}Y$, то $\mathbb{R}X \times \mathbb{R}Y \sim 0$ в $\mathbb{C}X \times \mathbb{C}Y$.

Более общая ситуация:

^{1*} Работа выполнена по теме государственного задания (№ 0729-2020-0055).

Пусть $p: E \rightarrow B$ – гладкое сюръективное отображение неособых вещественных алгебраических многообразий с ориентируемыми $\mathbb{R}B$ и $\mathbb{R}E$ и неособыми слоями. Тогда если $\mathbb{R}F \sim 0$ в $\mathbb{C}F$ для каждого слоя F , то $\mathbb{R}E \sim 0$ в $\mathbb{C}E$.

3. Основные результаты

3.1. Абсолютный случай

Пусть X – почти комплексное многообразие размерности $4n$, на котором действует гладкая инволюция conj . Следующая теорема сразу получается из результата Арнольда [2].

Теорема 2. Если $\mathbb{R}X \neq \emptyset$ ориентировано и $\mathbb{R}X \sim 0$ в X , то эйлерова характеристика $\chi(X_i) = 0$ для любой компоненты X_i множества $\mathbb{R}X$.

3.2. Относительный случай

Пусть B – почти комплексное многообразие размерности $2n$, на котором действует гладкая инволюция conj с $\mathbb{R}B \neq \emptyset$, и $A \subset B$ – его $(2n - 2)$ -мерное почти комплексное подмногообразие, инвариантное относительно conj . Пусть $\mathbb{R}A$ гомологично нулю в A , причём существует n -мерная ориентированная цепь с $\partial W = \mathbb{R}A$ и с $W \cap \text{conj}W = \emptyset$. Обозначим через ξ класс в $H_n(B)$, реализуемый циклом $W - \text{conj}W$. Пусть $B_1, \dots, B_l$ – все ориентируемые компоненты связности множества $\mathbb{R}B \setminus \mathbb{R}A$. Обозначим через b_j класс, определяемый в $H_n(\mathbb{R}B, \mathbb{R}A)$ множеством B_j , наделённым некоторой ориентацией.

Теорема 3. Если $\mathbb{R}A$ гомологично нулю в $\mathbb{R}B$, то существуют такие $x_1, \dots, x_l \in \mathbb{Z}$, что образ класса $x = \sum x_j b_j$ при граничном гомоморфизме $H_n(\mathbb{R}B, \mathbb{R}A) \rightarrow H_{n-1}(\mathbb{R}A)$ есть фундаментальный класс $[\partial W]$ многообразия $\mathbb{R}A$, наделённого комплексной W -ориентацией, причём $\sum x_j^2 \chi(B_j) = \xi^2/4$.

Приводятся вычисления правой части равенства теоремы 3 для конкретных гиперповерхностей. При $n = 1$ для кривой, разбивающей свою комплексификацию, теорема 3 даёт формулу Рохлина [1], имеющую форму, предложенную в [3].

Литература

1. Рохлин В.А. Комплексные топологические характеристики вещественных алгебраических кривых // Успехи мат. Наук. 1978. Т. 33. № 5. С. 77–89.
2. Арнольд В.И. О расположении овалов вещественных плоских алгебраических кривых, инволюциях четырехмерных гладких многообразий и арифметике целочисленных квадратичных форм // Функциональный анализ и его приложения. 1971. Т. 5. № 3. С. 1–9.
3. Звонилов В.И. Комплексные топологические характеристики вещественных алгебраических кривых на поверхности // Функциональный анализ и его приложения. 1982. Т. 16. № 3. С. 56–57.

НЕАЛЬТЕРНИРУЮЩИЕ ГАМИЛЬТОНОВЫ ФОРМЫ С КОЭФФИЦИЕНТАМИ В АЛГЕБРЕ РАЗДЕЛЕННЫХ СТЕПЕНЕЙ^{1*}

А.В. Кондратьева, М.И. Кузнецов

Нижегородский государственный университет им. Н.И. Лобачевского

В докладе представлены некоторые результаты общей теории неальтернирующих гамильтоновых алгебр Ли над совершенным полем характеристики два. Построен комплекс симметричных дифференциальных форм над алгеброй разделенных степеней. Дана классификация градуированных алгебр и найдены их фильтрованные деформации. Описаны классы эквивалентности неальтернирующих гамильтоновых форм с полиномиальными коэффициентами в разделенных степенях.

Ключевые слова: алгебры Ли, неальтернирующие гамильтоновы алгебры Ли, неальтернирующие гамильтоновы формы.

Хорошо известные алгебры Ли картановского типа, состоящие из векторных полей, сохраняющих специальную, гамильтонову или контактную форму, имеют аналог в характеристике p . Заметим, что алгебра функций должна быть заменена алгеброй разделенных степеней $O(F)$ (см. [1]), соответствующей некоторому обобщенному флагу F пространства E , $F: E = E_0 \supseteq E_1 \supseteq \dots$. Пусть $\{x_1, \dots, x_n\}$ – базис E , согласованный с F , тогда $O_n(F) = O(n; \mathbf{m})$, где $\mathbf{m} = (m_1, \dots, m_n)$ – n -набор высот переменных $x_1, \dots, x_n$ относительно флага F . В случае поля характеристики 2 можно построить большой класс простых гамильтоновых алгебр Ли, соответствующих неальтернирующим симметричным дифференциальным формам.

Класс неальтернирующих гамильтоновых алгебр Ли может представлять интерес по крайней мере по двум причинам. Во-первых, гамильтоновы алгебры Ли можно отнести к алгебрам картановского типа при $p = 2$, что имеет большое значение для классификации простых алгебр Ли малых характеристик. Во-вторых, эти алгебры Ли могут допускать невырожденные дифференцирования, что может представлять интерес для теории p -групп.

Неальтернирующие гамильтоновы алгебры над полем характеристики 2 были впервые построены в 1993 г. Lin Lei [2] как алгебры Ли полиномов от разделенных степеней с симметричной скобкой Пуассона $\{f, g\} = \sum_i \partial_i f \partial_i g$. В случае, когда высоты переменных равны 1, неальтернирующие гамильтоновы алгебры Ли изоморфны первой серии простых алгебр Ли, построенных И. Капланским [3]. В [4,5] вводятся симметричные дифференциальные формы в разделенных степенях и изучаются неальтернирующие гамильтоновы алгебры Ли, подобные гамильтоновым супералгебрам Ли нулевой характеристики относительно стандартных скобок Пуассона. Отметим, что классификация альтернирующих гамильтоновых форм над алгеброй срезуемых многочленов получена М.И. Кузнецовым, С.А. Кирилловым ([6]). Полная классификация над алгеброй разделенных степеней построена С.М. Скрыбиным ([7,8]).

В докладе обсуждается общая теория неальтернирующих гамильтоновых алгебр Ли в разделенных степенях над совершенным полем характеристики 2. Основные результаты можно найти в [9]. В [10] дана инвариантная конструкция комплекса симметричных дифференциальных форм в характеристике 2 и предложена программа исследования. На первом этапе авторы получили все инварианты симметрических гамильтоновых дифференциальных форм с постоянными коэффициентами относительно параболической подгруппы в $GL(V)$, соответствующей флагу F . В частности, было показано, что существует базис V , согласованный с флагом F , такой, что форма имеет матрицу $\text{diag}(M_0, \dots, M_0, M_1, \dots, M_1, 1_s)$, где

$$M_0 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, M_1 = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$$

Позднее, используя теорию усеченных коиндуцированных модулей ([11,12]), авторы доказали, что в случае, когда $n > 4$, или $n = 4$ и флаг F нетривиален, или $n = 2, 3$ и E_1 содержит

^{1*} Работа выполнена при финансовой поддержке РФФИ (проект № 18-01-00900) и Министерства науки и высшего образования Российской Федерации (проект № 0729-2020-0055).

неизотропный вектор относительно формы $\bar{\omega}$ на E , дуальной форме ω , каждая фильтрованная алгебра Ли, связанная с градуированной неальтернирующей гамильтоновой алгеброй, задается неальтернирующей гамильтоновой дифференциальной формой с полиномиальными коэффициентами в разделенных степенях.

Одним из основных результатов является доказательство эквивалентности неальтернирующей гамильтоновой формы ω с полиномиальными коэффициентами и форме $\omega(0)$ при условии, что канонический вид $\omega(0)$ содержит $dx_i^{(2)}$ или $dx_i dx_j + dx_j^{(2)}$ для некоторой переменной x_i высоты больше 1. Приведены некоторые результаты по проблеме классификации неальтернирующих гамильтоновых форм, для которых это условие не выполняется. В этом случае неальтернирующая гамильтонова форма ω может быть не эквивалентна $\omega(0)$.

Теорема. Пусть ω – неальтернирующая гамильтонова форма над алгеброй разделенных степеней $O_n(F)$, E^0 – подпространство всех изотропных векторов E относительно формы $\bar{\omega}$ на E , двойственной к $\omega(0)$.

(i) если $E_1 \not\subset E^0$, то ω эквивалентна $\omega(0)$;

(ii) если $E_1 \subset E^0$, то существует базис E , согласованный с F , такой, что ω эквивалентна

$$dx_1 dx_2 + \dots dx_{n-2} dx_{n-1} + dx_n^{(2)} + \sum_{j=1}^{n-1} b_{jn} dz_j dz_n, \text{ если } n = 2k + 1$$

и

$$dx_1 dx_2 + \dots dx_{n-1} dx_n + dx_n^{(2)} + \sum_{j \neq n-1} b_{j,n-1} dz_j dz_{n-1}, \text{ если } n = 2k$$

Более того, если $b_{jn} \neq 0$ ($b_{j,n-1} \neq 0$) для некоторого j , то ω не эквивалентна $\omega(0)$.

Задача классификации форм с полиномиальными (в разделенных степенях) коэффициентами полностью решена только в случае трех переменных [13]. Кроме того, найдены размерности всех простых неальтернирующих гамильтоновых алгебр Ли.

Литература

1. Кострикин А.И., Шафаревич И.Р. Градуированные алгебры Ли конечной характеристики // Изв. АН СССР. Сер.матем. 1969. Т.33, № 2. С. 251–322.
2. Lin Lei Non-alternating hamiltonian algebra $P(n,m)$ of characteristic two // Communications in Algebra. 1993. V.21, No.2. P. 399–411.
3. Kaplansky I. Some simple Lie algebras of characteristic 2 // Lecture Notes in Math. // 1982. V.933. P. 127–129.
4. Bouarrouj S., Grozman P., Lebedev A., Leites D. Divided power (co)homology. Presentation of simple finite dimensional modular superalgebras with Cartan matrix // Homology, Homotopy Appl. 2010. V.12, No. 1. P. 237–248.
5. Lebedev A. Analog of ortogonal, Hamiltonian, Poisson and contact Lie superalgebras in characteristic 2 // J. Nonlin. Math. Phys. 2010. V.17, No. 1. P. 399–411.
6. Кузнецов М. И., Кириллов С. А. Гамильтоновы дифференциальные формы над алгеброй разделенных степеней // Изв. Опросы. 1986. Т.41, № 2. С. 205–206.
7. Скрыбин С. М. Классификация гамильтоновых форм над разделенными степенями // Матем.сб. 1991. Т.69. С. 121–141.
8. Скрыбин С. М. Нормальные формы симплектических и контактных форм над алгебрами разделенных степеней. URL: <https://arxiv.org/abs/1906.11496v1> (дата обращения: 2019).
9. Kuznetsov M.I., Kondrateva A.V., Chebochko N.G. Non-alternating Hamiltonian Lie algebras in characteristic 2. I. URL: <http://arxiv.org/abs/1812.11213> (дата обращения: 2018).
10. Кузнецов М. И., Кондратьева А. В., Чебочко Н. Г. О гамильтоновых алгебрах Ли характеристики 2 // Матем. журнал. 2016. Т.16, № 2. С. 54–65.
11. Кузнецов М. И. Усеченные индуцированные модули над транзитивными алгебрами Ли характеристики p // Изв. АН СССР. Сер.матем. 1989. Т.53, № 3. С. 557–589.

12. Кузнецов М. И. Теорема вложения для транзитивных фильтрованных алгебр Ли характеристики p // Изв. вузов. Матем. 1991. Т.10. С. 43–45.
13. Kondrateva A. V. Non-alternating Hamiltonian Lie algebras in three variables // Lobachevskii Journal of Mathematics. 2021. V.42, No. 12. P. 2841–2853.

РЕАЛИЗАЦИЯ БИБЛИОТЕКИ ПОДДЕРЖКИ РАСПРЕДЕЛЁННОГО ГЛОБАЛЬНОГО АДРЕСНОГО ПРОСТРАНСТВА^{1*}

С.Ю. Коннов, Е.А. Козин

Нижегородский государственный университет им. Н. И. Лобачевского

Высокопроизводительные вычислительные системы отличаются архитектурой и способом разработки программного обеспечения. В данной работе внимание сконцентрировано на системах с распределенной памятью. Для уменьшения сложности разработки могут применяться различные инструменты и технологии. В качестве примера подобной технологии можно выделить использование распределённого глобального адресного пространства. В работе предлагается реализация распределённого глобального адресного пространства на основе MPI в виде библиотеки C++. Эффективность реализации демонстрируется на ряде классических примеров. На основе примеров дается рекомендация по использованию библиотеки при разработке программного обеспечения.

Ключевые слова: высокопроизводительные системы, параллельные вычисления, PGAS, MPI.

1. Введение

Область разработки реализаций алгоритмов, использующих все доступные мощности высокопроизводительных систем, является актуальной и интересной для исследований. Полноценное использование высокопроизводительных систем без использования механизмов распараллеливания вычислений невозможно. В зависимости от архитектуры вычислительной системы можно выделить несколько основных видов высокопроизводительных систем - системы с общей памятью, системы с распределённой памятью, а также их гибриды. Для каждого вида систем существуют свои технологии распределения вычислений.

В случае систем с общей памятью вычисления распределяют за счет использования потоков. Для таких систем существует множество технологий, из которых можно выделить OpenMP [1], TBB [2], Intel oneAPI² [3].

Для разработки программного обеспечения в системах с распределённой памятью часто используется MPI [4]. При использовании MPI разработчику необходимо как разработать последовательность действий каждого процесса с учетом балансировки нагрузки, так и обеспечить корректное распределение и обмен данных. Как следствие, сложность разработки распределенных приложений является существенной. Для её уменьшения могут применяться различные инструменты и технологии.

В качестве примера подобной технологии можно выделить использование распределённого глобального адресного пространства (PGAS) [4-6]. Существует несколько примеров технологий и инструментов, основанных на PGAS [7-13]. В рамках статьи рассматривается реализация распределённого глобального адресного пространства на языке C++ с использованием средств MPI [14]. Эффективность разработанной библиотеки показана на решении ряда классических примеров.

2. Постановка задачи

В рамках работы необходимо реализовать библиотеку, поддерживающую работу с распределенным адресным пространством. Выполнение работы предполагает реализацию основных

^{1*} Вычислительные эксперименты проводились с использованием ресурсов МСЦ РАН и СК «Лобачевский».

² Intel OneAPI также применяется для других архитектур, в том числе для графических ускорителей и гибридных систем.

компонент библиотеки, рассмотрение возможностей оптимизации её работы по времени и используемой памяти, а также реализацию ряда примеров. На основе данных примеров необходимо провести вычислительные эксперименты, оценить эффективность реализации и дать рекомендации по использованию библиотеки при разработке программного обеспечения.

3. Программная реализация

3.1. Основные этапы разработки

В ходе работы над библиотекой было выделено несколько ключевых компонентов: распределённый вектор, менеджер памяти, а также несколько дополнительных компонентов, использующих PGAS, таких как распределённая параллельная очередь и функции редукции данных.

Распределённый вектор предоставляет пользователю доступ к распределённому адресному пространству. Данный компонент является промежуточным звеном между пользователем и менеджером памяти. Вектор позволяет разбить память на кванты – неделимые блоки данных, состоящие из последовательно идущих элементов распределённого вектора.

Менеджер памяти контролирует распределение памяти между процессами и определяет протоколы взаимодействия процессов при обращении к распределённой памяти. В рамках данного компонента используется организация параллельных вычислений по схеме «мастер-рабочий». Процесс-мастер обрабатывает запросы на организацию передачи данных между процессами и регулирует пересылки данных в распределённой памяти. Процессы-рабочие отправляют запросы на осуществление деятельности с распределёнными данными процессу-мастеру, пересылают друг другу данные и определяют механизм работы с ними на одном процессе. В менеджере памяти также реализован ряд механизмов работы с памятью, в том числе механизм режимов работы с памятью и механизм кеширования данных.

3.2. Механизм режимов работы с памятью

Некоторые алгоритмы позволяют в ходе своей работы выделить несколько этапов вычислений. Этапы могут содержать как чтение и запись общих для параллельных процессов данных, так и выполнять только чтение. Было принято решение реализовать механизм режимов работы с памятью «чтение и запись» и «только чтение» (рис. 1). В первом режиме данные могут находиться одновременно только в одном экземпляре на одном процессе. Второй режим позволяет создавать копии одного и того же кванта памяти на разных процессах. Так как в этом случае запись в квант памяти запрещена, значения элементов в копиях остаются актуальными.

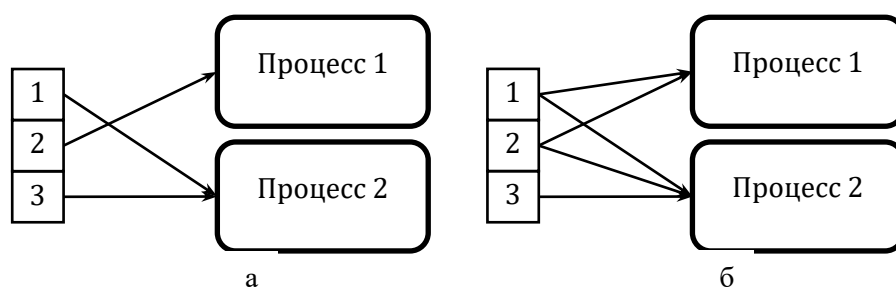


Рис. 1. Схемы распределения памяти в режимах работы «чтение и запись» (а) и «только чтение» (б)

3.3. Реализация кешей

Для оптимизации работы библиотеки по используемой памяти было решено ограничить количество одновременно хранящихся квантов в режиме «только чтение». Для каждого кванта, переходящего в данный режим, обозначается процесс-родитель, с которого он не может быть

удалён. В текущей реализации таким процессом считается тот процесс, на котором квант хранился при переходе из режима «чтение и запись» в режим «только чтение».

Изначально в кеше выделен пул свободных элементов, которые могут хранить индексы добавляемых в него квантов. При добавлении нового кванта в кеш берётся один элемент из данного пула. В случае, если кеш заполнен (в нём не осталось свободных элементов) и в него поступает новый квант, он вытеснит из кеша один из квантов. Если какой-либо квант, находящийся в кеше, переходит в режим работы «чтение и запись», он добавляется в список исключённых квантов. Элемент кеша, хранящий индекс данного кванта, освобождается и переходит в пул свободных элементов.

3.4. Сбор статистики

В рамках предлагаемой библиотеки реализован механизм сбора статистики об исполняемых приложениях. В текущей реализации возможен сбор следующих данных:

- 1) Время, в которое на конкретном процессе произошёл кеш-промах. Записывается в отдельный для каждого процесса файл;
- 2) Общее количество кеш-промахов отдельно для каждого процесса и для каждого распределённого вектора. Кроме того, для каждого распределённого вектора также запоминается и записывается в файл информация об итоговом числе кеш-промахов, равным сумме кеш-промахов на каждом процессе;
- 3) Общее количество вытеснений из кеша отдельно для каждого процесса и для каждого распределённого вектора. Также, как и для кеш-промахов, дополнительно записывается информация об итоговом суммарном количестве вытеснений из кеша.

4. Результаты вычислительных экспериментов

Эксперименты были проведены на вычислительных узлах со следующими характеристиками: каждый вычислительный узел имеет 2 процессора Intel(R) Xeon(R) CPU E5-2660 0, 2.20GHz (8 ядер на каждый процессор, всего 16 ядер), 64 GB RAM.

Так как в разработанной библиотеке используется организация параллельных вычислений по схеме «мастер-рабочий», запуск на одном процессе невозможен. Поэтому в примерах ускорение считается относительно времени работы в 2 процесса.

Были проведены следующие вычислительные эксперименты:

1. Сумма элементов вектора с использованием функции `parallel_reduce`.
2. Умножение матриц методом Фокса.

4.1. Сумма элементов вектора с использованием функции `parallel_reduce`

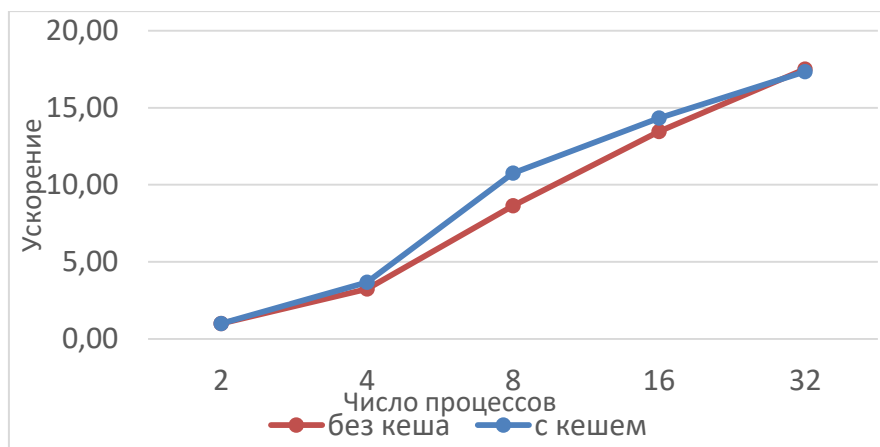


Рис. 2. Сравнение значений ускорения в случаях с использованием и без использования кеша, сумма элементов матрицы с использованием `parallel_reduce`

В данном примере элементы распределённого вектора перед переходом в режим «только чтение» были равномерно распределены по каждому процессу-работочему. Каждый процесс-работчий обращается только к своей порции элементов. Случаи, когда один квант находится на нескольких процессах, могут быть, только если элементы из данного кванта были распределены на разные процессы. Кеш не успевает наполниться и не вносит замедлений в ход работы программы. Соответственно, в случае с запусками с системой кешей и без неё дают практически идентичные результаты. В обоих случаях ускорение близко к линейному.

4.2. Умножение матриц методом Фокса

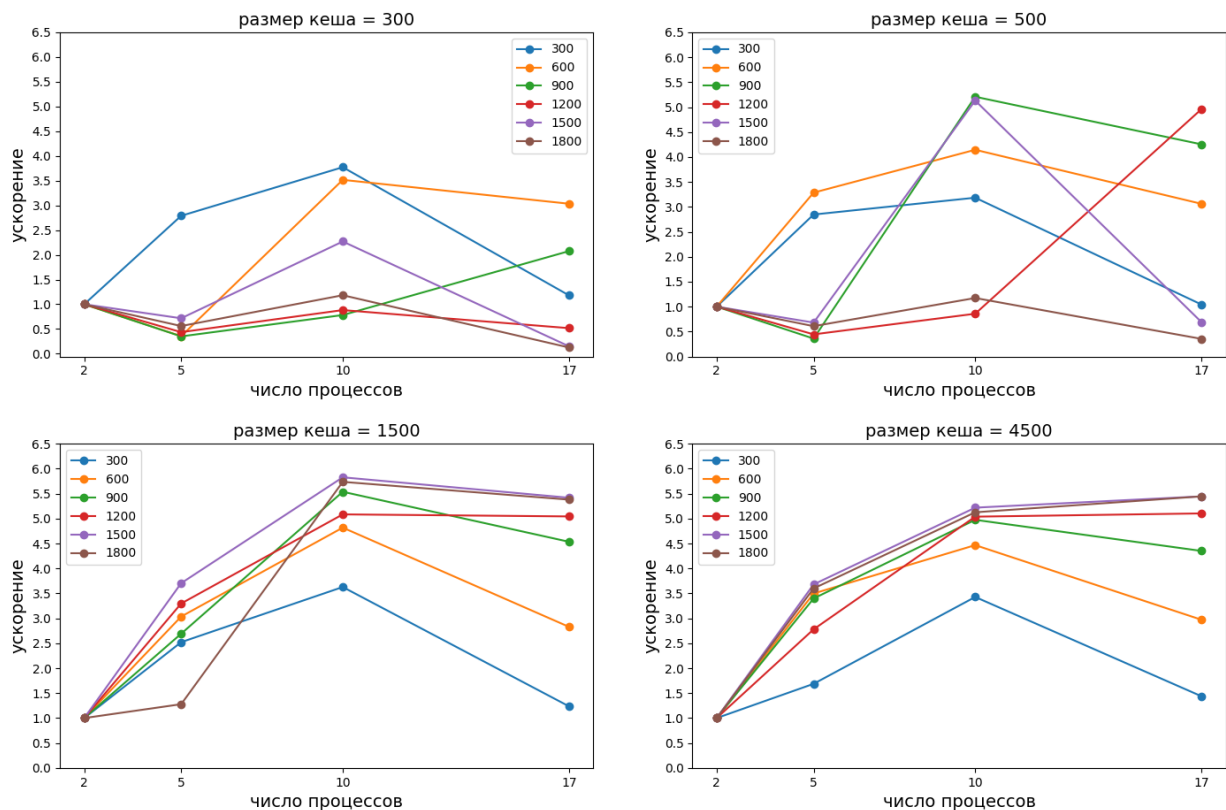


Рис. 3. Ускорение для умножения матриц методом Фокса по сравнению с запуском в 2 процесса при разных размерах кеша и количестве элементов в матрицах

В данном примере кванты могут перемещаться с одного процесса на другой в ходе работы программы. Соответственно, кеш может повлиять на время работы. Как видно из графиков, при уменьшении размера кеша ускорение значительно падает. Более того, на графике с размером кеша, равным 500 элементам, видно, что ускорение может значительно варьироваться при разном количестве элементов. Это связано с количеством кеш-промахов и количеством вытеснений из кеша. При увеличении размера кеша значения изменения ускорения стабилизируются. На графике с размером кеша = 4500 вытеснений из кеша нет, так как его размер достаточно большой, чтобы в нём смогли поместиться все копии квантов на каждом процессе.

Таким образом, в приложениях, использующих режим работы с памятью «только чтение», необходимо учитывать размер кеша, который ограничивает количество затрачиваемой во время работы программы памяти, но также и влияет на производительность. Целью дальнейших исследований является изучение применения технологий машинного обучения с целью составления расписания передачи данных между процессами, уменьшения числа вытеснений из кеша и количества обмена данных между процессами и увеличения производительности библиотеки.

5. Заключение

В ходе текущей работы над библиотекой были достигнуты следующие задачи:

1. Изучены существующие программные средства и технологии разработки параллельного программного обеспечения;
2. Реализована библиотека для поддержки работы с распределённым адресным пространством;
3. Реализован функционал кеширования для режима работы «только чтение»;
4. Реализован механизм сбора статистики по кеш-промахам и вытеснениям квантов из кеша;
5. Выполнены вычислительные эксперименты для ряда примеров использования библиотеки с целью проверки эффективности работы библиотеки. Проведено сравнение значения ускорения для случаев с использованием и без использования кеша. Из результатов экспериментов можно сделать вывод, что влияние кеша на ускорение заметно в тех случаях, когда данные активно перераспределяются в режиме работы «только чтение». В случае, когда данные заранее распределяются по процессам в начале работы программы, а затем практически не меняют своего расположения, кеш не влияет на время работы программы.

В дальнейшем планируется реализация следующих этапов работы:

1. Продолжение развития механизмов сбора статистики;
2. Исследование возможности использования методов машинного обучения для составления оптимального расписания передачи данных;
3. Реализация новых примеров использования библиотеки.

Литература

1. Dagum, L. OpenMP: an industry standard API for shared-memory programming. / L. Dagum, R. Menon // IEEE computational science and engineering. – 1998. – Vol. 5, no. 1. – PP. 46–55.
2. Willhalm, T. Putting intel® threading building blocks to work. / T. Willhalm, N. Popovici // Proceedings of the 1st international workshop on Multicore software engineering. – 2008. – PP. 3–4.
3. Reinders, J., Ashbaugh, B., Brodman, J., Kinsner, M., Pennycook, J., Tian, X. Data Parallel C++ // Apress Berkeley, CA. – 2008. 548 P.
4. MPI – the Complete Reference: the MPI core / M. Snir, W. Gropp, S. Otto et al. – MIT press, 1998. – Vol. 1.
5. Saraswat V. et al. The asynchronous partitioned global address space model //The First Workshop on Advances in Message Passing. – 2010. – PP. 1-8.
6. Almasi, G. PGAS (Partitioned Global Address Space) Languages. – 2011 / G. Almasi.
7. Mellor-Crummey J. et al. A new vision for Coarray Fortran //Proceedings of the Third Conference on Partitioned Global Address Space Programing Models. – 2009. – C. 1-9.
8. Deitz, S. J. Chamberlain B. L., Hribar M. B. Chapel: Cascade High-Productivity Language An Overview of the Chapel Parallel Programming Model / S. J. Deitz, B.L. Chamberlain, M. B. Hribar //Cray User Group. – 2006.
9. Charles, P. X10: an object-oriented approach to non-uniform cluster computing / P. Charles., C. Grothoff, V. Saraswat [et al] //Acm Sigplan Notices. – 2005. – T. 40. – №10. – PP. 519-538.
10. Saraswat, Vijay. X10 Language Specification. / Vijay Saraswat, Bard Bloom, Igor Peshansky [et al] – 2019.
11. Carlson, W. W. Introduction to UPC and language specification. / W. W. Carlson, J. M. Draper, D. E Culler [et al] – Technical Report CCS-TR-99-157, IDA Center for Computing Sciences – 1999. – T. 576.
12. Bonachea, D. et al. UPC language and library specifications, version 1.3. / D. Bonachea, G. Funck [et al]. – 2013.
13. Fuchs, T. A multi-dimensional distributed array abstraction for PGAS / T. Fuchs, K. Furlinger. //2016 IEEE 18th International Conference on High Performance Computing and Communications; IEEE 14th International Conference on Smart City; IEEE 2nd International Conference on Data Science and Systems (HPCC/SmartCity/DSS). – IEEE, 2016. – C. 1061-1068.

14. Репозиторий библиотеки поддержки распределённого адресного пространства на GitHub.
URL: https://github.com/skonnov/PGAS_support_library (Дата обращения: 30.05.2022).

О НЕКОТОРЫХ СТРАТЕГИЯХ ЧИСЛЕННОГО ИНТЕГРИРОВАНИЯ, ОСНОВАННЫХ НА ИСПОЛЬЗОВАНИИ КОНЕЧНО-РАЗНОСТНОЙ ФОРМУЛЫ

А.Г. Коротченко, В.М. Сморякова

Нижегородский государственный университет им. Н.И. Лобачевского

В статье рассматривается концепция одношагового выбора при построении стратегий численного интегрирования, основанных на использовании конечно-разностной формулы. Статья является продолжением работ авторов, посвященных данной тематике. Приводятся результаты вычислительных экспериментов по использованию улучшенных стратегий. Эксперименты проводятся на новом тестовом наборе.

Ключевые слова: концепция одношаговой оптимальности, численное интегрирование, конечно-разностные формулы.

1. Введение

Вопросы, связанные с построением алгоритмов, обладающих заданными свойствами (определяемыми критериями, оценивающими качество работы алгоритма) и предназначенными для решения задач различных классов, представляют интерес как с точки зрения теоретических исследований, так и с точки зрения практических приложений. С точки зрения теории такие исследования позволяют оценивать потенциальные возможности алгоритмов для соответствующих классов задач, а с точки зрения практики позволяют использовать их с учетом тех свойств, которыми они обладают в результате их построения. Данной проблеме посвящено достаточно много работ, укажем здесь лишь некоторые работы авторов [1-6], в которых содержится ряд ссылок на соответствующие статьи.

Одна из возможных схем построения такого рода алгоритмов основана на концепции одношагового выбора, состоящего в то, что результат работы алгоритма конструируется, с учетом критериев, оценивающих качество его работы, на каждом шаге алгоритма. Основным достоинством такой концепции является то, что она позволяет перестраивать модель вычислений при сохранении базовых свойств алгоритма.

В качестве одного из критериев, оценивающих качество работы алгоритма может выступать требование, связанное с оценкой точности решения задачи на шаге алгоритма. Еще одним критерием может быть требование простоты его реализации. Другими словами, задача построения алгоритма, по существу, является бикритериальной.

С учетом указанных критериев в [1, 2] реализована схема, основанная на концепции одношагового выбора при построении алгоритмов поиска наибольшего значения в классах многоэкстремальных функций. Здесь результатом работы алгоритма являются точки (точка), в которых следует вычислять значения функции на шаге алгоритма, а результат его работы зависит от параметров, с помощью которых задается класс функций. При этом указанные параметры могут меняться от шага к шагу.

В [5] описан один из адаптивных вариантов вычисления этих параметров с учетом информации о максимизируемой функции, получаемой в ходе реализации алгоритма.

В данной статье концепция одношагового выбора применяется для построения стратегий численного интегрирования систем обыкновенных дифференциальных уравнений, основанных на использовании конечно-разностной формулы. Статья является продолжением работ авторов [3, 4, 6].

2. Постановка задачи

Пусть требуется найти решение $Y = Y(x)$ (предполагается, что решение $Y(x)$ – единственно и четырежды непрерывно дифференцируемо на отрезке $[x_0, t]$) задачи Коши для системы обыкновенных дифференциальных уравнений:

$$\begin{aligned} Y' &= F(x, Y), Y(x_0) = Y_0, \\ Y &= Y(x) = (y^1(x), \dots, y^n(x)), x_0 \leq x < t, \end{aligned} \quad (1)$$

где значение параметра t не известно априори, а определяется в процессе интегрирования системы. В случае, когда система (1) линейна, $F(x, Y) = B(x)Y + b(x)$, где $B(x)$ – $(n \times n)$ -матрица, $b(x)$ – n -мерный вектор.

Для определения численных значений $Y_i = (y_1^i, \dots, y_n^i)$, $i = 1, 2, \dots$, решения системы (1) будем использовать метод, основанный на применении конечно-разностной формулы четвертого порядка:

$$Y_i = a_{i-3}Y_{i-3} + a_{i-2}Y_{i-2} + a_{i-1}Y_{i-1} + h_i a_i q_i, \quad (2)$$

здесь

$$\begin{aligned} q_i F &= (x_i, Y_i), Y_i = (y_1^i, y_2^i, \dots, y_n^i), \\ a_{i-3} &= 1 - a_{i-2} - a_{i-1}, a_{i-2} = \frac{-h_i^2 \psi_i^2}{h_{i-2}h_{i-1}(h_{i-1}^2 + h_{i-1}h_{i-2} + 2h_i h_{i-2} + 3h_i^2 + 4h_{i-1}h_i)}, \\ a_{i-1} &= \frac{(h_i + h_{i-1})^2 \psi_i^2}{h_{i-1}(h_{i-2} + h_{i-1})(h_{i-1}^2 + h_{i-1}h_{i-2} + 2h_i h_{i-2} + 3h_i^2 + 4h_{i-1}h_i)}, \\ a_i &= \frac{(h_i^2 + h_{i-1}h_i)\psi_i}{h_i(h_{i-1}^2 + h_{i-1}h_{i-2} + 2h_i h_{i-2} + 3h_i^2 + 4h_{i-1}h_i)}. \end{aligned}$$

при этом $\psi_i = h_{i-2} + h_{i-1} + h_i$, $h_i = x_i - x_{i-1}$ – шаг интегрирования.

Будем считать, что система (1) удовлетворяет условиям, при выполнении которых вектор $Y_i = (y_1^i, \dots, y_n^i)$ может быть определён из решения системы уравнений

$$Y_i = a_{i-3}Y_{i-3} + a_{i-2}Y_{i-2} + a_{i-1}Y_{i-1} + h_i a_i F(x_i, Y_i). \quad (3)$$

Если система (1) линейна, то (3) принимает вид

$$C_i Y_i = c_i, \quad (4)$$

где $C_i = E - h_i a_i B(x_i)$, $c_i = a_{i-3}Y_{i-3} + a_{i-2}Y_{i-2} + a_{i-1}Y_{i-1} + h_i a_i b(x_i)$, E – единичная матрица.

Будем считать, что система (3) может быть решена методом Ньютона, а система (4) – методом Гаусса.

В качестве основной характеристики формулы (2) будем рассматривать локальную ошибку, получаемую на i -ом шаге для каждой j -ой компоненты решения $Y_i = (y_1^i, \dots, y_n^i)$:

$$\begin{aligned} R_i^j &= \frac{h_i^6 + 2(h_{i-2} + 2h_{i-1})h_i^5 + 2h_{i-1}(h_{i-2} + 2h_{i-1})h_i^4}{3h_i^2 + 2(h_{i-2} + 2h_{i-1})h_i + (h_{i-1}^2 + h_{i-1}h_{i-2})} + \\ &+ \frac{2h_{i-1}(h_{i-2} + h_{i-1})(h_{i-2} + 2h_{i-1})h_i^3 + h_{i-1}^2(h_{i-2} + h_{i-1})^2 h_i^2}{3h_i^2 + 2(h_{i-2} + 2h_{i-1})h_i + (h_{i-1}^2 + h_{i-1}h_{i-2})} (y^{IV}(\Theta_i))^j, \end{aligned} \quad (5)$$

где $x_{i-2} - h_i \leq \Theta_i \leq x_{i-1} + h_i$ и $(y^{IV}(\Theta_i))^j$ – значение четвёртой производной от j -ой компоненты вектора решения $Y(x)$ в точке Θ_i . Локальная ошибка (5) возникает вследствие конечно-разностной аппроксимации производных системы дифференциальных уравнений.

Потребуем, чтобы в процессе интегрирования системы выполнялись условия:

$$\max_{1 \leq j \leq n} |R_i^j| \leq \varepsilon_i,$$

где ε_i – заданная точность на i -ом шаге интегрирования.

Пусть четвёртая производная от компонент вектора решения $Y(x)$ удовлетворяет на отрезке $[x_0, x_0 + z]$ условию $\max_{1 \leq j \leq n} |y^{IV}(x)^j| \leq K$, где $z > 0$, $K > 0$ – вещественные константы.

Тогда получаем, что ограничение на формулу (2), обусловленное точностью вычислений на i -ом шаге интегрирования, сводится к выполнению неравенства:

$$\begin{aligned} \varphi_i(h_{i-2}, h_{i-1}, h_i) &= h_i^6 + 2(h_{i-2} + 2h_{i-1})h_i^5 + 2h_{i-1}(h_{i-2} + 2h_{i-1})h_i^4 + \\ &+ 2h_{i-1}(h_{i-2} + h_{i-1})(h_{i-2} + 2h_{i-1})h_i^3 + h_{i-1}^2(h_{i-2} + h_{i-1})^2 h_i^2 - 3\Delta_i h_i^2 - \\ &- 2(h_{i-2} + 2h_{i-1})\Delta_i h_i - (h_{i-1}^2 + h_{i-1}h_{i-2})\Delta_i \leq 0 \end{aligned} \quad (6)$$

здесь $\Delta_i = 2\varepsilon_i K^{-1}$, $i = 1, 2, \dots$

Обозначим через $h_{2i-2} = \tau_{i-1}$, $h_{2i-1} = \tau_i$, $h_{2i} = \tau_i$ в функциях $\varphi_{2i}(h_{2i-2}, h_{2i-1}, h_{2i})$, определяемых в (6), и введём в рассмотрение следующие функции:

$$f_i(\tau_{i-1}, \tau_i) = 14\tau_i^5 + 12\tau_{i-1}\tau_i^4 + 3\tau_{i-1}^2\tau_i^3 - 8\Delta_i\tau_i - 3\Delta_i\tau_{i-1}, \quad (7)$$

при этом $\varphi_{2i}(h_{2i-2}, h_{2i-1}, h_{2i}) = \tau_i f_i(\tau_{i-1}, \tau_i)$, $i = 1, 2, \dots$

Тогда из выполнения неравенства $f_i(\tau_{i-1}, \tau_i) \leq 0$ при $\tau_i > 0$ и сделанных предположений относительно шагов интегрирования будет следовать выполнение соотношения (6).

Пусть $h_0 = \tau_0$, где τ_0 – заданное положительное число. Найдём решение системы (1) Y_{-1} , Y_{-2} , Y_{-3} в узлах интегрирования $x_{-1} = x_0 - h_0$, $x_{-2} = x_{-1} - h_0$, $x_{-3} = x_{-2} - h_0$, используя какой-либо одношаговый метод, например, метод Эйлера.

Определим процедуру выбора шагов интегрирования конечно-разностной формулы (2) на отрезке $[x_0, x_0 + z]$ при условии, что число узлов интегрирования априори не задано, а определяется в процессе интегрирования системы (1).

В случае, когда процесс интегрирования с использованием формулы (2) начинается с узла x_{2p} , $p \geq 1$, то при известном значении τ_{p+j-1} значение τ_{p+j} находим таким образом, чтобы выполнялось неравенство (7) при $i = p + j$, и полагаем $h_{2p+2j-1} = \tau_{p+j}$, $h_{2p+2j} = \tau_{p+j}$, где $h_{2p+2j-2} = \tau_{p+j-1}$, а $x_{2p+2j-1} = x_{2p} + h_{2p+2j-1}$, $x_{2p+2j} = x_{2p+2j-1} + h_{2p+2j}$, $j = 1, 2, \dots$

Если процесс интегрирования на отрезке $[x_{2p}, x_{2p} + z]$ заканчивается в узле x_{2m} , $m \geq 1$, то вектор $X(p, m) = (h_{2p+1}, h_{2p+2}, \dots, h_{2p+2m-1}, h_{2p+2m})$, получаемый с помощью описанной процедуры будем называть стратегией интегрирования или просто стратегией. Множество всех таких стратегий обозначим $\Omega_1(p, m)$.

Описанная процедура выбора шагов интегрирования основана на том, что номер завершающего узла интегрирования на каждом отрезке длины z является чётным. Данное обстоятельство не является существенным, т. е. указанный узел может быть и нечётным. В дальнейшем будем рассматривать случай, когда $p = 0$, что также не является существенным.

Будем считать, что процесс интегрирования с использованием формулы (2) заканчивается в узле x_{2m} . Таким образом, задание стратегии $X(0, m)$ определяет процедуру выбора шагов и узлов интегрирования на отрезке $[x_0, x_0 + z]$.

При этом будем считать, что процесс интегрирования на отрезке $[x_0, x_0 + z]$ заканчивается либо выполнением некоторого условия в узле x_{2m} , накладываемого на получаемое по формуле (2) решение системы (1), либо когда выполняется неравенство $x_{2m+2} > x_0 + z$.

Заметим, что если $x_2 > x_0 + z$, то в ограничении (6) следует уменьшить значение ε_1 .

3. Оптимальная стратегия

Рассмотрим одну из схем выбора τ_i , определяющих задание стратегии из множества $\Omega_1(0, m)$.

Определим стратегию $\bar{X}(0, m) = (\bar{h}_1, \bar{h}_2, \dots, \bar{h}_{2m-1}, \bar{h}_{2m}) \in \Omega_1(0, m)$ таким образом, что $\bar{h}_{2i-1} = \bar{\tau}_i$, $\bar{h}_{2i} = \bar{\tau}_i$, $i = 1, \dots, m$, где при заданном значении $\bar{\tau}_0$ величина $\bar{\tau}_i$ является решением задачи

$$\sum_{i=1}^m \tau_i \Rightarrow \max, \quad f_i(\bar{\tau}_{i-1}, \tau_i) \leq 0, i = 1, \dots, m. \quad (8)$$

Здесь функции $f_i(\tau_{i-1}, \tau_i)$ заданы в (7) соответственно.

Стратегию $\bar{X}(0, m)$ будем называть оптимальной стратегией на множестве $\Omega_1(0, m)$, если для любого $m = 1, 2, \dots$ последовательность $(\bar{\tau}_1, \dots, \bar{\tau}_m)$ является решением задачи

$$f_i(\tau_{i-1}, \tau_i) \leq 0, i = 1, \dots, m, \quad (9)$$

Такой выбор τ_i , $i = 1, \dots, m$, соответствует минимальному числу узлов интегрирования.

Отметим здесь, что последовательное решение локальных задач (8) дает решение интегральной задачи (9). Для решения задачи (8) на каждом шаге необходимо находить корень уравнения

$$f_i(\bar{\tau}_{i-1}, \tau_i) = 0, i = 1, \dots, m. \quad (10)$$

Для реализации оптимальной стратегии $\bar{X}(0, m)$ на отрезке $[x_0, x_0 + z]$ необходимо при заданном значении $\bar{t}_0$ находить $\bar{t}_i$, как единственный положительный корень уравнения (10), где $i = 1, \dots, m$, и нужно знать оценку четвертой от компонент решения системы дифференциальных уравнений (1). Будем находить указанные оценки, используя в качестве аппроксимации решения интерполяционный полином Лагранжа четвертой степени. Для этого нужно знать решение $Y(x_0 + z)$, которое будем определять по формуле (2), где решение Y_{-3}, Y_{-2}, Y_{-1} в узлах x_{-3}, x_{-2}, x_{-1} уже найдены с помощью какого-либо одношагового метода, как было указано выше. Заметим, что Y_0 - начальное условие в узле x_0 .

Если процесс интегрирования на отрезке $[x_0, x_0 + z]$ завершается выполнением неравенства $x_{2m+2} > x_0 + z$, то переходим к новому отрезку, где в качестве x_0 выбирается значение x_{2m} , и так далее, пока не выполнится условие окончания процесса интегрирования. Назовём описанную процедуру α алгоритмом.

4. Упрощенные стратегии

Рассмотрим более простую процедуру вычисления последовательности τ_i . Уравнение (10) можно представить в следующем виде

$$14\mu_i^5 + 12\mu_i^4 + 3\mu_i^3 - 8\mu_i\eta_i - 3\eta_i = 0, \quad (11)$$

где

$$\mu_i = \frac{\tau_i}{\tau_{i-1}}, \eta_i = \frac{2\varepsilon_i}{\tau_{i-1}^4 K}. \quad (12)$$

Уравнение (11) неявно задаёт функцию $\mu_i(\eta_i)$ при $\eta_i \geq 0$.

Тогда можно определить последовательность

$$\tau_i = \mu_i(\eta_i)\tau_{i-1}. \quad (13)$$

Указанная функция является монотонно возрастающей вогнутой функцией при $\eta_i \geq 0$. Учитывая данный факт, может быть предложен простой алгоритм $\alpha_1(W, \eta)$, основанный на алгоритме α , где W – множество характерных узлов, в которых формируется таблица значений функции $\mu_i(\eta_i)$ на отрезке $[0, \eta]$, учитывающих изменения текущего значения τ_{i-1} в формуле (13). Значения функции $\mu_i(\eta_i)$ в указанных узлах вычисляется путём решения уравнения (11), а в промежуточных точках отрезка $[0, \eta]$ приближённые значения этой функции вычисляются с помощью линейной интерполяции. Здесь $\mu_i(\eta)$ – величина, определяющая максимальное увеличение текущего значения τ_{i-1} в формуле (13). Другими словами, если значение η_i в (12) становится больше η , то тогда в (13) значение $\mu_i(\eta_i)$ полагаем равным $\mu_i(\eta)$.

Таким образом, используя сформированную таблицу, так как это было описано выше приходим к алгоритму $\alpha_1(W, \eta)$, в котором не требуется решать вспомогательную задачу на каждом этапе формирования τ_i . В алгоритме $\alpha_1(W, \eta)$ переход к новому отрезку длины z осуществляется так же, как и в алгоритме α .

Рассмотрим свойство последовательности τ_i , получаемой с помощью уравнений (10). Будем считать, что значение $\Delta_i = \Delta$ фиксировано на отрезке длины z , т. е. для каждого i -ого шага значение $\varepsilon_i = \varepsilon$, $\varepsilon > 0$. Пусть $\tilde{\tau} = \left(\frac{4}{7}\Delta\right)^{\frac{1}{4}}$, $\tilde{\tau} < z$. Тогда уравнение (10) при фиксированном τ_{i-1} имеет единственный положительный корень на отрезке $[0, \tilde{\tau}]$. Пусть g – отображение отрезка $[0, \tilde{\tau}]$ в себя, определяемое уравнением

$$14g^5(\tau) + 12\tau g^4(\tau) + 3\tau^2 g^3(\tau) - 8\Delta g(\tau) - 3\Delta\tau = 0, \quad (14)$$

т. е. $g(\tau)$ есть единственный положительный корень уравнения (14), где $\tau \in [0, \tilde{\tau}]$.

Можно показать, что отображение g на отрезке $[0, \tilde{\tau}]$ является сжимающим отображением и имеет единственную неподвижную точку $\tau^* = \left(\frac{11}{29}\Delta\right)^{\frac{1}{4}}$. Таким образом, определяется последовательность процессом интегрирования $\tau_i = g(\tau_{i-1})$, сходящаяся к неподвижной точке $\tau^* = \left(\frac{11}{29}\Delta\right)^{\frac{1}{4}}$.

Данный факт позволяет при выходе на неподвижную точку выбирать шаг равным $\tau^* = \left(\frac{11}{29}\Delta\right)^{\frac{1}{4}}$.

5. Вычислительный эксперимент

В рамках исследования был проведён вычислительный эксперимент, который заключался в запуске стратегий интегрирования, основанных на оптимальном выборе шага и сравнении с реализацией методов ode15s и ode45, предоставляемых средой MATLAB [8].

Для изучения использовался набор тестов, описанных в [7]. В данном тестовом наборе представлены системы дифференциальных уравнений, имеющих различный характер поведения решения: быстро осциллирующие, осциллирующие с быстрым затуханием, содержащие интервалы медленного и быстрого изменения компонент, имеющих быстро нарастающие или быстро затухающие компоненты.

Обозначим через α_2 – алгоритм, в котором применяются обе стратегии, описанные выше.

Будем сравнивать результаты, используя следующие критерии: число вычисления правых частей M и максимальную разницу между точным и полученным решением δ .

Приведем здесь на рис. 1 результаты, полученные для 1, 2, 6, 7, 10 и 23 тестов, описанных в [7]. Результаты работы алгоритмов ode15s, ode45 и α_2 изображены как круги, треугольники и квадраты соответственно. Для представления δ используется логарифмическая шкала. Начальный шаг равен 0,000001, относительная погрешность равна 0,01.

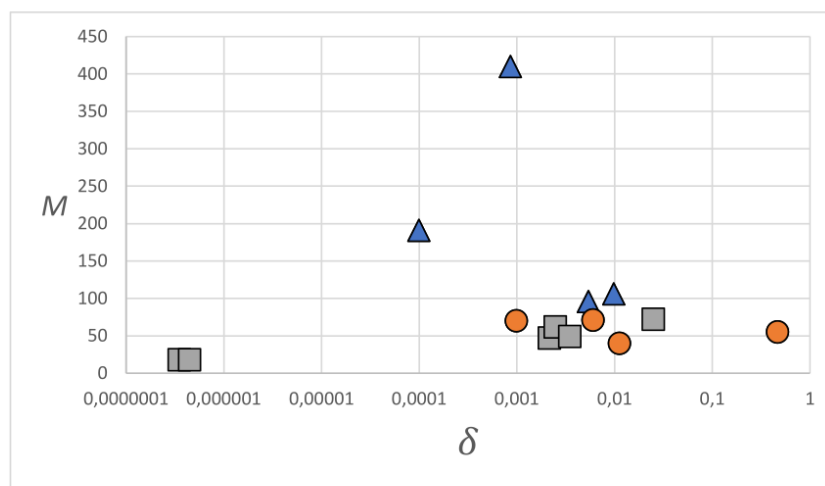


Рис. 1 Результаты эксперимента

Как было показано в [4] метод α_1 (оптимальная стратегия без каких-либо модификаций) применим тогда, когда нужно получить грубую оценку поведения решения на отрезке, не заданном априори, так как требует меньшее число вычислений, чем методы MATLAB. Как видно из рисунка, метод α_2 сохраняет указанное поведение. Таким образом, предложенные модификации могут быть использованы в качестве замены методы α_1 .

Литература

1. Коротченко А.Г., Сморякова В.М. Об оценке погрешности алгоритмов поиска экстремума в классах функций, определяемых кусочно-линейной мажорантой. Вестник Нижегородского университета им. Н.И. Лобачевского. 2013. № 3-1. С. 188-194.
2. Коротченко А.Г., Сморякова В.М. Об одном алгоритме поиска максимума в классе функций, определяемом кусочно-линейной мажорантой. Вестник Нижегородского университета им. Н.И. Лобачевского. 2014. № 4-1. С. 409-415.
3. Korotchenko A.G., Smoryakova V.M. On a method of construction of numerical integration formulas. AIP Conference Proceedings. 2016. vol. 1776., №9780735414389., pp. 090012-1-090012-4.

4. Korotchenko A.G., Smoryakova V.M. On a comparison of several numerical integration methods for ordinary systems of differential equations. *Lecture Notes in Computer Science*. 2020. vol. 2., №11974., pp. 406-412.
5. Коротченко А.Г., Сморякова В.М. Об оценке констант при задании классов функций, определяемых кусочно-линейной мажорантой // Проблемы теоретической кибернетики Материалы XIX международной конференции. Т. 1. 2021. С. 72-75.
6. Коротченко А.Г., Сморякова В.М. Об одной стратегии численного интегрирования, основанной на использовании конечно-разностной формулы. В сборнике: Математическое моделирование и суперкомпьютерные технологии. Труды XXI Международной конференции. Нижний Новгород, 2021. С. 174-179.
7. Арушанян О.Б., Залеткин С.Ф., Калиткин Н.Н. Тесты для вычислительного практикума по обыкновенным дифференциальным уравнениям. *Вычислительные методы и программирование*. 2002. Т. 3, С. 11-19.
8. Shampine, L.F.: The MATLAB ODE Suite. *SIAM Journal on Scientific Computing*, 1997. vol.18, pp. 1-122.

ПРИМЕНЕНИЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ КЛАССИФИКАЦИИ НЕМАРКИРОВАННЫХ ЭЛЕМЕНТОВ ПИТАНИЯ^{1*}

А.А. Коротышева¹, В.Р. Милов², Ю.С. Егоров², А.Ю. Кербенева², М.С. Дубов³

¹*Нижегородский государственный университет им. Н.И. Лобачевского,*

²*Нижегородский государственный технический университет им. Р.Е. Алексеева,*

³*ООО «Мабекс»*

В статье представлено описание модели для классификации немаркированных элементов питания. Обученная модель позволяет классифицировать тип элементов питания на основе рентгеновских снимков.

Ключевые слова: машинное обучение, рентгеновские снимки, нейронная сеть, элементы питания.

1. Введение

В условиях изменения мировой экологической политики все более актуальной становится проблема переработки твердых бытовых отходов (ТБО), а также таких опасных отходов, как электрохимические источники электрического тока, к которым относятся гальванические элементы питания и электрические аккумуляторы.

Несмотря на то, что элементы питания (ЭП) должны утилизироваться отдельно от остальных бытовых отходов, удаленность пунктов сбора вынуждает большинство жителей не накапливать отработавшие батарейки в домашних условиях, а выбрасывать их вместе с обычным мусором.

Попадая на полигон, металлический корпус батарейки начинает разлагаться, этикетка, как правило, стирается. В течение нескольких лет под воздействием осадков происходит процесс выщелачивания металлов (цинк, магний, кадмий, олово, свинец), которые попадают в почву и водоемы, а оттуда проникают в систему водоснабжения, что оказывает сильное негативное воздействие на окружающую среду и человека.

Целью работы является создание модели для классификации немаркированных элементов питания.

2. Анализ способа классификации элементов питания

2.1. Рентгенография

Рентгенография – представляет собой один из эффективных способов исследования внутренней структуры объектов. Через исследуемые объекты проходят рентгеновские лучи, в результате чего на специальной пленке или бумаге формируется проекционное изображение (рентгеновский снимок). Рентгеновские лучи – это электромагнитные волны, энергия фотонов которых лежит на шкале электромагнитных волн в диапазоне между ультрафиолетовым излучением и гамма-излучением, что соответствует длинам волн от $\sim 10^2$ до $\sim 10^{-3}$ нм [1].

Типовая рентгеновская установка состоит из источника и приемника, а исследуемый объект, располагается между ними [2]. Генерация рентгеновского излучения происходит в рентгеновской трубке – электровакуумном приборе с двумя электродами (катод и анод), на которые подается высокое напряжение. Катод обычно выполняется из вольфрама и конструктивно совмещается с нитью накаливания, служащей источником электронов, перемещаемых сильным электрическим полем в сторону массивного стального анода. Электроны, ударяясь об анод,

^{1*} Работа выполнена при поддержке Фонда содействия развитию малых форм предприятий в научно-технической сфере (договор № 57ГСИИИС12-D7/72200 от 21.12.2021 г.).

резко тормозятся, вследствие чего часть кинетической энергии электронов переходит в рентгеновское излучение, одновременно с этим анод разогревается, при этом его массивность позволяет отводить тепло на внешний радиатор.

Рентгеновская трубка помещается в защитный корпус, конструктивно имеющий отверстие для направленного выхода рентгеновского излучения. Основные характеристики рентгеновских трубок – это максимально допустимое напряжение, подаваемое на трубку и ее мощность.

Плоско панельный детектор рентгеновского излучения представляет из себя матрицу сенсоров, состоящую из фотодиодов, перед которыми располагается сцинтиллятор, как правило, на основе иодида цезия или окси сульфида гадолиния. TFT матрица плоско панельного детектора представляет собой двумерную решетку сенсоров, каждый из которых работает в режиме конденсатора и накапливает заряд. Считывание общей картины с матрицы происходит попеременным подключением каждого сенсора к считывающему устройству. При этом происходит построчное считывание изображения один раз одну строку сенсоров за один такт, как правило, сенсоры делятся на зоны в соответствии с ячейками памяти микросхемы декодера.

Чем темнее элемент на экране, тем меньше он задерживает излучение, таким образом, можно говорить, что его поглощающая способность больше. Например, стальные элементы задерживают больше излучения, чем пластиковые. При одинаковой толщине вещества с большим атомным весом сильнее поглощают рентгеновские лучи, как показано на рис. 1 [3]. В связи с этим, свинец, имеющий атомный номер 207, как правило, применяется для защиты от рентгеновского излучения. Влияет так же и толщина слоя.

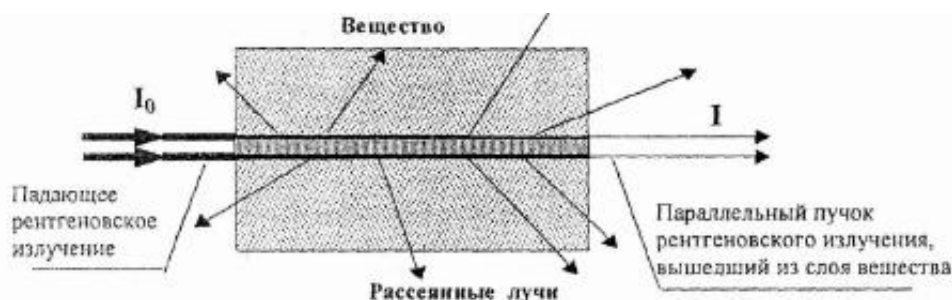


Рис. 1. Ослабление рентгеновского излучения

При увеличении напряжения на трубке энергия электронов возрастает, а излучение становится более проникающим. Излучение малой проникающей способности называют мягким, а большей – жестким.

На рис. 2 приведен пример рентгеновского снимка ЭП, выполненного на малой (40 kv), средней (60 kv) и высокой (80 kv) мощностях излучения.

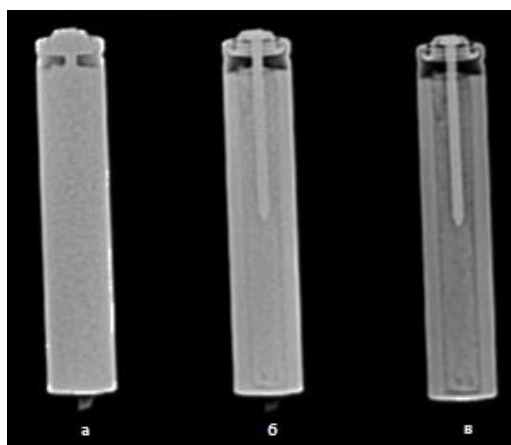


Рис. 2. Пример влияния мощности излучения:

а) малая; б) средняя; в) высокая

При малой мощности излучения лучи были поглощены образцом. Белый цвет изображения означает, что излучение не попало на датчик. При большой мощности излучения ЭП становится

ся прозрачным для излучения. Черный цвет на рентгеновском снимке означает, что предмет не задержал излучение. Восстановление белых, неосвещённых и черных пересечённых участков изображения в дальнейшем невозможно. В случае использования градаций серого становится возможным применение коррекции экспозиции, однако данное действие влечет за собой сброс всех калибровок оборудования, связанных с определением состава образца.

В качестве уникальных идентификаторов были выбраны следующие части ЭП: сепаратор, стержень, прокладка. При классификации ЭП за счет анализа контуров выполняется детекция групп пикселей, получивших промежуточные значения, относящих их к границам указанных частей ЭП.

2.2. Применение нейронных сетей

Распознавание образов на изображении – это задача компьютерного зрения, целью которой является отнесение объектов на изображении к определенному классу. Наиболее эффективный метод обработки изображений и выделения признакового описания – сверточные нейронные сети (англ. CNN, Convolutional Neural Network), состоящие из слоев свертки (англ. Convolutional Layer) и субдискретизации (англ. Pooling Layer). Сверточные слои и слои субдискретизации составляют основу различных архитектур глубоких нейронных сетей, обеспечивающих решение задачи распознавания образов, таких как VGG-16 и VGG-19, GoogLeNet с Inception модулями, ResNet, DenseNet, MobileNet и другие.

3. Разработка модели классификатора немаркированных ЭП

3.1. Описание исходных данных

Обучающие и тестовые выборки были собраны с помощью рентгеновского аппарата. Примеры изображений из датасета с соответствующими метками классов приведены на рис. 3.

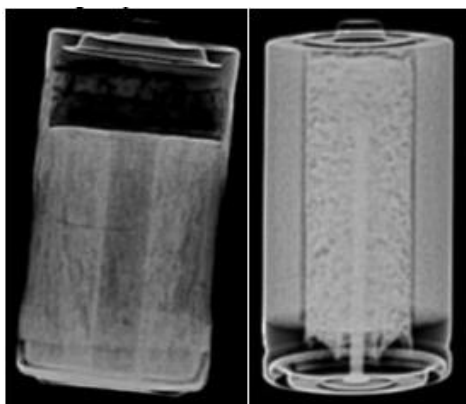


Рис. 3. Примеры изображений из датасета (слева – метка класса T1, справа – метка класса T2)

Для обучения нейросетевого классификатора было задано 5 классов, соответствующих типу состава ЭП: Zn-Mn (NH₄Cl), Zn-Mn (KOH), Ni-MH, Li-FeS₂, Ni-Cd. Ввиду редкости некоторых типов ЭП, в собранных данных наблюдалась проблема несбалансированности классов (англ. Imbalanced Data) [4], для решения которой были применены методы случайного удаления (англ. Random undersampling) для сокращения примеров мажоритарных классов и дублирования (Oversampling) для увеличения примеров миноритарных классов. В результате объем датасета составил 500 изображений.

Для предотвращения переобучения модели данные были разделены на обучающую (70% датасета), валидационную (20% датасета) и тестовую (10% датасета) выборки.

3.2. Описание архитектуры модели для классификации немаркированных ЭП

В работе была использована архитектура модифицированной MobileNet, позволяющая достигать той же точности распознавания, как исходная архитектура, но с еще большей скоростью работы [5].

MobileNetV2 имеет 2 типа сверточных блоков, называемых блоками с расширением (англ. Expansion Convolution Block) – с шагом 2 для снижения размерности и с шагом 1. Под шагами понимается значение параметра stride (шаг) в свертке. Каждый блок состоит из слоя точечной свертки размером 1x1, слоя глубинной свертки размером 3x3 с активацией ReLU и слоя свертки размером 1x1 с линейной активацией. Сверточный блок с шагом 1 также имеет соединение с последующей конкатенацией результатов аналогично остаточному блоку в архитектуре ResNet. Архитектура сверточных блоков MobileNetV2 представлена на рис. 4.

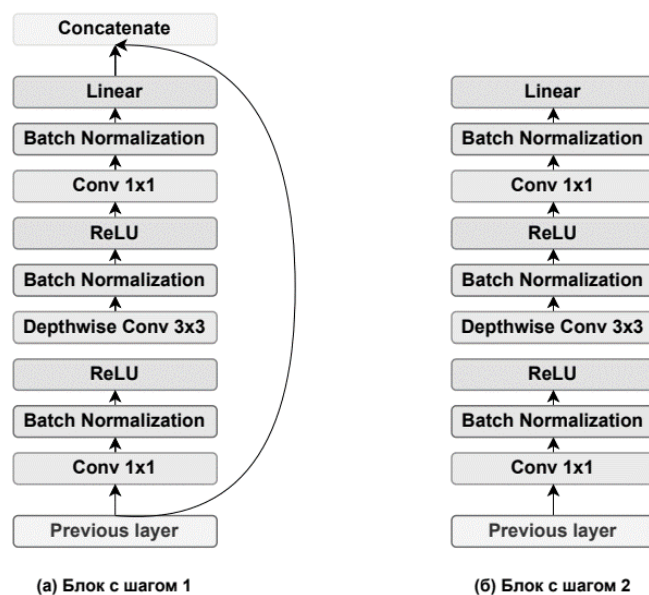


Рис. 4. Архитектура сверточных блоков с расширением MobileNetV2

Архитектура MobileNetV2 состоит из сверточных блоков с расширением и представлена на рис. 5.

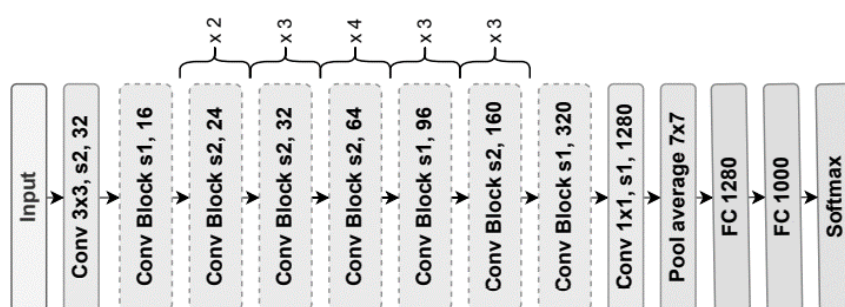


Рис. 5. Архитектура MobileNetV2

3.3. Тестирование классификатора ЭП

Для численной оценки качества классификатора ЭП вычислялась доля верно определенных типов ЭП на рентген снимках среди всех анализируемых рентген снимков (метрика Ассигасу). В результате тестирования, было получено значение Ассигасу=0,99.

Пример интерфейса разработанного прототипа ПО с функцией определения типа элемента питания по рентгеновскому снимку приведен на рис. 6.

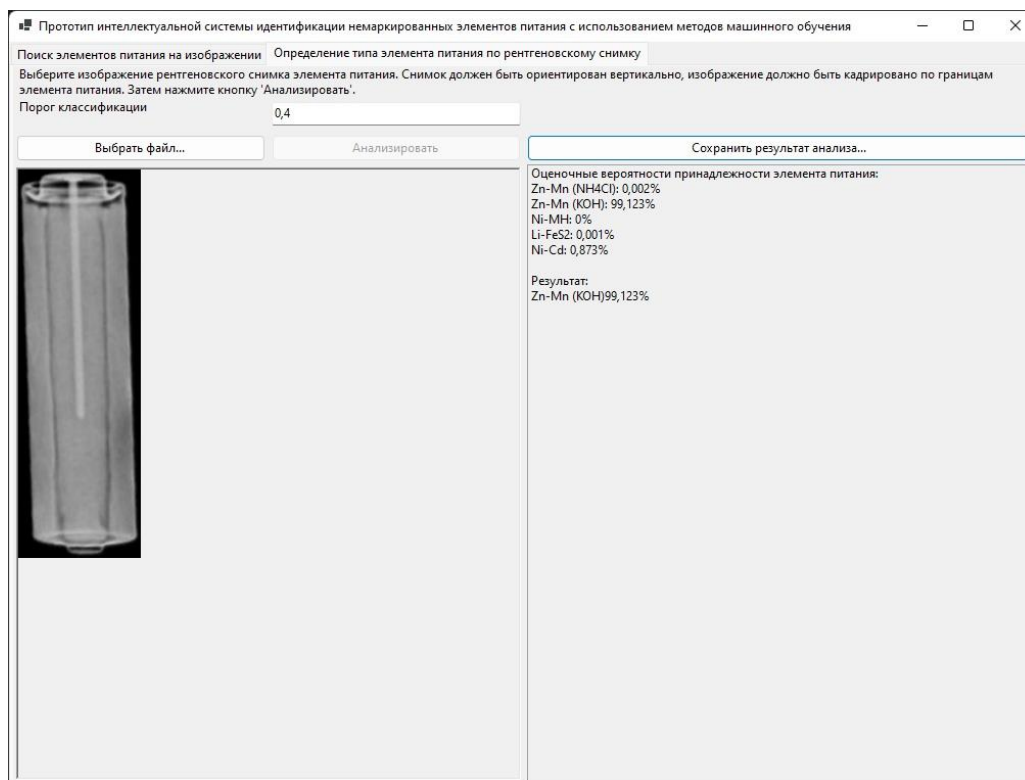


Рис. 6. Результат классификации ЭП по рентгеновскому снимку

4. Заключение

Разработанный классификатор немаркированных элементов питания характеризуется коэффициентом ошибок $K_{er}=0,01$ и может быть реализован в составе системы идентификации немаркированных элементов питания [6], которая, в свою очередь, станет основой программно-аппаратного комплекса для автоматизации сортировки опасных видов отходов, таких как ЭП.

Литература

1. Блохин М.А. Рентгеновское излучение // Физическая энциклопедия : [в 5 т.] / Гл. ред. А.М. Прохоров. – М.: Большая российская энциклопедия, 1994. Т. 4: Пойнтинга – Робертсона – Стримеры. С. 375–377. 704 с.
2. Схема работы рентгеновской установки. URL: http://oplib.ru/izobretatelstvo/view/166219_rentgenovskaya_dosmotrovaya_tehnika_v_rabote_tamozhennyh_organov (дата обращения: 24.02.2022).
3. Схема плоско панельного детектора. URL: <https://studfile.net/preview/1470150/page/3/> (дата обращения: 24.02.2022).
4. Махсотова, Ц. В. Исследование методов классификации при несбалансированности классов / Ц. В. Махсотова // Научный журнал. 2017. № 5(18). С. 35–36.
5. Sandler M., Howard A.G., Zhu M., Zhmoginov A., Chen L.-C. MobileNetV2: Inverted Residuals and Linear Bottlenecks // Computer Vision and Pattern Recognition. 2018. P. 4510–4520.
6. Блатов Р.И., Вострякова Е.А., Москвин А.С., Чупров Д. А., Егоров Ю.С., Коротышева А.А., Милов В.Р., Дубов М. С., Кербенева А.Ю. Свидетельство о регистрации программы для ЭВМ RUS 2022663863. Заявка № 2022662975 от 11.07.2022.

ОПТИМАЛЬНОЕ УПРАВЛЕНИЕ МАКСИМАЛЬНЫМ УКЛОНЕНИЕМ ВЫХОДА ПРИ ОГРАНИЧЕНИЯХ

Л.Н. Кривдина

Нижегородский государственный архитектурно-строительный университет

Рассматривается задача, в которой синтезируется оптимальное управление максимальным уклонением целевого выхода системы при ограничениях на вспомогательные выходы. Синтез таких управлений основан на применении аппарата линейных матричных неравенств. Приводятся результаты численного моделирования в задаче виброизоляции.

Ключевые слова: дискретный объект, максимальное уклонение выхода, оптимальное управление, линейные матричные неравенства.

1. Введение

Одной из задач теории управления является оптимальное управление максимальным уклонением выхода системы от его номинального значения. В работе [1] решалась задача для непрерывной системы, в которой был введен единый функционал при совместном влиянии начального и внешнего возмущений. В [2] было рассмотрено отдельно влияние начальных условий и отдельно – внешних возмущений в дискретном объекте. В статье [3] были синтезированы законы управления для дискретных систем без внешних возмущений с неизвестными начальными условиями. В [4] задача оптимального управления максимальным уклонением получила дальнейшее развитие: был введен критерий, являющийся максимальным уклонением выхода при внешнем и/или начальном возмущениях для линейной дискретной нестационарной системы, и были синтезированы оптимальные законы управления, в том числе и многокритериальные, на основе аппарата линейных матричных неравенств.

На практике могут возникать ситуации, когда существуют ограничения на часть выходов объекта. В соответствии с этим требуется оптимизировать основной критерий при условии имеющихся ограничений. В данной работе решается задача оптимального управления максимальным уклонением целевого выхода при условии ограничений на так называемые вспомогательные выходы. Приводятся результаты численного моделирования в задаче виброизоляции на бесконечном интервале времени.

2. Синтез оптимального управления

Пусть задана линейная дискретная нестационарная система следующими уравнениями

$$\begin{aligned}x(t+1) &= A(t)x(t) + B(t)v(t), \quad x(t_0), \\z(t) &= C(t)x(t), \quad t = t_0, t_0 + 1, \dots, t_0 + N_0, \quad N_0 \geq 1, \\z_i(t) &= C_i(t)x(t), \quad i = \overline{1, m},\end{aligned}\tag{1}$$

в которых $x(t) \in \mathbb{R}^{n_x}$ – состояние, $v(t) \in \mathbb{R}^{n_v}$ – возмущение, $z(t) \in \mathbb{R}^{n_z}$ – целевой выход, $z_i(t) \in \mathbb{R}^{n_{z_i}}$ – вспомогательные выходы, $A(t)$, $B(t)$, $C(t)$, $C_i(t)$ – заданные матричные функции соответствующих размерностей. Предполагается, что в общем случае начальное состояние $x(t_0)$ неизвестно и его влияние на динамику системы интерпретируется как начальное возмущение. Возьмем показатель, равный максимальному по времени значению отношения евклидовой нормы выхода к квадратному корню из суммы квадратов евклидовых норм возмущений и квадратичной формы начального состояния

$$\frac{\max_{t=t_0, \dots, t_0+N_0} |\xi(t)|}{\left(\sum_{k=t_0}^{t_0+N_0-1} |v(k)|^2 + x^T(t_0) R^{-1} x(t_0)\right)^{1/2}},\tag{2}$$

где $\xi(t)$ – выход (целевой или вспомогательный), $R = R^T > 0$ – заданная весовая матрица начального состояния и знаменатель не обращается в ноль. Задача состоит в том, чтобы минимизировать основной критерий $J_{0,v}$ (максимальное уклонение), являющийся точным гарантированным значением показателя (2) для целевого выхода $z(t)$, где

$$J_{0,v} = \max_{x(t_0), v(t_0), \dots, v(t_0+N_0-1)} \frac{\max_{t=t_0, \dots, t_0+N_0} |z(t)|}{\left(\sum_{k=t_0}^{t_0+N_0-1} |v(k)|^2 + x^T(t_0) R^{-1} x(t_0) \right)^{1/2}},$$

и ограничить этот же показатель для вспомогательных выходов $z_i(t)$

$$\frac{\max_{t=t_0, \dots, t_0+N_0} |z_i(t)|}{\left(\sum_{k=t_0}^{t_0+N_0-1} |v(k)|^2 + x^T(t_0) R^{-1} x(t_0) \right)^{1/2}} \leq \gamma_i,$$

где γ_i – заданные числа, $i = \overline{1, m}$.

Сформулируем теорему, основанную на теореме из [4].

Теорема 1. Максимальное уклонение целевого выхода при ограничениях на вспомогательные выходы находится в результате решения задачи полуопределенного программирования

$$J_{0,v}^2 = \min \gamma^2:$$

$$\begin{pmatrix} Y(t) & * & * \\ A(t)Y(t) & Y(t+1) & * \\ 0 & B^T(t) & I \end{pmatrix} \geq 0, \quad \begin{pmatrix} Y(t) & * \\ C(t)Y(t) & \gamma^2 I \end{pmatrix} \geq 0, \quad Y(t_0) \geq R,$$

$$\begin{pmatrix} Y(t_0 + N_0) & * \\ C(t_0 + N_0)Y(t_0 + N_0) & \gamma^2 I \end{pmatrix} \geq 0, \quad t = t_0, \dots, t_0 + N_0 - 1,$$

$$\begin{pmatrix} Y(t) & * \\ C_i(t)Y(t) & \gamma_i^2 I \end{pmatrix} \geq 0, \quad \begin{pmatrix} Y(t_0 + N_0) & * \\ C_i(t_0 + N_0)Y(t_0 + N_0) & \gamma_i^2 I \end{pmatrix} \geq 0, \quad i = \overline{1, m}$$

относительно неизвестных $Y(t_0 + 1), \dots, Y(t_0 + N_0)$.

Для устойчивого стационарного объекта (1) с нулевыми начальными условиями, в котором $A(t) \equiv A$, $B(t) \equiv B$, $C(t) \equiv C$, $C_i(t) \equiv C_i$ – заданные стационарные матрицы, $t_0 = 0$, максимальное уклонение целевого выхода при внешнем возмущении на бесконечном интервале определяется следующим образом

$$J_{v,\infty} = \max_{\{v(t)\}} \frac{\sup_{t \geq 0} |z(t)|}{\left(\sum_{k=0}^{\infty} |v(k)|^2 \right)^{1/2}}.$$

Теорема 2. Максимальное уклонение целевого выхода устойчивой стационарной дискретной системы (1) при ограничениях находится в результате решения задачи полуопределенного программирования

$$J_{v,\infty}^2 = \min \gamma^2: \quad \begin{pmatrix} Y & * & * \\ AY & Y & * \\ 0 & B^T & I \end{pmatrix} \geq 0, \quad \begin{pmatrix} Y & * \\ CY & \gamma^2 I \end{pmatrix} \geq 0, \quad (3)$$

$$\begin{pmatrix} Y & * \\ C_i Y & \gamma_i^2 I \end{pmatrix} \geq 0, \quad i = \overline{1, m}. \quad (4)$$

Рассмотрим теперь задачу управления. Пусть линейная дискретная стационарная система определяется уравнениями

$$\begin{aligned} x(t+1) &= Ax(t) + B_1 v(t) + B_2 u(t), \quad x(0) = 0, \\ z(t) &= Cx(t) + Du(t), \\ z_i(t) &= C_i x(t) + D_i u(t), \quad i = \overline{1, m}. \end{aligned} \quad (5)$$

Закон управления имеет вид линейной стационарной обратной связи $u(t) = \Theta x(t)$. Матрицы закона управления вычисляются как $\Theta = ZY^{-1}$ при решении линейных матричных неравенств (3) и (4), полученных из неравенств Теоремы 2 для системы (5) с заданными числами γ_i при замене в них соответствующих матриц на матрицы $A + B_2\Theta$, B_1 , $C + D\Theta$, $C_i + D_i\Theta$ и при введении переменных $Z = \Theta Y$.

3. Иллюстративный пример

Рассмотрим механическую систему с двумя степенями свободы (см. рис. 1), представляющую собой упругий объект, который моделируется тремя материальными точками 1, 2 и 3, связанными попарно друг с другом линейными упругим и диссипативным элементами. Материальная точка 1 находится внизу, представляет собой подвижное основание и связана с точкой 2 еще и управляемым элементом, который далее будем называть виброизолятором.

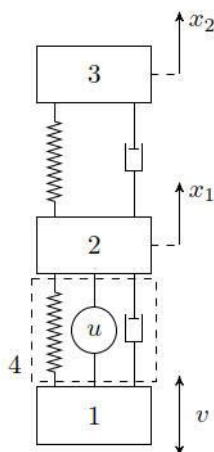


Рис. 1. Механическая система

Данная механическая система в безразмерных переменных и параметрах описывается следующими дифференциальными уравнениями

$$\begin{aligned}\ddot{x}_1 &= -2x_1 + x_2 - 2\beta\dot{x}_1 + \beta\dot{x}_2 + u + v, \\ \ddot{x}_2 &= x_1 - x_2 + \beta\dot{x}_1 - \beta\dot{x}_2 + v,\end{aligned}\quad (6)$$

где x_1 и x_2 — координаты материальных точек 2 и 3 относительно подвижного основания, β — заданный положительный параметр демпфирования, u — усилие, которое создает виброизолятор при смещении материальной точки 2 относительно точки 1, т.е. при его деформации, v — ускорение основания (материальной точки 1) с точностью до знака. Возьмем в качестве целевого выхода

$$z = |x_1|,$$

а в качестве вспомогательного выхода

$$z_1 = -x_1 - \beta\dot{x}_1 + u,$$

где $\beta = 0.1$. Функционал $J_{v,\infty}(\Theta)$, где $\Theta = (\Theta_1 \ \Theta_2 \ \Theta_3 \ \Theta_4)$ дает характеристику смещения материальной точки 1, а ограничение для вспомогательного выхода задает ограничение максимальной силы, противодействующей смещению упругого объекта относительно основания. Будем выбирать управление объектом (6) в виде стационарной линейной обратной связи по состоянию $u = \Theta x$ так, чтобы минимизировать критерий $J_{v,\infty}(\Theta)$ при ограничении для вспомогательного выхода z_1 .

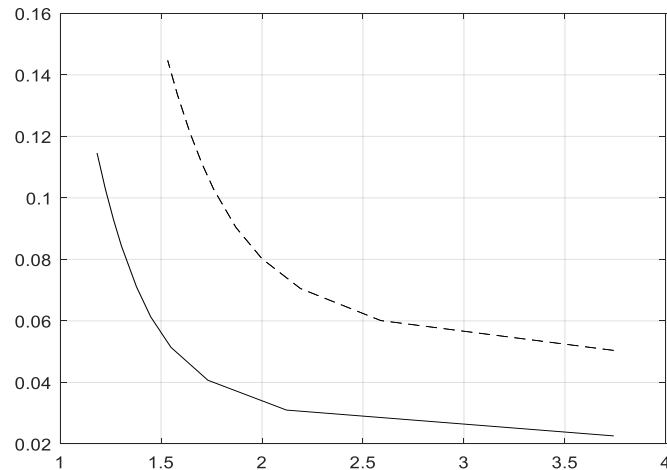


Рис. 2. График зависимости оптимального значения функционала от γ_1

На рисунке 2 представлены графики зависимости оптимального значения функционала от γ_1 для разных значений шага дискретизации h системы (6): сплошная линия – при $h = 0.2$ сек, пунктирная линия – при $h = 0.3$ сек. В частности, при $\gamma_1 = 1.7607$ получили значение функционала $J_{v,\infty}(\Theta) = 0.1028$ и матрицу параметров регулятора $\Theta = (-4.1322 \quad -0.8989 \quad -3.0375 \quad -0.4140)$ (при $h = 0.3$ сек).

4. Заключение

В данной статье решена задача синтеза оптимального управления максимальным уклонением целевого выхода при условии ограничений на вспомогательные выходы. Найдены оптимальные регуляторы (виброизоляторы) по состоянию для механической системы на бесконечном интервале времени.

Литература

1. Баландин Д.В., Бирюков Р.С., Коган М.М. Оптимальное управление максимальными уклонениями выходов линейной нестационарной системы на конечном интервале времени // *АиТ*. 2019. № 10. С. 37–61.
2. Amato F., Carbone M., Ariola M., Cosentino C. Finite-Time Stability of Discrete-Time Linear Systems // *IEEE Trans. Autom. Control*. 2005. V. 50. No. 5. P. 724-729.
3. Коган М.М., Кривдина Л.Н. Синтез многоцелевых линейных законов управления дискретными объектами при интегральных и фазовых ограничениях // *АиТ*. 2011. № 7. С. 83–95.
4. Баландин Д.В., Бирюков Р.С., Коган М.М. Минимаксное управление уклонениями выходов линейной дискретной нестационарной системы // *АиТ*. 2019. № 12. С. 3–23.

ИСПОЛЬЗОВАНИЕ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В СИСТЕМЕ УПРАВЛЕНИЯ ОСВЕЩЕНИЕМ УМНОГО ДОМА

Т.В. Куприянова¹, Д.И. Кислицын²

¹*Нижегородский государственный технический университет им. Р. Е. Алексеева*

²*Нижегородский государственный архитектурно-строительный университет*

В данной статье рассматривается возможность применения искусственного интеллекта в системе освещения умного дома. Проведен анализ и сравнение вариантов освещения, выявлена боль пользователя. Предлагается собственный вариант решения проблемы с применением разработанной умной лампы для smart home с интеграцией с искусственным интеллектом. А также был выбран наиболее подходящий метод машинного обучения для реализации предлагаемой умной системы освещения.

Ключевые слова: умный дом, интернет вещей, умная система освещения, искусственный интеллект, алгоритмы машинного обучения.

В настоящее время концепция Интернета вещей (Internet of Things, IoT) затрагивает многих людей. Проникновение IoT в различные сферы жизни неизбежно, поскольку использование этих технологий приносит удобство и экономию времени [1]. Поэтому умный дом (smart home) является одним из основных перспективных направлений развития в современном мире. А значит спрос на инновационные проекты продолжит расти.

Объем мирового рынка устройств для умного дома в 2020 году составил 801,5 млн штук, увеличившись на 4,5%. В компании IDC, которая занимается аналитикой, считают, что рынок оборудования для digital home останется растущим. Вплоть до 2025 года поставки устройств в умный дом будут в среднем расти на 12,2% в год и к концу периода должны превысить 1,4 млрд штук [2]. Помимо этого, использовать smart устройства экологично, так как они автоматически регулируют потребление электроэнергии.

Для создания умной системы в доме необходимо определиться с основными задачами, которые будут выполнять умные устройства. Наиболее распространенной задачей на начальном этапе проектирования умного дома является умное управление освещением.

В классическом понимании для управление умными лампами происходит с помощью голосового помощника или мобильного приложения, то есть при прямом действии пользователя. Это может быть неудобно для выполнения таких элементарных задач, как включение и выключение света, изменение яркости или режимов светодиодных ламп. Настоящая умная система освещения должна сама подстраиваться под потребителя и самостоятельно определять, когда пользователю необходимо включить или выключить свет, а когда изменить настройки.

Проблему ограничения технологии интернета вещей можно решить с помощью интеграции умной системы освещения с искусственным интеллектом (ИИ) и машинным обучением (МО).

Использование искусственного интеллекта в освещении для умного дома делает пребывание человека более комфортным и удобным, повысит качество жизни. Smart система будет сама принимать решения по управлению освещением без участия пользователя.

Целью проводимой научно-исследовательской работы является предложение варианта разработки умной системы освещения для умного дома с интеграцией с искусственным интеллектом. Для достижения поставленной цели планируется решение следующих задач:

- 1) изучение системы умного дома;
- 2) анализ и сравнение существующих умных ламп;
- 3) выявление боли при использовании системы освещения без искусственного интеллекта для отдельной категории пользователей, построение CJM;
- 4) выдвижение предложения по модернизации освещения и интеграции с ИИ;
- 5) разработка умной лампы;
- 6) ознакомление с сущностью и методами машинного обучения, выбор наиболее подходящих для данного проекта;

7) рассмотрение вариантов комбинации компонентов для разработки умной системы освещения с искусственным интеллектом и выбор наиболее подходящего;

8) построение схемы архитектуры умной системы освещения с ИИ.

Под понятием «умный дом» подразумевается система управления и автоматизации инженерными системами дома (освещение, отопление, вентиляция и др.). А также с помощью датчиков (движения, температуры, протечки воды и тд.), подключенных к «мозгу системы», выполняются функции сигнализации. «Мозг системы» обрабатывает данные с датчиков и отправляет сообщение на смартфон пользователя [3].

Раньше система освещения требовала от пользователя ручное управление с помощью переключателей. Сейчас же она стала умной, более удобной и комфортной для человека из-за простоты в использовании, долгосрочности, наличия широкого набора функций и высокой социальной эффективности. Умная система включает в себя не только умные лампочки, но и другие компоненты, а управляется с помощью голосовых команд или мобильного приложения.

Основным отличительным признаком smart систем освещения являются использующиеся умные лампы. Во всем мире существуют сотни производителей, готовых обеспечить потребителей smart лампами. Однако они имеют разный функционал и стоимость. Сравним преимущества и недостатки существующих готовых решений для выбора подходящей умной лампы, которая будет поддерживать управление с помощью ИИ (см. таблица 1) [4].

Таблица 1. Различия существующих умных ламп

Наименование	Производитель	Цена, руб.	Источник цены, ссылка и дата обращения	Уровень совместимости с умным домом	Управление с помощью сценариев	Диммирование света	Регулирование света с помощью ИИ
Smart LED Ceiling Lamp	Philips	6500	Яндекс Маркет https://clck.ru/322y2V (15.09.22)	Высокий	×	×	×
X-Light Plus	Meizu	2990	Store Xiaomi https://clck.ru/322y7Y (15.09.22)	Низкий	✓	✓	×
Aqara Smart Light Control	Xiaomi	3990	Яндекс Маркет https://clck.ru/322xyM (15.09.22)	Высокий	✓	×	×
Mi Night Lamp	Xiaomi	1490	Store Xiaomi https://clck.ru/322yEc (15.09.22)	Высокий	✓	✓	×
Frattina-C 97813	EGLO	17390	Яндекс Маркет https://clck.ru/enAr (08.04.22)	Средний	×	✓	×
Умная лампа с ИИ	DIY	≈5400		Высокий	✓	✓	✓

Как можно заметить из таблицы 1, самодельная (DIY) умная лампа отвечает всем необходимым требованиям, в отличие от готовых решений, которые зачастую дорогие, имеют функциональные ограничения и требуют человеческого вмешательства в управлении. При использовании готовых smart ламп некоторые потребители продолжают испытывать неудобства. Более подробно о проблемах и болях пользователей представлено на рисунках 1 и 2.

На них отображена CJM (Customer Journey Map, «карта путешествия клиента»), благодаря ей можно найти точки роста для умной системы освещения [5]. CJM иллюстрирует модель поведения учащегося. Он усердно готовится к экзаменам, много занимается за ПК в разное время суток и при любых погодных условиях. Его главной болью является ухудшение зрения. Пользователь зачастую забывает включать свет при отсутствии естественного источника света и сидит в темноте при работе за ПК. Из-за этого напрягается зрение, начинаются проблемы со здоровьем: зуд глаз, повышенное слезотечение, перегрузка нервной системы.

Для решения этой проблемы пользователь проходит несколько этапов. Первый этап – осознание проблемы, где учащийся задается вопросом о причине ухудшения его зрения и переходит ко второму этапу – поиску решения проблемы. На третьем этапе выбирается лампа из всего ассортимента с лучшим соотношением цена/качество. Далее на четвертом и пятом этапах происходит покупка и тестирование smart лампы. И, наконец, на шестом этапе пользователь делает вывод, соответствует ли новая умная лампа его ожиданиям, и решает ли она его проблему.

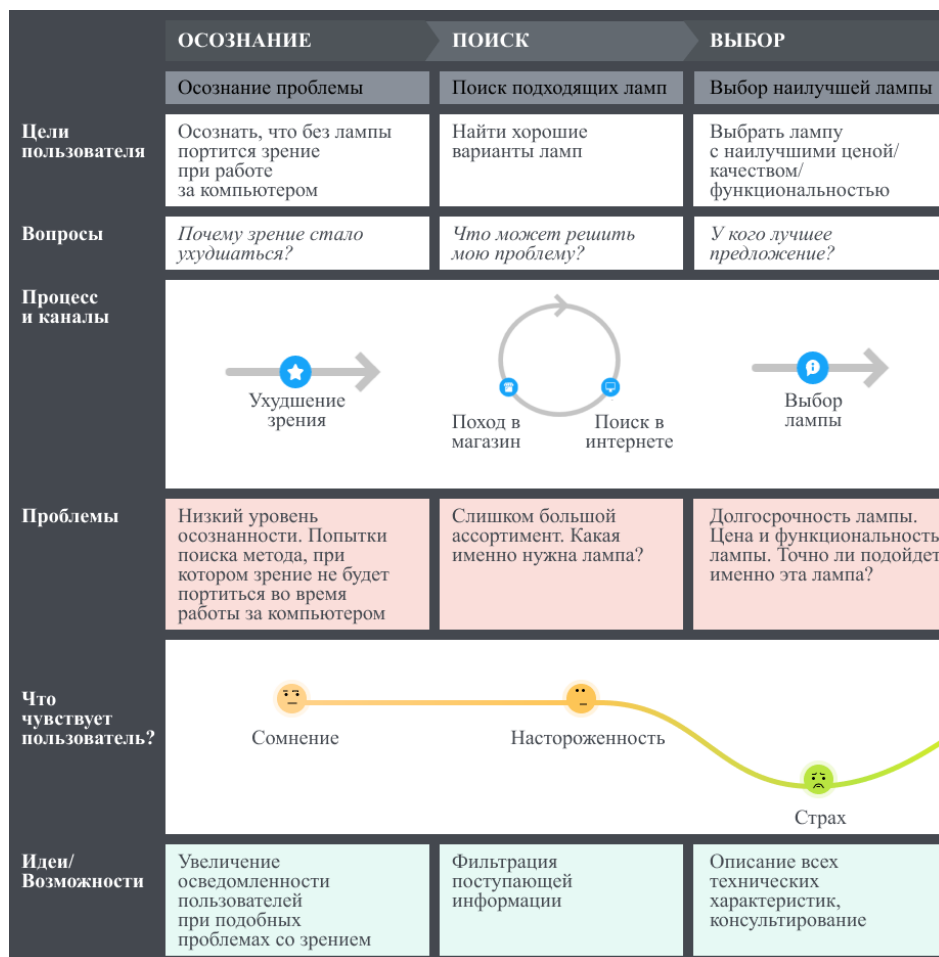


Рис. 1. CJM, график взаимодействия пользователя с умной лампой. Часть 1

На этапе вывода пользователь понял, что новой умной лампой пользоваться удобно, однако сохраняется частичное управление: с помощью мобильного приложения или голосовых команд и сценариев. Отсутствует возможность отслеживать время пребывания пользователя за рабочим местом, а также необходимы определенные действия, которые не всегда предпринимаются. Частичное решение проблемы не оправдало ожидания учащегося.

Полностью решить проблему поможет модернизация освещения, а именно его интеграция с искусственным интеллектом. Работа ИИ происходит следующим образом: создаются алгоритмы таким образом, чтобы искусственный интеллект мог принимать решения на основе собственного опыта или адекватно выходить из ситуаций, с которыми ранее не сталкивался [6].

Поскольку интернет вещей и искусственный интеллект постоянно эволюционируют, их связь увеличивает количество возможностей и решений в современном мире. Благодаря такому объединению можно быстро обрабатывать, анализировать данные, принимать решения и давать ответы в режиме реального времени на подключенные к IoT-устройства.

Исходя из анализа таблицы 1, готовые решения не могут взаимодействовать с ИИ, поэтому было принято решение разработать собственную умную лампу, поддерживающую интеграцию. Smart лампа может управляться с помощью сценариев и голосового помощника Алисы, мобильного приложения от платформы Blynk, поддерживает функции диммирования света и изменения эффектов свечения. Разработана с помощью программируемого микроконтроллера

ESP8266, адресной матрицы WS2812b и сенсорной кнопки TTP223, а также прочих деталей. Внешний вид разработанной умной лампы представлен на рисунке 3. Более подробная информация о причинах выбора данных компонентов, этапах и результате разработки умной лампы представлено в работе «Разработка умной лампы с интеграцией в экосистему Яндекса» научной конференции «Математика и ИТ – вместе в цифровое будущее» [7].

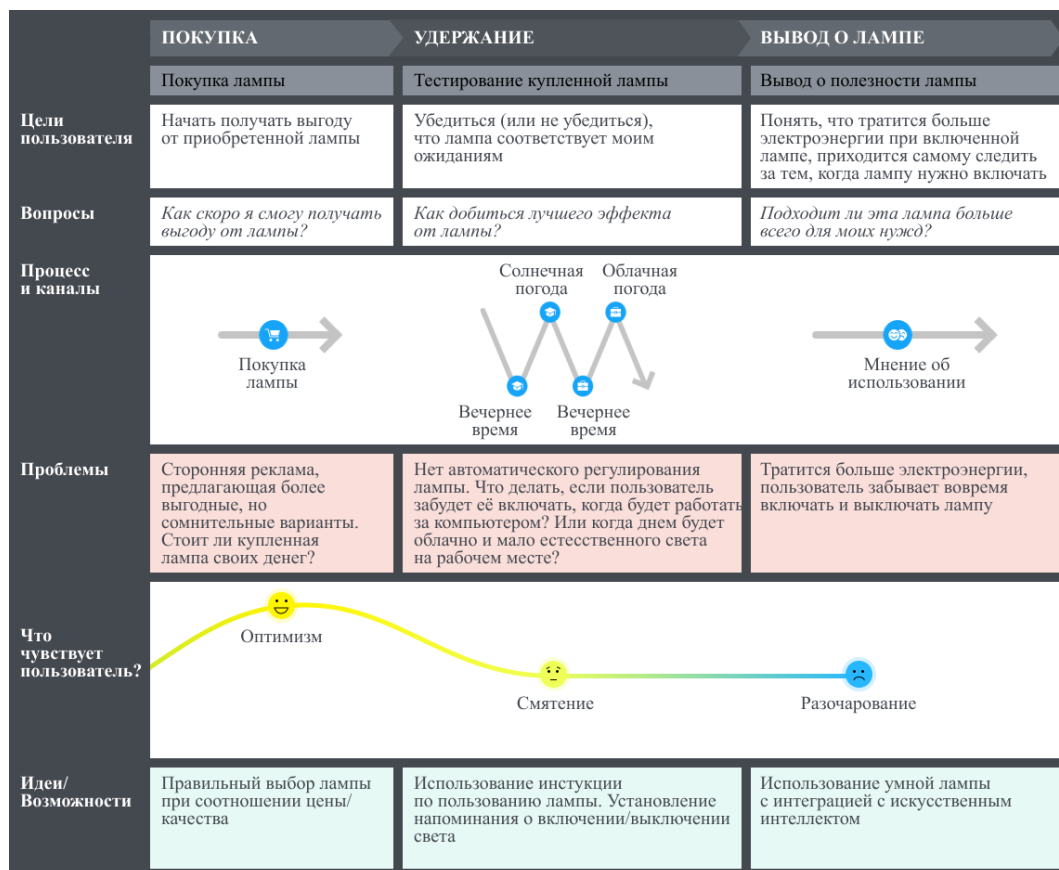


Рис. 2. CJM, график взаимодействия пользователя с умной лампой. Часть 2

Одной из форм искусственного интеллекта является машинное обучение, которое автоматизирует человеческое поведение, используя различные виды поведения ИИ. Существует множество методов МО. Для ознакомления с ними можно обратиться к шпаргалке по алгоритмам машинного обучения от Microsoft docs [8]. Исходя из информации из источника, можно сделать вывод, что каждый из методов машинного обучения подходит для решения конкретных задач. Чтобы понять, какой именно метод МО необходим для текущего проекта, нужно сначала определиться с деталями, с помощью которых система будет реализована.

Основными компонентами для работы будут являться умная лампа и ПК с необходимым софтом. При этом smart лампа и smart система находятся в постоянном контакте между собой и передают друг другу обратную связь. Дополнительными деталями могут выступать IP-камера, датчики движения и освещения. В рамках данной работы было рассмотрено четыре варианта реализации умной системы при использовании различных дополнительных компонентов.

Первый вариант предполагает использование всех дополнений: IP-камеры, датчиков движения и освещения. Датчик движения будет работать постоянно в дежурном режиме. При приближении человека происходит фиксация и передача сигнала системе. После получения данных включается видекамера, которая анализирует и отслеживает время нахождения пользователя за рабочим местом. Датчик освещенности передает текущий уровень освещенности системе, которая управляет умной лампой и регулирует яркость. В данном варианте появляется возможность адаптации smart системы под привычки пользователя. Однако умная система будет сложной, так как состоит из множества компонентов, а значит повышается вероятность ее сбоя.

Второй вариант предполагает дополнительное использование только датчиков движения и освещения. Датчик движения работает в дежурном режиме и фиксирует факт наличия пользователя, а датчик освещенности работает так же, как в первом варианте. Чтобы не перепутать приближение человека с приближением домашнего животного, задается конкретный объем приближающегося объекта. У датчика освещенности настраивается таймер, в течение которого считается, что пользователь находится за рабочим местом. Реализация такой системы простая, значительно дешевле, чем в первом варианте, а вероятность сбоя системы низкая. Однако данное решение сложно назвать умным, так как не учитывается, что пользователь может сидеть за компьютером разное время.

В третьем варианте в качестве дополнительных компонентов используется IP-камера и датчик освещенности. Веб-камера фиксирует пришедшего пользователя, далее анализирует его действия и отслеживает время пребывания человека на месте. Поскольку учащийся может садиться за рабочее место в разное время суток, датчик освещенности передает умной системе информацию о уровне освещения. И на основе этих данных регулируется smart лампа.

В четвертом варианте дополнительно используется только IP-камера, которая будет выполнять функции и датчиков движения и освещенности. Она дает больше представления, подтверждает присутствие или отсутствие пользователя на рабочем месте. По видео определяется появление человека, далее производится анализ его действий, отслеживание времени пребывания. В зависимости от предпочтений учащегося или степени освещенности умная система будет подстраивать smart лампу. Однако, возможно, что качество данных, касающихся уровня света без отдельного датчика освещенности, будет не настолько точным.

Исходя из описанных выше вариантов реализации умной системы, их достоинств и недостатков, был выбран четвертый вариант. Этот выбор объясняется тем, что для массовой реализации и производства данного проекта использование только IP-камеры будет удобнее и менее трудозатратно. Через видео будет определяться уровень освещенности, и, в зависимости от полученного значения, свет smart лампы будет корректироваться умной системой. Таким образом, умная система освещения будет состоять из ПК со специальным софтом, smart лампы и IP-камеры. На рисунке 3 представлена схема архитектуры умной системы освещения.

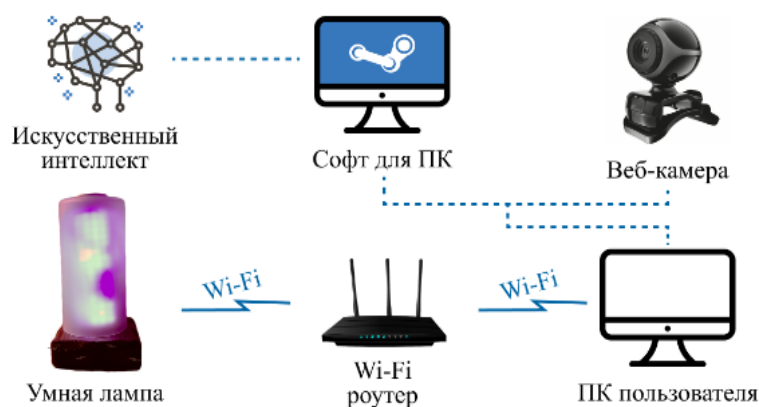


Рис. 3. Схема архитектуры умной системы освещения

После выбора набора компонентов для реализации проекта, исходя из данных с сайта Microsoft docs [8], можно определиться с методами машинного обучения. Так как к задачам smart системы относится обнаружение пользователя за рабочим местом, а также изменений в кадре (перемещение человека, изменение уровня освещенности), то будут использоваться два метода машинного обучения. Многоклассовая классификация будет отвечать за фиксацию человека и его дальнейший анализ, обработку, а с помощью метода обнаружения аномалий будут ловиться и отсеиваться всплески и другие выбросы, появляющиеся в кадре.

Перед процессом обучения алгоритмов МО, обозначим входные параметры для понимания рабочего места: уровень освещенности, детектирование пользователя, состояние лампы (включение/выключение, уровень яркости, режим), время суток и история браузера.

Во время процесса обучения алгоритма МО основной задачей является изучение предпочтений пользователя при управлении умной лампой. В начале обучения необходимо, чтобы учащийся сам включал/выключал лампу, менял яркость, режимы. Smart система подключается к ПК пользователя и, наблюдая за ним, создает примеры. В базе умной системы должно накопиться определенное количество опыта. При достижении конкретного порога, smart система переходит в режим работы, при котором использует накопленный опыт. Пользователю будет предлагаться решение прежде, чем он выполнит управление умной лампой сам. То есть учащийся садится за рабочее место, а smart система уже отработала весь алгоритм и включила свет. Также происходит запоминание действий пользователя и состояния smart лампы при определенных параметрах: уровень освещения, времени суток, истории браузера. Это необходимо для изучения предпочтений человека, предположения его вида деятельности.

При этом машинное обучение будет происходить в течение всей работы умной системы, без использования готовых шаблонов. Это необходимо из-за изменений условий использования рабочего места, привычек и самого пользователя. Умная система работает в том режиме, в котором научилась. При этом, если человеку не понравилось принятое решение умной системой, у него будет возможность вручную скорректировать лампу (увеличить яркость, переключить режим и тд.). Smart система имеет обратную связь с умной лампой и отслеживает ее состояние. Поэтому изменения, внесенные пользователем, будут замечены. Умная система производит корректировку примеров для обучения, а затем адаптирует свои решения в дальнейшем. Такие действия со стороны smart системы наиболее комфортны для пользователя.

Таким образом, в рамках данного проекта была начата работа по созданию умной системы освещения smart home с интеграцией с ИИ. Были выбраны необходимые для разработки методы машинного обучения, реализована умная лампа. В дальнейшем планируется продолжение работы над проектом, поскольку он является привлекательным для конечного потребителя за счет своей удобства, функциональности, пользы для здоровья, экологичности и социальной эффективности.

Литература

1. Куприянова Т.В., Кислицын Д.И. О проблемах развития интернета вещей в России // X Всероссийский фестиваль науки: сборник докладов. – Текст: электронный / Нижегород. гос. архитектур.- строит. ун-т; редкол.: А. А. Лапшин, И. С. Соболев, Д. В. Монич, А. А. Смыков [и др.]. – Н. Новгород: ННГАСУ, 2020. – 1 CD-ROM. – ISBN 978-5-528-00426-6.
2. Системы умного дома (мировой рынок). – Текст: электронный // Tadviser: сайт. – URL: <https://www.tadviser.ru/a/405044> (дата обращения: 29.09.2021).
3. Pegas27. Как выбрать устройства умного дома. – Текст: электронный // DNS клуб: сайт. – URL: <https://club.dns-shop.ru/blog/t-232-umnyii-dom/17780-kak-vyibrat-ustroistva-umnogo-doma/> (дата обращения: 16.07.2022).
4. Системы освещения в умном доме. Выбор, особенности, настройка. – Текст: электронный // Новостной портал Gadetpage: сайт. – URL: <https://gadgetpage.ru/smart-house/3044-sistemy-osveschenija-v-umnom-dome-vybor-osobennosti-nastrojka.html> (дата обращения: 14.09.22).
5. Разработка CJM. – Текст: электронный // Центр развития компетенций в бизнес-информатике высшей школы бизнеса: сайт. – URL: <https://hsbi.hse.ru/articles/razrabotka-cjm/> (дата обращения: 18.07.2022).
6. Искусственный интеллект и интернет вещей: в чем разница? – Текст: электронный // Яндекс Дзен: сайт. – URL: <https://zen.yandex.ru/media/funscience/iskusstvennyi-intellekt-i-internet-vescei-v-chem-raznica-5bc4d2989cc7e200a96d7c70> (дата обращения: 18.07.2022).
7. Куприянова Т.В., Кислицын Д.И. Разработка умной лампы с интеграцией в экосистему Яндекса // Математика и ИТ – вместе в цифровое будущее: сборник трудов Молодежной школы (Н. Новгород, 25–29 апреля 2022 г.). / Под ред. К.А. Баркалова, М.А. Богомоловой. – Нижний Новгород: Нижегородский госуниверситет им. Н.И. Лобачевского, 2022. – 151 с.
8. Igayhardt, cjgronlund, PeterCLu, v-kents, TomArcherMsft, j-martens, ФранческаЛаццери. Machine Learning Algorithm Cheat Sheet for Azure Machine Learning designer. – Текст: электронный // Microsoft docs: сайт. – URL: <https://docs.microsoft.com/en-us/azure/machine-learning/algorithm-cheat-sheet> (дата обращения: 21.07.2022).

ОБ ИДЕАЛАХ ЦЕНТРАЛЬНО СУЩЕСТВЕННЫХ КОЛЕЦ^{1*}

О.В. Любимцев¹, А.А. Туганбаев²

¹Нижегородский государственный университет им. Н.И. Лобачевского

²Национальный исследовательский университет «МЭИ»

Показано, что кольцо A с центром $Z(A)$ такое, что модуль $AZ(A)$ является существенным расширением модуля $Z(A)Z(A)$, не обязано быть квазиинвариантным справа, т.е. не все максимальные правые идеалы кольца A являются идеалами. В терминах центральной существенности получены достаточные условия того, что все максимальные правые идеалы являются идеалами, а минимальные правые идеалы лежат в центре кольца A . Существуют центрально существенные кольца, в которых не все замкнутые правые идеалы являются идеалами.

Ключевые слова: центрально существенное кольцо, максимальный правый идеал, минимальный правый идеал.

1. Введение

Мы рассматриваем только ассоциативные унитарные ненулевые кольца. Радикал Джекобсона кольца A обозначается через $J(A)$. Кольцо A называется *центрально существенным*, если либо A коммутативно, либо для каждого его нецентрального элемента a существуют такие ненулевые центральные элементы x, y , что $ax = y$. Ясно, что кольцо A с центром $Z(A)$ центрально существенно в точности тогда, когда модуль $A_{Z(A)}$ является существенным расширением модуля $Z(A)_{Z(A)}$. Глядя на определение центрально существенного кольца A может показаться, что такое кольцо, возможно, коммутативно. Действительно, A обладает многими свойствами коммутативных колец. Например,

- все идемпотенты кольца A центральны;
- если кольцо A полупервично, то кольцо A коммутативно;
- если A – центрально существенное локальное кольцо, то кольцо $A/J(A)$ – поле и поэтому коммутативно;
- если A – полуартиново справа или слева, центрально существенное кольцо, то факторкольцо $A/J(A)$ коммутативно;

Однако, центрально существенное кольцо A может быть весьма далеким от коммутативного кольца. Например,

- факторкольцо $A/J(A)$ кольца A по первичному радикалу может не быть центрально существенным и, в частности, полупервичное кольцо $A/J(A)$ может не быть коммутативным;
- факторкольца кольца A по идеалам, порожденным центральными идемпотентами, не обязательно центрально существенны;
- факторкольца кольца A не обязательно центрально существенны (см. два предыдущих пункта);
- существуют конечные некоммутативные центрально существенные групповые алгебры;
- существуют конечные некоммутативные центрально существенные внешние алгебры;
- существуют такие абелевы группы без кручения конечного ранга, что их кольца эндоморфизмов являются некоммутативными центрально существенными кольцами.

Подробнее см., монографию [1].

^{1*} Работа выполнена по теме государственного задания (№ 0729-2020-0055).

2. Минимальные и замкнутые идеалы центрально существенных колец

Предложение 1. Если A – центрально существенное кольцо и факторкольцо $A/J(A)$ коммутативно, то любой минимальный правый идеал кольца A лежит в центре $Z(A)$ кольца A . В частности, все минимальные правые идеалы кольца A являются идеалами и $\text{Soc}(A_A)$ лежит в центре кольца A .

Подмодуль X модуля M называется *замкнутым*, если $X = Y$ для любого подмодуля Y в M , являющегося существенным расширением модуля X .

Предложение 2. Существуют центрально существенные кольца, в которых не все замкнутые правые идеалы являются идеалами.

В связи с предложением 1 мы сформулируем открытый вопрос.

Открытый вопрос. Существуют ли центрально существенные кольца, в которых не все минимальные правые идеалы являются идеалами?

3. Максимальные идеалы центрально существенных колец

Предложение 3. Пусть A – центрально существенное кольцо с центром $Z(A)$. Если существует максимальный правый идеал M кольца A , не являющийся идеалом, то

$$Z(A) \cap \left(\bigcap_{n \geq 1} M^n \right) \neq 0.$$

Кольцо называется *квазиинвариантным справа (слева)*, если все максимальные правые (левые) идеалы этого кольца являются идеалами.

Предложение 4. Если кольцо A и факторкольцо $A/J(A)$ центрально существенны, то кольцо $A/J(A)$ коммутативно, а кольцо A квазиинвариантно справа и слева.

Предложение 5. Все полуартиновы справа или полусовершенные центрально существенные кольца являются квазиинвариантными справа.

Лемма 1. Пусть A – кольцо, $R = A[x]$ – кольцо многочленов и M – максимальный правый идеал кольца A . Тогда

a. $MR + xR$ – максимальный правый идеал кольца R .

b. Если $MR + xR$ – идеал кольца R , то M – идеал кольца A , фактор-кольца A/M и $R/(MR + xR)$ являются изоморфными телами, $(A/M)[x]$ – область главных односторонних идеалов, и имеется изоморфизм $\alpha: R/(MR + xR) \rightarrow A/M$, при котором $\alpha(f + (MR + xR)) = f_0 + M$, где f_0 – свободный член многочлена f из R .

c. Если A – тело и кольцо многочленов R квазиинвариантно справа, то кольцо A является полем.

d. Если кольцо многочленов R квазиинвариантно справа, то фактор-кольцо $A/J(A)$ коммутативно.

Лемма 2. Пусть A – центрально существенное кольцо. Тогда

a. Кольцо $A[x]$ – центрально существенно.

b. Если факторкольцо $A/J(A)$ некоммутативно, то $A[x]$ – центрально существенное кольцо, не являющееся квазиинвариантным справа (слева).

Из лемм 1, 2 и работы [2, Theorem 1.5] вытекает основной результат работы – теорема 1.

Теорема 1. Существуют центрально существенные кольца, в которых не все максимальные правые идеалы являются идеалами.

Литература

1. Туганбаев А.А. Центрально существенные кольца. Москва: Флинта, 2022, 160 с. ISBN 978-5-9765-5093-3.
2. Markov V.T., Tuganbaev A.A. Constructions of Centrally Essential Rings // Comm. Algebra. 2020. Vol. 48, No. 1. P. 198-203.

КОММУНИКАЦИОННЫЕ СХЕМЫ НА ОСНОВЕ СИНХРОНИЗАЦИИ СИСТЕМ СО СТРАННЫМИ ХАОТИЧЕСКИМИ И НЕХАОТИЧЕСКИМИ АТТРАКТОРАМИ^{1*}

Д.О. Любченко, О.Б. Исаева

*Саратовский национальный исследовательский госуниверситет им. Н.Г. Чернышевского
Саратовский филиал Института радиотехники и электроники им. В.А. Котельникова РАН*

Задачи, связанные с хаотической коммуникацией, представляют большой интерес с прикладной точки зрения. Коммуникация на основе генераторов гиперболического хаоса кажется одним из самых перспективных направлений. Еще одним интересным направлением является реализация передачи данных с помощью странного нехаотического аттрактора. В данной работе изучаются схемы связи на основе генераторов гиперболических хаотических и странных нехаотических сигналов. Мы предлагаем методы связи, основанные на синхронизации передатчика и приемника. Такой способ передачи обладает наибольшей устойчивостью к шуму по сравнению с классической коммуникацией, основанной на периодическом сигнале. В работе представлено численное моделирование и анализ этих схем.

Ключевые слова: хаотическая коммуникация, гиперболический хаос, странный нехаотический аттрактор.

1. Введение

Проблемы, связанные с хаотической коммуникацией, представляют большой интерес с прикладной точки зрения [1]. Хаотическая передача данных имеет ряд преимуществ: 1. передача сигнала происходит в широком диапазоне частот; 2. хаотический сигнал более устойчив к шуму, чем периодический; 3. безопасность передачи в том смысле, что для декодирования информации необходимо восстановить динамику хаотического передатчика. Наиболее перспективной представляется связь на основе гиперболического хаоса, поскольку Фурье образ такого сигнала не отличается от шума [2].

Еще одним интересным направлением в коммуникации является реализация передачи данных с помощью странного нехаотического аттрактора (СНА). Известно, что СНА возникает в системах с квазипериодическим воздействием на границе регулярной и хаотической динамики [3]. Он имеет фрактальную структуру в фазовом пространстве, хотя и стабилен по Ляпунову. Ранее считалось, что этот аттрактор характерен для таких динамических систем, как отображения, и может появиться только в вырожденных случаях. Однако более детальные исследования показали, что в некоторых системах СНА наблюдается в широком диапазоне значений параметров и может появляться в потоковых системах, что позволило наблюдать СНА на практике.

В данной работе исследуется возможность использования генераторов гиперболического хаоса и СНА в схемах связи. Мы предлагаем методы связи, основанные на синхронизации идентичных генераторов, используемых в качестве передатчика и приемника. Мы модулируем по фазе сигнал для передачи, который можно обнаружить стандартным способом.

2. Коммуникационные схемы

2.1. Схема на основе генератора гиперболического хаоса

В данной работе используется генератор гиперболического хаотического аттрактора, который может быть получен в реальной радиофизической системе [4]. Свяжем два аналогичных генератора способом описанным в [5]:

^{1*} Работа поддержана грантом Российского научного фонда № 21-12-00121, <https://rscf.ru/project/21-12-00121/>.

$$\left\{ \begin{array}{l} \ddot{x}_1 - \left(A \cos \frac{w_0 t}{T} - x_1^2 \right) \dot{x}_1 + w_0^2 x_1 = \varepsilon y_1 \cos(w_0 t + \rho(t)), \\ \ddot{y}_1 - \left(-A \cos \frac{w_0 t}{T} - y_1^2 \right) \dot{y}_1 + (2w_0)^2 y_1 = \varepsilon x_1^2, \\ \ddot{x}_2 - \left(A \cos \frac{w_0 t}{T} - x_2^2 \right) \dot{x}_2 + w_0^2 x_2 = \varepsilon S(t), \\ \ddot{y}_2 - \left(-A \cos \frac{w_0 t}{T} - y_2^2 \right) \dot{y}_2 + (2w_0)^2 y_2 = \varepsilon x_2^2, \\ S(t) = y_1 \cos(w_0 t + \rho(t)). \end{array} \right. \quad (1)$$

Первые два уравнения отвечают за передачу информации, $\rho(t)$ – полезная информация. На передатчике генерируется сигнал $S(t)$, который попадает на приемник, за который отвечают третье и четвертое уравнения системы (1). За счет этих уравнений можно восстановление динамику передатчика и в дальнейшем получить полезную информацию. На Рис. 1 представлены отправленная информация (синяя) и полученная информация (красная). Сначала информация не может быть расшифрована из-за расстройки генераторов, а после реализации полной синхронизации информация хорошо накладывается друг на друга (полученная информация была извлечена грубо для наглядности).

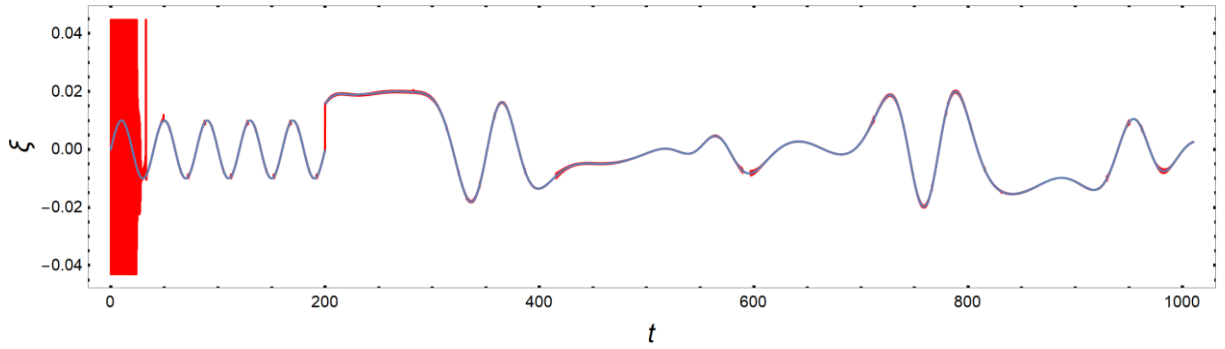


Рис. 1. Информация: (синий) полезная информация, шифруемая на передатчике; (красный) расшифрованная информация на приёмнике

2.2. Схема на основе генератора странного нехаотического аттрактора

Для генерации СНА используется система, реализующая аттрактор типа Ханта-Отта, состоящая из двух связанных неавтономных осцилляторов Ван-дер-Поля, находящихся под внешним квазипериодическим воздействием [6]. Аналогично свяжем два генератора:

$$\left\{ \begin{array}{l} \ddot{x}_1 - \left(A \sin \frac{w_0 t}{T} - x_1^2 \right) \dot{x}_1 + w_0^2 x_1 = \varepsilon y_1 \sin(w_0 t + \theta), \\ \ddot{y}_1 - \left(-A \sin \frac{w_0 t}{T} - y_1^2 \right) \dot{y}_1 + (2w_0)^2 y_1 = \varepsilon x_1 \sin(w_0 t + \rho(t)), \\ \ddot{x}_2 - \left(A \sin \frac{w_0 t}{T} - x_2^2 \right) \dot{x}_2 + w_0^2 x_2 = \varepsilon y_2 \sin(w_0 t + \theta), \\ \ddot{y}_2 - \left(-A \sin \frac{w_0 t}{T} - y_2^2 \right) \dot{y}_2 + (2w_0)^2 y_2 = \varepsilon S(t), \\ \theta = \frac{w_0 w}{T}, \\ S(t) = x_1 \sin(w_0 t + \rho(t)), \end{array} \right. \quad (2)$$

где $\rho(t)$ – полезная информация. Первые два уравнения задают передатчик, $S(t)$ – передаваемая информация. Два следующих уравнения задают приемник. Для реализации СНА, нужно взять $w = \frac{\sqrt{5}-1}{2}$ – обратное золотое сечение. На Рис. 2 представлены отправленная (синий) и полученная информация (красный). Наблюдается аналогичная ситуация: сначала информация не может быть расшифрована из-за расстройки генераторов, а после реализации полной синхронизации информация хорошо накладывается друг на друга.

зации информация хорошо накладывается друг на друга (полученная информация была извлечена грубо для наглядности).

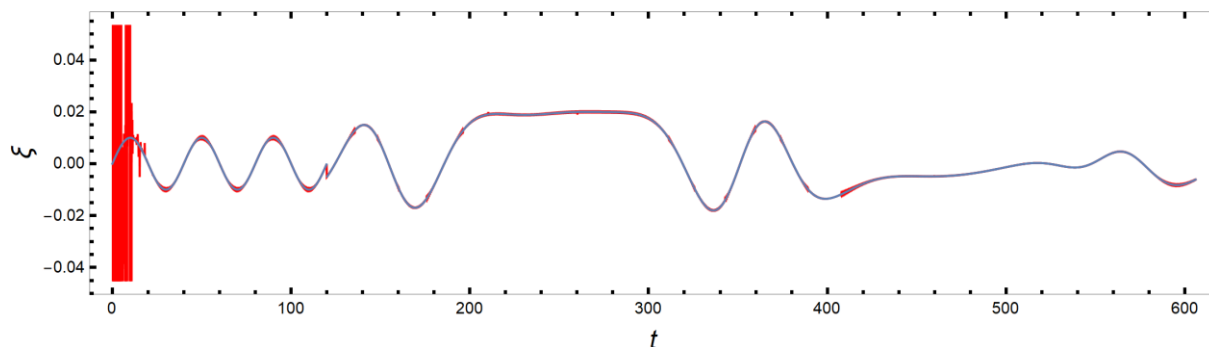


Рис. 2. Информация: (синий) полезная информация, шифруемая на передатчике; (красный) расшифрованная информация на приёмнике

Литература

1. Короновский А.А., Москаленко О.И., Храмов А.Е. О применении хаотической синхронизации для скрытой передачи информации. УФН, 2009, т. 179, н.12, с.1281–1310.
2. Дмитриев А.С., Панас А.И. Динамический хаос: новые носители информации для систем связи. М.: Изд. Физ.-мат. лит., 2002. 251 с.
3. Feudel U., Kuznetsov S., Pikovsky A. Strange Nonchaotic Attractors: Dynamics Between Order and Chaos in Quasiperiodically Forced Systems. World Scientific Publishing, 2006. 228 p. DOI: 10.1142/6006.
4. Kuznetsov, S.P. Example of a Physical System with a Hyperbolic Attractor of the Smale-Williams Type. Phys. Rev. Lett., 2005, vol. 95, 144101.
5. Isaeva O.B., Jalnine A.Yu., Kuznetsov S.P. Chaotic communication with robust hyperbolic transmitter and receiver. IEEE Xplore. Progress In Electromagnetics Research Symposium. Proceedings: St Petersburg, Russia, 22–25 May 2017, 2017. С. 3129–3136.
6. Jalnine A. Yu., Kuznetsov S. P. On the Realization of the Hunt–Ott Strange Nonchaotic Attractor in a Physical System. Technical Physics, 2007, Vol. 52, No. 4, pp. 401–408.

МЕТОДИКА УЛУЧШЕНИЯ ВИДИМОСТИ ТКАНЕЙ ГОЛОВНОГО МОЗГА В УСЛОВИЯХ МАССИВНОГО КРОВОТЕЧЕНИЯ ПО ДАННЫМ NIR-КАМЕРЫ И АЛГОРИТМОВ ШИАРЛЕТ-ПРЕОБРАЗОВАНИЯ ИЗОБРАЖЕНИЙ

А.В. Медиевский¹, А.Г. Зотин², К.В. Симонов³, М.А. Курако⁴, А.С. Кругляков³

¹*Красноярский государственный медицинский университет им. проф. В.Ф. Войно-Ясенецкого*

²*Сибирский государственный университет науки и технологий им. академика М.Ф. Решетнева*

³*Институт вычислительного моделирования СО РАН*

⁴*Сибирский федеральный университет*

Одним из самых опасных осложнений при операциях на головном мозге является кровотечение. Гемостаз может быть затруднен из-за недостаточной видимости, вызванной заполнением операционной раны кровью. Для восстановления видимости операционного поля предлагается использовать данные NIR-камеры и Shearlet-преобразование с алгоритмами цветового кодирования. При этом также предполагается проводить усиление яркостных характеристик изображений (кадров) и сегментацию интересующих биологических тканей.

Ключевые слова: эндоскопия, ближний инфракрасный диапазон, NIR, сегментация, shearlet-преобразование, цветовое кодирование, нейрохирургия, интраоперационное кровотечение.

1. Введение

Для скорейшего выздоровления пациентов в настоящее время повсеместно проводятся малоинвазивные операции. При таком способе размеры операционного поля могут не превышать нескольких сантиметров. Малый размер полости не позволяет в полной мере осуществлять хирургу необходимые манёвры с должной скоростью и в случае действий хирурга возможно массивное кровотечение, которое значительно закрывает имеющийся обзор с эндоскопа. Одной из основных проблем при кровотечении является сложность определения его источника. Вызвано это может быть плохой видимостью при заполнении кровью всего оперативного доступа. На данный момент не существует видеоэндоскопического оборудования, которое могло бы визуализировать интересующие ткани сквозь массивный объём излившейся крови. В таких случаях хирургу приходится работать практически вслепую. Для решения данной проблемы предлагается применение особого эндоскопического оборудования со встроенным сенсором ближнего инфракрасного диапазона. Предлагается обработка снимков при помощи набора алгоритмов для шумоподавления и улучшения яркостных характеристик, а также шиарлет-преобразования с применением цветового кодирования объектов интереса.

2. Материалы и методы

2.1. Постановка задачи

Кровь имеет красный цвет в связи с наличием в эритроцитах белка гемоглобина. Основными его функциями являются связывание и транспорт молекул кислорода. В зависимости от количества связанных молекул кислорода меняется спектр поглощаемого света. Нивелирование различий коэффициентов поглощения может быть достигнуто при использовании ближнего инфракрасного света (NIR) длинной волны 800 нм [1]. Лучи NIR способны проникать в ткани глубже, чем видимый свет [2], также NIR в большей степени поглощается гемоглобином, чем окружающими тканями [3]. Кровь имеет значительно выше коэффициент поглощения лучей NIR в диапазоне от 800 нм до 1050 нм в сравнении с коэффициентом поглощения того же спектра тканями, задействованными во время операции (рис. 1) [4]. Благодаря данному эффекту

имеется потенциальная возможность обнаружения интересных структур, несмотря на кровотечение в полости раны.

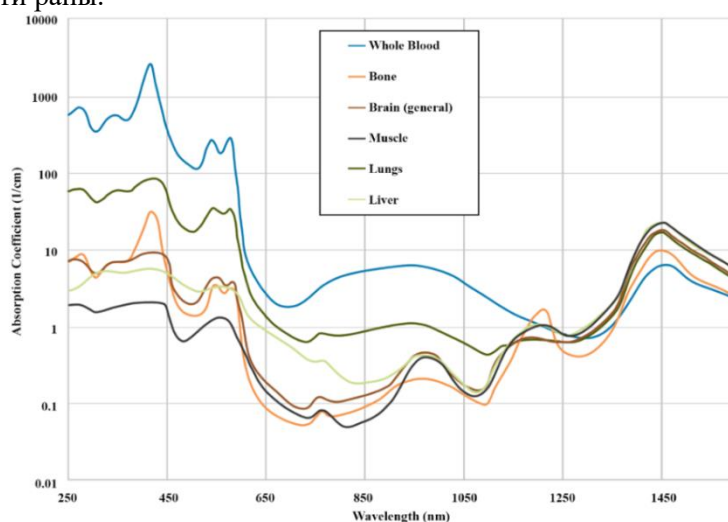


Рис. 1. Спектры коэффициентов поглощения для биологических тканей

Используя в виде источника освещения лучи инфракрасного спектра, изображение будет формироваться следующим образом: одна часть электромагнитного излучения поглотится элементами крови, а другая часть достигнет тканей операционной раны. Далее отражённые лучи вновь пройдут сквозь элементы крови. Те лучи, которые не испытают поглощение, попадут на линзу эндоскопа, что позволит сформировать двумерное изображение.

2.2. NIR-эндоскопия

Для освещения NIR лучами глубокорасположенных тканей предлагается использовать медицинский эндоскоп с двумя оптоволоконными пучками. Это позволит добиться минимальных размеров устройства и вывести изображения без задержки, свойственной для цифровых эндоскопов. В ходе эксперимента необходимо получить два вида изображения (в обычном спектре и в ИК) и осуществлять переключение между ними в автоматизированном режиме. Для решения данной задачи устанавливать две камеры на дистальный конец эндоскопа не рационально, поскольку это значительно увеличит размеры устройства. Поэтому в связи с вышеуказанными особенностями лучше использовать оптоволоконно/оптическую трубку. Малые размеры волоконно-оптического эндоскопа/ оптической трубки являются главным преимуществом при решении поставленной задачи. В [5] описана схема эндоскопа, диаметр наконечника которого не превышает 900 мкм. Это позволяет использовать его в комбинации с другим хирургическим инструментарием без необходимости расширять операционное поле. Указанная особенность важна при остановке кровотечения электрокоагулятором в труднодоступных местах.

Для проверки достоверности методики визуализации биологических тканей в момент кровотечения подготовлена модель операционного поля. С этой целью использовался гистологический препарат человеческого головного мозга, помещённый в пластиковую прозрачную чашку с венозной кровью. На рисунке 2 продемонстрированы отличия при съёмке интересных структур в диапазоне видимого излучения и в ближнем инфракрасном спектре. Применение камер NIR (рис. 2b) позволяет рассмотреть структуры, которые располагаются более глубоко, по сравнению с обычной съёмкой (рис. 2a). Полагаем, что у хирургов, несмотря на массивное кровотечение, появляется возможность оперативно получать изображение операционного поля и осуществить необходимые манипуляции.

В настоящее время существует несколько подходов, позволяющих нейрохирургу сохранить видимость и произвести гемостаз при развитии интраоперационного кровотечения. Операции с использованием эндоскопической аппаратуры позволяют применить такие методики, как dry-field maneuver [6, 7] и small-chamber irrigation technique (SCIT) [8]. Метод dry-field maneuver основан на аспирации ликвора из желудочков, что способствует выявлению и коагуляции источ-

ника кровотечения. В методе SCIT используется дополнительная жидкость для орошения операционного поля. Раствор вымывает эритроциты непосредственно перед камерой, что очищает поле зрения эндоскопа. В тоже время эффективное использование методов возможно только при работе с незначительными кровотечениями.

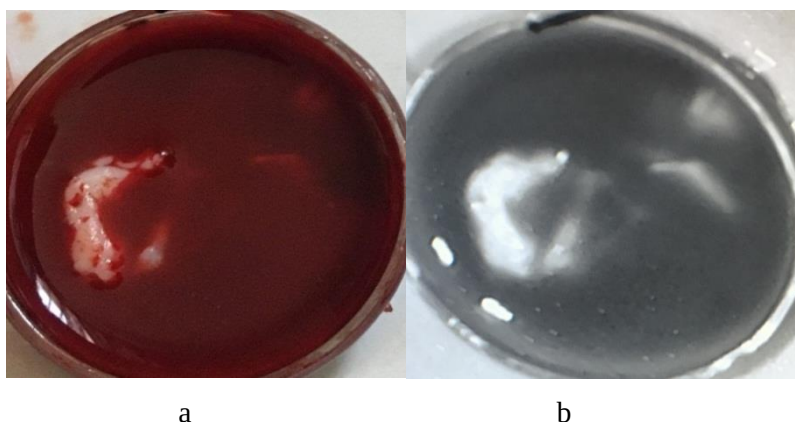


Рис. 2. Модели операционного поля при кровотечении: (а) съёмка в диапазоне видимого излучения; (б) съёмка в ближнем инфракрасном диапазоне

Отличительной чертой предлагаемого метода оперативной визуализации операционного поля, работающего в режиме реального времени, является отсутствие конкуренции с существующими подходами. Отметим, что в современных публикациях описывается применение ИК эндоскопа для различных задач. В тоже время применение его в момент кровотечения ещё никто не описывал. Данная особенность позволяет использовать разработанный метод визуализации в виде дополнения к имеющимся способам визуализации. Это объясняется тем, что наш метод работает обособленно и обрабатывает только входящее изображение. Поэтому не требуется от нейрохирургов вводить дополнительное оборудование в операционную рану.

2.3. Описание предлагаемого метода обработки

После получения снимков выполняется их обработка в рамках двух этапов: предварительная обработка и формирование цветового представления. Предварительная обработка включает шумоподавление и коррекцию яркостных характеристик. Для подавления шумов было решено использовать адаптивный медианный фильтр весовыми коэффициентами. Адаптация структуры ядра S_{xy} с размерностью 7×7 определяется на основе пороговых границ. Нижний и верхний пороги формируются на основании медианы и sample standard deviation по ядру: The lower and upper thresholds are formed based on median and sample standard deviation values in area of kernel:

$$T_{min}(S_{xy}) = \text{med}(Im_{S_{xy}}) - SD(x, y) \quad T_{max}(S_{xy}) = \text{med}(Im_{S_{xy}}) + SD(x, y).$$

$$SD(x, y) = \sqrt{\frac{1}{N^2 - 1} \sum_{(s,t) \in S_{xy}} (Im(s, t) - \text{avg}(Im_{S_{xy}}))^2}.$$

где $\text{med}()$ – функция расчета медианы; $\text{avg}()$ – функция определения среднеарифметического значения; $Im_{S_{xy}}$ – значения интенсивности канала изображения в соответствии с ядром S_{xy} размерностью $N \times N$ с координатами обрабатываемого пиксела (x, y) ;

Распределение весов задано линейной зависимостью – уменьшения коэффициента от радиуса. Для выполнения коррекции яркостных характеристик выбраны алгоритмы Balance Contrast Enhancement Technique (BCET), Гамма коррекция, Эквиализация гистограммы, а также алгоритмов на основе вариации Retinex [9]. При использовании вариаций Retinex формирование выходного изображения происходило с применением к отклику алгоритма BCET, растяжения диапазона с учетом среднеквадратического отклонения и линейного растяжения с множителем. В основе формирования цветового представления лежит использование шиарлет-преобразования и цветового кодирования объектов интереса. Алгоритм Fast Finite Shearlet Transform (FFST) [10] – «Быстрое дискретное шиарлет-преобразование» основан на дискретном быстром прямом и обратном преобразовании Фурье. Вейвлет Мейера используется в качестве

материнского вейвлета. Алгоритм FFST позволяет решать задачи построения контуров объекта интереса на изображении. В нем представление контуров формируется как сумма коэффициентов преобразования фиксированных значений параметров для последнего сдвига шкалы и различных значений параметров:

$$f_{cont} = \sum_{k=0}^{k_{max}} \sum_{m=0}^{m_{max}} sh_{\psi}(f(j^*, k, m)),$$

где sh_{ψ} коэффициент сдвига для шкалы j^* поворота (ориентации) k и смещения m , k_{max} – максимальное число поворотов, m_{max} – максимальная величина смещения.

Процедура цветового кодирования выполняется на основе алгоритмического подбора палитры в соответствии с выявленными морфологическими особенностями изучаемого изображения.

3. Экспериментальные исследования

Для оперативной адекватной оценки хирургом видимой операционной области разработана вычислительная методика обработки снимков. В методике заложены процедуры, призванные скорректировать яркостные характеристики, произвести шумоподавление, а также улучшить контрастность и детализацию тканей объектов интереса. Оценка эффективности работы предлагаемых алгоритмов осуществлялось посредством экспертного анализа. Все представленные варианты исходных снимков изучены специалистами и выполнены сравнения (рис. 3). Камера генерирует снимки размером 1920×1080 . Оценка наблюдаемости объектов выполнялась специалистом в процессе заполнения модели операционного поля кровью. При этом формировалась эталонное описание объектов интереса.

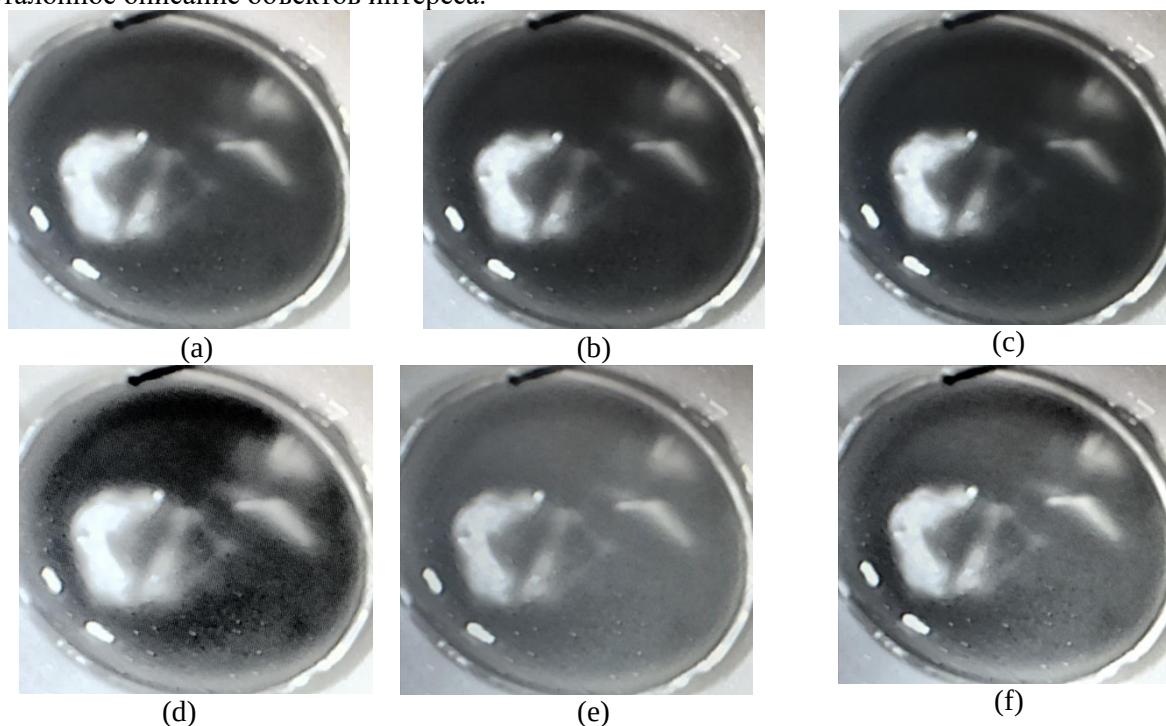


Рис. 3. Результаты применения методов контрастирования исходных изображений (a) BCET; (b) BCET mean100; (c) gamma correction 0.7 (d) Histogram equalization; (e) Retinex with BCET (f) Retinex SD3

На рисунке 3b представлен результат контрастирования исходного снимка, все визуализируемые объекты становятся пропорционально ярче. Повышается яркость исследуемой ткани мозга. Сохраняются все детали и в большей степени затемняется фон. Поэтому объект исследования воспринимается лучше. На рисунке 3c фон затемняется ещё сильнее в сравнении с рисунком 3b. Поэтому улучшается чёткость границ, но незначительно затемняются глубокие ткани мозга, которые необходимо выделять. Представленный на рисунке 3d вариант контрастиро-

вания в наибольшей степени подходит для визуализации глубоких тканей. Они становятся значительно ярче и при этом легко дифференцируются от фона. Однако, для визуализации выше лежащих тканей данный метод не подходит в виду сильного повышения яркостного параметра. Это приводит к гомогенности и неразличимости оттенков. В итоге, делает невозможным создание объёмного представления об оперируемой области.

На основании сравнительного анализа между использованными вариантами контрастирования можно выделить три наиболее успешных. Результаты отображены на рисунке 3b, рисунке 3с и рисунке 3d. Эти снимки отобраны для дальнейшей обработки (рис. 4) при помощи предлагаемой методики на основе модифицированного алгоритма шиарлет-преобразования и цветового кодирования объектов интереса [10].

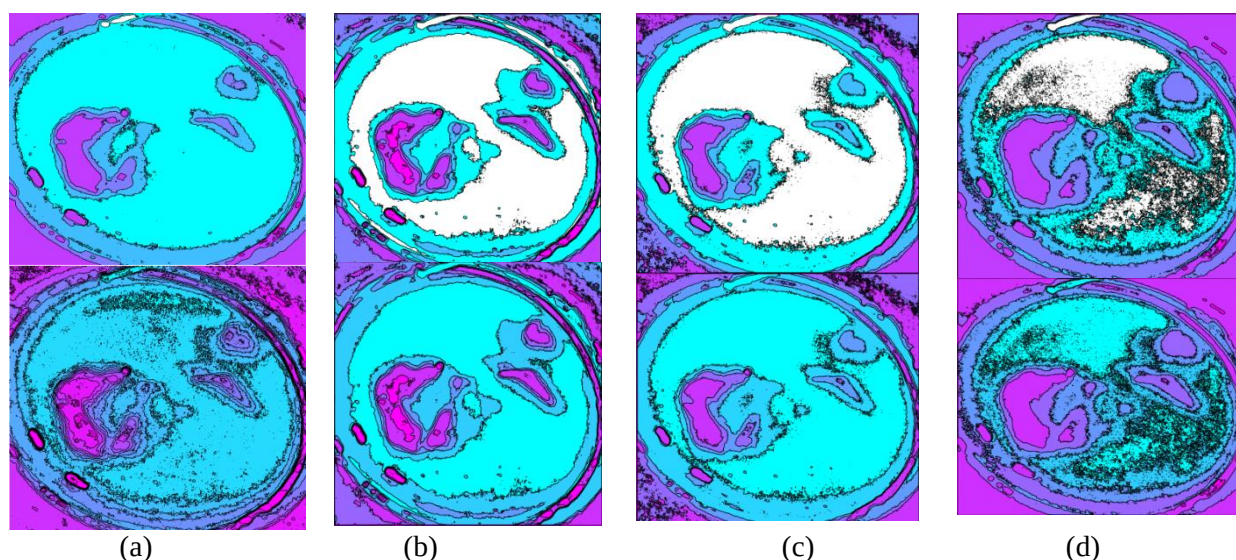


Рис. 4. Результаты применения методики шиарлет преобразования и цветового кодирования:
(а) обработка исходного изображения; (b) цветовое кодирование рис. 3b; (с) цветовое кодирование рис. 3с; (d) цветовое кодирование рис. 3d

На рисунке 4а представлена обработка исходного снимка на основе применения методики шиарлет-преобразования и цветового кодирования. Это позволяет получать искомое представление об объекте интереса. В тоже время отметим, что на рисунке 4а снимок первого ряда не достаточно корректно отображает морфологию биологических структур. Снимок второго ряда (рис. 4а) усложняет анализ морфологии нижележащих тканей. Поскольку их цветовой диапазон близок к цветовому диапазону фона представленного кровью. Поэтому нижележащие ткани и кровь становятся труднее различить между собой. Все остальные снимки первого и второго ряда, приведенные на рисунке 4b, рисунке 4с и рисунке 4d, имеют богатый цветовой градиент. Это позволяет детально визуализировать рельефность тканей мозга. Но наиболее адекватным в плане восприятия является изображение первого ряда на рисунке 4b. Данный вариант совместил все необходимые характеристики: более полный градиент, точность морфологии и хорошая контрастность с фоном.

В таблице 1 приведена сравнительная характеристика точности выделения объектов интереса на указанных выше снимках. В рамках этого этапа исследования нацеленного на определения объектов интересов использованы серии снимков, для которых эксперт сформировал координатное описание объектов интереса с выделением контуров. Данный этап исследования показал, что предлагаемый способ предварительной обработки снимков позволяет повысить точность определения положения искомых областей на 2.8-3.9%, и выделения соответствующих границ на 2.3-4.8%. При этом средняя точность определения контуров на модельных изображениях составляла 0.96 ± 0.01 по метрике Figure Of Merit и 0.95 ± 0.01 Dice similarity coefficient.

Таблица 1. Сравнительная характеристика точности выделения объектов интереса

	Рисунок 4а	Рисунок 4b	Рисунок 4с	Рисунок 4d
Figure Of Merit	0,84	0,96	0,90	0,93
Dice similarity coefficient	0,83	0,95	0,90	0,93

4. Заключение

Разработана вычислительная методика для визуализации операционного поля во время массивного кровотечения, которая способна функционировать в режиме реального времени. Предлагается методика использования NIR эндоскопа с алгоритмами контрастирования, шиар-лет преобразования и цветового кодирования в малоинвазивной хирургии. В рамках экспериментального исследования выполнено обоснование и тестирование методик и комбинации алгоритмов. Особенностью разработанного подхода является то, что он может применяться как самостоятельный метод визуализации тканей, так и совместно с уже существующими.

Литература

1. Dervieux E., Bodinier Q., Uhring W., Theron M. Measuring hemoglobin spectra: searching for carbamino-hemoglobin // Journal of Biomedical Optics. 2020. Vol. 25. P. 1–15.
2. Shourav M.K., Choi J., Kim J.K. Visualization of superficial vein dynamics in dorsal hand by near-infrared imaging in response to elevated local temperature // J. Biomed Opt. 2021. Vol. 26 (2). 026003.
3. Wang L., Leedham G., Gho S. Infrared imaging of hand vein patterns for biometric purposes // IET Computut. Vis. 2007. Vol. 1, № 3. P. 113–122.
4. Calabro K.W. Modeling Biological Tissues in LightTools. 2020.
5. Buenconsejo Andrea Louise et al. Submillimeter diameter rotary-pullback fiber-optic endoscope for narrowband red-green-blue reflectance, optical coherence tomography, and autofluorescence in vivo imaging // Journal of biomedical optics. 2019. Vol. 25(3). P. 1–7.
6. Turhan T. Dry-field maneuver for controlling the massive intraventricular bleeding during neuro-endoscopic procedures. Child's nervous system // Journal of the International Society for Pediatric Neurosurgery. 2018. Vol. 34(3). P. 541–545.
7. Oertel J., Linsler S., Csokonay A., Schroeder H., Senger S. Management of severe intraoperative hemorrhage during intraventricular neuroendoscopic procedures: the dry field technique // Journal of neurosurgery. 2018. Vol. 131(3). P. 931–935.
8. Manwaring J.C., El Damaty A., Baldauf J., Schroeder H.W. The small-chamber irrigation technique (SCIT): a simple maneuver for managing intraoperative hemorrhage during endoscopic intraventricular surgery // Neurosurgery. 2014. Vol. 10, Suppl. 3. P. 375–379.
9. Zotin A.G. Fast algorithm of image enhancement based on multi-scale retinex // Int. J. Reasoning-based Intelligent Systems. 2020. Vol. 12, № 2. P. 106–116.
10. Zotin A., Simonov K., Kapsargin F., Cherepanova T., Kruglyakov A., Cadena L. Techniques for Medical Images Processing Using Shearlet Transform and Color Coding. / In: Favorskaya M., Jain L. (eds). Computer Vision in Control Systems-4. Intelligent Systems Reference Library. 2018. Vol. 136. P. 223–259. Springer, Cham.

РАСПАРАЛЛЕЛИВАНИЕ ГИБРИДНОГО АЛГОРИТМА МУРАВЬИНОЙ КОЛОНИИ С ИЗМЕНЯЮЩИМИСЯ С ПОМОЩЬЮ ГЕНЕТИЧЕСКОГО АЛГОРИТМА ПАРАМЕТРАМИ^{1*}

И.И. Микулик, Е.А. Благовещенская

Петербургский государственный университет путей сообщения императора Александра I

В работе рассмотрена возможность параллельной реализации гибридного метода оптимизации, использующего эвристики муравьиного алгоритма и генетического, для решения задачи коммивояжера. Задача коммивояжера является NP-полной задачей и часто применяется для тестирования алгоритмов комбинаторной оптимизации. В качестве исследуемого алгоритма рассматривается алгоритм простой муравьиной колонии с изменяющимися с помощью генетического алгоритма параметрами. Показано, что распараллеливание данного метода приводит к увеличению скорости выполнения работы алгоритма.

Ключевые слова: распараллеливание, задача коммивояжера, методы оптимизации, муравьиные алгоритмы, генетический алгоритм, параллельные вычисления

1. Введение

Согласно эмпирическому закону Мура количество вычислительных узлов на микросхемах с каждым годом возрастает, следовательно, возрастают и вычислительные возможности устройств. Для эффективного использования вычислительных мощностей необходима реализация эффективных с точки зрения распараллеливания алгоритмов. Также распараллеливание является важной задачей для работы с суперкомпьютером – устройством, являющимся множеством вычислительных кластеров, соединенных высокоскоростной магистралью. Развиваются также и фундаментальные науки, которые вносят вклад в расширение возможностей прикладных наук. В работе [1] показано, что к задачам распараллеливания вычислительных процессов могут быть применены алгоритмы прямых разложений абелевых групп без кручения.

В работе рассматривается возможность параллельной реализации одной из модификаций алгоритма муравьиной колонии. Алгоритм муравьиной колонии – один из эмпирических методов оптимизации, использующийся для решения комбинаторных задач. Он имеет практическое применение для задач логистики [2,3], задач диспетчеризации [4], планирования маршрутов [5] и других. Алгоритм относится к классу так называемых биоинспирированных алгоритмов, поведение которых вдохновлено примерами из природы.

Общим недостатком методов класса муравьиных алгоритмов является чувствительность к параметрам, которые устанавливают баланс между расширением пространства поиска и эксплуатацией найденных решений. Оптимальные параметры могут быть специфичными для конкретной задачи и зависеть от требуемой точности решения. Поэтому некоторые работы посвящены динамической адаптации параметров муравьиного алгоритма [6]. В работе рассматривается один из таких способов, идеей которого является внедрение генетического алгоритма для обновления значений параметров.

2. Постановка задачи

Исследование проводилось на задаче коммивояжера (TSP). Задача является типовой и часто применяется для тестирования алгоритмов комбинаторной оптимизации. Поэтому существует большое количество исследований, с которыми можно сравнить полученные результаты. Задача состоит в нахождении кратчайшего пути, проходящего через все заданные города с ко-

^{1*} Исследование выполнено за счет гранта Российского научного фонда (проект № 22-21-00267)

нечным возвращением в исходный город. В терминах теории графов задача может быть сформулирована как нахождение кратчайшего гамильтонова цикла.

Задача является NP-полной и трансвычислительной, поэтому поиск точного решения задачи не всегда возможен из-за ограничений по времени и вычислительных ресурсов. В этом случае на практике часто используются приближенные алгоритмы, дающие субоптимальные решения. Алгоритмы различаются как по затрачиваемым ресурсам (время, память), так и по точности аппроксимации.

3. Описание алгоритма

Рассматриваемый алгоритм является гибридным, так как он основан на двух методах. В отличие от некоторых других гибридных алгоритмов [7] [8], предлагаемый алгоритм не является последовательным применением нескольких алгоритмов. Его идея состоит в том, чтобы применить генетический алгоритм для оптимизации параметров основного алгоритма оптимизации – муравьиной колонии. Сформулируем условия алгоритма.

Каждый муравей k имеет динамические параметры α_k, β_k, Q_k .

α_k – чувствительность. Определяет насколько чувствителен муравей к феромонам.

β_k – эвристическая чувствительность. Определяет насколько чувствителен муравей к эвристической информации.

Q_k – интенсивность феромона. Определяет количество феромона, которое будет отложено муравьем.

Другие статические общие параметры: n, ρ, τ_0 .

n – количество муравьев.

τ_0 – небольшое количество феромона. Определяет верхнюю границу количества феромона, который может быть отложен на дуге графа при инициализации. Этот параметр также является неизменяемым, так как участвует в работе алгоритма один раз.

ρ – скорость испарения феромона.

Каждый муравей выбирает дугу случайно, с вероятностью определяемой формулой

$$p_{ij}^k(t) = \frac{\tau_{ij}^{\alpha_k}(t) \eta_{ij}^{\beta_k}(t)}{\sum \tau_{iu}^{\alpha_k}(t) \eta_{iu}^{\beta_k}(t)}, \quad (1)$$

где $p_{ij}^k(t)$ – вероятность выбора дуги ij муравьем k на итерации t .

Каждый муравей откладывает феромон в соответствии с формулой:

$$\Delta \tau_{ij} = \frac{Q_k}{f(x_k)}, \quad (2)$$

где $\Delta \tau_{ij}$ – значение феромона; $f(x_k)$ длина пути x найденная муравьем k .

Алгоритм можно представить в виде последовательности действий:

Начало

Определить параметры: n, τ_0, ρ .

Для каждой дуги ij :

Присвоить случайное значение феромона $\tau_{ij} < \tau_0$.

Пока не выполнен критерий останова:

Для каждого муравья $k = 1..n$:

$x_k(t) = \emptyset$.

Пока полный путь не построен:

Выбрать следующую вершину j в соответствии с правилом (1)

$x_k(t) = x_k(t) \cup (i, j)$.

Для каждого ребра (i, j) :

Вычислить испарение и новую концентрацию феромона с помощью правила (2)

Создать новый набор параметров муравья с помощью оператора селекции (3)

Применить оператор кроссинговера для нахождения параметров

Применить оператор мутации для нахождения параметров

Установить новые параметры α_k, β_k, Q_k для муравьев.

Конец

Оператор селекции, который вероятностно определяет новый набор параметров задается следующим образом:

$$P(M_k) = \frac{f(x_k)}{\sum f(x_k)}, \quad (3)$$

где $P(M_k)$ – вероятность выбора k -го параметра муравья для дальнейшего его участия в работе алгоритма.

Оператором кроссинговера является побитовое сложение (исключающее «или») битовых представлений значений параметров. Оператор мутации случайным образом изменяет бит битового представления каждого параметра.

4. Параллельная реализация

Известно, что муравьиные алгоритмы имеют параллельную реализацию [9], так как поиск решения отдельного муравья одной итерации не зависит от результатов поиска других муравьев. Однако рассматриваемый метод является гибридным, а значит существует возможность появления новых зависимостей.

Генетический алгоритм также имеет параллельную реализацию. Однако распараллеливание генетического алгоритма представляет распараллеливание работы каждого оператора по отдельности. Таким образом, все вычислительные кластеры вынуждены ожидать завершения вычислений остальных кластеров при выполнении одного из генетических операторов, так как операторы применяются на всей популяции, а не на каждой особи отдельно.

Вычисление оператора селекции, реализованного в виде «колеса-рулетки», выполняется в два этапа. Первый, не распараллеливаемый, определяет длину «рулетки», то есть сумму фитнесс-функций всех особей ($\sum f(x_k)$). Второй этап – подсчет вероятности $P(M_k)$ по формуле (3), который может быть распараллелен.

Оператор кроссинговера складывает особым образом значения параметров попарно, что также может быть вычислено на отдельных независимых кластерах. Оператор мутации работает с каждым значением отдельно, поэтому тоже имеет параллельную реализацию. Таким образом, алгоритм имеет следующую параллельную модификацию:

Пусть m – количество вычислительных узлов, а $p \in P$ – отдельно взятый узел из множества вычислительных узлов P .

Начало

Определить параметры: n_p, τ_0, ρ .

Разбить множество всех ребер графа на множества w_p для каждого узла p

Для каждого вычислительного узла p :

Для каждой дуги $ij \in w_p$:

Присвоить случайное значение феромона $\tau_{ij} < \tau_0$.

Пока не выполнен критерий останова:

Для каждого вычислительного узла p :

Для каждого муравья $k = 1..n_p$:

$x_k(t) = \emptyset$.

Пока полный путь не построен:

Выбрать следующую вершину j в соответствии с правилом (1)

$x_k(t) = x_k(t) \cup (i, j)$.

Для каждого ребра (i, j) :

Вычислить испарение и новую концентрацию феромона с помощью правила (2)

Определить $\sum f(x_k)$.

Разбить множество всех $f(x_k)$ на m подмножеств

Для каждого вычислительного узла p :

Определить $P(M_k)$ в соответствии с правилом (3)

Определить пары для кроссинговера

Разбить множество пар на m подмножеств

Для каждого вычислительного узла p :

Применить оператор кроссинговера для нахождения параметров
Применить оператор мутации для нахождения параметров
Установить новые параметры α_k, β_k, Q_k для муравьев.
Конец

Таким образом, алгоритм может быть распараллелен по данным, но имеет ограничения в виде синхронизации данных на разных этапах его выполнения.

Литература

1. Благовещенская Е.А., Зуев Д.В., Кузнецова И.В., Тихомиров С.А. Приложения алгоритмов прямых разложений абелевых групп без кручения к задачам распараллеливания вычислительных и конструктивных процессов // Международные Колмогоровские чтения-XIV, посвященные 100-летию профессора ЗА Скопеца. 2017. С. 38–40.
2. Hassan M. et al. An improved ACS algorithm for the solutions of larger tsp problems //arXiv preprint arXiv:1304.3763. – 2013.
3. Batmetan J. R. et al. A multiple-objective ant colony algorithm for optimizing disaster relief logistics //Advanced Science Letters. – 2017. – Т. 23. – №. 3. – С. 2344-2347.
4. Liang D. et al. An efficient ant colony system approach for new energy vehicle dispatch problem //IEEE Transactions on Intelligent Transportation Systems. – 2019. – Т. 21. – №. 11. – С. 4784-4797.
5. Qian X., Zhong X. Optimal individualized multimedia tourism route planning based on ant colony algorithms and large data hidden mining //Multimedia Tools and Applications. – 2019. – Т. 78. – №. 15. – С. 22099-22108.
6. Esquinas-Requena J. L. et al. La anemia aumenta el riesgo de mortalidad debido a fragilidad y discapacidad en mayores: Estudio FRADEA //Atencion primaria. – 2020. – Т. 52. – №. 7. – С. 452-461.
7. Lee Z. J. et al. Genetic algorithm with ant colony optimization (GA-ACO) for multiple sequence alignment //Applied Soft Computing. – 2008. – Т. 8. – №. 1. – С. 55-78.
8. Yan-hua Z., Lei F., Zhi Y. Optimization of cloud database route scheduling based on combination of genetic algorithm and ant colony algorithm //Procedia Engineering. – 2011. – Т. 15. – С. 3341-3345.
9. Скобцов Ю.А., Федоров Е.Е. Метаэвристики: монография. Донецк: Издательство «Ноу-лидж», 2013. 426 с.

ОБЗОР МЕТОДОВ ОБЪЯСНЯЕМОГО ИСКУССТВЕННОГО ИНТЕЛЛЕКТ В ЗАДАЧАХ ОБРАБОТКИ ЭКГ

С.А. Насибли

Нижегородский государственный университет им. Н.И. Лобачевского

В данной работе сделан обзор наиболее популярных методов ХАИ, а также их применение в задачах и доменах из медицинской сферы. Наиболее часто используемыми методами оказались локальные post-hoc методы различных спецификаций: model-agnostic (LIME, SHAP), и model-aware (Input-gradient, integrated gradient, LRP, также называемые saliency maps методы). В обзоре указан растущий интерес и объем исследований в ХАИ и особенно их частое применение в медицинской сфере, и частности в задачах по обработке и интерпретации ЭКГ данных. Данная область является молодым и стремительно растущим и есть ещё пространство для развития существующих систем и создания новых методов. Делаются попытки для разработки новых интерфейсов ХАИ для автоматической интерпретации ЭКГ, чтобы улучшить работу интерпретатора ЭКГ и повысить точность диагностики.

Ключевые слова: Объяснимый искусственный интеллект, глубокое обучение, машинное обучение, электрокардиограмма

1. Введение

В данной работе сделан общий обзор различных методов объясняющего искусственного интеллекта ХАИ и использование искусственного интеллекта (АИ) в интерпретации ЭКГ и возможности улучшения необъяснимости АИ (ХАИ) при обработке ЭКГ. При разработке систем ИИ необходимо учитывать многие принципы (права человека, благополучие, информационное агентство, эффективность, прозрачность, подотчетность, осведомленность о неправомерном использовании и компетентность), чтобы гарантировать надежность и применимость ИИ. Современные алгоритмы интерпретации ЭКГ позволяют обнаруживать различные типы сердечных заболеваний. Однако есть некоторые проблемы и недостатки, которые необходимо решить, например, проблема необъяснимости и взаимодействия между человеком и ИИ для принятия клинических решений. Эти проблемы создают возможности для разработки надежного ХАИ для автоматической интерпретации ЭКГ с высокой производительностью и высоким уровнем достоверности. В будущем должны быть разработаны новые интерфейсы ХАИ, которые будут способствовать большей прозрачности логики принятия решений алгоритма, что может позволить пользователям откалибровать свое доверие и использование системы ИИ.

2. Объяснимый искусственный интеллект в медицине

Объяснимый искусственный интеллект - достаточно новая и развивающаяся область машинного обучения, которая посвящена тому, чтобы обучать модели машинного обучения с сохранением их эффективности, но при условии того, что они будут более интерпретируемыми, чем модели текущие (изначальные). Это включает в себя как и новые методы обучения моделей, так и объяснения уже построенных обычных моделей (например, POST-HOC explanation).

2.1. Методы объяснимого искусственного интеллекта

Не секрет, что последний прогресс в области искусственного интеллекта тесно связан с развитием глубоких нейронных сетей. К сожалению, они являются черными ящиками (black-box) и никто не понимает, как они работают и почему они выдают такие предсказания, которые дают. Их пользователи просто видят их достаточно высокую эффективность на конкретных задачах, на конкретных метриках, однако в то же время процесс принять решение остается загадкой. Чаще всего люди, которые используют эти модели, даже люди которые их обучают, не

очень понимают, как и где заканчиваются пределы возможностей этих моделей, почему она принимает такие решения, которые она принимает? Как это все можно отлаживать (дебажить)? Как можно изменить/обучить модель, чтобы оно работало лучше? ХАИ задается такими вопросами [4, 5], чтобы создавать новые процессы обучения, объяснять уже готовые модели, чтобы люди, которые используют модели, тренируют их, понимали, почему они работают, могли им доверять.

Можно построить некую иерархию для post-hoc explanations методов. Post-hoc explanations можно разделить на два больших класса, это global expansions и local expansions.

Post-hoc explanations:

- Global explanations (Activation Maximization), глобальная интерпретация моделей заключается в том, что мы интерпретируем в общем модели, в общем их стратегии, без привязки к конкретным примерам. Одним из самых наверное популярных и известных алгоритмов [10] для глобальной интерпретации модель является метод активации максимизации (Activation Maximization, AM), а точнее это целое семейство методов.
- Local explanations, Локальные объяснения:
 - Model-agnostic (Lime, SHAP) [7, 15]. LIME [7] Local interpretable model-agnostic, локальные интерпретируемые не зависящие от модели объяснения, создают объяснение для определенных входных данных, аппроксимируя модель локально интерпретируемой, например, линейной моделью. SHAP [15] SHapley Additive exPlanations, Аддитивные объяснения Шапли: теоретико-игровой подход к объяснениям машинного обучения.
 - Model-aware (Input Gradient, Integrated Gradients, LRP) [16-17]. Model-aware методы же, в отличие от model-agnostic, используют внутреннюю структуру модели машинного обучения, то есть вычислительный граф. Их достоинства в том, что они работают быстрее и чаще дают более точные интерпретации.

2.2. Объяснимый искусственный интеллект в задачах обработки ЭКГ

В данной работе представлены определенные промежуточные результаты в обзоре методов ХАИ в контексте использования в задачах ЕСГ (ЭКГ). Заложено начало в исследовании применения ХАИ в рамках задачи ЭКГ, приложены систематические карты исследования искусственного интеллекта (ИИ) в здравоохранении, в рамках задач классификаций, визуализаций результатов классификаций на основе ХАИ. Особо важную роль в сфере здравоохранения имеют четыре основных этических вопроса, в различных опубликованных исследованиях, - это прозрачность, справедливость и беспристрастность, подотчетность и ответственность, а также конфиденциальность и безопасность. Поскольку индустрия искусственного интеллекта постоянно расширяется, этика искусственного интеллекта в здравоохранении будет вызывать все большую озабоченность общества. Кроме того, в области ИИ может не хватать информации, ясности и структуры. Машинное обучение и глубокое обучение играют жизненно важную роль в классификации различных аритмий с использованием данных электрокардиографии (ЭКГ). В особенности модели глубокого обучения похожи на черный ящик, без объяснения прогнозируемых результатов, что часто требуется в клинической медицине. Это ограничивает применение глубокого обучения в реальных системах здравоохранения.

В данной работе делается попытка резюмировать промежуточный результат исследований в области использования искусственного интеллекта (AI) в интерпретации ЭКГ и возможности улучшения необъяснимости AI (ХАИ) при чтении ЭКГ. При разработке систем ИИ необходимо учитывать многие принципы (права человека, благополучие, информационное агентство, эффективность, прозрачность, подотчетность, осведомленность о неправомерном использовании и компетентность), чтобы обеспечить надежность и применимость ИИ. Современные алгоритмы интерпретации ЭКГ позволяют обнаруживать различные типы сердечных заболеваний. Однако есть некоторые проблемы и недостатки, которые необходимо решить, например, проблема объяснимости и взаимодействия между человеком и ИИ для принятия клинических решений. Эти проблемы создают возможности для разработки надежного ХАИ для автоматической интерпретации ЭКГ с высокой производительностью и высоким уровнем достоверности.

Всего было проанализировано 37 статей, и в результате было получено, что на текущий момент самыми популярными методами в области обработки ЭКГ данных являются LIME и SHAP локальные model-agnostic методы.

В прил. 1 один приведены некоторые примеры работ [18-29], в которых используются различные модели для обработки и интерпретации ЭКГ сигналов XAI методами.

В приложении 2 представлены график различных аспектов исследований и вклада по годам с 2019 по 2020, начало 2021. Можно заметить, что касательно вопроса о заинтересованных сторонах и этических вопросов, то также интересно увидеть их эволюцию с точки зрения внимания со стороны академической литературы. На рисунке представлена пузырьковая диаграмма, которая показывает значительный рост, начиная с 2019 года. Легко увидеть, что количество статей об использовании XAI алгоритмов в автоматической интерпретации ЭКГ данных увеличилось вдвое с 2019 по 2020 год.

Автоматизация интерпретации ЭКГ может включать ML/DL, чтобы обеспечить автоматическую интерпретацию ЭКГ. Ранее были некоторые проблемы, такие как прозрачность алгоритма, что означало, что предсказание необъяснимо. Исследователи показали, что алгоритмы ЭКГ, которые просто представляют решение без какого-либо объяснения, могут вызвать у читателя предвзятость автоматизации. Предвзятость автоматизации - это когда читатель просто принимает машину без критического обзора [22]. Следовательно, есть очевидные возможности для улучшения автоматической интерпретации ЭКГ путем предоставления пользовательского интерфейса XAI во время интерпретации ЭКГ. Применение XAI может улучшить интерпретацию ЭКГ и, как следствие, улучшить принятие решений в кардиологии. Существуют различные методы разработки интерфейса XAI [14-29], такие как отображение внимания / тепловой карты для подходов DL или правила для символических подходов. Важно подчеркнуть, какие функции были важны и учитывались при прогнозировании, что в результате могло бы повлиять на доверие переводчика и уменьшить систематическую ошибку при интерпретации ЭКГ. Более того, XAI может решить не технологические проблемы, такие как внедрение ИИ врачами, потому что ИИ часто считается «черным ящиком», и только ~50% врачей знакомы с технологиями ИИ [25]. Исследования указывают на возможность разработки новых интерфейсов XAI для автоматической интерпретации ЭКГ, чтобы улучшить работу интерпретатора ЭКГ и повысить точность диагностики. Эти новые интерфейсы следует разработать и протестировать, чтобы продемонстрировать преимущества XAI.

Литература

1. The Artificial Intelligence Channel. The Great AI Debate - NIPS2017 - Yann LeCun. [YouTube channel]. Retrieved August 29, 2021, from <https://www.youtube.com/watch?v=93Xv8vJ2acI>.
2. Ribeiro, M. T., Singh, S., Guestrin, C. 2016. "Why should I trust you?": explaining the predictions of any classifier. In Proceedings of the 22nd SIGKDD International Conference on Knowledge Discovery and Data Mining, 1135-1144.
3. Arrieta, Alejandro Barredo, et al. "Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI". Information Fusion 58 (Jun 2020): 82-115.
4. Gunning, D., & Aha, D. (2019). DARPA's Explainable Artificial Intelligence (XAI) Program. AI Magazine, 40(2), 44-58. <https://doi.org/10.1609/aimag.v40i2.2850>
5. Vaishak Belle, Ioannis Papantonis. "Principles and Practice of Explainable Machine Learning" (2020) arXiv preprint: arXiv:2009.11698.
6. Samhita, L., & Gross, H. J. (2013). The "clever Hans phenomenon" revisited. communicative & integrative Biology, 6(6), e27122.
7. Ribeiro, M.R., Singh, S., & Guestrin C. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining (pp. 1135-1144).
8. Everingham, M, van Gool, L, Williams, CKI, Winn, J & Zisserman, A 2010, 'The PASCAL Visual Object Classes (VOC) Challenge', International Journal of Computer Vision, vol. 88, no. 2, pp. 303-338.

9. Lapuschkin, S., Binder, A., Montavon, G., Samek, W., & Müller, K. R. Unmasking Clever Hans predictors and assessing what machines really learn. *Nature communications*, 10(1), 1-8.
10. Erhan, D., Bengio, Y., Courville, A., & Vincent, P. Visualizing Higher-Layer Features of a Deep Network, 2009. University of Montreal, 134(3), 1.
11. Nguyen, A., Dosovitskiy, A., Yosinski, J., Brox, T., & Clune, J. (2016) Synthesizing the preferred inputs for neurons in neural networks via deep generator networks. In *Advances in neural information processing systems* (pp. 3387-3395).
12. Olah, C., Mordvintsev, A., & Schubert, L. (2017). Feature visualization, 2(11), e7.
13. Carter, S., Armstrong, Z., Schubert, L., Johnson, I., & Olah, C. Activation atlas. *Distill*, 4(3), e15.
14. Olah, C., Mordvintsev, A., & Schubert, L. (2017). Feature visualization. *Distill*, 2(11), e7.
15. Scott M Lundberg and Su-In Lee. "A unified approach to interpreting model predictions". In: *Advances in neural information processing systems*. 2017, pp4765-4774.
16. Sturmfels, P., Lundberg, S., & Lee, S. I. (2020). Visualizing the impact of feature attribution baselines, *Distill*, 5(1), e22.
17. Montanov, G., Binder, A., Lapuschkin, S., Samek, W., & Muller K. R. (2019). Layer-wise relevance propagation: an overview. In *Explainable AI: interpreting, explaining and visualizing deep learning* (pp. 193-209). Springer, Cham.
18. Jo, Y. Y., Kwon, J. M., Jeon, K. H., Cho, Y. H., Shin, J. H., Lee, Y. J., ... & Oh, B. H. (2021). Detection and classification of arrhythmia using an explainable deep learning model. *Journal of Electrocardiology*, 67, 124-132.
19. Costa Cortéz, N. A., Fernández Cortés, J., Couso Blanco, I., & Sánchez Ramos, L. (2020). Graphical analysis of the progression of atrial arrhythmia using recurrent neural networks. *International Journal of Computational Intelligence Systems*, 13 (1).
20. Raza, A., Tran, K. P., Koehl, L., & Li, S. (2021). Designing ECG Monitoring Healthcare System with Federated Transfer Learning and Explainable AI. *arXiv preprint arXiv:2105.12497*.
21. Taniguchi, H., Takata, T., Takechi, M., Furukawa, A., Iwasawa, J., Kawamura, A., ... & Tamura, Y. (2021). Explainable Artificial Intelligence Model for Diagnosis of Atrial Fibrillation Using Holter Electrocardiogram Waveforms. *International Heart Journal*, 62(3), 534-539.
22. Uddin, M. Z., & Soylu, A. (2021). Human Activity Recognition Using Wearable Sensors, Discriminant Analysis, and Long Short-Term Memorybased Neural Structured Learning.
23. Hossain, M. S., Muhammad, G., & Guizani, N. (2020). Explainable AI and mass surveillance system-based healthcare framework to combat COVID-19 like pandemics. *IEEE Network*, 34(4), 126-132.
24. Neves, I., Folgado, D., Santos, S., Barandas, M., Campagner, A., Ronzio, L., ... & Gamboa, H. (2021). Interpretable heartbeat classification using local model-agnostic explanations on ECGs. *Computers in Biology and Medicine*, 133, 104393.
25. Rashed-Al-Mahfuz, M., Moni, M. A., Lio, P., Islam, S. M. S., Berkovsky, S., Khushi, M., & Quinn, J. M. (2021). Deep convolutional neural networks based ECG beats classification to diagnose cardiovascular conditions. *Biomedical engineering letters*, 11(2), 147-162.
26. Dwivedi, P., Khan, A. A., Mugde, S., & Sharma, G. (2021, July). Diagnosing the major contributing factors in the classification of the fetal health status using cardiotocography measurements: An AutoML and XAI approach. In *2021 13th International Conference on Electronics, Computers and Artificial Intelligence (ECAI)* (pp. 1-6). IEEE.
27. Jo, Y. Y., Cho, Y., Lee, S. Y., Kwon, J. M., Kim, K. H., Jeon, K. H., ... & Oh, B. H. (2021). Explainable artificial intelligence to detect atrial fibrillation using electrocardiogram. *International Journal of Cardiology*, 328, 104-110.
28. Dave, D., Naik, H., Singhal, S., & Patel, P. (2020). Explainable ai meets healthcare: A study on heart disease dataset. *arXiv preprint arXiv:2011.03195*.
29. Vaccari, I., Orani, V., Paglialonga, A., Cambiaso, E., & Mongelli, M. (2021). A Generative Adversarial Network (GAN) Technique for Internet of Medical Things Data. *Sensors*, 21(11), 3726.

ОПТИМИЗАЦИЯ СЛОЕВ НЕЙРОННЫХ СЕТЕЙ НА АРХИТЕКТУРЕ ARM

А.А. Оболенский, А.Ю. Нестеров, И.Б. Мееров

Нижегородский государственный университет им. Н.И. Лобачевского

В работе рассматриваются способы ускорения вычислений при выводе искусственных нейронных сетей в программном комплексе OpenVINO на процессорах архитектуры ARM. На основе экспериментов с сетью `pnfnet-f0` показано, что использование параллелизма и подходящих аппроксимаций на слое GELU позволяют значительно ускорить вычисления при сохранении точности работы сети.

Ключевые слова: машинное обучение, оптимизация, ARM.

1. Введение

Использование глубоких нейронных сетей позволило достичь достаточно больших успехов в классификации изображений, распознавании речи, детектировании объектов. Со временем появилась потребность запускать вывод (inference) используемых для решения пользовательских задач нейронных сетей на достаточно слабом аппаратном обеспечении, где для работы не предполагается использование мощных графических ускорителей. Использование таких систем для запуска тяжеловесных нейронных сетей возможно только при использовании соответствующих методов, направленных на оптимизацию производительности кода.

Приемы оптимизации кода существенно зависят от особенностей используемой архитектуры. В данной работе рассматривается вывод нейронных сетей на архитектуре ARM. Она широко используется в потребительском сегменте смартфонов, а также выходит на широкий рынок ноутбуков и персональных компьютеров. Ядра ARM Cortex-M спроектированы для энергоэффективных встраиваемых устройств. Микроконтроллеры и другие устройства на базе этих процессоров часто используются для запуска вывода нейронных сетей.

В настоящее время разработаны программные продукты, которые позволяют запускать вывод нейронных сетей, используя аппаратные возможности и особенности архитектуры ARM, а также поддерживают механизмы, оптимизированные для выполнения этой задачи. В данной работе рассматривалась возможность применения различных оптимизаций для ускорения вывода двух нейронных сетей с помощью программного продукта OpenVINO Toolkit [1] на процессорах с ядрами ARM Cortex-M. Рассматриваются подходы к оптимизации производительности, приводится оценка их эффективности.

2. Цель работы

Основная цель работы заключается в изучении приемов оптимизации кода для вывода искусственных нейронных сетей на процессорах с ядрами ARM Cortex M.

Достижение этой цели требует решение следующих задач:

1. Выбрать перспективные для оптимизации слои путем проведения послойного анализа производительности вывода сети.
2. Проанализировать реализацию выбранных для оптимизации слоев.
3. Выбрать методы оптимизации производительности кода.
4. Провести вычислительные эксперименты и сделать выводы относительно эффективности проведенной работы, оценить улучшение производительности, эффективность использования железа и время работы.

3. Исследование методов оптимизации и проведение экспериментов

3.1. Программно-аппаратное окружение

Все запуски производились на Raspberry Pi 4B.

CPU: Broadcom BCM2711 (Quad core Cortex-A72 (ARM v8) 64-bit SoC @ 1.5 GHz)

RAM: 4 GB LPDDR4-3200

OS: Ubuntu 20.10 aarch64

Практическая часть работы проводится с использованием продукта OpenVINO Toolkit [1], в котором реализован программный модуль OpenVINO ARM CPU plugin [2], позволяющий запускать вывод нейронных сетей на процессорах на базе архитектуры ARM.

Для проведения экспериментов использованы следующие нейронные сети, выбранные из проекта Open Model Zoo [3]: image-retrieval-0001 [4], nfnet-f0 [5].

3.2. Эксперимент с квантизацией

Одним из подходов к оптимизации, который используется повсеместно, является квантизация. Квантизация [6,7] – это метод для уменьшения размера нейронной сети в оперативной памяти. Суть метода заключается в том, что все множество весов сети разбивается на количество интервалов, соответствующее количеству целых чисел, которые могут быть представлены в N бит памяти. В данной работе используется 8 бит (256 интервалов). В этом случае каждому интервалу ставится в соответствие 8-битное целое число, т.е. исходным 32-битным числам с плавающей запятой ставятся в соответствие 8-битные целые числа. Это сокращает затраты памяти в 4 раза.

Квантизация позволяет сократить размер нейронной сети в оперативной памяти в два раза, а также ускорить вычисления в результате использования более быстрых операций с целыми числами. При этом происходит пересчет весов в каждом интервале в соответствии с конфигурацией интервала. Очевидным недостатком этого метода является потеря точности весов. Вопрос о том, как потеря точности весов сказывается на итоговой точности работы сети, является открытым и требует исследования в каждом конкретном случае.

Для экспериментов с квантизацией была выбрана сеть image-retrieval-0001 [4]. Это сеть, решающая задачу поиска изображений на основе содержимого (content based image retrieval) [8]. На ней были произведены первоначальные запуски на архитектуре x86_64, где сеть работает корректно. При запуске на ARM64 возникает ошибка, связанная с тем, что слой L2-нормализации не квантизован.

Этот слой является предпоследним. Была проанализирована его реализация, которая обладает следующими параметрами:

Вход: 2 массива чисел типа float (in, sum), значения типа float, ограничений нет.

Выход: 1 массив чисел типа float, значения из отрезка [0, 1].

Алгоритм: $out[i] = in[i] / \sqrt{\max(sum[i], eps)}$

В ходе исследования было обнаружено, что при переходе от FP16 к INT8 значительно теряется точность операции нормализации, которая существенно зависит от операций над числами с плавающей запятой (деление и sqrt). Заметим, что операцию isqrt (вычисление обратного квадратного корня) можно реализовать методом Ньютона или алгоритмом Джарвиса на вычитаниях, но это имеет смысл, когда доступна более высокая точность (например, INT16).

В данной ситуации выглядит целесообразным включить возможность использования слоя нормализации в точности FP16. Это возможно, так как данная сеть работает в режиме смешанной точности (mixed precision) режиме (есть слои, использующие FP16 и INT8). Другие возможности связаны с поиском подходящей аппроксимации исходного выражения, которая все еще является достаточно точной, но при этом проще вычисляется, либо с использованием другого способа нормализации.

На данном направлении добиться существенных улучшений не удалось, поэтому мы рассмотрели другие возможности для ускорения вычислений.

3.3. Эксперимент с оптимизацией слоя GELU

Указанные ниже преобразования были выполнены для другой сети из проекта Open Model Zoo – nfnet-f0 [5] – классификационной сети на базе набора данных ImageNet.

Для этой сети был произведен послойный анализ производительности, чтобы выделить слой, который необходимо оптимизировать (таблица 1).

Таблица 1. Общее время работы всех слоев GELU в сети nfnet-f0 и общее время работы самой сети. Времена приведены в секундах

Эксперимент	Время, с
Слой v7::Gelu	813
Общее время	4259
% времени работы слоя Gelu относительно общего времени	19.09%

В качестве целевого для оптимизации был выбран слой Gaussian Error Linear Unit (GELU) [9], который реализует функцию активации. Слой GELU можно рассматривать как более гладкий аналог широко распространенного слоя ReLU.

Рассмотрим возможные варианты реализации слоя GELU.

$$GELU(x) = x P(X \leq x) = x\Phi(x) = x \cdot \frac{1}{2} \left[1 + \operatorname{erf} \left(\frac{x}{\sqrt{2}} \right) \right], \text{ если } X \sim N(0, 1)$$

Также известны следующие аппроксимации:

- использующая гиперболический тангенс:

$$0.5x \left(1 + \tanh \left[\frac{\sqrt{2}}{\pi} (x + 0.44715x^3) \right] \right);$$

- использующая сигмоид:

$$\sigma(1.702x).$$

Известно, что на практике для реализации слоя GELU часто используются аппроксимации, поскольку $\operatorname{erf}()$ – достаточно сложная для вычисления функция. В программном модуле OpenVINO ARM CPU plugin используются исходная формула с $\operatorname{erf}()$, а также аппроксимация с использованием гиперболического тангенса. По умолчанию для данной сети используется второй вариант.

Рассмотрим шаги, которые были предприняты для оптимизации производительности:

1. Аппроксимации (попробовать известные аппроксимации для функции GELU).
2. Параллелизм (максимально использовать ядра процессора).
3. Использование ARM интринсиков (intrinsics): ARM Neon (рассмотреть вариант использования доступных векторных инструкций архитектуры ARM).

В работе выполнялись модификации кода слоя GELU в программном модуле OpenVINO ARM plugin. В качестве базовой версии использовалась реализация из OpenVINO, основанная на аппроксимации с гиперболическим тангенсом.

В начале рассмотрим использование аппроксимации, которая использует сигмоид. На выборке из 1900 изображений из ImageNet получены следующие результаты по точности и времени работы сети (таблица 2).

Таблица 2. Время работы и top-1, top-5 точность на выборке из 1900 изображений при использовании аппроксимации с гиперболическим тангенсом и при использовании аппроксимации с сигмоидом. Время работы вывода сократилось на 6.8%, при этом точность изменилась незначительно

	tanh	Sigmoid
Точность (Top-1)	82.41%	82.36%
Точность (Top-5)	94.67%	94.67%
Время работы, с	18152	16907

Рассмотрим возможность ускорения вычислений с использованием параллелизма. В исходном коде OpenVINO ARM plugin данная возможность не была использована, поэтому мы опробовали различные технологии. В начале был проведен анализ в отдельном приложении-бенчмарке, содержащем только вычисления в рамках слоя GELU. При количестве потоков,

равном 4, и количестве элементов массива, равном 50 млн, получились следующие результаты (таблица 3).

Таблица 3. Время работы слоя GELU на массиве из 50 млн. элементов при использовании разных технологий параллелизма (количество потоков равно 4)

	Время выполнения GELU, с	Ускорение выполнения GELU
Последовательная версия	4.166	-
Распараллеливание с использованием OpenMP	1.091	3.82
Распараллеливание с использованием TBB	1.104	3.77
Распараллеливание с использованием std::thread	1.076	3.87

Далее ускоренный код был интегрирован в OpenVINO и были получены следующие результаты по точности и времени работы сети с ускоренной реализацией слоя GELU (таблица 4).

Таблица 4. Время работы на выборке из 1900 изображений при использовании аппроксимации с гиперболическим тангенсом и при использовании разных технологий параллелизма

	Время выполнения GELU, с	Общее время выполнения, с	% времени выполнения GELU
Последовательная версия	812.938	4258.98	19.09%
Распараллеливание с использованием OpenMP	214.899	3639.62	5.90%
Распараллеливание с использованием TBB	245.727	3600.78	6.82%

В данном случае не удалось протестировать параллелизм с помощью технологии TBB, т.к. ее поддержка находится на стадии внедрения в OpenVINO ARM plugin.

В целом результаты показывают, что распараллеливание вычислений позволяет добиться существенного ускорения вывода. Связано это с тем, что основная формула слоя GELU не зависима по данным от соседних индексов и имеет идеальный потенциал для распараллеливания.

Далее была рассмотрена возможность ускорения вычислений в результате использования более быстрых аппроксимаций гиперболического тангенса. Было найдено несколько вариантов реализации векторной версии гиперболического тангенса на архитектуре ARM:

- ARM Compute Library [10] (реализация выдает некорректный результат для данного примера).
- NCNN by Tencent [11] (тесты завершились успешно, именно эта реализация рассматривалась в дальнейших экспериментах).

Реализация [11] была интегрирована в OpenVINO toolkit, были получены следующие результаты экспериментов (таблица 5).

Таблица 5. Время работы и top-1, top-5 точность на выборке из 1900 изображений при использовании аппроксимации с гиперболическим тангенсом (оригинальная версия и версия с использованием векторных инструкций ARM Neon) и при использовании аппроксимации с сигмOIDом

	tanh	sigmoid	tanh-optimized
Точность Top-1	82.41%	82.36%	82.52%
Точность Top-5	94.67%	94.67%	94.67%
Время работы, с	18152	16907	17777

Результаты показывают, что при сохранении достаточной точности вычислений использование аппроксимации гиперболического тангенса позволяет ускорить расчеты в сравнении с оригинальной реализацией из OpenVINO toolkit, однако применение другой аппроксимации (сигмOID) приводит к лучшим результатам.

В дальнейшем планируется изучить следующие возможности:

- Профилирование слоя GELU в поисках дополнительных возможностей для оптимизации.
- Использование аппроксимации на основе предвычисленной таблицы значений для подсчета гиперболического тангенса.
- Тестирование использованных оптимизаций гиперболического тангенса на примерах других сетей.

Заключение

В рамках данной работы проведен ряд экспериментов, направленных на улучшение производительности вывода нейронных сетей на архитектуре ARM в продукте OpenVINO. Идея с применением квантизации для слоя L2-нормализации не оправдала себя из-за серьезной потери точности после преобразования весов. Оптимизация слоя GELU дала положительные результаты в первую очередь при использовании параллелизма. Был исследован потенциал использования векторных инструкций для ARM, что в комбинации с другими методами также приводит к хорошим результатам. В дальнейшем планируется продолжение поиска путей ускорения вычислений с сохранением точности работы сети.

Литература

1. OpenVINO™ Toolkit. URL: <https://github.com/openvinotoolkit/openvino>.
2. OpenVINO™ Inference Engine ARM CPU plugin. URL: https://github.com/openvinotoolkit/openvino_contrib/tree/master/modules/arm_plugin
3. Open Model Zoo. URL: https://github.com/openvinotoolkit/open_model_zoo
4. Модель image-retrieval-0001 от Intel. URL: https://github.com/openvinotoolkit/open_model_zoo/tree/master/models/intel/image-retrieval-0001.
5. Модель nfnet-f0. URL: https://github.com/openvinotoolkit/open_model_zoo/tree/master/models/public/nfnet-f0.
6. Jacob B. et al. Quantization and training of neural networks for efficient integer-arithmetic-only inference //Proceedings of the IEEE conference on computer vision and pattern recognition. – 2018. – С. 2704-2713.
7. Krishnamoorthi R. Quantizing deep convolutional networks for efficient inference: A whitepaper //arXiv preprint arXiv:1806.08342. – 2018.
8. Chen W. et al. Deep image retrieval: A survey //ArXiv. – 2021.
9. Hendrycks D., Gimpel K. Gaussian error linear units (gelus) //arXiv preprint arXiv:1606.08415. – 2016.
10. Tanh vector implementation in ARM Compute Library. URL: <https://github.com/ARM-software/ComputeLibrary/issues/77#issuecomment-298895287>.
11. Tanh vector implementation by Tencent in NCNN. URL: https://github.com/Tencent/ncnn/blob/9391fae741a1fb8d58cdfdc92878a5e9800f8567/src/layer/arm/neon_mathfun_fp16s.h#L484.

ПРИМЕНЕНИЕ НЕЙРОННЫХ СЕТЕЙ ДЛЯ ПОСТРОЕНИЯ КЛАССИФИКАЦИОННОЙ МОДЕЛИ ОБРАБОТКИ МЕДИЦИНСКИХ ДАННЫХ

Ю.С. Половинкина, И.М. Дорофеев

Северный (Арктический) федеральный университет имени М.В. Ломоносова

В статье описываются этапы разработки классификационной модели для обработки медицинских данных пациентов по типам травм. Проводится выбор архитектуры нейронной сети, исследуются способы предварительной обработки специализированных текстовых медицинских данных для повышения точности разработанной модели.

Ключевые слова: нейронная сеть, классификатор травм пациентов, обработка текстовых данных

Классификация данных о различных аспектах травмы пациента играет важную роль в прогнозировании и профилактике травм у различных групп населения. Исследуя лист учета травм, специалист может выделить причину, травмирующий фактор, место получения травмы и другие параметры, на основании которых можно сформулировать рекомендации для проведения соответствующих профилактических мероприятий. При этом если требуется выделить параметры травм из нескольких тысяч листов учета, то сделать это вручную практически невозможно. Для решения данной проблемы разработана и реализована классификационная модель данных пациентов по типам травм. Модель может применяться в лечебных учреждениях для выявления наиболее распространенных типов травм и разработки комплекса мер для их предотвращения.

С 01.01.2015 в г. Шенкурске (Россия, Архангельская область) начал функционировать регистр травм, построенный на основе модели регистра травм г. Харстад (Норвегия). Цель создания регистра – мониторинг состояния травматизма и разработка профилактических мероприятий. Источник информации о травмированном – лист учета травм, который содержит разделы для сбора информации о пациенте и обстоятельствах травмы. Внесение данных в лист учета травм, перенос их в сводную таблицу и дальнейшая классификация проводились вручную. Запуском проекта в России занимались ряд норвежских и российский медицинских ВУЗов при участии Национального Института Общественного Здравоохранения Норвегии и Шенкурской ЦРБ [1].

Для разработки классификационной модели взяты данные о травмах в формате таблицы MS Excel, каждая запись которой состоит из текстовых полей, кратко описывающих обстоятельства травмы (таблица 1).

Таблица 1. Пример текстовых столбцов

PlaceofInjury	WhatWasDoing	WhatWentWrong	HowYouWereHurt	Diagnosis
д.Часовенская дома	шла по комнате	подскользнулась на полу	упала на левую руку	Перелом шейки левого плеча.

Также в записи присутствуют кодировки, классифицирующие различные признаки травмы. Все записи составлены на основе листа учета травм (таблица 2).

Таблица 2. Пример кодировочных столбцов

PlaceType	Action1	Factor1	MechAcc	Factor2	MechInj	Factor3	BodyPart
101	40	7	44	1121	9	202	600

Лист учета травм заполняется медицинским работником в лечебном учреждении путем опроса пациента или сопровождающих его лиц. В предоставленных данных содержалось достаточно много пропусков, так как не всегда известны все требуемые признаки травмы. Решае-

мая задача классификации подразумевает получение по пяти текстовым полям таблицы восьми кодировок различных признаков травм.

Исходная таблица содержала 5398 записей о травмах. Также использовались предоставленные классификаторы в виде файлов MS Word. В работе использовался классификатор факторов травмы. Он содержал около 450 уникальных классов в записях о травмах, что порождает избыточное число кластеров. Для получения приемлемой точности при обучении разрабатываемой нейронной сети количество классов необходимо было сократить. Каждый код данного классификатора фактора травмы состоит из четырех цифр (таблица 3).

Таблица 3. Пример классификационных групп фактора травмы

Кодировка в классификаторе	Описание класса
042	Лестница в пешеходной зоне снаружи
0420	Лестница в пешеходной зоне снаружи, из цемента
0421	Лестница в пешеходной зоне снаружи, из камня
0422	Лестница в пешеходной зоне снаружи, из дерева

Для повышения продуктивности обучения четвертая цифра в коде не учитывалась. Это позволило сократить количество классов примерно до 220. Классификаторы места травмы, действий пострадавшего до травмы, механизма несчастного случая, механизмов получения травмы и травмированных частей тела содержали 65, 45, 35, 35 и 55 уникальных кластеров соответственно, поэтому изменения в соответствующих столбцах таблицы не проводились.

При первоначальной предобработке все текстовые данные приводились к единому формату, текст переводился в нижний регистр, исключались знаки препинания. Некоторые коды классификации также необходимо было привести к одному виду (например, коды 7, 07, 007 и 0007). Из-за ручного внесения данные таблицы содержали различные ошибки. Например, в тексте встречались опечатки, в кодировочных столбцах появлялись значения, которых нет в классификаторах, имелись данные, которые не несут смысловой нагрузки для классификации (например, «шла в аптеку» и «шла»).

Для обработки текстовых данных обычно применяются рекуррентные или сверточные нейронные сети. В ходе исследования лучшие результаты показала рекуррентная сеть, составленная из GRU и LSTM слоев. Затем в построенной сети проводился подбор набора гиперпараметров. В результате построенная итоговая нейронная сеть содержит по 80 нейронов в каждом слое. Архитектура сети приведена на рисунке 1.

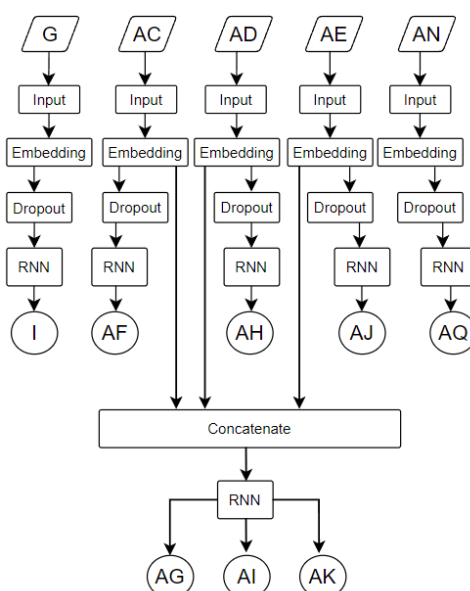


Рис 1. Архитектура нейронной сети

В кодировках, которые классифицируются по фактору травмы, первоначальная средняя точность была ниже 60%. Для повышения точности в соответствующих столбцах проводилась дополнительная предварительная обработка как текстовых, так и кодировочных данных для

уменьшения количество классов. В текстовых данных для этого проведена лемматизация и семантический анализ, а для кодировочных данных укрупнены группы в классификаторах. В процессе лемматизации слова, стоящие в разных падежах и наклонениях, записываются в одинаковой начальной форме, что снизило количество уникальных слов. Также при анализе данных выявлены слова, встречающиеся в таблице один или два раза, которые можно отбросить. В среднем все это позволило сократить количество уникальных слов в текстовых столбцах в 2,86 раза. При проведении семантического анализа выявлялись синонимичные слова. При этом обрабатывались только глаголы и существительные, несущие основную смысловую нагрузку, например, «поскользнулся на лестнице», «упал со стула», «укусило животное». Таким образом, слова собирались в массивы и объединялись в группы (таблица 4).

Таблица 4. Пример объединения слов в синонимичные группы в масштабе классификации

Обобщающее слово	Слова, замененные на обобщающее
резать	строгать стругать вырезать резать разрезать порезаться порезать ранить пораниться разделять

Количество уникальных существительных сократилось в 1,5 раза, а глаголов в 1,32 раза. При переработке исходных классификаторов сокращено количество классов в них путем объединения близких групп. В классификационных столбцах AG, AI, AK (факторы травмы 1, 2 и 3 соответственно) после отбрасывания четвертой цифры было 230, 230 и 238 уникальных классов соответственно. Путем объединения схожих групп количество классов сократилось до 185, 192 и 201 соответственно (таблица 5).

Таблица 5. Результаты работы нейросети до и после преобработки данных и изменения архитектуры сети

Сравнительная характеристика	До преобразований	После преобразований
Точность столбца I	0.7289	0.7648
Точность столбца AF	0.7412	0.8007
Точность столбца AG	0.5967	0.6519
Точность столбца AH	0.7279	0.7685
Точность столбца AI	0.6109	0.6519
Точность столбца AJ	0.7411	0.7896
Точность столбца AK	0.5685	0.6333
Точность столбца AQ	0.8611	0.8722

В дальнейшем для повышения точности и работоспособности модели предполагается провести существенную переработку имеющихся классификаторов, так как типичные травмы населения и их факторы в России отличаются от взятых за основу. Также могут сильно отличаться травмы населения в мегаполисах и сельской местности. Предполагается также разработать приложение для использования в медицинских учреждениях. Перед внедрением необходимо будет дополнительно собрать данные о травмах по образцу листа учета травм. После дообучения модели на локальных данных приложение сможет показывать наиболее вероятные травмы различных групп населения. Для усовершенствования гиперпараметров модели нейросети предполагается воспользоваться инструментом KerasTuner.

Литература

1. Т.Н. Унгурияну, А.В. Кудрявцев, В.Г. Анфимов, Б. Юттерштад, А.М. Гржибовский Первый в России муниципальный регистр травм: создание, логистика и роль в городской программе профилактики травматизма // Экология человека. 2017. Том 24, № 3. С. 56–64. DOI. 10.33396/1728-0869-2017-3-56-64.
2. Антонио Джулли, Суджит Пал Библиотека Keras – инструмент глубокого обучения. Москва. Оформление, перевод, издание ДМК Пресс, 2018. 298 с.
3. Франсуа Шолле Глубокое обучение на Python. Санкт-Петербург. Перевод, издание ООО Издательство «Питер». 2018. 400 с.

ИССЛЕДОВАНИЕ АЛГОРИТМА РЕШЕНИЯ ОПТИМИЗАЦИОННОЙ ЗАДАЧИ РАСПРЕДЕЛЕНИЯ КОЕЧНОГО ФОНДА В СЕТИ МЕДИЦИНСКИХ УЧРЕЖДЕНИЙ РЕГИОНА

Е.В. Пройдакова, З.М. Гитлина, А.Е. Котова

Нижегородский государственный университет им. Н.И. Лобачевского

Недавняя пандемия SARS-CoV-2 (COVID-19) показала важность рационального перераспределения медицинских ресурсов, в том числе больничных коек при ограниченном финансировании. Для этого требуются методы, позволяющие оперативно определить оптимальный состав коечного фонда для каждого медицинского учреждения. В работе предложено описание одного из таких методов и проведен анализ влияния основных параметров на решение оптимизационной задачи распределения больничных коек по сети медицинских учреждений региона.

Ключевые слова: сеть медицинских учреждений, коечный фонд, точность измерений, оптимизационная задача, основные параметры решения, алгоритм решения.

1. Введение

Данная статья является непосредственным продолжением цикла публикаций [1-2], в которых была построена математическая модель типового медицинского учреждения (больницы) в виде управляющей кибернетической системы.

По итогам предыдущих работ мы уже получили готовую математическую модель функционирования типового медицинского учреждения в виде управляющей системы и возможность генерировать статистику любого конечного объема для основных показателей функционирования больницы. В настоящей статье будет поставлена и решена задачи оптимизации функционирования сети медицинских учреждений в одном отчетном году, на основе выбранного критерия эффективности и с учетом ошибок результатов измерений.

Перейдем к постановке и решению задачи оптимального распределения коек между больницами. Для этого примем в качестве управляющих переменных количество коек $x_1(t), \dots, x_n(t)$, для каждого из медицинских учреждений M_j , $j \in J = \{1, 2, \dots, n\}$ в отчетном году $t \in T = \{1, 2, \dots, k\}$. Далее запишем систему ограничений на эти переменные.

- Ограничения на количество $x_j(t)$ коек по медицинскому учреждению M_j :

$$x_{j,\min}(t) \leq x_j(t) \leq x_{j,\max}(t), \quad j = 1, 2, \dots, n \quad (1)$$

- Ограничение на общее количество коек по всем медицинским учреждениям:

$$\sum_{j=1}^n x_j(t) \leq \left(\sum_{j=1}^n x_j(t) \right)_{\max} \quad (2)$$

- Ограничения на общее количество $Q^{(\varepsilon)}(\omega, t) = \sum_{j=1}^n [q_j^{(u, \varepsilon)}(\omega, t)x_j(t)]$ пролеченных больных по всем медицинским учреждениям

$$\sum_{j=1}^n Q_j(\omega, t) \leq \sum_{j=1}^n q_j^{(u, \varepsilon)}(\omega, t)x_j(t) \leq \left(\sum_{j=1}^n [q_j^{(u, \varepsilon)}(\omega, t)x_j(t)] \right)_{\max} \quad (3)$$

- Ограничение на общее количество $G^{(\varepsilon)}(\omega, t) = \sum_{j=1}^n [g_j^{(u, \varepsilon)}(\omega, t)x_j(t)]$ умерших больных по всем медицинским учреждениям:

$$\sum_{j=1}^n g_j^{(u, \varepsilon)}(\omega, t)x_j(t) \leq \sum_{j=1}^n G_j(\omega, t) \quad (4)$$

- Ограничение на общее количество $L^{(\varepsilon)}(\omega, t) = \sum_{j=1}^n [l_j^{(u, \varepsilon)}(\omega, t)x_j(t)]$ койко-дней по всем медицинским учреждениям в отчетном году t

$$\sum_{j=1}^n l_j^{(u, \varepsilon)}(\omega, t)x_j(t) \leq \left(\sum_{j=1}^n x_j(t)\right)_{\max} * 365 \quad (5)$$

- Если число m определяет общее количество статей затрат каждого медицинского учреждения, то ограничения по стоимости каждой статьи затрат с номером i в каждом отчетном году всеми медицинскими учреждениями можно записать в виде

$$\sum_{j=1}^n a_j^{(u, \varepsilon)}(\omega, t)x_j(t) \leq A_{i, \text{план}}(t), \quad i = 1, 2, \dots, m \quad (6)$$

С учетом введенных ограничений можно формулировать различные оптимизационные задачи [1], [2]. Будем считать распределение коек $x_1^*(t), x_2^*(t), \dots, x_n^*(t)$ оптимальным, если выполняются ограничения (1) - (6) и условие (7):

$$a^{(\varepsilon)}(\omega; x_1^*(t), x_2^*(t), \dots, x_n^*(t)) = \min \left\{ a^{(\varepsilon)}(\omega; x_1(t), x_2(t), \dots, x_n(t)) : (x_1(t), x_2(t), \dots, x_n(t)) \in X^{(n)}(t) \right\}. \quad (7)$$

В приведенной формулировке на решение оптимизационной задачи влияют такие параметры, как точность измерений (внешняя среда) $\varepsilon = \varepsilon(t)$, ограничение снизу на число коек $x_{j, \min}$, $j \in J = \{1, 2, \dots, n\}$ и конкретная реализация нормального закона. Напомним физический смысл этих параметров, определенных ранее в [1], [2].

Внешняя среда в течение отчетного года $t \in T = \{1, 2, \dots, k\}$ влияет на точность $\varepsilon = \varepsilon(t)$ измерений всех параметров и результатов работы медицинского учреждения M_j , $j \in J = \{1, 2, \dots, n\}$. Математическим описанием блока внешней среды является некоторое число $\varepsilon(t)$, для которого справедливо соотношение $0 < \varepsilon(t) < 1$.

Величина $x_{j, \min}$ определяет минимально возможное число коек для медицинского учреждения M_j , $j \in J = \{1, 2, \dots, n\}$ и задается для отчетного года $t \in T = \{1, 2, \dots, k\}$ условиями функционирования сети из n медицинских учреждений.

В нашем случае действительное число ω представляет собой значением случайной величины с равномерным законом распределения на отрезке $[0; 1]$. В свою очередь ω порождает реализацию случайной величины $\xi = \xi(\omega)$, распределенной по нормальному закону $N(a, \sigma)$ с параметрами a , σ . Значение параметров a , σ находятся следующим образом. Задавая точность $\varepsilon(t) = \varepsilon$ такое, что $0 < \varepsilon < 1$, мы учитываем, что вычисления проводятся с конкретной ошибкой. Тогда, используя свойства закона распределения Гаусса и в том числе «правило трёх сигм», можно положить математическое ожидание a равным фактическому значению показателя для медицинского учреждения M_j , при заданном управлении $u(t)$ в отчетном году t , а среднеквадратическое отклонение σ положить равным $\varepsilon a / 3$. Таким образом, генерируя значения случайной величины $\xi = \xi(\omega)$ при помощи метода имитационного моделирования, как значения нормальной случайной величины с известными a и σ , мы можем получать выборку необходимого объема для основных параметров функционирования медицинского учреждения: числа пролеченных или умерших больных, для количества койко-дней, для стоимостей независимых статей затрат и одной укрупненной статьи, а также для их плотностей [3].

2. Численный эксперимент

Для выявления влияния перечисленных выше параметров на решение оптимизационной задачи распределения больничных коек по сети медицинских учреждений региона был поставлен численный эксперимент.

Решение оптимизационной задачи проводилось в Microsoft Excel 2021 с использованием надстройки «Поиск решения» методом общего понижающего градиента (ОПГ). При проведении эксперимента были рассмотрены данные по пяти медицинским учреждениям г. Нижнего Новгорода № 34, № 24, № 14, № 11, № 37 региона за 9 лет, с 2007 по 2015 годы, $t \in \{1, 2, \dots, 9\}$.

Фактическое распределение коек для больниц определялось следующим набором значений: $x_1^{(u)}(t) = 50$, $x_2^{(u)}(t) = 50$, $x_3^{(u)}(t) = 50$, $x_4^{(u)}(t) = 50$, $x_5^{(u)}(t) = 25$ [4].

Эксперимент каждый раз проводился при пяти значениях точности измерений: $\varepsilon \in \{0.01, 0.02, 0.025, 0.029, 0.1\}$ и при пяти различных реализациях нормальной случайной величины $\xi(\omega)$, порожденной соответствующими значениями $\omega \in \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5\}$.

Ниже, в качестве примера, приведем некоторые результаты исследования.

В таблицах 1 – 4 представлено сравнение оптимального числа коек для пяти больниц, при критерии оптимальности «минимизация статей затрат», ограничениях снизу на число коек, равных нулю ($x_{j,\min} = 0$, $j \in J = \{1, \dots, 5\}$) и десяти ($x_{j,\min} = 10$, $j \in J = \{1, \dots, 5\}$), при значениях параметров $\varepsilon = 0.01$, $\omega \in \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5\}$ в отчетных годах $t \in \{4, 7, 8, 9\}$.

Далее будем полагать, что рассчитанное алгоритмом количество коек не различимо, если разница между полученными значениями не превышает 10%.

Цветовая шкала, принятая в таблицах 1-4:

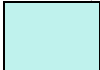
	– оптимальное количество коек, вычисленное при разных ограничениях снизу, совпадает;
	– оптимальное количество коек, вычисленное при разных ограничениях снизу, отличается менее чем на 10%;
	– оптимальное количество коек, вычисленное при разных ограничениях снизу, отличается ровно на 10%;
	– оптимальное количество коек, вычисленное при разных ограничениях снизу, отличается более чем на 10%.

Таблица 1. Результат оптимизации при $\varepsilon_1 = 0.01$, $t = 4$

Больница	№ 34	№ 24	№ 14	№ 11	№ 37
$x_{j,\min} = 0$ $\omega = \omega_1$	54	47	53	47	23
$x_{j,\min} = 10$ $\omega = \omega_1$	54	47	53	47	23
$x_{j,\min} = 0$ $\omega = \omega_2$	54	48	52	47	23
$x_{j,\min} = 10$ $\omega = \omega_2$	54	48	52	47	23
$x_{j,\min} = 0$ $\omega = \omega_3$	59	49	48	46	23
$x_{j,\min} = 10$ $\omega = \omega_3$	59	49	48	46	23
$x_{j,\min} = 0$ $\omega = \omega_4$	56	46	51	47	23
$x_{j,\min} = 10$ $\omega = \omega_4$	56	46	51	47	23
$x_{j,\min} = 0$ $\omega = \omega_5$	56	46	50	47	23
$x_{j,\min} = 10$ $\omega = \omega_5$	56	46	50	47	23

По результатам, представленным в таблице 1, можно сделать вывод о том, что в 2010 на оптимальное количество коек в больнице не влияет ограничение снизу $x_{j,\min}$.

Таблица 2. Результат оптимизации при $\varepsilon_1 = 0.01$, $t = 7$

Больница	№ 34	№ 24	№ 14	№ 11	№ 37
$x_{j,\min} = 0$ $\omega = \omega_1$	41	58	69	51	6
$x_{j,\min} = 10$ $\omega = \omega_1$	39	76	75	20	11
$x_{j,\min} = 0$ $\omega = \omega_2$	45	53	68	56	3
$x_{j,\min} = 10$ $\omega = \omega_2$	41	80	71	19	11
$x_{j,\min} = 0$ $\omega = \omega_3$	44	62	68	49	2
$x_{j,\min} = 10$ $\omega = \omega_3$	39	77	75	17	11
$x_{j,\min} = 0$ $\omega = \omega_4$	43	67	68	45	2
$x_{j,\min} = 10$ $\omega = \omega_4$	40	79	72	17	12
$x_{j,\min} = 0$ $\omega = \omega_5$	43	69	67	43	3
$x_{j,\min} = 10$ $\omega = \omega_5$	39	71	74	23	11

По результатам таблицы 2 можно сделать вывод о том, что в 2013 отчетном году ограничение снизу на количество коек $x_{j,\min}$ существенно влияет на нахождение оптимальных значений числа коек для каждой из больниц.

Таблица 3. Результат оптимизации при $\varepsilon_1 = 0.01$, $t = 8$

Больница	№ 34	№ 24	№ 14	№ 11	№ 37
$x_{j,\min} = 0$ $\omega = \omega_1$	210	0	0	0	0
$x_{j,\min} = 10$ $\omega = \omega_1$	173	10	10	10	10
$x_{j,\min} = 0$ $\omega = \omega_2$	209	0	0	0	0
$x_{j,\min} = 10$ $\omega = \omega_2$	172	10	10	10	10
$x_{j,\min} = 0$ $\omega = \omega_3$	209	0	0	0	0
$x_{j,\min} = 10$ $\omega = \omega_3$	171	10	10	10	10
$x_{j,\min} = 0$ $\omega = \omega_4$	208	0	0	0	0
$x_{j,\min} = 10$ $\omega = \omega_4$	171	10	10	10	10

$x_{j,min} = 0$ $\omega = \omega_5$	207	0	0	0	0
$x_{j,min} = 10$ $\omega = \omega_5$	170	10	10	10	10

По результатам таблицы 3 видно, что в 2014 году ограничение снизу $x_{j,min}$ оказывало сильное воздействие. Оптимальные значения при любом наборе параметров различаются больше, чем на 10%. Во всех случаях одинаковая тенденция распределения оптимального числа коек между больницами: в больнице №34 большое количество коек, а значения в больницах №24, №14, №11, №37 совпадают с ограничением снизу ($x_{j,min}$).

Таблица 4. Результат оптимизации при $\varepsilon_1 = 0.01$, $t = 9$

Больница	№ 34	№ 24	№ 14	№ 11	№ 37
$x_{j,min} = 0$ $\omega = \omega_1$	149	4	55	0	0
$x_{j,min} = 10$ $\omega = \omega_1$	141	12	42	10	10
$x_{j,min} = 0$ $\omega = \omega_2$	150	2	55	0	0
$x_{j,min} = 10$ $\omega = \omega_2$	142	10	42	10	10
$x_{j,min} = 0$ $\omega = \omega_3$	151	0	55	0	0
$x_{j,min} = 10$ $\omega = \omega_3$	140	12	42	10	10
$x_{j,min} = 0$ $\omega = \omega_4$	150	1	55	0	0
$x_{j,min} = 10$ $\omega = \omega_4$	141	10	42	10	10
$x_{j,min} = 0$ $\omega = \omega_5$	151	0	52	4	0
$x_{j,min} = 10$ $\omega = \omega_5$	140	11	41	11	10

По результатам таблицы 4 можно сделать вывод о том, что в 2015 отчетном году ограничение снизу $x_{j,min}$ оказывает существенное влияние. Мы видим, что оптимальное распределение коек во всех случаях имеет одинаковую тенденцию: в больнице №34 самое большое количество коек, в больнице №14 значение примерно равно одной трети от значения в больнице №34, а в больницах №24, №11 и №37 значения очень близки или достигают ограничения снизу $x_{j,min}$.

3. Заключение

По результатам численного исследования, в том числе таблицам 1 – 4, можно сделать вывод о том, что на оптимальное распределение коечного фонда в рассматриваемой сети медицинских учреждений влияют все рассматриваемые факторы.

При этом присутствует существенное влияние ограничения снизу $x_{j,min}$ на оптимальное распределение коек по сети медицинских учреждений, а также, что данная зависимость носит нелинейный характер.

Также выявлено влияние точности и найдена верхняя граница ε^* при которой оптимизационная задача всегда имеет решение. При расчете оптимальных значений методом ОПГ такая

верхняя граница $\varepsilon^* = 0.029$, а при расчёте симплекс методом она равна $\varepsilon^* = 0.025$ для любых наборов исследуемых параметров.

Дополнительные исследования позволили провести сравнение оптимального распределения коек, полученного при решении задачи методом общего понижающего градиента и симплекс-методом. По результатам вычислений можно сделать вывод о том, что выбор метода оказывает существенное влияние на результаты оптимизации. В зависимости от отчетного года, оптимальное число коек по больницам было близко (или даже совпадало) во всех исследуемых случаях или же, наоборот, существенно различалось каждый раз для разных больниц. Поэтому не представляется возможным однозначно выбрать один из перечисленных методов для дальнейших вычислений. Как возможный вариант, предлагается проводить расчеты обоими методами, а затем находить среднее между полученными оптимальными значениями коек для каждого медицинского учреждения.

Литература

1. Пройдакова Е.В., Гитлина З.М., Котова А.Е. Математическое моделирование и оптимизация процесса функционирования сети медицинских учреждений региона // Математическое моделирование и суперкомпьютерные технологии. 2021. С. 276-281.
2. Федоткин М.А., Пройдакова Е.В. Оптимизация динамики функционирования сети больниц с учетом ошибок результатов наблюдений // Математическое моделирование и суперкомпьютерные технологии. 2020. С.380-385.
3. Федоткин М.А., Пройдакова Е.В., Эделева А.Н. Математические и инструментальные методы построения модели экономики функционирования больницы // Вестник Нижегородского университета им. Н.И. Лобачевского. – 2019. – №4. – с. 56-66.
4. Федоткин М.А. Нетрадиционные проблемы математического моделирования экспериментов. – М.: Физматгиз, 2018. – С. 107.

ДИНАМИКА РЕКУРРЕНТНОЙ СПАЙКОВОЙ СЕТИ В ЗАДАЧАХ ВЫЧИСЛИТЕЛЬНОЙ НЕЙРОНАУКИ^{1*}

М.М. Пугавко¹, О.В. Масленников^{1,2}, В.И. Некоркин^{1,2}

¹*Федеральный исследовательский центр Институт прикладной физики РАН,*

²*Нижегородский государственный университет им. Н.И. Лобачевского*

В работе представлено исследование искусственной спайковой нейронной сети в процессе решения нескольких задач вычислительной когнитивной нейронауки, которая обучена с помощью методов машинного обучения. Исследована динамика сети и установлены основные механизмы решения каждой рассматриваемой задачи. Установлена иерархическая структура сети после обучения, в которой нейроны разделены в соответствии с разными задачами и формируют отдельные кластеры. Выявлено влияние различных кластеров на решение задач.

Ключевые слова: спайковая нейронная сеть, нелинейная динамика, машинное обучение, когнитивная нейронаука.

1. Введение

В области исследования нейронных сетей наблюдается тенденция к объединению нейронауки и когнитивистики с системами машинного интеллекта [1, 2]. Все больше работ направлены на использование методов машинного обучения для построения моделей с помощью автоматизированного подбора параметров, которые позволяют модели генерировать целевой отклик. Модели, построенные таким образом, можно рассматривать как динамические системы, к которым применимы методы нелинейной динамики. В рамках данного подхода особенности выполнения нейросетью различных задач может отражаться в определенном «ландшафте» фазового пространства, который задается особыми траекториями, такими как устойчивые или неустойчивые состояния равновесия, метастабильные состояния, предельные циклы и др. [3, 4, 5]. В данной работе мы рассматриваем в качестве целевых задач несколько задач когнитивной нейронауки, которые инспирированы реальными экспериментами [6-11]. Большинство работ в данном направлении не учитывают ключевую особенность, а именно, генерация потенциалов действия или спайков. В исследовании использована спайковая модель нейрона с адаптивным порогом возбуждения.

2. Когнитивные задачи

В настоящей работе в качестве целевых функций рассматривается класс задач, инспирированных реальными биологическими экспериментами в области когнитивной нейронауки. Всего рассмотрено 6 типов задач. Каждая задача разделена на две фазы. Первая фаза называется этапом стимула, в который сеть получает на вход внешние стимулы. Вторая фаза называется этапом отклика, в который сеть генерирует отклик-решение, соответствующее определенной целевой задаче. Пары стимул-отклик используются во время обучения. В качестве первой задачи рассмотрим задачу принятия решения в форме двухальтернативного выбора (DM) [6]. В нашей постановке задача состоит в том, что сеть должна сравнивать среднее значение зашумленного входного стимула с некоторым пороговым значением (см. рис. 2(а, ж)). В задаче контекстного принятия решения (CtxDM) [7] на вход сети подается два стимула; при активации первого контекстного входа сеть должна сравнивать с порогом первый стимул, а при активации второго – другой стимул (см. рис. 2(б, з)). Следующие три задачи относятся к одному типу Go/No-Go [8, 9, 10]. Первая задача данного типа (GO) заключается в том, что на вход подается сигнал на этапе стимула, а на этапе отклика сеть должна повторить данный сигнал (см. рис. 2(в, и)). Вторая

^{1*} Построение спайковой сети выполнено при поддержке РФФИ (проект 21-72-00142). Исследование динамики модели выполнено при поддержке РФФИ (проект 19-12-00338).

задача данного типа (GODI) заключается в том, что на вход подается короткий стимул, который сеть должна запомнить, а после периода задержки сгенерировать отклик, который равен по значению входному стимулу (см. рис. 2(г, й)). Последняя задача данного типа (GoRt) является задачей на время реакции. В течение этапа стимула на вход не подается стимул, а во время отклика на вход поступает стимул, который сеть должна максимально быстро повторить (см. рис. 2(д, к)). Последняя задача из нашего набора (Romo) [11] заключается в том, что на вход подается два стимула, которые разделены задержкой. Сеть должна запомнить первый стимул и сравнить его со вторым и принять решение о том, какой стимул больше (см. рис. 2(е, л)). Каждая задача может подаваться на один из двух входов, что увеличивает количество задач до 12.

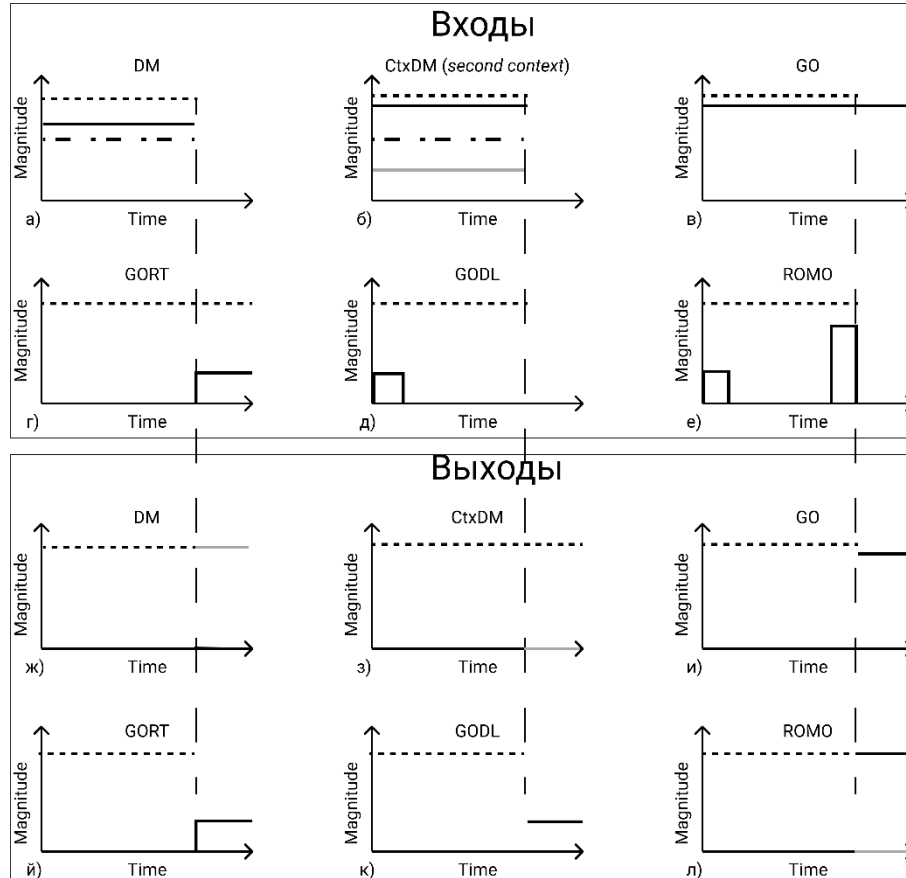


Рис. 1. Входы и целевые выходы. Пунктирной линией отображен сигнал фиксации. Линия точка-тире обозначает порог для задач DM и CtxDM. Черная линия обозначает первый выход или вход, серая линия обозначает второй выход или вход

Для всех задач длительность фазы отклика задается случайно и равномерно. Для задач CtxDM, DM, Go длительность этапа стимула выбрана из распределения $U(300, 1800)$ мс. Для задач GoDI и Romo длительность каждого стимула задается распределением $U(200, 600)$ мс, а период задержки после стимула для данных задач – $U(200, 1700)$ мс. Для задач DM, CtxDM и Romo значение стимула задается равномерным распределением $U(0, 1)$, а для задач Go/No-Go $U(\{0, 1/7, 2/7, \dots, 1\})$.

3. Модель и обучение

Архитектура сети представлена на рис. 2. Сеть имеет входной слой, который можно разделить на три части. К первой относится сигнал фиксации u_{fix} , который равен единице во время фазы получения стимулов и нулю во время фазы отклика. Ко второй части относятся два содержательных стимула, которые моделируют сенсорные стимулы двух различных модальностей $\mathbf{u}_{mod} = (u_{mod1}, u_{mod2})^T$ [12]. К третьей части относится вектор, который кодирует тип по-

даваемой задачи $\mathbf{u}_{tasks} = (u_1, u_2, \dots, u_{N_{task}})^T$, где N_{task} – количество задач, которых в нашем случае было 12. Таким образом входной вектор описывается следующим образом

$$\mathbf{u} = (u_{fix}, \mathbf{u}_{mod}, \mathbf{u}_{tasks})^T$$

Центральным элементом нашей модели является слой $N = 256$ рекуррентно связанных в соответствии с матрицей W_{rec} нейронов. Сигналы с входного слоя поступают на нейроны через матрицу W_{in} . Активность сети снимается с помощью матрицы W_{out} и вектора смещений $\mathbf{b}^{out}$.

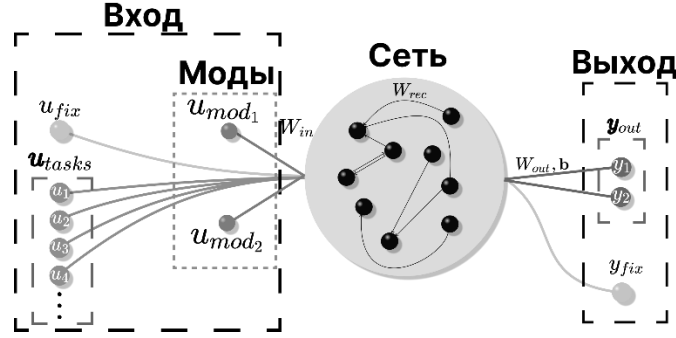


Рис. 2. Схематичное представление сети

В качестве элементов нейронной сети мы выбрали модель нейрона AdEx, которая является модификацией стандартной LIF модели нейронов с добавлением адаптивной переменной и экспоненциальной составляющей в правой части. Данная модель выбрана из-за наличия адаптивной переменной, которая позволяет кратковременно запоминать информацию, чтобы выполнять задачи с задержкой. Также данная модель нейрона демонстрирует более биологически релевантную активность. Активность нейронов в сети определяется системой уравнений

$$\begin{cases} I_j^{in}(n+1) = \sum_{i=1}^{N_{in}} w_{ji}^{in} u_i, \\ I_j^{rec}(n+1) = \sum_{l=1}^N w_{jl}^{rec} z_l(n), \\ v_j(n+1) = v_j(n) + \frac{\Delta t}{\tau_m} (-v_j(n) + \theta \exp(\frac{v_j(n) - v_{th}}{\theta}) \\ + I_j(n) - a_j(n)), \\ I_j(n+1) = I_j(n) \left(1 + \frac{\Delta t}{\tau_s}\right) + I_j^{in}(n+1) + I_j^{rec}(n+1), \\ a_j(n+1) = a_j(n) + \frac{\Delta t}{\tau_a} (a_j^{current} v_j(n) - a_j(n)), \end{cases} \quad (1)$$

где Δt – шаг симуляции, равный 1мс, w_{ij}^{in} – элемент матрицы входных весов, w_{kl}^{rec} – элемент матрицы связей между нейронами, $n \in \mathbb{Z}$ – дискретное время. Выход сети определяется уравнением

$$y_j(n+1) = \kappa y_j(n) + \sum_{i=1}^N w_{ji}^{out} z_i(n) + b_j^{out}, \quad (2)$$

где $\kappa = \exp(-\Delta t / \tau_{out})$ – параметр фильтра с константой времени $\tau_{out} = 50\text{мс}$, $z(n) = H(v(n) - v_{th})$ – спайковая переменная. При достижении мембранным потенциалом значения v_{th} происходит сброс мембранного потенциала на значение v_{reset} и увеличение адаптационной переменной $a \rightarrow a + a_s$. Все параметры сети представлены в Таблице 1.

Для построения и обучения использована библиотека «Norse» [13]. Обучение производится с помощью метода обратного распространения ошибки во времени (BPTT) с заменой производной на псевдопроизводную в соответствии с методом SuperSpike [14]. В вычислениях использован метод Adam, который позволяет повысить эффективность обучения. Во время обучения задачи подавались случайным образом. Функцию стоимости мы определили в виде среднеквадратичной ошибки с регуляризацией по средней частоте

$$E = MSE + \lambda E_r,$$

где среднеквадратичная ошибка определяется стандартным образом $MSE = \frac{1}{N_{out} N_{batch} N_{step} \sum_{il} m_{il}(n)} \sum_{i=1}^{N_{out}} \sum_{l=1}^{N_{batch}} \sum_{n=1}^{N_{step}} m_{il}(n) (y_{il}(n) - \hat{y}_{il}(n))^2$, где $m_{il}(n) = 0.1$, когда в сеть поступают стимулы, $m_{il}(n) = 1$, когда сеть должна генерировать отклик. Регуляризация

по средней частоте определяется как $E_r = \sum_{i=1}^N f_i H(f_i - f_{th})$, где средняя частота $f_i = \frac{1}{N_{batch} N_{time}} \sum_{l=1}^{N_{batch}} \sum_{n=1}^{N_{time}} z_{il}(n)$,

Таблица 1. Параметры сети

Параметр	τ_a	τ_m	τ_s	V_{th}	V_{reset}	θ	a_s	$a_{current}$
Значение	2с	10мс	5мс	0,45мВ	0мВ	0,5мВ	0,02мА	4нс

4. Результаты

В процессе обучения структура связей сети перестраивается, что приводит к изменению поведения каждого элемента в сети. Установлено, что наблюдается разделение нейронов сети на различные кластеры, которые специализируются на решении отдельных задач. В качестве общего анализа разбиения на кластеры использован метод иерархической кластеризации с методом минимизации дисперсии Уорда. Данный метод выбран из-за близкого расположения кластеров. Установлено появление трех групп нейронов. Самая малочисленная группа является достаточно активной во время выполнения всех задач. Вторая группа также активна во время выполнения всех задач, но средние частоты генерации спайков сравнительно низки. Третья группа нейронов имеет высокую активность и в данной группе присутствует специализация нейронов на выполнение различных задач.

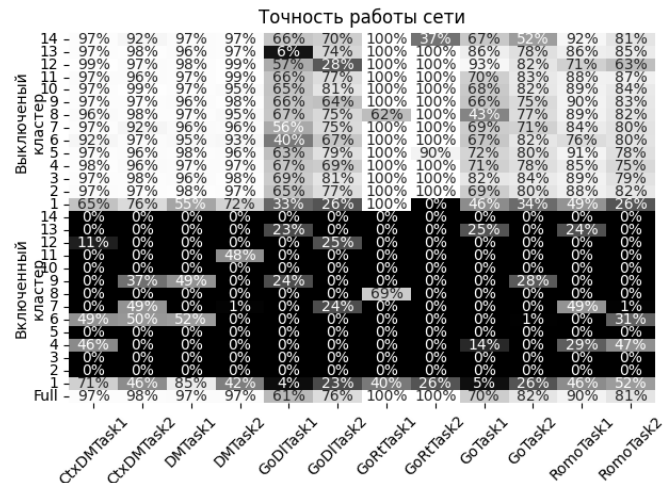


Рис. 3. Точность работы сети в зависимости от отключенных кластеров для разных задач

Кластеризация выполнена методом k-средних. Установлено два важных результата. Во-первых, существуют кластеры, которые влияют на выполнение всех задач, и существуют кластеры, которые влияют в большей степени на выполнение только определенных задач (см. рис. 3). Анализ произведен путем отключения или включения отдельных кластеров в сети.

Искусственная нейронная сеть после обучения представляет собой динамическую систему с многомерным фазовым пространством. На основе метода главных компонент получены проекции траекторий исходной динамической системы на низкоразмерные подпространства. Анализ траекторий в процессе выполнения задачи двухальтернативного выбора без контекста или с контекстом показал, что траектории разделяются на два потока, каждый из которых соответствует одной из двух категорий (выше или ниже порога среднее значение входного стимула) (см. рис. 3(а, б)). Анализ траекторий во время выполнения задач типа Go/No-Go показал, что в пространстве главных компонент траектории равномерно разделены в зависимости от входного стимула. Задача Romo и задача типа Go/No-Go с задержкой выполняется за счет существования метастабильных состояний, в которых система находится во время задержки после получения входного стимула. Достижение того или иного метастабильного состояния определяется величиной входного стимула. Во время фазы отклика траектория сети быстро переходит в область фазового пространства, которая соответствует определенной категории задачи, а затем после непродолжительного нахождения в данной области траектория уходит из данной области.

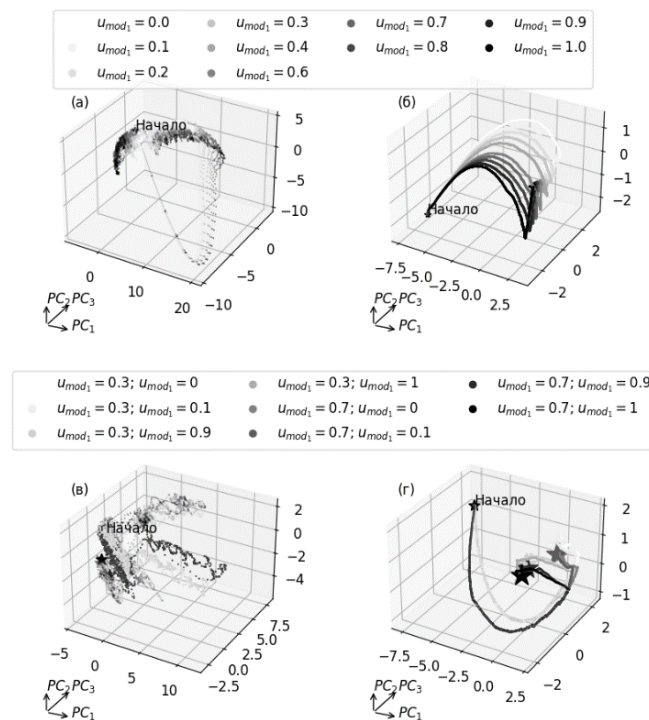


Рис. 4. Проекция в пространство трех главных компонент для задачи DM (а, б) и Romo (в, г). На графиках (а) и (в) отображены проекции для переменной v , а на графиках (б) и (г) отображены проекции для переменной a

5. Заключение

В работе представлена рекуррентная спайковая нейронная сеть, обученная решать ряд целевых задач вычислительной когнитивной нейронауки. Проанализирована динамика система с помощью анализа главных компонент и выявлены динамические механизмы выполнения нейросетью различных целевых функций. Установлено формирование функциональных кластеров после обучения и определена их роль при выполнении различных задач.

Литература

1. Maslennikov, O. V., Pugavko, M. M., Shchapin, D. S., & Nekorkin, V. I. (2022). Non-linear dynamics and machine learning of recurrent spiking neural networks. Phys. Usp. Accepted.
2. Richards, B.A., Lillicrap, T.P., Beaudoin, P., Bengio, Y., Bogacz, R., Christensen, A., Clopath, C., Costa, R.P., de Berker, A., Ganguli, S. and Gillon, C.J., 2019. A deep learning framework for neuroscience. Nature neuroscience, 22(11), pp.1761-1770.
3. Maslennikov, O. V., & Nekorkin, V. I. (2020). Stimulus-induced sequential activity in supervisedly trained recurrent networks of firing rate neurons. Nonlinear Dynamics, 101(2), 1093-1103.
4. Pugavko, M. M., Maslennikov, O. V., & Nekorkin, V. I. (2020). Dynamics of spiking map-based neural networks in problems of supervised learning. Communications in Non-linear Science and Numerical Simulation, 90, 105399.
5. Saurabh Vyas, Matthew D Golub, David Sussillo, and Krishna V Shenoy. Computation through neural population dynamics. Annual Review of Neuroscience, 43:249–275, 2020
6. Britten K. H., Shadlen M. N., Newsome W. T., Movshon J. A., The analysis of visual motion: a comparison of neuronal and psychophysical performance, Journal of Neuroscience 12 (12) (1992) 4745–4765.

7. Mante V., Sussillo D., Shenoy K. V., Newsome W. T., Context-dependent computation by recurrent dynamics in prefrontal cortex, *nature* 503 (7474) (2013) 78–84.
8. Funahashi S., Bruce C. J., Goldman-Rakic P. S., Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex, *Journal of neurophysiology* 61 (2) (1989) 331–349.
9. Brücke C., Kempf F., Kupsch A., Schneider G.-H., Krauss J. K., Aziz T., Yarrow K., Poghosyan A., Brown P., Kühn A. A., Movement-related synchronization of gamma activity is lateralized in patients with dystonia, *European Journal of Neuroscience* 27 (9) (2008) 2322–2329.
10. Brücke C., Bock A., Huebl J., Krauss J. K., Schöonecker T., Schneider G.-H., Brown P., Kühn A. A., Thalamic gamma oscillations correlate with reaction time in a go/nogo task in patients with essential tremor, *Neuroimage* 75 (2013) 36–45.
11. Romo R., Brody C.D., Hernández A., Lemus L., Neuronal correlates of parametric working memory in the prefrontal cortex, *Nature* (6735) (1999) 470–473.
12. Yang G. R., Joglekar M. R., Song H. F., Newsome W. T., Wang X.-J., Task representations in neural networks trained to perform many cognitive tasks, *Nature neuroscience* 22 (2) (2019) 297–306.
13. Pehle C., Pedersen J. E., Norse - A deep learning library for spiking neural networks, documentation: <https://norse.ai/docs/>.
14. Zenke F., Ganguli S., Superspike: Supervised learning in multilayer spiking neural networks, *Neural computation* 30 (6) (2018) 1514–1541.

ИССЛЕДОВАНИЕ СИНАПТИЧЕСКИХ СВЯЗЕЙ В МОДЕЛЯХ НЕЙРОНОВ ХОДЖКИНА-ХАКСЛИ И РОВАТА-СЕЛЬВЕРСТОНА

Д.С. Пушкарев¹, С.Ю. Маковкин¹, И.А. Кастальский^{1,2,3,5}, С.Ю. Гордлеева^{1,4,5}

¹Нижегородский государственный университет им. Н.И. Лобачевского

²Федеральный исследовательский центр «Институт прикладной физики РАН»

³Самарский государственный медицинский университет

⁴Автономная некоммерческая организация высшего образования «Университет Иннополис»

⁵Балтийский федеральный университет имени Иммануила Канта

Аннотация: в данной работе проведено исследование колебательных режимов одиночных нейронов Ходжкина-Хаксли и Ровата-Сельверстоны под действием внешнего тока. Исследованы режимы синхронизации двух нейронов Ходжкина-Хаксли и двух нейронов Ровата-Сельверстоны. Рассчитаны бифуркационные диаграммы активности обеих динамических систем.

Ключевые слова: Нейронная сеть, Синхронизация, Тормозный нейрон, Возбуждающий нейрон, Бифуркация.

Изучение нелинейных процессов в системах мозга с позиций колебательно-волнового подхода в настоящее время является одним из актуальных направлений изучения механизмов генерации и обработки сигналов в мозге. Эти исследования могут помочь в создании подходов к формированию ассоциативной памяти и распознаванию образов с помощью них. Для создания алгоритма работы ассоциативной памяти необходимо использовать модели нейронов головного мозга млекопитающих. В данной работе будет использоваться нейрон Ходжкина-Хаксли, как биологически релевантная модель, а также его редуцированная модель Ровата-Сельверстоны.

Целью данной работы является исследование и сравнение эффектов синхронизации нейронов Ходжкина-Хаксли [1] и нейронов Ровата-Сельверстоны [2] на воздействие внешним электрическим током. Динамика нейронов описывается системой дифференциальных уравнений.

Нейрон Ходжкина-Хаксли:

$$\begin{aligned}\frac{dV}{dt} &= 1000 \frac{1}{C} (g_{Na} m^3 h (E_{Na} - V) + g_K n (E_K - V) + g_{leak} (E_{leak} - V) + I_{app} + I_{syn}) \\ \frac{dm}{dt} &= 1000 (\alpha_m (1 - m) - \beta_m m) \\ \frac{dn}{dt} &= 1000 (\alpha_n (1 - n) - \beta_n n) \\ \frac{dh}{dt} &= 1000 (\alpha_h (1 - h) - \beta_h h)\end{aligned}$$

Нейрон Ровата-Сельверстоны:

$$\begin{aligned}\tau_m \frac{dV}{dt} &= I_{fast}(V_i) - I_{syn}(V_i, V_k^0) - W_i - I_{app} \\ \tau_w(V_i) \frac{dW_i}{dt} &= W_\infty(V_i) - W_i\end{aligned}$$

Алгоритмы для расчета динамики систем были реализованы на высокоуровневом языке программирования Julia. Интегрирование уравнений выполнялось с помощью метода Рунге-Кутты 4 порядка с фиксированным временным шагом.

При исследовании одиночных (изолированных) нейронов на них подавался постоянный ток смещения различной амплитуды. Были определены средние периоды автоколебаний, обнаружены значения параметров, соответствующие бифуркации Андронова-Хопфа, построены графики зависимости средней частоты от управляющего тока смещения (рисунки 1 и 2).

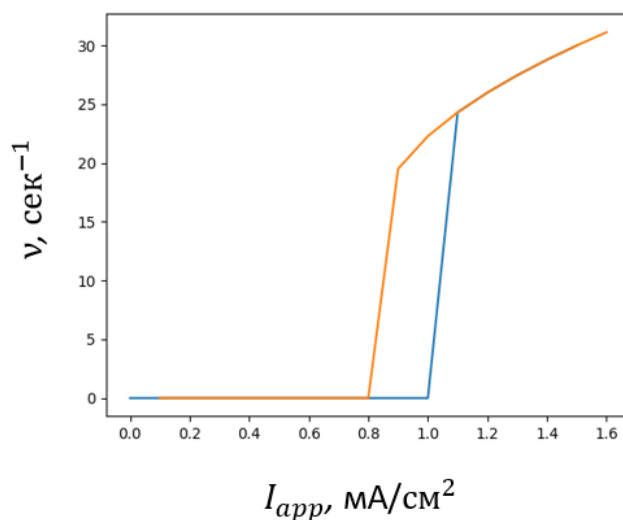


Рис. 1. Зависимость средней частоты ν от тока смещения I_{app} для нейрона Ходжкина-Хаксли

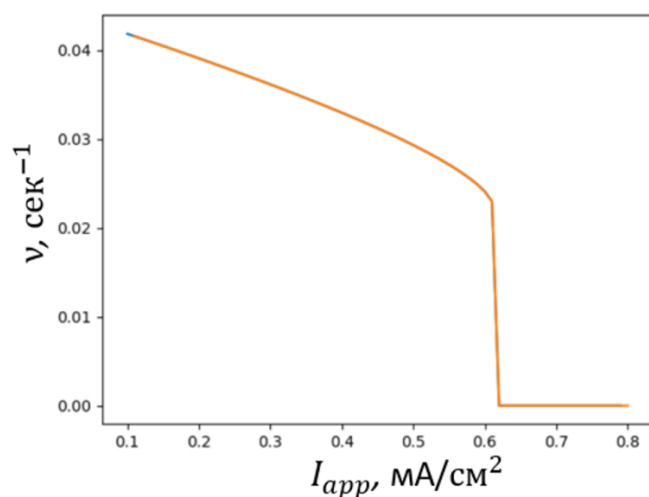


Рис. 2. Зависимость средней частоты ν от тока смещения I_{app} для нейрона Ровата-Сельверстона

Таким образом, было обнаружено, что в системе Ходжкина-Хаксли существует гистерезис, а в системе Ровата-Сельверстона его нет. При увеличении и уменьшении I_{app} устойчивый предельный цикл в динамической системе рождается и разрушается по разным сценариям. Устойчивый предельный цикл рождается через бифуркацию Андронова-Хопфа, а разрушается через седло-узловую бифуркацию на цикле.

Далее было исследовано взаимодействие системы из двух нейронов Ходжкина-Хаксли, соединенных синаптической связью. Была выявлена область параметров, соответствующая синхронизации нейронов при возбуждающей (рисунок 3) и тормозной синаптической связи (рисунок 4). Управляемый нейрон находился в возбудимом режиме с наличием устойчивого фокуса в фазовом пространстве. На вход подавался внешний ток смещения I_{app} .

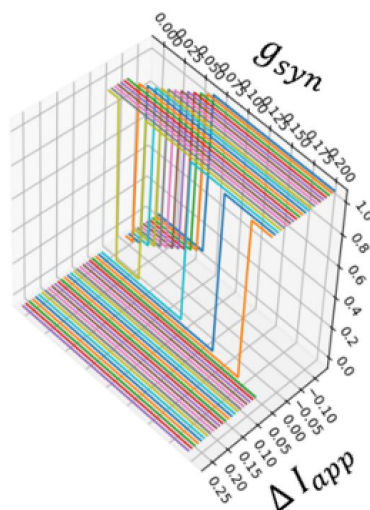


Рис. 3. Область синхронизации нейронов при возбуждающей синаптической связи в зависимости от параметров g_{syn} и ΔI_{app} (1 – нейроны синхронизированы, 0 – синхронизации нет)

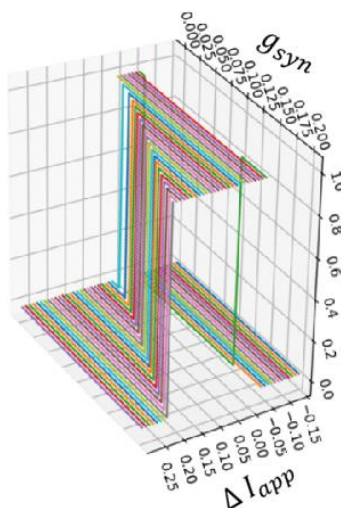


Рис. 4. Область синхронизации нейронов при тормозной синаптической связи в зависимости от параметров g_{syn} и ΔI_{app} (1 – нейроны синхронизированы, 0 – синхронизации нет)

Также, с помощью синаптической связи были соединены два нейрона Ровата-Сельверстона. Разность фаз считалась как разность между временными значениями спайков постсинаптического и пресинаптического нейронов. Обнаружены диапазоны синхронизации в фазе и в противофазе.

Таким образом, было исследовано поведение двух систем связанных нейронов Ходжкина-Хаксли и Ровата-Сельверстона. На основе данных исследований возможно построение нейронной сети для моделирования ассоциативной памяти [3-10].

Литература

1. Hodgkin A. L., Huxley A. F. A quantitative description of membrane current and its application to conduction and excitation in nerve //The Journal of physiology. – 1952. – Т. 117. – №. 4. – С. 500-544.
2. Rowat, P. F., & Selverston, A. I. (1993). Modeling the gastric mill central pattern generator of the lobster with a relaxation-oscillator network. Journal of Neurophysiology, 70(3), 1030–1053.
3. Simonov, A., Kastalskiy, I., & Kazantsev, V. (2012). Pattern retrieval in a three-layer oscillatory network with a context dependent synaptic connectivity. Neural Networks, 33, 67-75.

4. Gordleeva, S. Y., Lotareva, Y. A., Krivonosov, M. I., Zaikin, A. A., Ivanchenko, M. V., & Gorban, A. N. (2019, October). Astrocytes organize associative memory. In *International Conference on Neuroinformatics* (pp. 384-391). Springer, Cham.
5. Gordleeva, S. Y., Tsybina, Y. A., Krivonosov, M. I., Ivanchenko, M. V., Zaikin, A. A., Kazantsev, V. B., & Gorban, A. N. (2021). Modeling working memory in a spiking neuron network accompanied by astrocytes. *Frontiers in Cellular Neuroscience*, 15, 631485.
6. Makovkin, S. Y., Shkerin, I. V., Gordleeva, S. Y., & Ivanchenko, M. V. (2020). Astrocyte-induced intermittent synchronization of neurons in a minimal network. *Chaos, Solitons & Fractals*, 138, 109951.
7. Gordleeva, S. Y., Ermolaeva, A. V., Kastalskiy, I. A., & Kazantsev, V. B. (2019). Astrocyte as spatiotemporal integrating detector of neuronal activity. *Frontiers in physiology*, 10, 294.
8. Gordleeva, S. Y., Lebedev, S. A., Rumyantseva, M. A., & Kazantsev, V. B. (2018). Astrocyte as a detector of synchronous events of a neural network. *JETP Letters*, 107(7), 440-445.
9. Tsybina, Y., Kastalskiy, I., Krivonosov, M., Zaikin, A., Kazantsev, V., Gorban, A. N., & Gordleeva, S. (2022). Astrocytes mediate analogous memory in a multi-layer neuron–astrocyte network. *Neural Computing and Applications*, 34(11), 9147-9160.
10. Makovkin, S., Kozinov, E., Ivanchenko, M., & Gordleeva, S. (2022). Controlling synchronization of gamma oscillations by astrocytic modulation in a model hippocampal neural network. *Scientific reports*, 12(1), 1-15.

РЕШЕНИЕ ЗАДАЧ ДИСКРЕТНОЙ ОПТИМИЗАЦИИ ПРИ ПОМОЩИ МОДЕЛЬНЫХ ЯЗЫКОВ ПРОГРАММИРОВАНИЯ

В.В. Разин

Нижегородский государственный университет имени Н.И. Лобачевского

Задачи дискретной оптимизации являются крайне актуальными и важными задачами, охватывающими сферы производства, экономики и бизнеса. В рамках данной работы демонстрируются решения задач дискретной оптимизации при помощи модельного языка программирования AMPL, а также предоставляется расширение для данного модельного языка программирования, позволяющее использовать логические функции.

Ключевые слова: дискретная оптимизация, модельный язык программирования, AMPL.

1. Введение

В 1939 г. Л. В. Канторовичем представлен ряд производственно-экономических задач оптимизации, которые могут быть строго математически описаны [1], в последствии развитые и предшествующие созданию линейного программирования [2]. С 1942 г. начинают активно проводиться исследования по линейному и нелинейному программированию [3].

К середине XX в. сформировались задачи оптимизации двух типов: максимизации прибыли и минимизации расходов [4–5].

Общего алгоритма решения данных задач, на текущий момент не существует, однако их можно свести к определенному классу задач, решение которых уже известно. Эту операцию можно выполнять вручную, но в общем случае это затрачивает много времени, которое хотелось бы минимизировать, не теряя при этом правильности решения. Сократить данное время можно используя высокоуровневые языки программирования [6]. Для этого необходимо на выбранном языке описать математическую модель, алгоритм решения самой задачи по заданной модели и получение конкретного результата. Однако, существуют модельные языки программирования [7], непосредственно предназначенные для решения различных математических задач, где данный процесс сводится к описанию математической модели исходной задачи на выбранном модельном языке программирования и выбору определенного программного модуля, позволяющему ее решить [8].

2. Решение задач дискретной оптимизации

В рамках данной работы демонстрируются решения различных задач дискретной оптимизации при помощи модельного языка программирования AMPL [9].

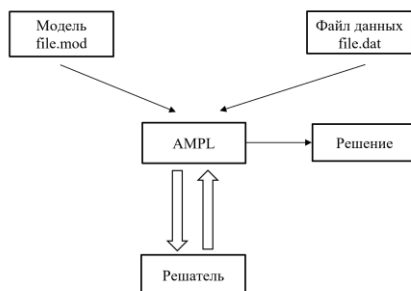


Рис. 1. Схема работы системы моделирования AMPL

2.1. Задача о максимизации прибыли

Сталелитейный завод обрабатывает алюминий и медь. Обработка металлов занимает разное время, за час завод может обработать 200 тонн алюминия и 140 тонн меди. За тонну обработанного алюминия завод получает 25\$, а за обработанную тонну меди 30\$. В неделю завод может обрабатывать до 6000 тонн алюминия и 4000 тонн меди. Сколько тонн алюминия и меди нужно обработать за 40 часов, чтобы получить максимальную прибыль?

Рис. 2. Формулировка задачи 2.1

```
var xAl;  
var xCu;  
s.t. limitAl: 0 <= xAl <= 6000;  
s.t. limitCu: 0 <= xCu <= 4000;  
s.t. Time: xAl / 200 + xCu / 140 <= 40;  
maximize Profit: xAl * 25 + xCu * 30;
```

Рис. 3. Запись математической модели задачи 2.1 на AMPL

Решается задача с формулировкой на Рис. 2. Подав на вход модель, описанную на Рис. 3, получаем решение: заводу необходимо обработать 6000 тонн алюминия и 1400 тонн меди.

2.2. Задача о «восьми ферзях»

На доске 8 на 8 надо расставить 8 ферзей так, чтобы они «не били» друг друга (не было более одного ферзя на одной диагонали, вертикали или горизонтали).

Рис. 4. Формулировка задачи 2.2

```
param n >= 0, default 8;  
set N := 1..n;  
var x{N,N} binary;  
s.t. cols {j in N} : sum{i in N} x[i,j] <= 1;  
s.t. rows {i in N} : sum{j in N} x[i,j] <= 1;  
s.t. SW {i in N, j in N} :  
sum{h in N : h+j<=n and h < i} x[i-h,j+h] + sum{h in N :  
h < j and h+i<=n} x[i+h,j-h] <= 1;  
s.t. NW {i in N, j in N} :  
sum{h in N : h < i and h < j} x[i-h,j-h] + sum{h in N :  
h+j<=n and h+i<=n} x[i+h,j+h] <= 1;  
maximize queens : sum{i in N, j in N} x[i,j];
```

Рис. 5. Запись математической модели задачи 2.2 на AMPL

Решается задача с формулировкой на Рис. 4. Подав на вход модель, описанную на Рис. 5, получаем одну из возможных расстановок ферзей, изображенную на Рис. 6.

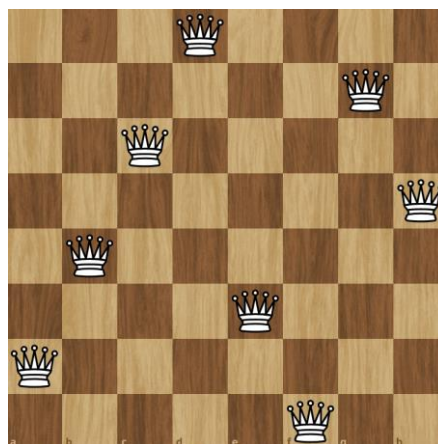


Рис. 6. Одна из возможных расстановок ферзей (решение задачи 2.2)

2.3. Задача о «волке, козе и капусте»

Необходимо переправить через реку волка, козу и капусту. При этом лодка вмещает в себя только один из вышеуказанных объектов (не считая перевозчика). Если волк и коза останутся на одном берегу, то волк съест козу, а если коза и капуста, то коза съест капусту.

Рис. 7. Формулировка задачи 2.3

```
set K = 0..8;
set T = 1..8;
var W {K} binary;
var G {K} binary;
var C {K} binary;
s.t. col1: W[1]+C[1]+G[1]<= 1 + W[0]+C[0]+G[0] and
W[1]+C[1]+G[1]<=1 + W[2]+C[2]+G[2];
s.t. col3: W[3]+C[3]+G[3]<= 1 + W[2]+C[2]+G[2] and
W[3]+C[3]+G[3]<=1 + W[4]+C[4]+G[4];
s.t. col5: W[5]+C[5]+G[5]<= 1 + W[4]+C[4]+G[4] and
W[5]+C[5]+G[5]<=1 + W[6]+C[6]+G[6];
s.t. col7: W[7]+C[7]+G[7]<= 1 + W[6]+C[6]+G[6] and
W[7]+C[7]+G[7]<=1 + W[8]+C[8]+G[8];
s.t. wolf1: W[1]>=W[0] and W[1]<=W[2];
s.t. wolf3: W[3]>=W[2] and W[3]<=W[4];
s.t. wolf5: W[5]>=W[4] and W[5]<=W[6];
s.t. wolf7: W[7]>=W[6] and W[7]<=W[8];
s.t. cabbage1: C[1]>=C[0] and C[1]<=C[2];
s.t. cabbage3: C[3]>=C[2] and C[3]<=C[4];
s.t. cabbage5: C[5]>=C[4] and C[5]<=C[6];
s.t. cabbage7: C[7]>=C[6] and C[7]<=C[8];
s.t. goat1: G[1]>=G[0] and G[1]<=G[2];
s.t. goat3: G[3]>=G[2] and G[3]<=G[4];
s.t. goat5: G[5]>=G[4] and G[5]<=G[6];
s.t. goat7: G[7]>=G[6] and G[7]<=G[8];
s.t. WG1: W[1]!=G[1] and W[1]!=0;
s.t. GC1: G[1]!=C[1] and G[1]!=0;
s.t. WG2: W[2]!=G[2] and W[2]!=1;
s.t. GC2: G[2]!=C[2] and G[2]!=1;
s.t. WG3: W[3]!=G[3] and W[3]!=0;
s.t. GC3: G[3]!=C[3] and G[3]!=0;
s.t. WG4: W[4]!=G[4] and W[4]!=1;
s.t. GC4: G[4]!=C[4] and G[4]!=1;
s.t. WG5: W[5]!=G[5] and W[5]!=0;
s.t. GC5: G[5]!=C[5] and G[5]!=0;
s.t. WG6: W[6]!=G[6] and W[6]!=1;
s.t. GC6: G[6]!=C[6] and G[6]!=1;
s.t. WG7: W[7]!=G[7] and W[7]!=0;
s.t. GC7: G[7]!=C[7] and G[7]!=0;
s.t. WG8: W[8]!=G[8] and W[8]!=1;
s.t. GC8: G[8]!=C[8] and G[8]!=1;
let W[0] := 0;
let G[0] := 0;
let C[0] := 0;
maximize Right : sum {i in T} (W[i]+G[i]+C[i]);
```

Рис. 8. Запись математической модели задачи 2.3 на AMPL

Решается задача с формулировкой на Рис. 7. Подав на вход модель, описанную на Рис. 8, получаем решение данной задачи.

3. Поддержка логических функций

Как можно заметить, что описание модели для решения задачи 2.3 достаточно объемное, однако в языке AMPL на данный момент отсутствует поддержка логических функций, использование которых могло бы существенно упростить описание подобного типа задач.

В связи с этим была написана программа [10], позволяющая обеспечить поддержку логических функций в AMPL.

Необходимо на вход подать лишь формулу логической функции и на ее выходе получить полный необходимый набор ограничений. Для преобразования заданной логической функции в набор конкретных ограничений необходимо произвести непосредственное чтение логической функции, синтаксический разбор полученного выражения, приведение распознанной логической функции к конъюнктивной нормальной форме и по ней запись набора ограничений на языке AMPL. Чтение логической функции можно производить непосредственно из файлов моделей или принимать ручной ввод пользователя из консоли. Синтаксический разбор и вычисление логической функции при разных значения параметров ее переменных основан на преобразовании исходного выражения в обратную польскую запись и реализации стековой машины для осуществления вычисления значений заданной логической функции. После удачного перевода выражения в обратную польскую запись и реализации стековой машины при ее помощи строится таблица истинности заданной функции. По данной таблице можно построить конъюнктивную нормальную форму логической функции, а уже по конъюнктивной нормальной форме описать набор ограничений, используя знания о синтаксисе языка AMPL, описывающих данную логическую функцию.

4. Результаты

В ходе решения различных задач дискретной оптимизации при помощи использования модельного языка AMPL было выявлено, что есть тип задач, для решения которых целесообразно использовать логические функции, однако их поддержка отсутствует в AMPL. Поэтому как вариант решения данной проблемы была написана программа [10], которая осуществляет перевод записи заданной логической функции в код AMPL.

5. Заключение и выводы

Использование модельных языков программирования позволяет в некоторых случаях существенно сократить время для решения различных математических задач. Однако в них отсутствуют полезные для решения задач компоненты, которые могут упростить описание математических задач.

В ходе данной работы была написана программа, позволяющая использовать логические функции в AMPL. В будущем планируется реализация других полезных компонентов в AMPL, а также других языках программирования.

Литература

1. Канторович Л. В. Математические методы организации и планирования производства. – 1939.
2. Палий И. А. Линейное программирование. – 2019.
3. Карманов В. Г. Математическое программирование. – Физматлит, 2005.
4. Хачатуров В. Р. и др. Комбинаторные методы и алгоритмы решения задач дискретной оптимизации большой размерности // ВР Хачатуров, ВЕ Веселовский, АВ Злотов и др. – М.: Наука. – 2000.
5. Рассказова М. Н., Зарипова Р. Х. Решение одной задачи формирования производственной программы с использованием дискретной оптимизации // Динамика систем, механизмов и машин. – 2016. – Т. 4. – №. 1. – С. 97-103.

6. Мельцов И. А. Высокоуровневые языки программирования: сравнительная характеристика // Экономика и предпринимательство. – 2020. – №. 3. – С. 993-995.
7. Белякова Ю. В. Автоматическое построение ограничений в модельном языке программирования с шаблонами функций и автовыводом типов.
8. Gay D. M. Hooking your solver to AMPL. – Technical Report 93-10, AT&T Bell Laboratories, Murray Hill, NJ, 1993, revised, 1997.
9. Fourer R., Gay D. M., Kernighan B. W. AMPL: A mathematical programming language. – Murray Hill, NJ : AT & T Bell Laboratories, 1987.
10. Razin V. AMPL logic functions [Электронный ресурс] – URL: <https://github.com/razinw/AMPLlogicfunctions> (дата обращения 07.07.2022).

ОЦЕНКА ЭФФЕКТИВНОСТИ МОДЕЛИ ОБНАРУЖЕНИЯ КОМПЬЮТЕРНЫХ АТАК В ОТНОШЕНИИ АТАКИ НУЛЕВОГО ДНЯ

Д.А. Рыболовлев, В.Е. Салимоненко

Академия ФСО России, Орел

В работе обсуждаются результаты практического эксперимента по сравнению сигнатурного и эвристического анализаторов компьютерных атак в отношении ранее неизвестных атак. На частном примере искусственной деградации базы решающих правил моделируется ситуация, когда одна из реализаций SQL инъекции становится атакой нулевого дня. Отмечается невозможность обнаружения такой атаки сигнатурной системой защиты и успешное обнаружение атаки моделью машинного обучения типа «случайный лес». Подчеркивается принципиальное преимущество несигнатурных систем обнаружения атак в отношении новых атак.

Ключевые слова: информационная безопасность, система обнаружения атак, машинное обучение, набор данных, сетевой трафик, компьютерная атака, атака нулевого дня, сигнатурный анализатор, эвристический анализатор.

1. Введение

Известным подходом к обнаружению компьютерных атак является применение сигнатурных анализаторов сетевого трафика, эффективность которых определяется полнотой базы решающих правил (сигнатур) обнаружения известных вредоносных информационных воздействий. Недостатком такого подхода является практически нулевая вероятность обнаружения сигнатурными анализаторами новых (ранее неизвестных) компьютерных атак [1].

По этой причине повышенный интерес для исследователей представляют эвристические методы, позволяющие при корректной формализации задачи на основе некоторого набора признаков отнести наблюдаемую последовательность событий безопасности (состояний контролируемого объекта, пакетов сетевого трафика и т.п.) к классу нормальных или вредоносных (аномальных) [2].

При построении эвристических анализаторов систем обнаружения сетевых атак широкое применение находят современные методы машинного обучения, позволяющие выявлять ранее неизвестные закономерности в анализируемых данных. Обобщенная схема реализации технологии машинного обучения с учителем включает этапы сбора и предобработки исходных данных для обучения (размеченного сетевого трафика), отбора признаков и оценки их значимости, обучения модели и подбора гиперпараметров, апробации и оценки качества классификации [3]. Последующее практическое применение полученного решения подразумевает использование предобученной модели машинного обучения для обнаружения сетевых компьютерных атак в защищаемой сети.

Отличительной особенностью эвристического анализатора на основе машинного обучения по сравнению с сигнатурным является возможность обнаруживать ранее неизвестные компьютерные атаки за счет обобщающей способности модели после обучения на прецедентах. В настоящем исследовании указанная гипотеза проверяется на одном из частных примеров: моделируется ситуация, когда конкретная атака типа «SQL инъекция» становится атакой нулевого дня (0-day), и оценивается способность сигнатурного и эвристического анализаторов обнаружить такую атаку.

2. Исходные данные для эксперимента

В эксперименте класс рассматриваемых компьютерных атак был ограничен только веб-атаками по причине широкого распространения этого класса атак, доступности инструментов генерации, относительной простоты реализации тестовой инфраструктуры.

За основу объекта защиты было выбрано «чертовски уязвимое веб-приложение» DVWA (Damn Vulnerable Web Application)¹. Такое решение позволило атаковать объект защиты в реальном времени, обеспечило наличие известных уязвимостей и средств их эксплуатации для сбора обучающего трафика, а также доступа к объекту защиты для подключения, сбора и анализа данных.

В качестве сигнатурного анализатора компьютерных атак выбрано средство защиты WAF ModSecurity² – один из наиболее распространенных межсетевых экранов веб-приложений с открытым исходным кодом. Использовалась база сигнатур Core Rule Set (CRS)³ версии 3.3.2, одна из наиболее используемых открытых баз решающих правил, которая охватывает широкий перечень сетевых атак и по умолчанию идет в поставке с ModSecurity.

В качестве эвристического анализатора была выбрана синтезированная авторами модель типа «случайный лес» [4]. Обучение модели производилось на сбалансированной и предварительно обработанной подвыборке веб-атак WebAttacks набора данных CICIDS2017⁴.

Набор данных CICIDS2017 подготовлен Канадским институтом кибербезопасности по результатам анализа сетевого трафика в изолированной среде, в которой были смоделированы действия 25 легальных пользователей, а также вредоносные действия нарушителей. Набор объединяет более 50 Гб «сырых» данных в формате PCAP и включает 8 предварительно обработанных файлов в формате CSV, содержащих размеченные сетевые сессии с выделенными признаками в разные дни наблюдения. Сетевые сессии относятся к одному из 15 классов: классу нормального трафика («BENIGN») или одному из 14 классов атак («DoS Hulk», «PortScan», «DDoS», «DoS GoldenEye», «FTP-Patator», «SSH-Patator», «DoS slowloris», «DoS Slowhttptest», «Bot», «Infiltration», «Heartbleed», «Web Attack – Brute Force», «Web Attack – XSS», «Web Attack – SQL Injection») [5, 6].

В подвыборке WebAttacks содержатся 4 класса: «BENIGN» (5087 записей), «Web Attack – Brute Force» (1507 записей), «Web Attack - Sql Injection» (21 запись), «Web Attack - XSS» (652 записи). Каждая запись в наборе данных WebAttacks представляет собой сетевую сессию и характеризуется 84 признаками, например, IP-адресами источника и приёмника потока данных («SourceIP» и «DestinationIP»), скоростью потока данных («Flow Bytes/s») и др. Из признаков пространства были исключены признаки «Flow ID», «Source IP», «Source Port», «Destination IP», «Destination Port», «Protocol», «Timestamp» в силу относительной лёгкости их подделки злоумышленником и предположении, что признаки «формы» (соответствующие статистикам сетевого трафика) являются более значимыми для общего случая.

Авторы настоящего исследования не имели возможности разработать атаку нулевого дня, но на практике возможно смоделировать такую ситуацию для сигнатурных средств защиты. Идея состоит в том, чтобы «откатить назад» («деградировать») базу решающих правил «назад по истории» и исключить одну из сигнатур, чтобы соответствующая атака стала атакой нулевого дня.

Для этого был проанализирован публичный репозиторий базы правил CRS⁵, и обнаружен запрос от сообщества (issue #2211) на включение новой сигнатуры в базу с тегом «False Negative – Evasion». Заголовок сообщения: «Drop keyword not blocked for SQL injection»⁶. Суть сообщения авторам проекта CRS заключалась в том, что одна из атак типа «SQL инъекция» не обнаруживалась WAF ModSecurity. Ошибка состояла в том, что для SQL инструкции DROP отсутствовал шаблон поиска в правиле 942360. Следующий запрос не блокировался до обновле-

¹ Damn Vulnerable Web Application. <https://github.com/digininja/DVWA>

² ModSecurity. <https://github.com/SpiderLabs/ModSecurity>

³ OWASP ModSecurity Core Rule Set. <https://coreruleset.org>

⁴ Intrusion Detection Evaluation Dataset (CICIDS2017). <https://www.unb.ca/cic/datasets/ids-2017.html>

⁵ OWASP ModSecurity Core Rule Set (Official Repository). <https://github.com/coreruleset/coreruleset>

⁶ <https://github.com/coreruleset/coreruleset/issues/2211>

ния набора правил: «<https://some.web.site/index.html?q='drop table users;--'>» (далее – тестовая SQL инъекция нулевого дня).

В используемом в эксперименте наборе правил CRS версии 3.3.2 правило 942360 «SQL Injection» было изменено – искусственно была выполнена замена правила на предыдущую версию (более раннюю редакцию). Таким образом смоделирована ситуация, когда одна из атак стала реализацией 0-day уязвимости – после изменения в базе правил CRS стала отсутствовать информация об одной реализации SQL инъекции.

Для устранения дисбаланса обучающей выборки в отношении атаки типа «SQL инъекция» набор данных был дополнен примерами SQL атак, реализованными генератором sqlmap¹.

3. Результаты эксперимента

Как и предполагалось, после модификации базы решающих правил («откат назад» к старой редакции правила 942360) сигнатурный классификатор WAF ModSecurity перестал обнаруживать реализацию тестовой SQL инъекции нулевого дня в отношении объекта защиты DVWA.

При этом обученная модель типа «случайный лес» успешно обнаруживала атаку – эксплуатацию искусственной 0-day уязвимости. Важно отметить, что конкретная реализация тестовой SQL инъекции нулевого дня не предъявлялась модели на этапе обучения. И обнаружение ранее неизвестной атаки стало возможным благодаря обобщающей способности модели.

4. Заключение

На узком классе веб-атак система обнаружения компьютерных атак на основе применения методов машинного обучения продемонстрировала преимущество по сравнению с сигнатурным анализатором WAF ModSecurity в отношении ранее неизвестных вредоносных воздействий.

Искусственное удаление одной сигнатуры из базы решающих правил привело к тому, что сигнатурный классификатор перестал обнаруживать соответствующую реализацию тестовой SQL инъекции нулевого дня. При этом эвристический анализатор на основе модели машинного обучения типа «случайный лес», обученной на публичном наборе данных CICIDS2017, за счет обобщающей способности модели оказался способным обнаруживать ранее неизвестную атаку.

Проведенный эксперимент подтверждает главное преимущество несигнатурных анализаторов: возможность обнаруживать атаки, которые ранее не встречались в обучающем наборе данных, т.е. новые атаки.

В выбранных условиях эксперимента синтезированная модель машинного обучения представляет собой самостоятельное решение, не требующее наличия сигнатурного анализатора. Однако в общем случае для реализации защиты от широкого перечня атак с максимальным качеством эвристический анализатор следует использовать как дополнение к сигнатурному анализатору. Предлагаемый подход к созданию гибридных систем обнаружения компьютерных атак является направлением дальнейших исследований авторов.

Литература

1. Talabis M., McPherson R., et al. Information Security Analytics. Waltham: Elsevier, 2015. 166 p.
2. Рыболовлев Д. А., Горюнов М.Н., Рыболовлев А.А. Оценка применимости методов машинного обучения для обнаружения компьютерных атак // Информационные системы и технологии. 2020. № 6. С. 103–111.
3. Chio C. Machine Learning and Security: Protecting Systems with Data and Algorithms. O'Reilly Media, 2018. 383 p.
4. Горюнов М. Н., Мацкевич А. Г., Рыболовлев Д. А. Синтез модели машинного обучения для обнаружения компьютерных атак на основе набора данных CICIDS2017 // Труды ИСП РАН. 2020. Том 32, вып. 5. С. 81-94.

¹ Sqlmap. <https://sqlmap.org>

5. Sharafaldin I., Lashkari A.H., Ghorbani Ali A. Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization // Proceedings of the 4th International Conference on Information Systems Security and Privacy (ICISSP), 2018. P. 108-116.
6. Gharib A., Sharafaldin I., Lashkari A.H., Ghorbani A.A. An Evaluation Framework for Intrusion Detection Dataset // International Conference on Information Science and Security (ICISS), 2016. DOI: 10.1109/ICISSEC.2016.7885840.

МАТЕМАТИЧЕСКИЕ МОДЕЛИ ПОВЕДЕНИЯ МАЛОГО БИЗНЕСА В УСЛОВИЯХ ВЫПЛАТ ПО КРЕДИТУ

В.П. Савельев, Н.И. Сулягина

Нижегородский государственный инженерно-экономический университет

Социально-экономическое развитие сельских территорий невозможно без устойчивого развития малого бизнеса региона. Поэтому, разнообразные подходы к изучению функционирования предприятий малого бизнеса актуальны и имеют как теоретическую, так и практическую значимость. Построенные ранее математические модели развития предприятий [1-3], оказывающих услуги населению сельских территорий, позволяют исследовать условия существования и устойчивости малого бизнеса. В данной статье построены и исследованы модели поведения малого бизнеса в условиях получения займа и необходимостью делать регулярные выплаты по кредиту. Получены условия существования устойчивых состояний равновесия.

Ключевые слова: выплаты по кредиту, предприятия малого бизнеса, динамическая система, состояние равновесия, линеаризация, фазовый портрет.

1. Линейная модель

Аналогично построенным ранее моделям предполагаем, что трудоспособное население региона состоит из двух частей: первая часть работает в сфере производства товаров и услуг для населения рассматриваемого региона (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 предоставленных услуг населению региона предприятиями малого бизнеса, $y(t)$ – объем произведенных товаров, выполненных работ и оказанных услуг предприятиями, работающими на внешний сервис.

Предполагается, что предприятия внешнего сервиса реализуют товары и услуги не только вне данного региона. В этом случае динамику изменения состояния $(x(t), y(t))$ можно задать с помощью линейной системы дифференциальных уравнений:

$$\begin{cases} \dot{x} = -ax + by - p, \\ \dot{y} = R + dx - cy. \end{cases} \quad (1)$$

где (в единицу времени) R – доход предприятий внешнего сервиса от продажи товаров и услуг вне данного региона, ax – оплата труда работников внутреннего сервиса за вычетом их затрат на свой продукт, by – доход предприятий малого бизнеса от домашних хозяйств работников внешнего сервиса, cy – оплата труда работников внешнего сервиса за вычетом их затрат на свой продукт, dx – доход предприятий внешнего сервиса от домашних хозяйств работников малого бизнеса, p – уровень выплат по кредиту (все параметры положительны).

При естественном предположении, что $a > d, c > b$, суммы денег $(a - d)x, (c - b)y$ тратятся на покупку товаров и услуг вне данного региона. Находим состояние равновесия системы (1):

$$x^* = \frac{bR - pc}{ac - bd}, \quad y^* = \frac{aR - dp}{ac - bd}. \quad (2)$$

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 $(0, 0)$:

$$\begin{cases} \dot{X} = -aX + bY, \\ \dot{Y} = dX - cY. \end{cases} \quad (3)$$

Из уравнения $\lambda^2 + (a + c)\lambda + ac - bd = 0$ находятся собственные значения матрицы:

$$\lambda_{1,2} = \frac{-(a+c) \pm \sqrt{(a+c)^2 - 4(ac-bd)}}{2} = \frac{-(a+c) \pm \sqrt{(a-c)^2 + 4bd}}{2}. \quad (4)$$

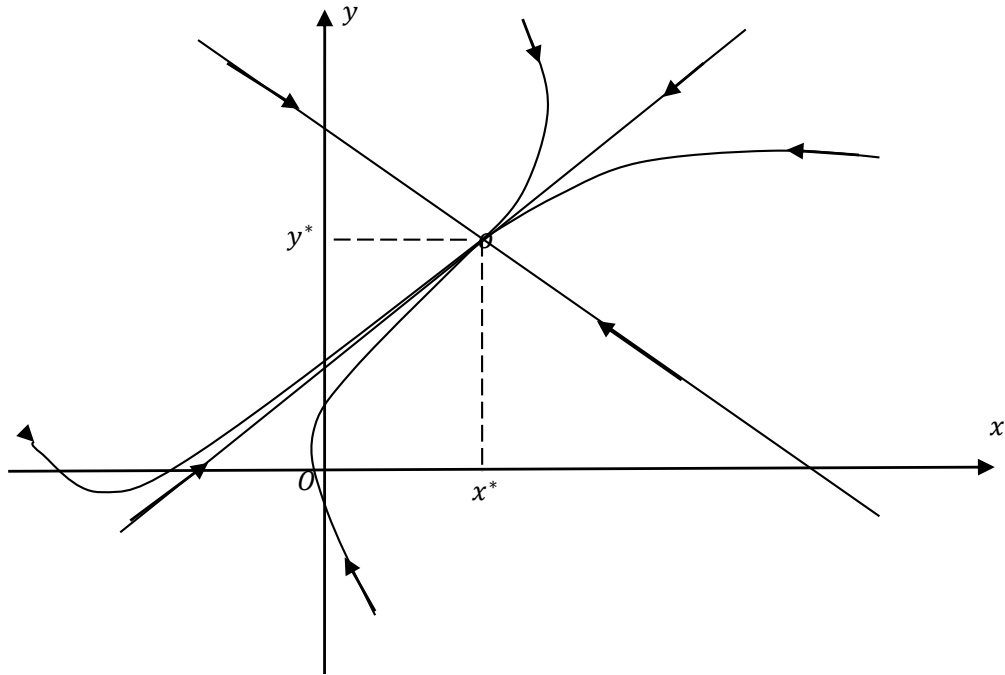


Рис. 1. Фазовый портрет системы (1) для случая $x^* > 0, y^* > 0$

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).

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 $y < \frac{p}{b}$.

Заметим, что в зависимости от значений параметров R, p, b, c баланс $x^* = \frac{bR-pc}{ac-bd}$ системы предприятий малого бизнеса может быть как положительным, так и отрицательным. То есть предприятия малого бизнеса могут устойчиво работать даже в условиях постоянной кредитной задолженности.

2. Нелинейная модель

В линейной модели (1) предполагалось, что доход by системы предприятий малого бизнеса зависит лишь от объема y произведенных товаров и услуг предприятиями внешнего сервиса. 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, модель динамики регионального бизнеса принимает вид:

$$\begin{cases} \dot{x} = -ax + bxy - p, \\ \dot{y} = R + dx - cy - bxy, \end{cases} \quad (5)$$

При выполнении условия

$$(R - p) - ac > 2\sqrt{bcp(a - d)} \quad (6)$$

динамическая система (5) имеет два состояния равновесия с координатами

$$x_1 = \frac{(R-p)b-ac+\sqrt{[(R-p)b-ac]^2-4bcp(a-d)}}{2b(a-d)}, \quad y_1 = \frac{(R-p)b+ac-\sqrt{[(R-p)b-ac]^2-4bcp(a-d)}}{2bc}, \quad (7)$$

$$x_2 = \frac{(R-p)b-ac-\sqrt{[(R-p)b-ac]^2-4bcp(a-d)}}{2b(a-d)}, \quad y_2 = \frac{(R-p)b+ac+\sqrt{[(R-p)b-ac]^2-4bcp(a-d)}}{2bc}. \quad (8)$$

Заметим, что оба состояния равновесия находятся в правой полуплоскости. В окрестности первого состояния равновесия линеаризованная [4] система имеет вид:

$$\begin{cases} \dot{X} = \frac{(R-p)b-ac-\sqrt{[(R-p)b-ac]^2-4bcp(a-d)}}{2c}X + \frac{(R-p)b-ac+\sqrt{[(R-p)b-ac]^2-4bcp(a-d)}}{2(a-d)}Y, \\ \dot{Y} = -\frac{(R-p)b+ac-2cd-\sqrt{[(R-p)b-ac]^2-4bcp(a-d)}}{2c}X - \frac{(R-p)b+ac-2cd+\sqrt{[(R-p)b-ac]^2-4bcp(a-d)}}{2(a-d)}Y. \end{cases} \quad (9)$$

где $X = x - x_1, Y = y - y_1$.

Собственные значения матрицы линейной системы находятся из уравнения $\lambda^2 + A\lambda + B = 0$, где $A = \frac{(c+d-a)[(R-p)b-ac-\sqrt{[(R-p)b-ac]^2-4bcp(a-d)}]}{2(a-d)c} + \frac{\sqrt{[(R-p)b-ac]^2-4bcp(a-d)}}{a-d} + c$, $B = \sqrt{[(R-p)b-ac]^2-4bcp(a-d)}$.

Коэффициент B положителен, а коэффициент A будет положительным при дополнительном условии, что $c + d \geq a$, которое можно считать вполне разумным.

Тогда собственные числа матрицы либо оба действительные и отрицательные, либо комплексно-сопряженные с отрицательной действительной частью. Соответствующее состояние равновесия является либо устойчивым узлом, либо устойчивым фокусом.

В окрестности второго состояния равновесия линеаризованная система имеет вид:

$$\begin{cases} \dot{X} = \frac{(R-p)b-ac+\sqrt{[(R-p)b+ac]^2-4bcp(a-d)}}{2c}X + \frac{(R-p)b-ac-\sqrt{[(R-p)b+ac]^2-4bcp(a-d)}}{2a}Y, \\ \dot{Y} = -\frac{(R-p)b+ac+\sqrt{[(R-p)b+ac]^2-4bcp(a-d)}}{2c}X - \frac{(R-p)b+ac-\sqrt{[(R-p)b+ac]^2-4bcp(a-d)}}{2a}Y. \end{cases} \quad (10)$$

Уравнение для собственных значений матрицы системы (10) имеет вид: $\lambda^2 + C\lambda + D = 0$, где $D = -\sqrt{[(R-p)b-ac]^2-4bcp(a-d)}$.

Поскольку $D < 0$, собственные числа оба действительные и разных знаков, то есть, второе состояние равновесия является седлом. Соответствующий устойчивому узлу фазовый портрет системы (5) представлен на рисунке 2.

Фазовое пространство делится на две области кривой Γ , составленной из ω —сепаратрис S_1 и S_2 седла $M_2(x_2, y_2)$.

Если начальное состояние системы (5) расположено справа от кривой Γ , то фазовая траектория приводит систему в устойчивый узел $M_1(x_1, y_1)$. Если же начальное состояние системы расположено слева от кривой Γ , то фазовая траектория приводит систему в левую полуплоскость, что соответствует увеличению задолженности предприятий малого бизнеса и возможному банкротству. Это означает, что устойчивость работы предприятий малого бизнеса может быть гарантирована лишь при достаточно большом значении начального состояния x_0 .

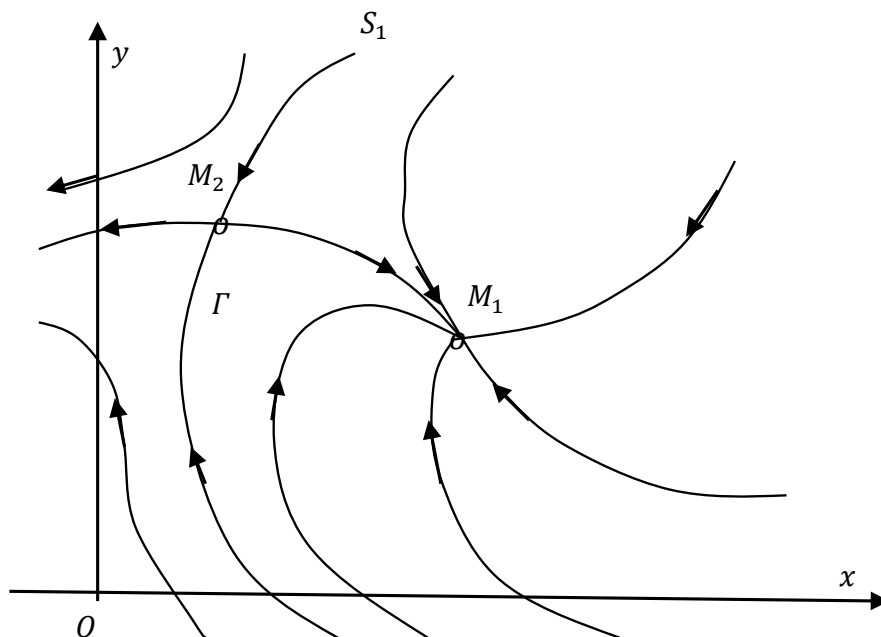


Рис. 2. Фазовый портрет системы (5) с устойчивым узлом

Заметим, что при увеличении параметра p состояния равновесия сближаются и происходит седло-узловая бифуркация: при равенстве $(R - p)b = ac + 2\sqrt{abc p}$ образуется так называемый седло – узел. Затем оба состояния равновесия исчезают, и все фазовые траектории системы пересекают справа налево ось Oy .

Литература

1. Савельев В.П., Сутягина Н.И. Простые модели динамики малого бизнеса // Математическое моделирование и суперкомпьютерные технологии : Труды XX Международной конференции, Нижний Новгород, 23–27 ноября 2020 года. Нижний Новгород: Национальный исследовательский Нижегородский государственный университет им. Н.И. Лобачевского, 2020. 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 // Математическое моделирование и суперкомпьютерные технологии : Труды XXI Международной конференции, Нижний Новгород, 22–26 ноября 2021 года. Нижний Новгород: Национальный исследовательский Нижегородский государственный университет им. Н.И. Лобачевского, 2021. С. 298 – 300.
4. Андронов А. А., Леонтович Е. А. Гордон И. И., Майер А. Г. Качественная теория динамических систем второго порядка. М.: Наука, 1966.

ПОСТРОЕНИЕ ГИБРИДНЫХ СИСТЕМ ОБНАРУЖЕНИЯ КОМПЬЮТЕРНЫХ АТАК НА ОСНОВЕ ОБЪЕДИНЕНИЯ СИГНАТУРНОГО И ЭВРИСТИЧЕСКОГО МЕТОДОВ

В.Е. Салимоненко

Академия ФСО России, Орел

В работе рассматривается практический подход по объединению сигнатурного и эвристического анализаторов в единой системе обнаружения компьютерных атак. Предлагается использовать ранее накопленные данные о срабатываниях сигнатурных модулей при формировании обучающей выборки для модели машинного обучения. Обсуждаются частные подзадачи процесса внедрения и факторы, снижающие итоговое качество модели обнаружения атак.

Ключевые слова: система обнаружения атак, сигнатурный анализатор, эвристический анализатор, машинное обучение, компьютерная атака.

1. Введение

В настоящее время при решении задачи обнаружения компьютерных атак широкое применение находят сигнатурные анализаторы (например, Snort¹ или Suricata²), качество работы которых определяется полнотой базы решающих правил. Преимуществами таких систем являются относительная простота работы и интерпретируемость результатов (всегда возможно установить, какой конкретно пакет или группа пакетов вызвали срабатывание сенсора), высокая производительность в сравнении с другими методами обнаружения атак.

Вместе с тем сигнатурные системы обнаружения атак (СОА) имеют свои недостатки. Одним из принципиальных ограничений таких СОА является неспособность обнаруживать ранее неизвестные, новые типы атак [1]. Кроме того, известна проблема снижения производительности сигнатурной СОА при увеличении количества сигнатур в базе решающих правил.

Одним из перспективных подходов к обнаружению новых атак является применение моделей машинного обучения, которые за счет своей обобщающей способности могут детектировать ранее неизвестные виды атак [2]. Актуальной представляется задача построения гибридных СОА, объединяющих в себе возможности сигнатурных и эвристических методов для повышения качества обнаружения как известных, так и неизвестных компьютерных атак.

В настоящем исследовании предлагается подход к построению гибридных СОА, основанный на использовании результатов срабатываний сигнатурного анализатора при формировании обучающего набора данных для модели машинного обучения (эвристического анализатора) и объединении двух этих анализаторов в единую СОА.

2. Внедрение машинного обучения в СОА

Один из возможных вариантов внедрения машинного обучения в существующую сигнатурную СОА представлен на рис. 1. Отличительной особенностью такого подхода является обогащение обучающей (дообучающей) выборки примерами срабатываний (класс «атака») от используемых в системе средств защиты (сигнатурная СОА, межсетевой экран веб-приложений WAF, web application firewall, и др.) и дальнейшее обучение (дообучение) эвристического классификатора на этих данных.

¹ <https://www.snort.org/>

² <https://suricata.io/>

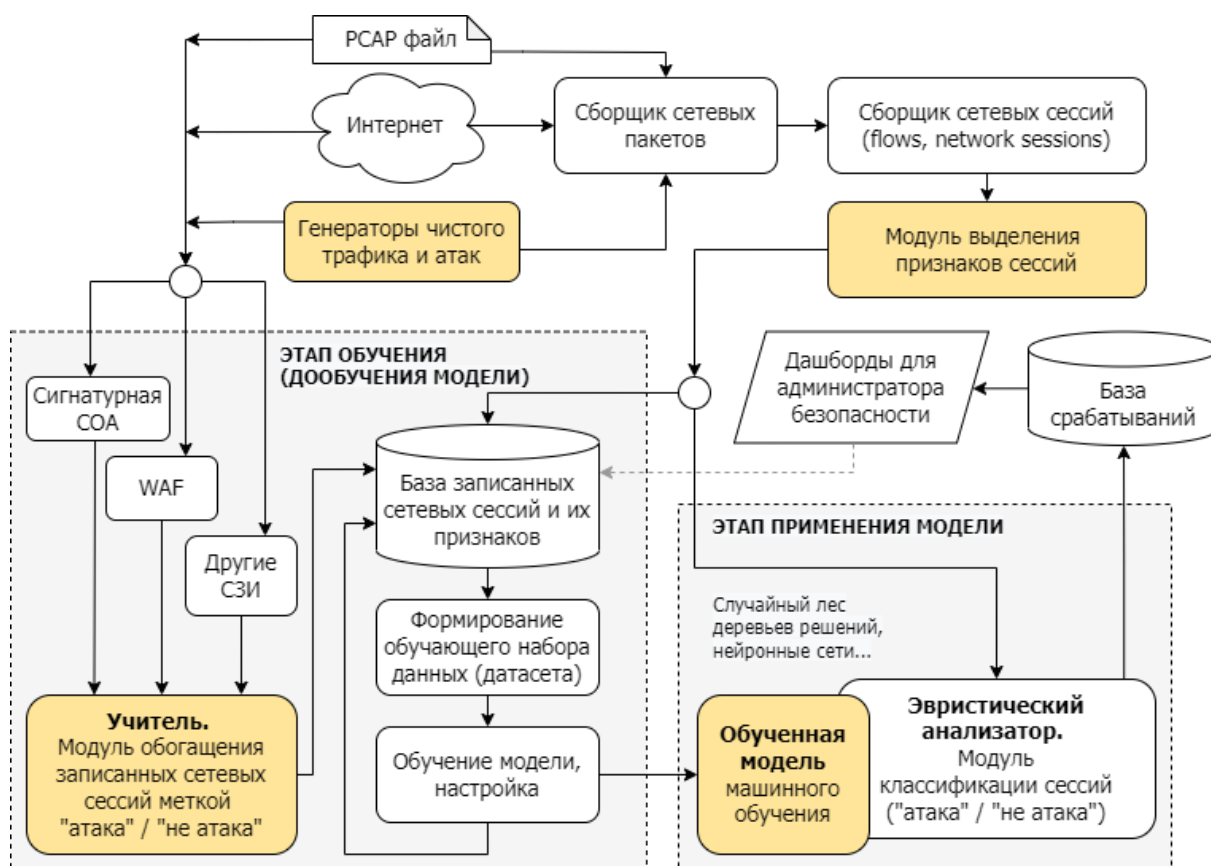


Рис. 1. Схема гибридной системы обнаружения атак

Сетевые пакеты, выделенные из поступающего сетевого трафика, направляются на модуль сбора и выделения сетевых сессий. Поступающие пакеты группируются в сессии по правилу так называемой «пятерки» (5-tuple) параметров: *src_IP* – IP-адрес источника, *src_port* – порт источника, *dst_IP* – IP-адрес получателя, *dst_port* – порт получателя, *proto* – протокол.

Далее записанная сетевая сессия поступает на вход модуля выделения признаков сессий. Примерами таких признаков могут служить: средняя длина пакетов сессии, скорость потока в байтах в секунду, максимальная длина пакета сессии, межпакетная задержка и её среднеквадратическое отклонение, суммарный размер пакетов сессии и др. Задача выбора оптимального признакового пространства является одной из наиболее сложных и важных при построении модели машинного обучения.

С выхода данного модуля векторы признаков сессий поступают непосредственно на вход эвристического анализатора и/или записываются в базу сетевых сессий и их признаков. При этом, если на одном из пакетов сессии срабатывает сигнатурный анализатор (одно из используемых в системе средств защиты), то модуль обогащения добавляет к сессии метку «атака» и один или несколько признаков, соответствующие идентификатору сенсора и идентификатору сработавшего правила. Иначе, при закрытии сессии (по поступающему пакету «FIN», по таймеру или др.) она помечается меткой «чистый трафик».

После разметки записанных сетевых сессий требуется произвести оценку качества собранных данных и их балансировку. Предполагается, что полученные на «реальном» сетевом трафике атаки будут соответствовать лишь малой части базы решающих правил сигнатурного анализатора. Поэтому для улучшения качества модели машинного обучения требуется дополнение обучающей выборки искусственно полученными данными. Такие данные (другие виды атак, чистый трафик) могут быть получены от вспомогательного модуля – генератора атак и чистого трафика. Кроме того, возможен вариант дополнения полученной обучающей выборки данными публичных наборов (например, CICIDS2017¹, ISCXIDS2012², UNSW-NB15¹ и др.).

¹ <https://www.unb.ca/cic/datasets/ids-2017.html>

² <https://www.unb.ca/cic/datasets/ids.html>

При записи чистого трафика важным является отсутствие в сети вредоносного трафика, иначе обученная на таком сетевом трафике модель машинного обучения будет показывать неудовлетворительное качество.

Следуя этой схеме, на выходе формируется размеченный набор данных, полученный в защищаемой инфраструктуре сети на «реальном» сетевом трафике, что является одним из условий для получения высоких метрик качества обнаружения атак моделью машинного обучения [1].

Далее, после обучения модели становится возможным её использование для обнаружения атак в защищаемой сети. Обязательным является этап оценки качества модели и сравнения полученных показателей с показателями качества сигнатурной СОА. При неудовлетворительных значениях необходим пересмотр признакового пространства, повторное обучение модели, настройка гиперпараметров модели. Также обязательным является дальнейшее дообучение модели при обновлении базы решающих правил сигнатурного анализатора и при обновлении базы генератора атак.

3. Заключение

Внедрение машинного обучения в сетевые системы обнаружения компьютерных атак является трудоёмким и сложным процессом, требующим наличия высокого уровня экспертизы у разработчиков, высокопроизводительной инфраструктуры для экспериментов, понимания поставленной задачи как научно-исследовательской с возможным отрицательным результатом.

Предложенный вариант внедрения машинного обучения в сигнатурную СОА позволит разработчикам таких систем использовать уже накопленные размеченные данные для обучения классификатора.

Отдельного внимания после обучения потребуют вопросы оценки качества собранных наборов данных (баланс классов, разнообразие наблюдаемых атак), синтезированной модели (устранение переобучения) и выбранного признакового пространства.

При построении эвристических анализаторов следует учитывать множество факторов, которые могут отрицательно сказываться на качестве обнаружения атак. Такими факторами могут выступать: появление новых пользователей в сети, новых сетевых служб, изменение конфигураций сетевого оборудования, изменения в инфраструктуре сети и др. Для предотвращения деградации качества эвристического анализатора в связи с перечисленными факторами требуется постоянное дообучение модели.

Литература

1. Горюнов М.Н., Мацкевич А.Г., Рыболовлев Д.А. Синтез модели машинного обучения для обнаружения компьютерных атак на основе набора данных CICIDS2017 // Труды ИСП РАН. 2020. Том 32, вып. 5. С. 81–94.
2. Sarker, I.H., Kayes, A.S.M., Badsha, S. et al. Cybersecurity data science: an overview from machine learning perspective. J Big Data 7, 41 (2020). DOI: 10.1186/s40537-020-00318-5.

¹ <https://research.unsw.edu.au/projects/unsw-nb15-dataset>

ИССЛЕДОВАНИЕ РЕАКЦИИ ГЛУБОКИХ СЛОЕВ ОБУЧЕННЫХ СЕТЕЙ ГЛУБОКОГО ОБУЧЕНИЯ НА СЕМАНТИЧЕСКИ НЕЗНАЧИМЫЕ ИЗМЕНЕНИЯ В ДАННЫХ^{1*}

Я.А. Середя, И.Д. Никоноров

¹*Центр доверенного искусственного интеллекта ИСП РАН*

Была исследована устойчивость реакции нейронов, кодирующих высокоуровневые признаки, в ситуациях, когда семантика изображения (его высокоуровневые признаки) не меняются, но происходят малозначимые попиксельные изменения. Для исследования были взяты популярные модели сетей глубокого обучения. Было показано, что сохранение высокой активации высокоуровневого нейрона даже при незначительных изменениях (например, при сдвиге контента изображения на 5 пикселей), не гарантировано.

Ключевые слова: искусственный интеллект, нейронные сети, аугментации изображений, глубокое обучение.

1. Введение

Одним из предметов интереса в биологии мозга является поиск связи между тем, какие нейроны наиболее сильно активируются в данный момент, и некоторым конкретным стимулом, находящимся во входных данных. На основе этого строится вывод о специализациях нейронов. В качестве примера можно привести концепцию «нейрона бабушки» [1]. Это гипотетический нейрон, представляющий сложную, но специфическую концепцию или объект. Он активируется, когда человек «видит, слышит или иным образом разумно различает» конкретное явление, в данном случае бабушку.

Для глубоких нейросетей предполагается, что нейроны глубоких слоев кодируют высокоуровневую семантику изображения. В ходе данного исследования был рассмотрен вопрос о том, в какой мере сохраняется высокая активация этих нейронов при семантически незначительных изменениях в изображениях [2].

2. Постановка эксперимента

Для проведения исследования была создана методика замера активации, состоящая в проверке «выживаемости» самых активированных изображением нейронов на глубоком слое при не меняющих высокоуровневую семантику изображения изменениях. Для каждого слоя вычислялся квантиль уровня 0.999 индексов самых высокоактивированных изображением нейронов: сначала в оригинальных изображениях, а затем в изображениях после аугментаций. Полученные значения сравнивались между собой, выделялось число индексов нейронов, совпавших у исходного изображения и у измененного, т.е. число «выживших индексов».

Для рассмотрения были взяты две известные предварительно обученные сети: ResNet34 [3] и VGG16 [4]. В качестве входных данных был использован набор CelebA [5], состоящий из более чем 200 000 изображений знаменитостей размером 178×218, из которого для экспериментов было отобрано 10 тестовых изображений.

^{1*} Работа выполнена при поддержке гранта для исследовательских центров в области искусственного интеллекта, предоставленного Аналитическим центром Правительства РФ в соответствии с договором о субсидии (идентификатор соглашения 000000D730321P5Q0002) и соглашением с Институтом системного программирования имени Иванникова Российской Академии наук от 2 ноября 2021 г. № 70-2021-00142.

3. Результаты экспериментов

В экспериментах было рассмотрено 3 вида семантически незначимых преобразований изображений, а именно: фильтр blur, сдвиг изображения, а также поворот изображения. При заданных размерах изображений такие преобразования никак не отражаются на его высокоуровневой семантике. Для человека изображения визуально меняются слабо. Диаграммы результатов, приведённые в данном разделе, строились только для одного глубокого слоя сети. Ниже будут приведены примеры изображений из тестируемого набора до и после аугментаций.



Рис. 1. Слева оригинальное изображение, справа - после поворота



Рис. 2. Слева оригинальное изображение, справа – изображение со сдвигом



Рис. 3. Слева оригинальное изображение, справа - изображение после применения фильтра blur

3.1. Размытие изображения

Размытие является семантически незначимым преобразованием и для человека изображение визуально меняется слабо. На рис. 4 показана диаграмма результатов одного из экспериментов, отражающая соотношение числа выживших индексов к не выжившим.



Рис. 4. Диаграмма процентного соотношения выживших индексов к не выжившим при наложении фильтра blur

Рис. 4 демонстрирует, что при достаточно слабом размытии, в среднем, около 15 процентов индексов не выживают даже на глубоком слое.

3.2. Сдвиг изображения

На рис. 5 представлена диаграмма одного из результатов эксперимента по сдвигу изображения на заданное число пикселей.

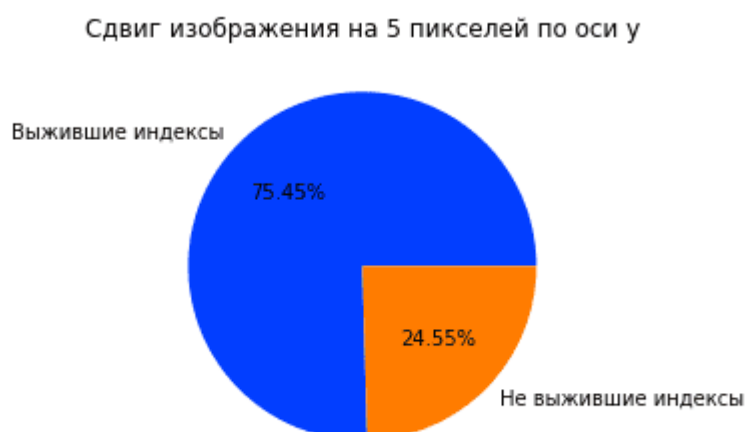


Рис. 5. Диаграмма процентного соотношения выживших индексов к не выжившим при сдвиге

На рис. 5 показано, что даже при сдвиге на 5 пикселей, что не отображается на высокоуровневой семантике, 1/4 индексов не выживают на глубоком слое.

3.3. Поворот изображения

Ещё одной серией преобразований был поворот изображения, результаты для одного из вариантов данного типа преобразования представлены ниже, на рис. 6.



Рис. 6. Диаграмма процентного соотношения выживших индексов к не выжившим при повороте

Результаты данной серии экспериментов стоит выделить отдельно. Как можно заметить из диаграммы, при такой аугментации более 83 процентов индексов не выживают. Для демонстрации результатов размытия и сдвига изображений была задействована только сеть ResNet34, так как в этих экспериментах не было качественных отличий по сравнению с сетью VGG16, однако, в случае с поворотом изображения, отличия между сетями присутствуют, диаграмма для этого случая представлена ниже.



Рис. 7. Диаграмма процентного соотношения выживших индексов к не выжившим при повороте, сеть VGG16

На диаграмме можно увидеть, что несмотря на то, что число не выживших индексов достаточно велико, их доля составляет менее 50 процентов от общего числа, что заметно меньше, чем в аналогичном эксперименте с сетью ResNet34.

4. Заключение

Результаты выживаемости самых активированных изображением нейронов во всех случаях имеют сходную качественную картину:

- Чем менее глубокий слой, тем хуже выживаемость высокой активации (замеряемой по описанной методике). Этот результат ожидаем: в самом деле, чем ближе нейрон ко входному слою, тем менее абстрактные признаки он должен кодировать.
- Однако, и в глубоких слоях выживаемость неполная, что является менее ожидаемым результатом.

В экспериментах с глубокими свёрточными сетями было показано, что нейроны глубоких слоев свёрточных сетей не обладают гарантированной инвариантностью к семантически незначимым изменениям в данных, в том числе к малым сдвигам, и особенно поворотам.

Литература

1. Gross C. G. Genealogy of the “grandmother cell” //The Neuroscientist. – 2002. – Т. 8. – №. 5. – С. 512-518.
2. Van Dyk D. A., Meng X. L. The art of data augmentation //Journal of Computational and Graphical Statistics. – 2001. – Т. 10. – №. 1. – С. 1-50.
3. He K. et al. Deep residual learning for image recognition //Proceedings of the IEEE conference on computer vision and pattern recognition. – 2016. – С. 770-778.
4. Simonyan K., Zisserman A. Very deep convolutional networks for large-scale image recognition //arXiv preprint arXiv:1409.1556. – 2014.
5. Liu Z. et al. Large-scale celebfaces attributes (celeba) dataset //Retrieved August. – 2018. – Т. 15. – №. 2018. – С. 11.

АЛГОРИТМ ГЛОБАЛЬНОЙ ОПТИМИЗАЦИИ, ИСПОЛЬЗУЮЩИЙ ДЕРЕВЬЯ РЕШЕНИЙ ДЛЯ ВЫЯВЛЕНИЯ ЛОКАЛЬНЫХ ЭКСТРЕМУМОВ^{1*}

Д.И. Силенко, И.Г. Лебедев

Нижегородский государственный университет им. Н. И. Лобачевского

В работе рассматривается решение задач многомерной глобальной оптимизации с применением деревьев решений для выявления областей притяжения локальных минимумов. Локальный метод используется для локального уточнения и призван ускорить сходимость алгоритма. Искомая функция удовлетворяет условию Липшица с неизвестной константой. Для подтверждения теории были проведены вычислительные эксперименты, демонстрирующие ускорение при решении серии тестовых задач.

Ключевые слова: глобальная оптимизация, локальная оптимизация, многоэкстремальные функции, деревья решений.

1. Введение

При поиске глобального минимума функции можно придерживаться различных алгоритмов. Поскольку в реальных задачах глобальной оптимизации каждое вычисление значения функции представляет собой весьма трудоемкую процедуру, необходимо уменьшить количество таких операций. Этого можно добиться целенаправленным выбором вариантов в процессе поиска оптимального решения. На этой идее основывается алгоритм глобального поиска (АГП) [1]. В данной работе мы объединим АГП и метод Хука-Дживса (метод локальной оптимизации) с помощью деревьев решений, с целью уменьшения количества проводимых итераций, а вместе с этим и количества вычислений целевой функции. В разделе «Эксперименты» можно найти результаты сравнения такого подхода и стандартного АГП.

2. Постановка задачи

Рассмотрим задачу поиска глобального минимума функции $\varphi(y)$ в гиперинтервале $D = \{y \in R^N: a_i \leq y_i \leq b_i, 1 \leq i \leq n\}$. При этом будем предполагать, что функция удовлетворяет условию Липшица с априори неизвестной константой L .

$$\varphi(y^*) = \min(\varphi(y)): D = \{y \in R^N: a_i \leq y_i \leq b_i, 1 \leq i \leq n\}, \quad (1)$$

$$|\varphi(y_1) - \varphi(y_2)| \leq L \|y_1 - y_2\|, y_1, y_2 \in R^N, 0 < L < \infty \quad (2)$$

Численное решение задачи (1) сводится к построению оценки $y_k^* \in D$, отвечающей некоторому понятию близости к точке y^* на основе конечного числа k вычислений значений оптимизируемой функции. Оптимизируемая функция $\varphi(y)$ может быть задана не аналитически, а алгоритмически, как результат работы некоторой подпрограммы или библиотеки.

Поиск глобального оптимума сводится к построению некоторой сетки в области параметров, и выбору наилучшего значения функции на данной сетке. Вычислительные затраты на решение задачи растут экспоненциально с ростом размерности. В ННГУ им. Н.И. Лобачевского под руководством проф. Р.Г. Стронгина разработан эффективный подход к решению задач глобальной оптимизации. Решение многомерных задач сводится к решению серии вложенных задач меньшей размерности.

^{1*} Работа выполнена при финансовой поддержке Министерства науки и высшего образования РФ (проект № 0729-2020-0055) и научно-образовательного математического центра “Математика технологий будущего” (соглашение № 075-02-2021-1394).

3. Описание алгоритмов

3.1. Многомерный алгоритм глобального поиска

При решении многомерных задач применяется редукция размерности с использованием кривых Пеано. Они позволяют свести многомерную задачу минимизации в области D к одномерной задаче минимизации на отрезке $[0, 1]$

$$\varphi(y(x^*)) = \min\{\varphi(y(x)): x \in [0, 1]\}, \quad (3)$$

где функция удовлетворяет уже равномерному условию Гельдера

$$|\varphi(y_1) - \varphi(y_2)| \leq H|y_1 - y_2|^{\frac{1}{N}}, y_1, y_2 \in [0, 1]. \quad (4)$$

Поэтому, не ограничивая общности, можно рассматривать минимизацию одномерной функции $f(x) = \varphi(y(x))$, $x \in [0, 1]$, удовлетворяющей условию Гельдера.

На начальной итерации алгоритма проводятся испытания на границах отрезка $[0, 1]$: $x^0 = 0$ и $x^1 = 1$. Далее на каждой итерации мы имеем набор интервалов, ограниченных точками испытания. Для каждого интервала вычисляется характеристика, а в интервале с наибольшей характеристикой проводится новое испытание. Итерации повторяются, пока не достигнут один из критериев остановки [2].

Схема алгоритма:

1. Упорядочить точки по координате:

$$0 = x_0 < \dots < x_i < \dots < x_k = 1. \quad (5)$$

2. Вычислить оценку μ для неизвестной константы Липшица L ,

$$\mu = \max \left\{ \frac{|z_i - z_{i-1}|}{(x_i - x_{i-1})^{1/N}}, i = 1, \dots, k \right\}. \quad (6)$$

3. Для каждого (x_{i-1}, x_i) , $1 \leq i \leq k$, вычислить характеристику $R(i)$,

$$R(1) = 2\Delta_1 - 4\frac{z_1}{\mu r}; R(k) = 2\Delta_k - 4\frac{z_{k-1}}{\mu r}; R(i) = \Delta_i + \frac{(z_i - z_{i-1})^2}{z^2 \mu^2 \Delta_i} - 2\frac{(z_i + z_{i-1})}{r\mu}, \quad (7)$$

где $1 < i < k$ Δ_i – корень степени N из длины интервала, $r > 1$, – параметр метода.

4. Определить номер t :

$$R(t) = \max\{R(i): 1 \leq i \leq k\}. \quad (8)$$

Провести очередное испытание в точке из интервала (x_{t-1}, x_t)

$$x^{k+1} = \frac{x_t + x_{t-1}}{2}, t = 1, t = k; x^{k+1} = \frac{x_t + x_{t-1}}{2} - \text{sign}(z_t - z_{t-1}) \frac{1}{2r} \left[\frac{|z_t - z_{t-1}|}{\mu} \right]^N. \quad (9)$$

Алгоритм прекращает работу, если выполняется условие $\Delta_t \leq \varepsilon$ хотя бы для одного номера t ; здесь $\varepsilon > 0$ есть заданная точность. В качестве оценки глобально-оптимального решения задачи (3) выбираются значения

$$f_k^* = \min_{1 \leq i \leq k} f(x^i), x_k^* = \arg \min_{1 \leq i \leq k} f(x^i). \quad (10)$$

Обоснование данного способа организации вычислений см. в [12]. Модификации, учитывающие наличие ограничений-неравенств в задаче, а также информация о производной целевой функции, представлены в [3, 4].

3.2 Метод Хука-Дживса

Метод Хука-Дживса – это комбинация исследующего поиска по направлениям и поиска по образцу [6, 7].

Суть исследующего поиска: задается величина шага. Если значение целевой функции в пробной точке не превышает значение в исходной, то шаг – успешный. Иначе, необходимо попробовать сместиться в другое направление. Когда перебор всех N координат закончится, то считаем найденную точку базовой и заканчиваем исследующий поиск. Поиск по образцу реализует шаг из базовой точки вдоль прямой, соединяющей ее с предыдущей базовой точкой. Величина этого шага определяется задаваемым заранее параметром.



Рис. 1. Пример итераций алгоритма Хука-Дживса

3.3 Деревья решений

Дерево решений – это метод автоматического анализа больших массивов данных, который применяется в машинном обучении. Дерево решений является бинарным и его можно использовать как для решения задач классификации, так и для регрессионных задач.

В работе реализация деревьев решений была взята из библиотеки OpenCV (библиотека алгоритмов компьютерного зрения, обработки изображений и численных алгоритмов общего назначения с открытым кодом). Подробнее с деревьями решений можно ознакомиться в [8].

3.4 Объединение метода Хука-Дживса и АГП в случае многомерных задач

Необходимо понять: запускать локальный метод из текущей точки или нет. Для этого можно проверить соседние точки и если среди них нет ни одной такой, значение функции в которой меньше, чем в текущей, то можно предположить, что мы находимся в области локального спуска и запустить локальный метод. Использование кривых Пеано здесь нецелесообразно, т.к. вид функции после отображения может сильно измениться, а кроме того, мы можем потерять информацию о взаимном расположении точек исходной функции. Поэтому мы будем работать именно с исходными точками.

После проведения определенного числа испытаний используем все накопившиеся точки для тренировки дерева решений. Далее построим равномерную сетку с определенным шагом по каждой из размерностей, а ее узлы подадим на вход предсказанию по дереву, чтобы получить значения функции в них. Теперь найдем ближайшую, с точки зрения евклидова расстояния, точку к исходной и определим ее соседей перебором индексов. Если нашли значение функции меньшее, чем в текущей точке, то продолжаем работу АГП; если нашли такое же значение функции, то необходимо аналогично проверить такую точку; иначе, запускаем локальный метод.

Модифицированная версия алгоритма глобального поиска выглядит следующим образом. Вначале выполняются 4 этапа классического алгоритма, затем:

5. Если $k < 100 * N$, то продолжаем работу АГП.
6. Если используем деревья решений не впервые, то переходим к пункту 13.
7. Инициализируем дерево решений и тренируем его.
8. Строим равномерную сетку, откладывая по каждой координате $\lfloor \sqrt[N]{300} \rfloor$ точек.
9. Предсказываем значения функции в узлах сетки с помощью дерева решений.
10. Находим ближайшую с точки зрения Евклидова расстояния точку к текущей.
11. Обходим соседей.
12. Если не нашли точек со значением функции меньше, то запускаем метод Хука-Дживса.
13. Если запускаем алгоритм не впервые, то накапливаем массив из $100 * N$ точек.
14. Все $100 * N$ точек обрабатываем аналогично пунктам 7-12, но без пункта 8.

4. Программная реализация

Все полученные результаты были добавлены в систему globalizer [10]. Была реализована объединенная схема с запуском локальных методов, для чего была внедрена система построения дерева решений, равномерной сетки и прочих сопутствующих вычислений. Для всего этого были добавлены новые методы в уже существующие классы: в классе TMethod метод UpdateStatus теперь выполняет проверку о запуске локального метода, а метод LocalS производит запуск выбранного локального метода. Новый метод UpdateStatusDecisionTreesMultiDims реализует всю логику, связанную с деревьями решений. Ниже приведена блок-схема того, как данная схема была объединена с АГП.

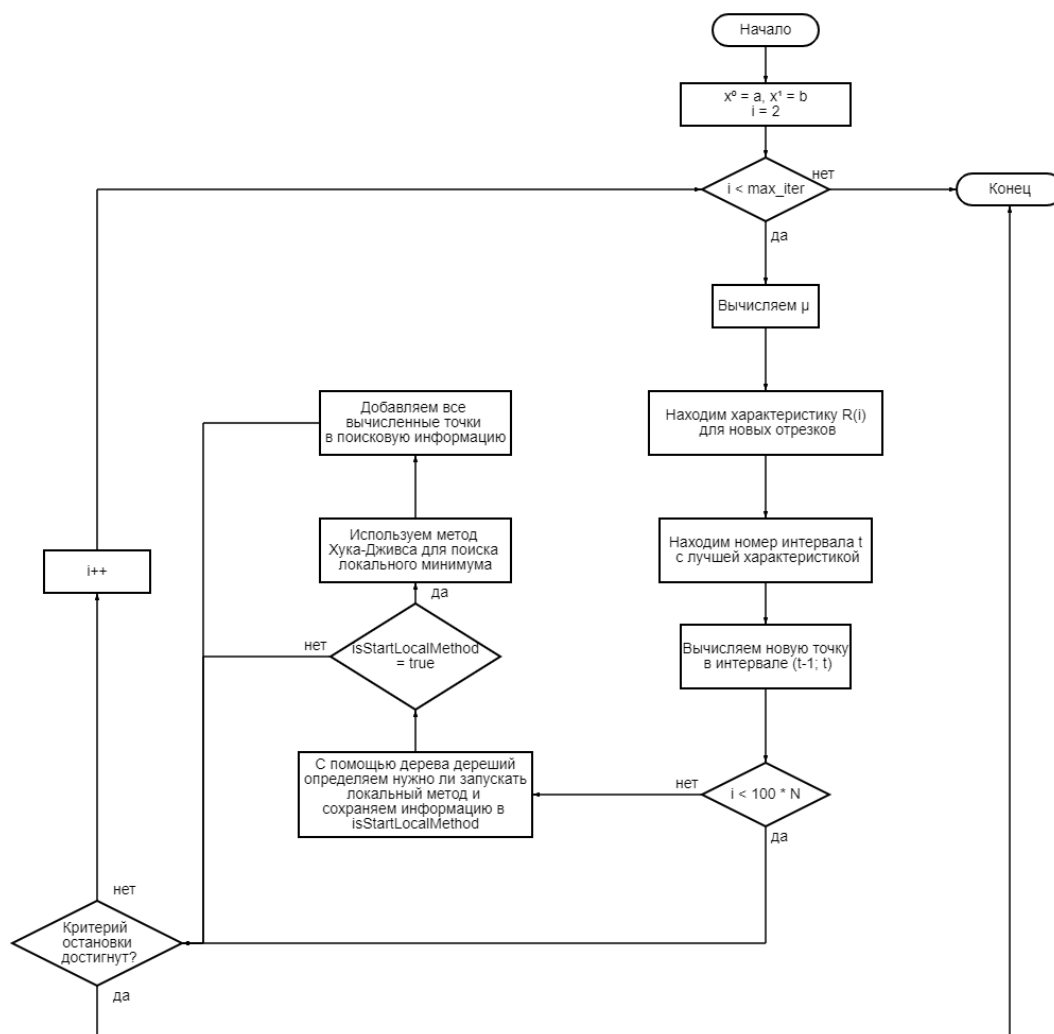


Рис. 2. Блок-схема объединения АГП и метода Хука-Дживса с использованием дерева решений

5. Эксперименты

Вычислительные эксперименты проводились на персональном компьютере со следующими характеристиками: Операционная система Windows 11, Процессор: Intel(R) Core™ i5-8250U CPU @ 1.60 GHz, 12 Gb RAM.

В работах [5, 9] описан GKLS-генератор, позволяющий порождать задачи многоэкстремальной оптимизации с заранее известными свойствами: количеством локальных минимумов, размерами их областей притяжения, точкой глобального минимума, значением функции в ней и т.п.

В таблице 1 представлено среднее количество итераций, и среднее число точек испытаний, проводимое двумя методами: алгоритмом глобального поиска (АГП) и объединением АГП и локального метода с использованием деревьев решений (Деревья решений) на задачах разной размерности (N) и разного класса (Класс задачи). Значение в скобках (если оно есть) показывает количество не решенных задач.

Таблица 1. Среднее число итераций и среднее число испытаний, проводимое разными алгоритмами

N	Класс задачи	Среднее число итераций		Среднее число точек испытаний	
		АГП	Деревья решений	АГП	Деревья решений
2	Simple	2350.0	247.4	2348.0	390.0
	Hard	4732.3	353.6	4730.3	1101.1
3	Simple	2115.5	579.5	2113.5	2441.4
	Hard	5347.4	588.1	5345.4	2532.7
4	Simple	12168.9	768.2	12166.9	3582.2
	Hard	25636.5	960.0	52634.5	5171.6
5	Simple	20633.5	978.0	20631.5	4840.4
	Hard	161094.9 (4)	1217.0	161092.9 (4)	7115.5

Итак, ускорение по числу итераций неоспоримое, причем на задачах любой размерности. Среднее число точек испытаний тоже, практически везде, уменьшилось, но эти значения могут быть сильно выше, чем среднее число итераций. Это связано с тем, что в точках испытаний учитываются и те точки, которые в процессе своей работы ставил локальный метод, а их может быть довольно много.

6. Заключение

В результате работы удалось реализовать схему объединения АГП и метода локальной оптимизации при использовании деревьев решений. Такой подход позволяет ускорить сходжение к локальному минимуму, если это целесообразно, что позволяет уменьшить общее число проводимых итераций при работе АГП. Вычислительные эксперименты подтвердили, что это дает заметное ускорение на задачах разной размерности.

Литература

1. Стронгин Р.Г., Гергель В.А., Гришагин В.А., Баркалов К.А. Параллельные вычисления в задачах глобальной оптимизации. М.: Издательство Московского университета, 2013. 280 с.
2. Шефов К. С., Степанова М. М. Реализация и применение параллельного алгоритма глобального поиска минимума к задаче оптимизации параметров молекулярно-динамического потенциала ReaxFF: URL: http://crm.ics.org.ru/uploads/crmissues/crm_2015_3/15750.pdf, 2015 (дата обращения: 13.10.2020).
3. Gergel V.P. A global optimization algorithm for multivariate functions with lipschitzian first derivatives / Gergel V.P. // Journal of Global Optimization. – 1997. – Vol. 10, No. 3. – P. 257-281.
4. Barkalov K.A. A global optimization technique with an adaptive order of checking for constraints / Barkalov K.A., Strongin R.G. // Computational Mathematics and Mathematical Physics. – 2002. – Vol. 42, No. 9. – P. 1289-1300.
5. Gaviano, M. Software for generation of classes of test functions with known local and global minima for global optimization/ M. Gaviano, D. Lera, D. E. Kvasov, Y. D. Sergeyev // ACM Transactions on Mathematical Software. – 2003. – Vol. 29. – P. 469-480.
6. Химмельблау Д. Прикладное нелинейное программирование // М.: Мир – 1975 – 536 с.
7. Nelder J., Mead R. A simplex method for function minimization // Computer Journal – 7(4) – 1965 – P. 308-313.

8. OpenCV (Open Source Computer Vision) documentation. URL: https://docs.opencv.org/4.x/dc/dd6/ml_intro.html (дата обращения: 10.07.2022).
9. Сергеев, Я.Д. Диагональные методы глобальной оптимизации / Я.Д. Сергеев, Д.Е. Квасов – М.: Физматлит, 2008. – 352 с.
10. Sysoyev, A., Barkalov, K., Sovrasov, V., Lebedev, I., Gergel, V. Globalizer – A parallel software system for solving global optimization problems. Lecture Notes in Computer Science. Volume 10421 LNCS, 2017, P. 492-499.
11. Sergeyev, Ya.D. Global search based on efficient diagonal partitions and a set of Lipschitz constants / Ya.D. Sergeyev, D.E. Kvasov // SIAM Journal on Optimization. – 2006. Vol. 16, No. 3. – P. 910–937.
12. Стронгин Р.Г. Параллельные вычисления в задачах глобальной оптимизации / Стронгин Р.Г., Гергель В.П., Гришагин В.А., Баркалов К.А. – М.: Издательство Московского университета, 2013. 280 с.

ОПТИМАЛЬНОЕ ОЦЕНИВАНИЕ МНОЖЕСТВ ДОСТИЖИМОСТИ В ЗАДАЧАХ С ПАРАМЕТРИЧЕСКОЙ НЕОПРЕДЕЛЕННОСТЬЮ

М.С. Сорокина

Нижегородский госуниверситет им. Н.И. Лобачевского

Приводится способ оптимального оценивания множеств достижимости для систем, содержащих параметрическую неопределенность. Для получения такой оценки применяется оптимальный наблюдатель.

Ключевые слова: множества достижимости, эллипсоидальные множества достижимости, оптимальный наблюдатель, матричные неравенства.

1. Введение

Рассматривается линейная нестационарная система, включающая в себя параметрическую неопределенность, при неточно известном действующем возмущении. Матричная функция, характеризующая эту неопределенность, является нестационарной и, в общем случае, может зависеть от состояния системы. Требуется определить, в какую область фазового пространства могут попасть траектории систем при наложенных ограничениях и различных возмущениях. Приведен способ оценки эллипсоидального множества достижимости с использованием дифференциального линейного матричного неравенства.

2. Оценивание множеств достижимости

Для системы, подверженной возмущениям, множество достижимости характеризует область, в которую под воздействием возмущения приходит система, и позволяет оценить точность попадания системы в заданное конечное состояние [1]. Также, при добавлении в систему управления, оно дает возможность понять, возможно ли привести систему из имеющегося состояния в заданное. Множество достижимости в данной задаче представляет собой эллипсоид с матрицей, являющейся решением линейного матричного дифференциального или разностного неравенства [2].

2.1. Постановка задачи

Рассмотрим систему линейных дифференциальных уравнений[3]:

$$\begin{cases} \dot{x} = [A(t) + F(t)\Omega(t)E(t)]x + B(t)v \\ y(t) = C(t)x(t) + D(t)v(t) \end{cases}$$

где x – состояние объекта, v – возмущение, действующее на объект, $\Omega(t)$ – матричная функция, характеризующая неопределенные параметры в системе:

$$\Omega^T(t)\Omega(t) \leq I$$

Начальное состояние системы $x(t_0) = x_0$ и возмущение удовлетворяют неравенству:

$$x_0^T R^{-1} x_0 + \int_{t_0}^T v^T(t)v(t)dt \leq 1$$

2.2. Результат

Дифференциальное матричное линейное неравенство относительно матриц $X = X^T > 0$ и $Z = XL$ (L – матрица параметров наблюдателя) и функции $\xi(t)$ для замкнутой системы принимает вид:

$$\begin{pmatrix} \dot{X} + A^T X + XA - ZC - C^T Z^T + \xi(t)E^T E & XB - ZD & XF \\ B^T X - D^T Z^T & -I & 0 \\ F^T X & 0 & -\xi(t)I \end{pmatrix} \leq 0$$

Минимизацию размера эллипсоида $x^T X x \leq 1$ обеспечим, добавив еще одно линейное матричное неравенство:

$$\begin{pmatrix} X & I \\ I & Y \end{pmatrix} \geq 0$$

где Y – матрица вспомогательного эллипсоида $x^T Y^{-1} x \leq 1$, $Y = Y^T > 0$.

Поиск оптимального наблюдателя [4] будет сводиться к следующему: нахождение матриц X, Y, Z и функции $\xi(t)$, позволяющих минимизовать след матрицы Y . После этого матрица оптимального наблюдателя будет найдена следующим образом:

$$L = X^{-1}Z$$

Множество достижимости будет найдено из следующего матричного неравенства [5]:

$$\dot{X} + Y A^T + Y A + X B B^T X + \xi^{-1} X F F^T X + \xi E^T E \leq 0, \quad X(t_0) = R^{-1}$$

Литература

1. Черноусько Ф.Л. Оценивание фазового состояния динамических систем. Метод Эллипсоидов – М.: Наука. Главная редакция физико-математической литературы. 1988. – 320 с.
2. Chernousko F.L., Ovseevich A.I. Properties of the Optimal Ellipsoid Approximating the Reachable Sets of Uncertain Systems // Journal of Optimization Theory and Applications. 2004. V. 120. No. 2. P. 223-246.0.
3. Баландин Д.В., Коган М.М. Синтез законов управления на основе линейных матричных неравенств. М: Наука, 2007.
4. Квакернаак Х., Сиван Р. Линейные оптимальные системы управления. М.: Мир, 1977.
5. Баландин Д.В., Бирюков Р.С., Коган М.М. Эллипсоидальные множества достижимости линейных нестационарных систем в задачах оценивания и управления // Дифференциальные уравнения. 2019. Т. 55, № 11. С. 1485 – 1498.

ДУОПОЛИЯ ХОТЕЛЛИНГА НА ПЛОСКОСТИ С ИНФОРМАЦИОННЫМИ РАССЫЛКАМИ В ЕВКЛИДОВОЙ МЕТРИКЕ

Е.В. Ушакова

Омский государственный университет им. Ф.М. Достоевского

Исследуется модель дуополии Хотеллинга на плоскости с информационными рассылками для двух игроков. Размещение продавцов фиксировано. Каждый из них одновременно выбирает уровень цен и уровень информационных рассылок потребителям. Потребители, расположенные внутри круга равномерно, выбирают, у кого приобрести товар, учитывая стоимость и транспортные издержки, определяемые Евклидовой метрикой. Функция рекламных расходов рассматривается в двух вариантах – квадратичной, и логарифмической.

Ключевые слова: дуополия Хотеллинга, информационные рассылки, Евклидова метрика.

Теоретико-игровые модели размещения игроков (продавцов) в городе, модели ценообразования, а также модели, изучающие эффективность рекламы, ставшие в последствии классическими, появились в начале 19 века [2], [3], [4].

Основоположителем исследований, посвященных информативной рекламе в олигополии является Баттерс [2]. В его работе моделью города является единичный отрезок. Тироль [6], Гроссман и Шапиро [3] изучали рекламу в контексте “круга” Салопа [5], в которой фирмы, равномерно расположены на дуге окружности, транспортные издержки потребителей описывались линейной функцией.

В данном исследовании фирмы располагаются на плоскости. Полное описание модели с информационными рассылками представлено в [1]. Предполагаем, что жители расположены внутри круга равномерно, с заданной плотностью распределения.

Продавцы реализуют однотипный товар. Игроки одновременно выбирают цену на свой товар и уровень рекламы – долю от общего числа потенциальных покупателей, которые получают информационные рассылки: $\psi_i, i = 1, 2$.

Следуя Баттерсу [2], рассматриваем информационную рекламу, когда продавцы сообщают потребителям стоимость товара и своё местоположение. Покупатели принимают решение о приобретении товара, только в том случае, если они получили рекламное сообщение. Если они информированы только об одной фирме, то приобретают товар у неё, если же получили сообщения от обеих фирм, то принимают решение о том, какой приобрести товар, минимизируя свои расходы, в которые входит стоимость единицы товара и транспортные расходы. В [1] рассмотрены случаи, когда транспортные расходы представлены метрикой Манхэттена и квадратичной функцией. В данной работе в качестве базовой – рассматривается Евклидова метрика.

Предполагаем, что игроки расположены в точках: $(-x_1, 0), (0, x_2)$, тогда расстояние от потенциального покупателя до фирмы определяется по формуле:

$$\rho_i(x, y) = \sqrt{(x - x_i)^2 + y^2}.$$

Функции затрат на рекламу: $A(\psi_i), i = 1, 2$.

Функции спроса потребителей: $D_1(\psi_1, \psi_2), D_2(\psi_1, \psi_2)$.

Функции прибыли фирм:

$$\begin{aligned} H_1(\psi_1, \psi_2, p_1, p_2) &= p_1 D_1(\psi_1, \psi_2) - A(\psi_1) \\ H_2(\psi_1, \psi_2, p_1, p_2) &= p_2 D_2(\psi_1, \psi_2) - A(\psi_2) \end{aligned}$$

Каждая фирма стремится максимизировать прибыль:

$$\begin{cases} \max_{p_1, \psi_1} H_1(\psi_1, \psi_2, p_1, p_2) \\ \max_{p_2, \psi_2} H_2(\psi_1, \psi_2, p_1, p_2) \end{cases}$$

В данной работе найдено равновесие для симметричного расположения игроков для логарифмической и квадратичной функции. Также найдено численное решение при различных уровнях предельных расходов на рекламу.

Литература

1. Ушакова, Е. В. Дуополия Хотеллинга на плоскости с информационными рассылками / Е. В. Ушакова // Южно-Сибирский научный вестник. – 2022. – № 2(42). – С. 66-70.
2. Butters G. (1977) Equilibrium distributions of sales and advertising prices/ Butters G // Review of Economic Studies, 44(3), 465-491.
3. Grossman G. (1984) Informative Advertising with Differentiated Products / Grossman G., Shapiro C. // Review of Economic Studies, 51, 63-81.
4. Hotelling, H. (1929) Stability in Competition // The Economic Journal. Mar. Vol. 39 (153). P. 41-57.
5. Salop S. (1979) Monopolistic Competition with Outside Goods/ Salop S. The Bell Journal of Economics, 10(1), 141-156.
6. Tirole J. The theory of industrial organization. – MIT press, 1988.

АЛГОРИТМИЧЕСКОЕ УПРАВЛЕНИЕ С ДООБСЛУЖИВАНИЕМ НЕОРДИНАРНЫМИ ПУАССОНОВСКИМИ ПОТОКАМИ МАЛОЙ ИНТЕНСИВНОСТИ

А.М. Федоткин

Нижегородский государственный университет им. Н.И. Лобачевского

В работе предложен метод построения адекватной математической модели реальной задачи управления конфликтными транспортными потоками малой интенсивности на изолированном перекрёстке. В модели рассмотрена система, которая управляет потоками неоднородных требований с помощью циклического алгоритма с продлением и дообслуживанием. При этом каждый поток является неординарным пуассоновским с заданной интенсивностью и в любой вызывающий момент по потоку может поступить одна, две или три заявки. Основные результаты работы относятся к изучению предельных свойств многомерной случайной последовательности со счётным числом состояний.

Ключевые слова: неординарный пуассоновский поток, алгоритмическое управление, экстремальная стратегия обслуживания, случайная последовательность, предельное распределение.

В работах [1-2] исследована математическая модель реальной системы циклического управления конфликтными транспортными потоками малой интенсивности на изолированном перекрёстке. Рассмотрим теперь систему управления конфликтными неординарными пуассоновскими потоками Π_1 и Π_2 [2] с помощью циклического алгоритма с продлением и дообслуживанием. При $j = 1, 2$ поток Π_j является неординарным пуассоновским с интенсивностью λ_j для вызывающих моментов. В каждый вызывающий момент по потоку Π_j малой интенсивности может поступить одна, две или три заявки соответственно с вероятностью p_j , q_j или s_j , где $p_j + q_j + s_j = 1$. Обозначим через $\eta_j(t)$ число автомобилей, поступающих в систему массового обслуживания за промежуток времени $[0, t)$. Модель потока $\{\eta_j(t): t \geq 0\}$ – это неординарный пуассоновский поток, для которого [2] при $m = 0, 1, 2, \dots$ Вероятности

$$\varphi_j(m; t) = \mathbf{P}(\eta_j(t) = m) = e^{-\lambda_j t} \sum_{g=0}^{\lfloor \frac{m}{2} \rfloor} \sum_{h=0}^{\lfloor \frac{m-2g}{3} \rfloor} p_j^{m-2g-3h} q_j^g s_j^h \frac{(\lambda_j t)^{m-g-2h}}{g!h!(m-2g-3h)!}. \quad (1)$$

Соотношение (1) позволяет найти основные числовые характеристики случайной величины $\eta(t)$.

Лемма 1. *Справедливы равенства*

$$\begin{aligned} M\eta_j(t) &= \lambda_j t (2s_j + q_j + 1), D\eta_j(t) = \lambda_j t (8s_j + 3q_j + 1), \\ Ka\eta_j(t) &= \frac{26s_j + 7q_j + 1}{\sqrt{\lambda_j t} (8s_j + 3q_j + 1)^{3/2}}, Ea\eta_j(t) = \frac{(80s_j + 15q_j + 1)}{\lambda_j t (8s_j + 3q_j + 1)^2} \end{aligned}$$

где $M\eta_j(t)$ – математическое ожидание, $D\eta_j(t)$ – дисперсия, $Ka\eta_j(t)$ – коэффициент асимметрии и $Ea\eta_j(t)$ – эксцесс случайной величины $\eta(t)$ для всех $t > 0$.

Множество состояний обслуживающего устройства, реализующего циклический алгоритм с продлением и дообслуживанием, будем обозначать через $\Gamma = \{\Gamma^{(1)}, \Gamma^{(2)}, \Gamma^{(3)}, \Gamma^{(4)}\}$, где минимальная длительность каждого состояния $\Gamma^{(e)}$ при $e \in M = \{1, 2, 3, 4\}$ равна T_e . При состоянии $\Gamma^{(2j-1)}$ обслуживается только поток Π_j с пропускной способностью $l_j = \lfloor \hat{\mu}_j T_{2j-1} \rfloor$, где $\hat{\mu}_j$ – интенсивность обслуживания заявок этого потока. Состояние $\Gamma^{(2j)}$, $j = 1, 2$ соответствует процессу переналадки обслуживающего устройства после обслуживания соответствующего потока. В это время дообслуживаются только заявки потока Π_j , которые уже начали обслуживаться в

предыдущем состоянии. Считаем, что $T_{2j-1} > T_{2j}$. Пропускную способность в состоянии $\Gamma^{(2j)}$ обозначаем $l_j' = [\hat{\mu}_j T_{2j}]$, $l_j' \leq l_j$. Смена текущего состояния обслуживающего устройства или его продление принимается в случайные моменты времени $\tau_0, \tau_1, \dots$. Пусть случайный элемент $\Gamma_i \in \Gamma$ есть состояние обслуживающего устройства на промежутке времени $[\tau_i, \tau_{i+1})$ при $i \in I = \{0, 1, \dots\}$. Будем предполагать, что $\tau_{i+1} = \tau_i + f(\Gamma_i)$, $i \in I$, где функция $f(\Gamma^{(e)}) = T_e$, $e \in M$. Так как потоки Π_1 и Π_2 не могут обслуживаться одновременно на любом промежутке $[\tau_i, \tau_{i+1})$, то они являются конфликтными [1].

На содержательном уровне процесс обслуживания требований и управление потоками можно описать следующим образом. Обозначим через $h_1 \in \{0, 1, \dots, l_1 - 1\}$ величину порога потока Π_1 и через $h_2 \in \{0, 1, \dots, l_2 - 1\}$ величину порога потока Π_2 . Обслуживание первого потока происходит в состоянии $\Gamma^{(1)}$, при этом второй поток не обслуживается. Если $\alpha_{2,i} + \eta_{2,i}$ не достигнет величины h_2 к концу текущего интервала $[\tau_i, \tau_{i+1})$, то обслуживающее устройство продлевает состояние $\Gamma^{(1)}$ на такой же промежуток времени, иначе переходит в следующее состояние $\Gamma^{(2)}$.

Пусть теперь обслуживание второго потока происходит в состоянии $\Gamma^{(3)}$, при этом первый поток не обслуживается. Аналогично, если по первому потоку Π_1 величина очереди меньше порогового значения h_1 , то обслуживающее устройство продлевает состояние $\Gamma^{(3)}$ на такой же промежуток времени, иначе переходит в состояние $\Gamma^{(4)}$. Из остальных состояний обслуживающее устройство может перейти только в следующее состояние, то есть из $\Gamma^{(2)}$ в $\Gamma^{(3)}$ и из $\Gamma^{(4)}$ в $\Gamma^{(1)}$. Состояния $\Gamma^{(2)}$ и $\Gamma^{(4)}$ соответствуют периоду дообслуживания только заявок первого потока и только заявок второго потока. При $h_1 = 0$ и $h_2 = 0$ приведенный выше алгоритм управления превращается в циклический на множестве состояний $\{\Gamma^{(1)}, \Gamma^{(2)}, \Gamma^{(3)}, \Gamma^{(4)}\}$. Величина порога h_j ограничена сверху пропускной способностью l_j . Это сделано потому, что в случае выполнения неравенства $h_j \geq l_j$ система не сможет справиться с обслуживанием потока Π_j . Тогда по этому потоку начнет возрастать количество требований в очереди и управление будет неэффективно.

Введем теперь следующие случайные величины, с помощью которых обозначим характеристики конфликтной управляемой системы обслуживания на промежутках времени $[\tau_i, \tau_{i+1})$, $i \in I = \{0, 1, \dots\}$:

– $\eta_{j,i} \in X = \{0, 1, \dots\}$ – количество поступивших в систему требований потока Π_j за промежутки времени $[\tau_i, \tau_{i+1})$;

– $\alpha_{j,i} \in X$ – количество находящихся в очереди ожидания требований потока Π_j в момент времени τ_i ;

– $\xi_{j,i} \in \{0, l_j', l_j\}$ – максимальное количество требований потока Π_j , которое может быть обслужено за промежуток времени $[\tau_i, \tau_{i+1})$;

– $\xi'_{j,i} \in Y_j = \{0, 1, \dots, l_j\}$ – реально обслуженное количество требований потока Π_j за промежуток времени $[\tau_i, \tau_{i+1})$.

Заметим, что соотношение (1) и определение случайной величины $\eta_{j,i}$ позволяют вычислить следующие условные вероятности:

$$\mathbf{P}(\eta_{j,i} = m | \Gamma_i = \Gamma^{(e)}) = \varphi_j(m; T_e), \text{ при } m = 0, 1, 2, \dots \text{ и } T_e \in \{T_1, T_2, T_3, T_4\}. \quad (3)$$

Условные вероятности $\mathbf{P}(\xi_{j,i} = b | \Gamma_i = \Gamma^{(e)}) = \beta_j(b; \Gamma^{(e)})$ для случайной величины $\xi_{j,i}$ будем задавать следующим равенством:

$$\mathbf{P}(\xi_{j,i} = b | \Gamma_i = \Gamma^{(e)}) = \beta_j(b; \Gamma^{(e)}) = \begin{cases} 1 \text{ при } b = 0, e \in M \setminus \{2j-1, 2j\}, j \in \{1, 2\}; \\ 1 \text{ при } b = l_j, e = 2j-1, j \in \{1, 2\}; \\ 1 \text{ при } b = l_j', e = 2j, j \in \{1, 2\}; \\ 0, \text{ в остальных случаях.} \end{cases} \quad (4)$$

При $\Gamma^{(e)} \in \Gamma$, $x_1 \in X$, $x_2 \in X$, $m_1 \in X$, $m_2 \in X$ определим теперь функцию

$$U(\Gamma^{(e)}, x_1, x_2, m_1, m_2) = \begin{cases} \Gamma^{(e+1)}, e = 2; \\ \Gamma^{(1)}, e = 4; \\ \Gamma^{(1)}, e = 1, x_2 + m_2 < h_2; \\ \Gamma^{(2)}, e = 1, x_2 + m_2 \geq h_2; \\ \Gamma^{(3)}, e = 3, x_1 + m_1 < h_1; \\ \Gamma^{(4)}, e = 3, x_1 + m_1 \geq h_1. \end{cases}$$

Вид этой функции и определения случайных величин $\alpha_{j,i}$, $\eta_{j,i}$, h_j , позволяют отобразить на рисунке 1 граф переходов состояний обслуживающего устройства (ОУ).

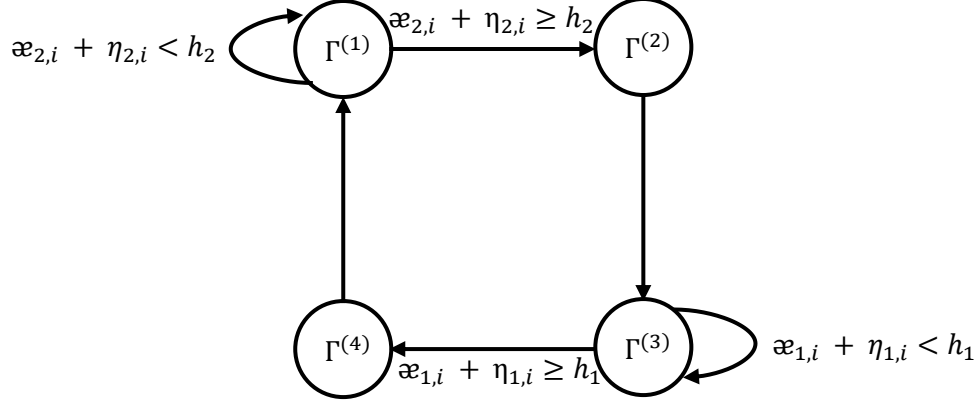


Рис.1. Граф переходов состояний ОУ

Пусть применяется экстремальная стратегия [3], при которой реально обслуживается максимально возможное число требований. В этом случае при $i \in \{0, 1, \dots\}$ имеем равенство

$$\xi'_{j,i} = W_j(\Gamma_i, \alpha_{j,i}, \eta_{j,i}) = \min\{\alpha_{j,i} + \eta_{j,i}, \xi_{j,i}\}, \text{ если } \Gamma_i = \Gamma^{(2j-1)}, \quad (5)$$

и равенство вида

$$\xi'_{j,i} = W_j(\Gamma_i, \alpha_{j,i}, \eta_{j,i}) = \min\{\alpha_{j,i}, \xi_{j,i}\}, \text{ если } \Gamma_i = \Gamma^{(2j)}. \quad (6)$$

Так как объём накопителей очередей неограничен, то при $i \in \{0, 1, \dots\}$

$$\alpha_{j,i+1} = \alpha_{j,i} + \eta_{j,i} - \xi'_{j,i}. \quad (7)$$

Используя равенства (5) и (6) для $\xi'_{j,i}$ непосредственно из (7) окончательно получаем, что динамика величины очереди по потоку Π_j определяется соотношением

$$\alpha_{j,i+1} = V_j(\Gamma_i, \alpha_{j,i}, \eta_{j,i}, \xi_{j,i}) = \max\{0, \alpha_{j,i} + \eta_{j,i} - \xi_{j,i}\}, \text{ если } \Gamma_i = \Gamma^{(2j-1)}, \quad (8)$$

и соотношением

$$\alpha_{j,i+1} = V_j(\Gamma_i, \alpha_{j,i}, \eta_{j,i}, \xi_{j,i}) = \eta_{j,i} + \max\{0, \alpha_{j,i} - \xi_{j,i}\}, \text{ если } \Gamma_i = \Gamma^{(2j)}. \quad (9)$$

Математической моделью система управления конфликтными неординарными пуассоновскими потоками Π_1 и Π_2 малой интенсивности с помощью циклического алгоритма с продлением и дообслуживанием является многомерная случайная последовательность вида

$$\{(\Gamma_i, \alpha_{1,i}, \alpha_{2,i}, \xi'_{1,i-1}, \xi'_{2,i-1}), i \in I\},$$

где $\Gamma_{i+1} = U(\Gamma_i, \alpha_{1,i}, \alpha_{2,i}, \eta_{1,i}, \eta_{2,i})$, $\alpha_{j,i+1} = V_j(\Gamma_i, \alpha_{j,i}, \eta_{j,i}, \xi_{j,i})$, $\xi'_{j,i} = W_j(\Gamma_i, \alpha_{j,i}, \eta_{j,i})$ при $j = 1, 2$.

Справедлива следующая теорема.

Теорема 1. Многомерная случайная последовательность

$$\{(\Gamma_i, \alpha_{1,i}, \alpha_{2,i}, \xi'_{1,i-1}, \xi'_{2,i-1}), i \in I\}$$

для которой выполняются приведенные выше соотношения (1) – (9), с заданным начальным распределением вектора $(\Gamma_0, \alpha_{1,0}, \alpha_{2,0}, \xi'_{1,-1}, \xi'_{2,-1})$ на пространстве $\Gamma \times X \times X \times Y_1 \times Y_2$ является однородной многомерной марковской цепью.

Литература

1. Федоткин А. М., Федоткин М. А. Анализ и оптимизация выходных процессов при циклическом управлении конфликтными транспортными потоками Гнеденко-Коваленко. Автоматика и телемеханика. РАН, 2009, № 12, с. 92 – 108.
2. Федоткин М. А., Федоткин А. М., Кудрявцев Е. В. Динамические модели неоднородного потока транспорта на магистралях. Автоматика и телемеханика. РАН, 2020, № 8, с. 149 – 164.

СТАБИЛИЗАЦИЯ ПО ИЗМЕРЯЕМОМУ ВЫХОДУ ПЕРЕВЕРНУТОГО МАЯТНИКА ПРИ ОГРАНИЧЕНИЯХ НА ФАЗОВЫЕ ПЕРЕМЕННЫЕ

А.А. Федюков

Нижегородский государственный университет им. Н.И. Лобачевского

Рассматривается задача о поиске закона управления по измеряемому выходу, который обеспечивает стабилизацию линейной динамической системы и выполнение заданных ограничений на фазовые переменные. Подход к решению задачи основан на применении метода квадратичных функций Ляпунова и аппарата линейных матричных неравенств. Получены достаточные условия для нахождения матрицы параметров обратной связи. Приведен подход, который позволяет получить оценку для области притяжения состояния равновесия замкнутой системы.

Ключевые слова: стабилизация, линейные матричные неравенства, оценка для области притяжения.

1. Введение

В работе [1] рассмотрена и решена задача поиска управления в виде линейной обратной связи по состоянию, которое обеспечивает стабилизацию линейной динамической системы при ограничениях на фазовые и управляющие переменные. В работе [2] рассмотрен вопрос о построении оценки для области притяжения состояния равновесия замкнутой управлением системы. Но в реальной ситуации полная информация о состоянии системы обычно недоступна измерению. Как правило, доступно измерению лишь часть фазовых переменных или их линейная комбинация. В связи с этим возникает нетривиальная задача стабилизации линейных динамических систем с помощью управления по измеряемому выходу.

Задача стабилизации линейной динамической системы с помощью управления в виде линейной обратной связи по измеряемому выходу является одной из важных и сложных теоретических проблем в теории управления [3–5]. Актуальность и важность этой задачи связана с ее практической значимостью. Вместе с тем до настоящего времени не найден алгоритм решения, который позволил бы однозначно найти матрицу параметров закона управления или однозначно определить, что такой матрицы не существует. Это связано с тем, что решение задачи поиска закона управления по измеряемому выходу сводится к решению систем билинейных матричных неравенств. Как отмечают авторы в работе [6], проблема решения билинейных матричных неравенств, в отличие от решения системы линейных матричных неравенств, относится к числу NP-трудных задач.

Заметим, что более сложным является случай, когда закон управления по измеряемому выходу, должен обеспечить не только асимптотическую устойчивость состояния равновесия замкнутой системы, но и обеспечить выполнение заданных требований к функционированию системы. Например, ограничений на фазовые переменные объекта. При этом актуальным остается вопрос об области притяжения состояния равновесия замкнутой управлением системы при ограничениях на фазовые переменные.

В данной работе решена задача стабилизации с помощью управления по измеряемому выходу неустойчивого объекта – перевернутого маятника, при ограничениях на фазовые переменные. В терминах линейных матричных неравенств получены достаточные условия для нахождения матрицы параметров обратной связи. Приведен подход, который позволяет получить оценку для области притяжения состояния равновесия замкнутой системы. Численные эксперименты, проведенные в пакете для инженерных расчетов Matlab [7], согласуются с теоретическими результатами.

2. Постановка задачи

Для линейного управляемого объекта с не измеряемым состоянием

$$\dot{x} = Ax + Bu, \quad x(0) = x_0, \quad (1)$$

$$z_i = C_i x, \quad i = \overline{1, N},$$

$$y = Cx, \quad (2)$$

требуется построить управление в виде линейной обратной связи по измеряемому выходу

$$u = \Theta y, \quad (3)$$

которое будет обеспечивать асимптотическую устойчивость замкнутой системы (1)-(3) и выполнение при заданных значениях γ_i ограничений

$$\max_{t \geq 0} |z_i(t)| \leq \gamma_i, \quad i = \overline{1, N}. \quad (4)$$

Здесь $x \in R^n$ – состояние системы, $u \in R^l$ – управление, $l < n$, $z_i \in R^{m_i}$ – управляемые выходы системы, $y \in R^k$ – измеряемый выход системы, $k < n$; A , B , C и C_i – заданные матрицы соответствующих размеров. Предполагаем, что пара (A, B) – стабилизируема, а пара (A, C) – детектируема.

3. Достаточные условия существования закона управления

Получены достаточные условия для нахождения матрицы параметров обратной связи.

Теорема 1. Пусть взаимнообратные матрицы $X = X^T > 0$, $Y = Y^T > 0$ и величины $\gamma_i > 0$, $i = \overline{1, N}$ удовлетворяют системе линейных матричных неравенств

$$W_{B^T}^T (YA^T + AY) W_{B^T} < 0, \quad W_C^T (A^T X + XA) W_C < 0,$$

$$C_i^T C_i \leq \gamma_i^2 X, \quad i = \overline{1, N},$$

где $B^T W_{B^T} = 0$, $C W_C = 0$. Тогда все траектории замкнутой системы (1)-(3) с начальными условиями $x(0) \in E(X)$, $E(X) = \{x: x^T X x \leq 1\}$ удовлетворяют ограничениям (4). Матрицу параметров закона управления для линейной динамической системы с ограничениями на фазовые переменные находим как решение относительно неизвестной матрицы Θ линейного матричного неравенства

$$\Psi + P^T \Theta^T Q + Q^T \Theta P < 0,$$

где $\Psi = YA^T + AY$, $P = CY$, $Q = B^T$.

4. Оценка для области притяжения

Приведен подход, который позволяет уточнить множество начальных состояний, для которых полученный закон управления стабилизирует замкнутую систему при ограничениях $\max_{t \geq 0} |z_i(t)| \leq \gamma_i$, $i = \overline{1, N}$. Доказана теорема.

Теорема 2. Пусть матрица $X_i = X_i^T > 0$ удовлетворяет системе линейных матричных неравенств

$$A^T X_i + X_i A + C^T \Theta^T B^T X_i + X_i B \Theta C < 0, \quad C_i^T C_i \leq \gamma_i^2 X_i.$$

Тогда все траектории замкнутой системы с начальными условиями $x(0) \in E(X_i)$, $E(X_i) = \{x: x^T X_i x \leq 1\}$ удовлетворяют ограничению $\max_{t \geq 0} |z_i(t)| \leq \gamma_i$.

Показано, что если каждая из матриц X_i , $i = \overline{1, N}$, имеет минимальный след и является решением данной системы линейных матричных неравенств для значения γ_i , тогда множество

$\bigcap_{i=1}^N E(X_i)$ является оценкой для области притяжения состояния равновесия замкнутой системы при ограничениях $\max_{t \geq 0} |z_i(t)| \leq \gamma_i, i = \overline{1, N}$.

5. Результаты численного моделирования

Рассмотрим управляемый перевернутый маятник

$$\ddot{\varphi} - \varphi = u \quad (5)$$

при ограничении на φ – угол отклонения звена маятника от вертикали

$$\max_{t \geq 0} |\varphi(t)| \leq 0.1. \quad (6)$$

Предполагаем, что φ – угол отклонения звена маятника, и $\dot{\varphi}$ – скорость изменения угла маятника измерению не доступны, но доступна измерению их линейная комбинация

$$y = \varphi + 10\dot{\varphi}.$$

Численное решение получено в пакете Matlab. Найдено управление

$$u = -49.9677y, \quad (7)$$

которое обеспечивает асимптотическую устойчивость замкнутой системы (5), (7) и выполнение ограничения на фазовую переменную (6).

Литература

1. Баландин Д.В., Коган М.М. Синтез линейных законов управления при фазовых ограничениях // Автоматика и телемеханика, 2009. № 6. С 48-57.
2. Баландин Д.В., Федюков А.А. Стабилизация линейных динамических объектов по измеряемому с ошибкой состоянию при ограничениях на фазовые и управляющие переменные // Известия Российской академии наук. Теория и системы управления, 2021. № 5. С. 5–17.
3. Баландин Д.В., Коган М.М. Синтез законов управления на основе линейных матричных неравенств. – М.: Физматлит, 2007.
4. Syrmos, V. L., Abdallah C.T., Dorato P., Grigoriadis K. Static Output Feedback.: A Survey // Automatica, 1997. V.33, No. 2. P. 125-137. DOI: 10.1016/S0005-1098(96)00141-0.
5. Sadabadi, M., Peaucelle D. From Static Output Feedback to Structured Robust Static Output Feedback: A Survey // Annual Reviews in Control, Elsevier. 2016. V.42. P. 11-26. DOI: 10.1016/j.arcontrol.2016.09.014.
6. Toker O., Ozbay H. On the NP-hardness of solving bilinear matrix inequalities and simultaneous stabilization with static output feedback // In IEEE American Control Conference. Seattle, Washington: Institute of Electrical and Electronics Engineers, 1995. P. 2525-2526. DOI: 10.1109/ACC.1995.532300.
7. Gahinet P., Nemirovski A., Laub A.J., Chilali M. The LMI Control Toolbox. For Use with Matlab. User's Guide. Natick, MA: The MathWorks, Inc. 1995. 227 p.

ДИАГНОСТИКА ПЕРЕМЕЖАЮЩЕЙСЯ ФАЗОВОЙ СИНХРОНИЗАЦИИ С УЧЕТОМ МУЛЬТИСТАБИЛЬНОСТИ В АНСАМБЛЕ ОДНОНАПРАВЛЕННО СВЯЗАННЫХ СИСТЕМ РЕССЛЕРА^{1*}

В.А. Ханадеев, А.Д. Плотникова, О.И. Москаленко

*Саратовский национальный исследовательский госуниверситет им. Н.Г. Чернышевского,
Региональный научно-образовательный математический центр
«Математика технологий будущего»*

Исследована возможность применения количественной меры мультистабильности для диагностики перемежаемости на границе фазовой синхронизации. Предложен метод выделения характерных фаз поведения связанных систем, основанный на использовании меры мультистабильности. На примере однонаправленно связанных систем Ресслера, находящихся в режиме ленточного хаоса, показана его эффективность.

Ключевые слова: мультистабильность, перемежаемость, фазовая синхронизация, однонаправленно связанные системы, ламинарные фазы поведения, вероятность наблюдения ламинарной фазы.

1. Введение

Перемежаемость является распространенным явлением как в физических, так и биологических системах [1]. Данное явление подразумевает чередование двух различных режимов и может наблюдаться на границах различных типов хаотической синхронизации. Различают перемежающиеся полную синхронизацию, синхронизацию с запаздыванием, обобщенную синхронизацию, фазовую синхронизацию, синхронизацию временных масштабов и т.д. [2-6]. Для диагностики перемежающегося поведения, в частности, на границе обобщенной синхронизации, традиционно используют метод вспомогательной системы [4,7]. Однако недавно было обнаружено, что перемежающееся поведение на границе этого режима может зависеть от выбора начальных условий для ведомой и вспомогательной систем, т.е. наблюдается мультистабильность [8].

Так как различные типы синхронного поведения, наблюдающиеся в однонаправленно и взаимно связанных хаотических системах, являются частными проявлениями режима синхронизации временных масштабов [9], можно предположить, что мультистабильность будет иметь место и на границах других типов хаотической синхронизации. В настоящей работе предложен метод диагностики перемежающегося поведения на границе фазовой синхронизации с учетом наличия в данном случае мультистабильности. Метод основан на рассмотрении относительно большого ансамбля связанных систем и вычислении вероятности наблюдения ламинарной фазы поведения в зависимости от времени.

2. Объект и методы исследования

В качестве объекта исследования выбран ансамбль однонаправленно связанных систем Ресслера, находящихся в режиме ленточного хаоса:

^{1*} Работа выполнена при финансовой поддержке Совета по грантам Президента Российской Федерации для государственной поддержки молодых российских ученых – докторов наук (проект № МД-18.2022.1.2).

$$\begin{aligned}
\dot{x}_d^i &= -\omega_d y_d^i - z_d^i \\
\dot{y}_d^i &= \omega_d x_d^i + a y_d^i \\
\dot{z}_d^i &= p + z_d^i (x_d^i - c) \\
\dot{x}_r^i &= -\omega_r y_r^i - z_r^i + \varepsilon (x_d^i - x_r^i) \\
\dot{y}_r^i &= \omega_r x_r^i + a y_r^i \\
\dot{z}_r^i &= p + z_r^i (x_r^i - c),
\end{aligned} \tag{1}$$

где индексы d и r относятся к ведущей и ведомой системам, соответственно, индекс $i = 1 \dots N$ соответствует рассматриваемой паре осцилляторов, $a = 0.15$, $p = 0.2$, $c = 10$, $\omega_d = 0.93$ и $\omega_r = 0.95$ – управляющие параметры, значения которых выбраны фиксированными, что соответствует случаю относительно слабой расстройки между взаимодействующими системами, ε – параметр связи. Начальные условия для ведущей системы всегда выбирались фиксированными, а для ведомых систем – всегда различными. Как показали проведенные расчеты, независимо от выбора начальных условий ведомых систем при выбранных значениях управляющих параметров в случае относительно малой расстройки режим фазовой синхронизации в системе (1) наблюдается при $\varepsilon \geq 0.041$, в случае относительно большой расстройки – при $\varepsilon \geq 0.118$.

Для диагностики фазовой синхронизации в системе (1) были введены в рассмотрение фазы хаотических сигналов и найдены разности между фазами каждой пары взаимодействующих систем. Фазы вводились в рассмотрение как углы в полярной системе координат на плоскости (x^i, y^i) [2], а разность $\Delta\varphi^i$ между ними была сведена к диапазону шириной 2π , аналогично тому, как это делалось в работе [10]. По полученным разностям были рассчитаны временные зависимости вероятности наблюдения ламинарной фазы (фазы синхронного поведения). Расчет производился по формуле

$$p = \frac{m}{N}, \tag{2}$$

где m – число пар, в которых диагностируются ламинарные фазы, N – общее число пар осцилляторов в ансамбле.

3. Результаты исследования

На рисунке 1 приведены результаты расчёта вероятности нахождения ламинарных фаз от времени для случая малой расстройки для $N=100$ систем (1). Также на рисунке 1 приведены временные зависимости разности фаз в том же временном промежутке. Видно, что по полученным данным в обоих случаях можно легко определить начало и конец ламинарных фаз. В случае временной зависимости разности фаз ламинарное поведение наблюдается в диапазоне $[-4; 0]$, а турбулентное – когда значения разности фаз находятся за пределами этого диапазона. При рассмотрении количественной характеристики степени мультистабильности ламинарная фаза присутствует при вероятности близкой к 1, а турбулентная фаза – при резком снижении значения вероятности до величин меньше 1. Очевидно, что минимальные значения вероятности приходятся на середину турбулентной фазы, а при приближении или удалении вероятность начинает расти и уменьшаться, соответственно.

Подобные зависимости позволяют рассчитать количественные характеристики перемежаемости, распределения длительностей ламинарных фаз при фиксированных значениях управляющих параметров и зависимости средних длительностей ламинарных фаз от параметра связи, для режима перемежающейся фазовой синхронизации. Как показывают произведенные расчеты, вышеуказанные характеристики, рассчитанные обоими методами, с высокой степенью точности совпадают друг с другом, что позволяет говорить о независимости характеристик перемежаемости от учета наличия мультистабильности вблизи границы фазовой синхронизации.

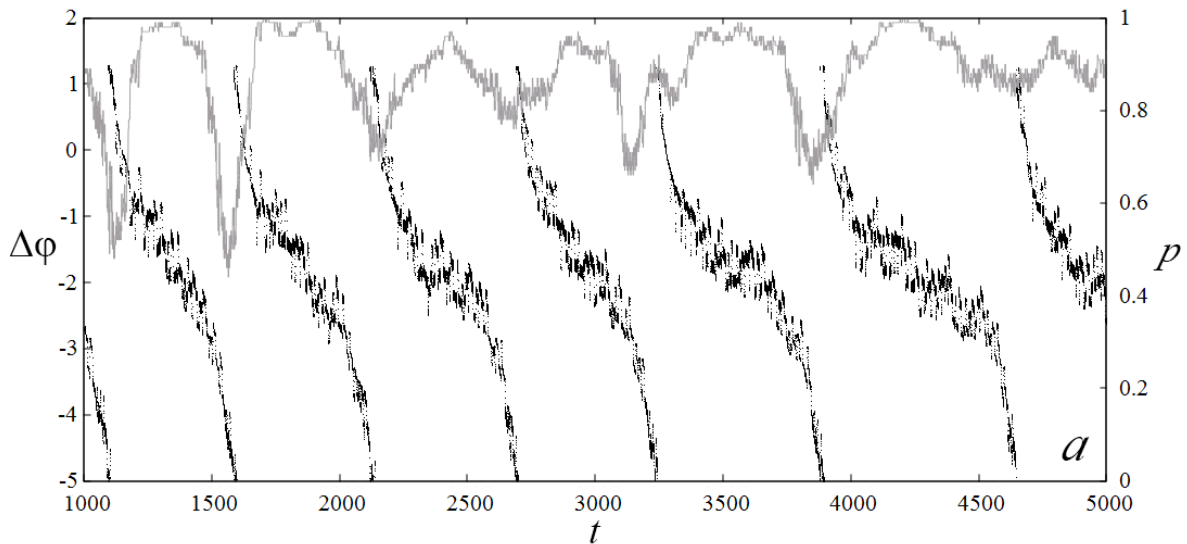


Рис. 1. Временные зависимости разности фаз $\Delta\varphi$ (черные кривые) одной пары систем Ресслера и вероятности нахождения ламинарных фаз p (серые линии) для ансамбля однонаправленно связанных систем Ресслера при малой расстройке $\omega_d = 0.93$ и силе связи $\varepsilon = 0.03$

4. Заключение

Таким образом, в настоящей работе введена в рассмотрение количественная характеристика степени мультистабильности. На примере ансамбля однонаправленно связанных систем Ресслера показана возможность определения участков ламинарных и турбулентных фаз с помощью предложенного метода. Результаты, полученные на основе анализа временных зависимостей разности фаз и вероятности нахождения ламинарных фаз, оказались схожими друг с другом.

Несмотря на то, что в рамках данной работы результаты проиллюстрированы только на одной системе, можно ожидать, что они будут справедливы и для других систем, способных демонстрировать перемежающееся поведение вблизи границы фазовой синхронизации.

Литература

1. Берже П., Помо И., Видаль К. Порядок в хаосе. Мир, М., 1991.
2. Pikovsky A.S., Osipov G.V., Rosenblum M.G. et al. Attractor-repeller collision and eyelet intermittency at the transition to phase synchronization // Phys. Rev. Lett. 1997. Vol. 79, No. 1. P. 47. DOI: <https://doi.org/10.1103/PhysRevLett.79.47>.
3. Boccaletti S., Valladares D.L. Characterization of intermittent lag synchronization // Phys. Rev. E. 2000. Vol. 62, No. 5. P. 7497. DOI: <https://doi.org/10.1103/PhysRevE.62.7497>.
4. Hramov A.E., Koronovskii A.A. Intermittent generalized synchronization in unidirectionally coupled chaotic oscillators // Europhys. Lett. 2005. Vol. 70, No. 2. P. 169. DOI: [10.1209/epl/i2004-10488-6](https://doi.org/10.1209/epl/i2004-10488-6).
5. Hramov A.E., Koronovskii A.A., Kurovskaya M.K. et al. Ring Intermittency in Coupled Chaotic Oscillators at the Boundary of Phase Synchronization // Phys. Rev. Lett. 2006. Vol. 97. P. 114101. DOI: <https://doi.org/10.1103/PhysRevLett.97.114101>.
6. Zhuravlev M.O., Koronovskii A.A., Moskalenko O.I. et al. Ring intermittency near the boundary of the synchronous time scales of chaotic oscillators // Phys. Rev. E. 2011. Vol. 83. P. 027201. DOI: <https://doi.org/10.1103/PhysRevE.83.027201>.
7. Abarbanel H.D.I., Rulkov N.F., Sushchik M.M. Generalized synchronization of chaos: The auxiliary system approach // Phys. Rev. E. 1996. Vol. 53, iss. 5. P. 4528-4535. DOI: [10.1103/PhysRevE.53.4528](https://doi.org/10.1103/PhysRevE.53.4528). DOI: <https://doi.org/10.1103/PhysRevE.53.4528>.

8. Moskalenko O.I., Koronovskii A.A., Selskii A.A. et al. On multistability near the boundary of generalized synchronization in unidirectionally coupled chaotic systems // CHAOS. 2021. Vol. 31, No. 8. P. 083106. DOI: <https://doi.org/10.1063/5.0055302>.
9. Hramov A.E., Koronovskii A.A. An approach to chaotic synchronization // CHAOS. 2004. Vol. 14, No. 3. P. 603. DOI:10.1063/1.1775991.
10. Журавлев М.О., Куровская М.К., Москаленко О.И. Метод выделения ламинарных и турбулентных фаз в перемежающихся временных реализациях систем, находящихся вблизи границы фазовой синхронизации // Письма в ЖТФ. 2010. Т. 36, № 10. С. 31-38.

ОБРАБОТКА АНАЛОГОВОЙ ИНФОРМАЦИИ СПАЙКОВОЙ НЕЙРОН-АСТРОЦИТАРНОЙ СЕТЬЮ^{1*}

Ю.А. Цыбина^{1,2}, И.А. Кастальский^{1,3,4}, А.А. Заикин^{1,2,5}, С.Ю. Гордлеева^{1,6}

¹ННГУ им. Н.И. Лобачевского,

²ФГАОУ ВО Первый МГМУ им. И.М. Сеченова,

³Институт прикладной физики РАН,

⁴ФГБОУ ВО СамГМУ Минздрава России

⁵University College London, United Kingdom,

⁶Университет Иннополис

Обработка информации с помощью спайковых нейронных сетей (SNN) является одним из важнейших приложений современной нейронауки. Известно, что в живом мозге астроциты влияют на синаптическую передачу, высвобождая глутрансмиттеры, которые модулируют возбудимость нейронов и, следовательно, скорость их активации. В отличие от электрических спайков, форма которых неизменна и определяется свойствами самого нейрона, реакция астроцита в ответ на синхронную активность нейронных популяций пропорциональна входному стимулу. В представленном исследовании данное свойство астроцитов было использовано в качестве механизма обработки полутонковых изображений. Согласно полученным результатам, разработанная модель SNN, двунаправленно взаимодействующая с сетью астроцитов, позволяет воспроизводить даже сильно зашумленные изображения, закодированные в уровнях активации астроцитов в течение нескольких секунд. Данный эффект памяти обусловлен только лишь действием астроцитов, а время хранения определяется характерным временным масштабом их активации.

Ключевые слова: астроцит, нейронная сеть, нейрон-астроцитарное взаимодействие, обработка сигналов

1. Введение

Интерес исследователей к моделированию нейронных сетей постоянно развивается в сторону биологической правдоподобности. Считается, что математические модели, которые учитывают сложные взаимодействия между нейронами и астроцитами для обработки информации в мозге, в перспективе будут определять будущие тенденции в области вычислительных нейронаук.

В детализированных моделях астроцитов было показано, что они могут иметь как регулярную, так и хаотическую динамику [1], при этом они также модулируют синаптическую передачу между нейронами. Действительно, только за последние несколько лет накопилось достаточно доказательств того, что астроциты активно участвуют в различных когнитивных функциях в норме, таких как: восприятие сенсорных стимулов [2,3], обработка информации [4], формирование памяти [5], а также в некоторых патологических процессах [6,7]. Различные исследования показали, что увеличение внутриклеточной концентрации ионов кальция (Ca^{2+}) в астроцитах коры возникает в ответ на стимулы [8,9]. Эти ответные сигналы возникают с задержкой в несколько секунд по сравнению с активностью окружающих нейронов [9]. Плавная зависимость Ca^{2+} активности от входных сигналов нейронной сети позволяет астроцитам в разной степени участвовать в регуляции синаптической передачи.

В данной работе механизм астроцитарной регуляции активности SNN применяется для кодирования визуальных стимулов и распознавания полутонковых изображений. Был выявлен эффект кратковременной памяти, проявляющийся в латентности распознавания образов, определяемый интегрирующей ролью астроцитов сети. Также было показано, что включение в модель SNN глиальных клеток приводит к улучшенной обработке изображений с помощью SNN.

^{1*}Работа поддержана грантами РФФИ № 21-72-10129 и № 22-12-00216.

2. Модель и архитектура нейрон-астроцитарной сети

Нейрон-астроцитарная сеть состоит из SNN (151×151 нейронов, соединенных случайными возбуждающими синаптическими связями) и астроцитарной сети (50×50 астроцитов, равномерно распределенных напротив SNN). Динамика трансмембранного потенциала единичного нейрона описывается моделью Ижикевича [10]. Астроциты соединены с четырьмя ближайшими соседями через щелевые контакты, которые проницаемы для молекул инозитол 1,4,5-трифосфата (IP_3) и ионов Ca^{2+} . Для моделирования динамики внутриклеточной концентрации Ca^{2+} в астроците использовалась модель Уллаха [11]. Каждый астроцит взаимодействует с ансамблем из $4 \times 4 = 16$ нейронов с перекрытием в одну строку и один столбец. Входной сигнал подается на слой нейронов. В зависимости от амплитуды входного сигнала нейроны срабатывают с различной частотой (частота пропорциональна амплитуде входного сигнала). Частота импульсов нейронов преобразуется в амплитуду Ca^{2+} импульса в астроците, а астроциты в свою очередь модулируют синаптическую передачу.

3. Результаты

В начале симуляции модель нейрон-астроцитарной сети обучалась на изображение в оттенках серого. Для этого в качестве входного сигнала нейронному слою подавалось неискаженное изображение в течение 100 мс, амплитуда входного сигнала была равномерно распределена в диапазоне [0-8 мкА] в зависимости от интенсивности пикселя изображения (амплитуда 0 мкА соответствовала белому пикселю, а 8 мкА черному). После обучения сетевой модели спустя 1.1 с (необходимое время для формирования кальциевых событий в астроцитах модели), в качестве теста предъявлялись 4 вида изображений: известное сети с добавлением 80% гауссова шума, известное сети с добавлением 40% «Соль и перец» шума, равномерный шум и другое изображение, на которое сеть не обучалась. Выходной паттерн декодировался как средняя частота нейронных спайков во временном окне 500 мс с момента начала предъявления тестового изображения.

Предложенная в данной работе модель нейрон-астроцитарной сети способна распознавать и эффективно восстанавливать искаженное тестовое изображение. В первом и втором тестах, когда подавалось известное сети зашумленное изображение, система значительно уменьшала уровень шума. Предъявление равномерного шума или неизвестного сети тестового изображения приводит к неспецифическому или химероподобному нейронному ответу.

4. Заключение

В данной работе была предложена модель SNN, взаимодействующая с сетью астроцитов, способная хранить изображения в оттенках серого и обладающая возможностью их надежного извлечения из памяти. Хранение паттернов управляется сигналами астроцитов и осуществляется в масштабе времени кальциевых событий.

Изображения в оттенках серого использовались для стимуляции двухслойной нейрон-астроцитарной сети. Соответствующая активация астроцитарного слоя позволяет системе сохранять изображения в виде уровней активации астроцитов во время Ca^{2+} событий. Кроме того, различные уровни Ca^{2+} определяли разную силу модуляции синаптических связей в нейронном слое. Следовательно, в нейронном слое изображения появлялись в виде паттернов активности с разной скоростью срабатывания.

Разработанная модель позволяет выполнять довольно эффективное извлечение изображения из памяти, даже если предъявлялся сильно зашумленный паттерн. Полученные результаты согласуются с концепцией рабочей памяти, впервые предложенной в нашей недавней работе [12], где модель также реализована на основе SNN, сопровождаемой сетью астроцитов.

Литература

1. V.V. Matrosov, S.Y. Gordleeva, N.A. Boldyreva, E. Ben-Jacob, V.B. Kazantsev, and M. De Pittà, Emergence of regular and complex calcium oscillations by inositol 1,4,5-trisphosphate signaling in astrocytes, ch. 6 Computational Glioscience. Springer Series in Computational Neuroscience. M. De Pittà. and H. Berry, Eds., Springer, Cham, 2019, pp. 151-176.
2. J.L. Stobart, K.D. Ferrari, M.J. Barrett, C. Glück, M.J. Stobart, M. Zuend, and B. Weber, Cortical circuit activity evokes rapid astrocyte calcium signals on a similar timescale to neurons, vol. 98, no. 4 Neuron, 2018, pp. 726-735.
3. J. Lines, E.D. Martin, P. Kofuji, J. Aguilar, and A. Araque, Astrocytes modulate sensory-evoked neuronal network activity, vol. 11 Nat. Commun., 2020, pp. 1-12.
4. K.V. Kastanenka, R. Moreno-Bote, M. De Pittà, G. Perea, A. Eraso-Pichot, R. Masgrau, K.E. Poskanzer, and E. Galea, A roadmap to integrate astrocytes into Systems Neuroscience, vol. 68, no. 1 Glia, 2019, pp. 5-26.
5. M. Santello, N. Toni, and A. Volterra, Astrocyte function from information processing to cognition and cognitive impairment, vol. 22, no. 2 Nat. Neurosci., 2019, pp. 154-166.
6. H.J. Whitwell, M.G. Bacalini, O. Blyuss, S. Chen, P. Garagnani, S.Y. Gordleeva, S. Jalan, M.V. Ivanchenko, O.I. Kanakov, V. Kustikova, I.P. Mariño, I.B. Meyerov, E. Ullner, C. Franceschi, and A. Zaikin, The Human Body as a Super Network: Digital Methods to Analyze the Propagation of Aging, vol. 12 Front. Aging Neurosci., 2020, pp. 136.
7. S.Y. Gordleeva, O.I. Kanakov, A. Zaikin, M.V. Ivanchenko, and C. Franceschi, Brain aging and garbage cleaning. Modelling the role of sleep, glymphatic system and microglia senescence in the propagation of inflammaging, vol. 42, no. 5 Semin. Immunopathol., 2020, pp. 647-665.
8. J. Schummers, H. Yu, and M. Sur, "Tuned responses of astrocytes and their influence on hemodynamic signals in the visual cortex," Science, vol. 320, no. 5883, pp. 1638-1643, June 2008.
9. K. Sonoda, T. Matsui, H. Bito, and K. Ohki, Astrocytes in the mouse visual cortex reliably respond to visual stimulation, vol. 505, no. 4 Biochem. Biophys. Res. Commun., 2018, pp. 1216-1222.
10. E.M. Izhikevich, Simple model of spiking neurons, vol. 14. no. 6 IEEE Trans. Neural Netw., 2003, pp. 1569-1572.
11. G. Ullah, P. Jung, and A.H. Cornell-Bell, Anti-phase calcium oscillations in astrocytes via inositol (1,4,5)-trisphosphate regeneration, vol. 39, no. 3 Cell Calcium, 2006, pp. 197-208.
12. S.Y. Gordleeva, Y.A. Tsybina, M.I. Krivonosov, A.A. Zaikin, M.V. Ivanchenko, V.B. Kazantsev, and A.N. Gorban, Modelling working memory in spiking neuron network accompanied by astrocytes, vol. 15 Front. Cell. Neurosci., 2021, pp. 86.

АНТАГОНИСТИЧЕСКАЯ ИГРА С FUZZY-2 ПЛАТЕЖНОЙ МАТРИЦЕЙ

В.Г. Чернов

Владимирский государственный университет им. А.Г. и Н.Г. Столетовых

Рассматривается метод нахождения наилучших стратегий в антагонистической игре с платежной матрицей, элементы которой представлены нечеткими множествами (числами) второго порядка Fuzzy-2.

Ключевые слова: платежная матрица, нечеткое множество второго порядка, нечеткое число, функция принадлежности.

1. Введение

Классическая антагонистическая игра может быть представлена тройкой $G=\{A,B,L(A,B)\}$, где $A = \{a_i: i = \overline{1,M}\}$, $B = \{b_j: j = \overline{1,N}\}$ – множества стратегий игроков; $L(A,B)$ – платежная матрица.

Достаточно часто при формализации ситуации, требующей принятия решений, антагонистической игрой недостаточно информации для задания точных значений элементов платежной матрицы. В этих условиях в большом количестве работ, например [1,2], предлагается для элементов платежной матрицы использовать нечеткие числа.

В настоящее время известны два типа нечетких множеств: нечеткие множества первого порядка (Fuzzy-1) и второго порядка (Fuzzy-2). Если рассматривать нечеткие числа как частный случай нечетких множеств, определенных на множестве вещественных чисел, то можно говорить о том, что упомянутые выше исследования, как и многие другие, рассматривают ситуацию, когда элементы платежной матрицы – множества типа Fuzzy-1. В этом случае для элементов платежной матрицы необходимо задать четкие значения точек, начиная с которых степень принадлежности (значения функций принадлежности) могут быть отличными от нуля. При задании таких точек возможны ошибки из-за не включения значений, расположенных около границ функций принадлежности (ФП) и находящихся под сомнением. Такая ситуация возможна из-за недостатка информации, например, когда возникает новая ситуация, требующая принятия решения. Размытие границ ФП является переходом от Fuzzy-1 к Fuzzy-2.

Отметим, что нечеткие множества второго порядка были предложены Л. Заде в 1975г.

2. Постановка и решение задачи

Пусть имеется некоторая ситуация, требующая принятия решения, формальной моделью которой может быть антагонистическая игра. Будем считать, что выполняется предположение, о так называемом «общем знании», которое гласит: игра со всеми правилами известна игрокам и каждый из них знает, что все участники осведомлены о том, что известно остальным партнерам по игре, и такое положение сохраняется до конца игры [3,4]. Элементы платежной матрицы заданы в виде нечетких множеств(чисел) второго порядка с треугольными функциями принадлежности (ФП-2) (рис.1).

Согласно [5] нечеткое множество второго порядка определяется следующим образом:

$$A = \{(x, y), \mu_A(x, w)) | \forall x \in X, \forall w \in [0,1]\},$$

где X – универсальное множество; $\mu_A(x, w)$ – множество функций принадлежности $\mu_A(x)$, характеризующих степень принадлежности элементов x и w (третье измерение, характеризующее вторичную функцию принадлежности) множеству A .

Тогда платежная матрица приобретает вид

$$\tilde{M} = \|\tilde{m}_{ij}\|, \tilde{m}_{ji} = \{UMF_{ij}(z), LMF_{ij}(z)\},$$

где $UMF_{ij}(z), LMF_{ij}(z)$ – верхняя, нижняя функции принадлежности (рис.1); $z \in [z_{min}, z_{max}]$; $[z_{min}, z_{max}]$ – область определения элементов платежной матрицы. В дальнейшем обозначения

$UMF_{ij}(z), LMF_{ij}(z)$ будем использовать не только для функций принадлежности, но и для самих нечетких множеств.

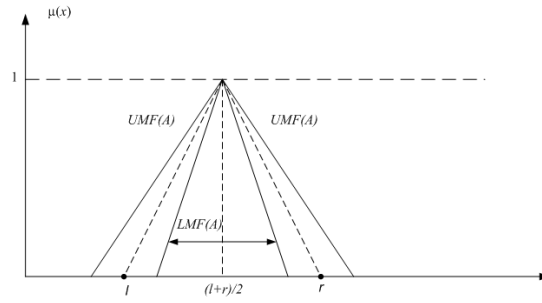


Рис. 1. Функция принадлежности нечеткого множества второго порядка (Fuzzy-2)

Предлагаемый метод решения антагонистической игры с Fuzzy-2 платежной матрицей основан на предположении, что при заданном виде нечеткости элементов платежной матрицы у игроков может быть по крайней мере одна чистая стратегия, дающая наилучший результат. В этом случае необходимо определить интегральную оценку последствий выбора игроком какой-то стратегии при любом выборе другого участника. Такую оценку можно получить, если рассматривать любую строку(столбец) как совокупность нечетких множеств(чисел) и применить к ним преобразование FztoTriangle[6].

Применительно к Fuzzy-2 элементам платежной матрицы это преобразование имеет свою специфику. Прежде всего это относится к необходимой для преобразования FztoTriangle координате центра тяжести

$$CG_i = \frac{\sum_j z_{ij} UMF(z_{ij}) + \sum_j z_{ij} LMF(z_{ij})}{\sum_j UMF(z_{ij}) + \sum_j LMF(z_{ij})}. \quad (1)$$

При выполнении преобразования FztoTriangle координата центра тяжести должна рассчитываться для каждой строки(столбца) по всему множеству их элементов. Можно показать, что если элементы платежной матрицы имеют симметричные функции принадлежности (рис.1), то координата центра тяжести строки(столбца) может вычисляться как среднее значение координат центров тяжести нечетких элементов строки(столбца), вычисленных по соотношению (1).

Если не использовать понижение порядка нечеткого множества, то в результате преобразования FztoTriangle будет также получено нечеткое множество второго порядка. При этом левая граница эквивалентной функции принадлежности $UMF_i^3(z)$ определяется как

$$Uz_i^1 = \min_j \{z_i^1 / UMF_{ij}(z)\},$$

где Uz_i^1 – левая граница носителя эквивалентного нечеткого множества $UMF_i^3(z)$; z_i^1 – левая граница носителей UMF_{ij} ; $Uz_i^2 = \max_j \{z_i^2 / UMF_{ij}(z)\}$. Аналогично для $LMF_i^3(z)$

$$Lz_i^1 = \min_j \{z_i^1 / LMF_{ij}(z)\},$$

$$Lz_i^2 = \max_j \{z_i^2 / LMF_{ij}(z)\}.$$

Координата максимума соответствующих эквивалентных множеств рассчитывается из известного соотношения для центра тяжести треугольника $CG = \frac{1}{3}(z_1 + z^* + z_2)$, где z_1, z_2 – координаты основания, z^* – координата вершины.

Тогда $Lz_i^* = 3LCG_i - (Lz_i^1 + Lz_i^2)$, $Uz_i^* = 3UCG_i - (Uz_i^1 + Uz_i^2)$, при этом возможно, что $Lz_i^* = Uz_i^*$ или $Lz_i^* \neq Uz_i^*$. В последнем случае координату максимума эквивалентного нечеткого множества второго порядка будем рассчитывать как

$$z_i^* = \frac{Lz_i^* + Uz_i^*}{2}.$$

Для второго игрока выполняем аналогичные преобразования для столбцов платежной матрицы. В результате получим интегральные оценки последствий возможных решений игроков в виде нечетких множеств второго порядка. Очевидно, что для выбора наилучшего решения необходимо сравнить полученные игроками интегральные оценки. Методов сравнения нечетких чисел первого порядка достаточно много [7–9]. Вопрос о сравнении нечетких чисел второго порядка практически не рассматривался.

Если рассматривать нечеткие множества $UMF_{ij}(z), LMF_{ij}(z)$ как отдельные, то можно применить для сравнения точечные оценки, предложенные R.R Yager [10,11], значения которых зависят от вида функции принадлежности, положения на области определения и от ширины носителя нечеткого числа. В рассматриваемом случае, т.к. оценки имеют одинаковый вид функций принадлежности существенными будут значения последних параметров. В результате возможные решения для первого игрока будут оцениваться точечными оценками S_i^L, S_i^U , для второго – S_j^L, S_j^U . В качестве итоговых оценок предлагаются

$$S_i = \frac{S_i^L + S_i^U}{2}, S_j = \frac{S_j^L + S_j^U}{2}.$$

Если предположить, что игрок A выигрывает, B проигрывает, то для первого наилучшей будет стратегия $a_i^* \rightarrow \max_i(S_i)$, для второго – $b_j^* \rightarrow \min_j(S_j)$. Нечеткая цена игры будет определяться как пересечение результатов наилучших решений игроков.

Литература

1. Butnariu D. Fuzzy games: a description of the concept // D. Butnariu / Fuzzy Sets and System. – 1978.– №1.– P. 181-19.
2. Krishnaveni G., Ganesan K. A new approach for the solution of fuzzy games/ G. Krishnaveni, K.Ganesan // National Conference on Mathematical Techniques and its Applications (NCMTA 18) IOP PublishingI OP Conf. Series: Journal of Physics: Conf. Series 1000. 2018.– 012017 doi :10.1088/1742-6596/1000/1/012017.
3. Myerson R.B. Game theory: analysis of conflict / R.B. Myerson. – London. Harvard :Harvard University Press, 1997, – 584 p.
4. Geanakoplos J. Common Knowledge. Handbook of GameTheory.v.2. / J. Geanakoplos.– ed. R. Aumann and S. Hart: Elsevier Science B.V.1994, P.1438-1496.
5. Mendel J.M., John R.I. Bob Tept-2 Fuzzy Sets Made Simple//IEEE Transaction on fuzzy systems.2002. Vol. 10. № 2. April
6. В.Г. Чернов Нечеткая модель принятия решений в условиях неопределенности, основанная на преобразовании FztoTriangle// Вестник Ижевского государственного технического университета, «Интеллектуальные системы в производстве». 2022. т.20. №1.
7. Rao P.P.B., Shankar N.R. Ranking generalized fuzzy numbers using area, mode, spread and weight. International Journal of Applied Science and Engineering, 2012. №10. v.1. p.41-57.
8. Воронцов Я.А., Матвеев М.Г. Методы параметризованного сравнения нечетких и трапециевидных чисел. Вестник ВГУ, Серия Системный анализ и информационные технологии. 2014. №2. с.90-97.
9. Ухоботов В. И., Стабулит И. С., Кудрявцев К. Н. Сравнение нечетких чисел треугольного типа. Вестник Удмуртского университета. Математика. Механика. Компьютерные науки, 2019. Т. 29, Вып.2. с.197-210. doi: 10.20537/vm190205
10. Yager R.R. Multiple-objective decision – making using a fuzzy sets // International Journal. Man - Machine Studies. 1977. Vol.9. No. 4. P.375-382.
11. Yager R.R. Multicriteria decisions with soft: an application of fuzzy set and possibility theory //Fuzzy Mathematics. 1982. Vol.2. No. 2. Pt.1. P.21-28; Vol.2. No 3. Pt.2, P.7-16.

ПОИСК ЛОКАЛЬНЫХ ЭКСТРЕМУМОВ ФУНКЦИЙ С ИСПОЛЬЗОВАНИЕМ МЕТОДОВ АНАЛИЗА ДАННЫХ

Д.Е. Шапошников, Е.Р. Тренина

Нижегородский государственный университет им. Н.И. Лобачевского

Данная статья описывает новый подход к решению задачи поиска локальных экстремумов многоэкстремальных функций. Для формирования и сохранения множества точек локальных минимумов используются генетические алгоритмы и метод кластерного анализа при предположении интерактивного участия исследователя. Описывается реализация данного подхода и результаты работы алгоритмов на примере двумерных многоэкстремальных функций.

Ключевые слова: локальная оптимизация, локальные экстремумы, генетические алгоритмы, кластерный анализ.

1. Введение

Рассматривается задача поиска множества локальных экстремальных точек функции с использованием интерактивных поисковых методов оптимизации с возможностью пошагового анализа исследователем и управления приостановкой или завершением работы и анализом результата. Такими возможностями, в частности, обладают широко распространённые эвристические методы, например, эволюционно-генетические алгоритмы. В дальнейшем для определённости будем говорить про задачу минимизации.

Традиционно считается, что ответ, полученный при решении задачи оптимизации общего вида, концептуально должен быть лучшим из возможных локальных, то есть глобальным минимумом. В процессе поиска экстремальных точек локальные минимумы обычно рассматриваются как проблема: они во многих случаях затрудняют поиск глобального минимума из-за необходимости нерациональных и излишних действий. Такой вывод сделан в предположении, что локальные минимумы значительно отличаются от глобальных. Однако зачастую возникают ситуации, когда не требуется поиска точного глобального решения (достаточно «хорошего» локального), поскольку значительное время, затраченное на его нахождение, будет потрачено нерационально.

В качестве примера можно рассмотреть практическую пользу от анализа локальных минимумов, которая хорошо просматривается при обучении нейронных сетей. Теоретически, целью обучения нейросети является ее сходимость к глобальному минимуму, когда сеть оптимизирует свою способность классифицировать обучающую выборку. Однако в исследовании замечен факт, что несмотря на наличие большого количества локальных минимумов при увеличении параметров сети, нет заметной разницы значений функции потерь в этих точках, то есть все минимумы являются хорошим решением. Это означает, что значения функции в точках локальных минимумов несильно отличаются от глобального. Также отмечено, что при нахождении глобального минимума нейросеть может быть перетренирована, то есть менее эффективна.

Целью данной работы является разработка программы поиска и фиксации локальных экстремумов. Основная задача – предложить новый подход к решению задачи поиска локальных экстремумов с использованием генетических алгоритмов и методов кластерного анализа. В докладе рассматривается разработка данного подхода, его реализация в виде программного кода на языке Python с использованием существующих библиотек, а также практические примеры и результаты.

2. Постановка задачи

Рассмотрим традиционную задачу безусловной оптимизации [1] как задачу поиска точки минимума функции $f(x)$, определённой на множестве допустимых значений D :

$$\begin{aligned} f(x) &\rightarrow \min, x \in D \subseteq R^n \text{ или} \\ x^* &= \arg \min_{x \in D \subseteq R^n} (f(x)). \end{aligned} \quad (1)$$

Будем считать, что точка x^* является точкой глобального минимума, или глобальным решением задачи (1), если

$$f(x^*) \leq f(x), \quad \text{при всех } x \in D. \quad (2)$$

Также будем называть точку x^0 точкой локального минимума или локальным решением задачи (1), если

$$f(x^0) \leq f(x), \quad \text{при всех } x \in X \cap U_\varepsilon(x^0), \quad (3)$$

где $U_\varepsilon(x^0) = \{x \in R^n \mid \|x - x^0\| \leq \varepsilon\}$ – шар радиуса $\varepsilon > 0$ с центром в x^0 .

Если неравенство в (2) или (3) выполняется как строгое при $x \neq x^*$ и при $x \neq x^0$ соответственно, то будем называть x^* и x^0 точками строгого минимума в локальном или глобальном смысле.

Будем считать, что точки локального минимума (или локальных минимумов) представляют собой интерес с точки зрения предметной области. Уже было сказано о примере с обучением искусственной нейронной сети. Можно также привести немало примеров возникновения необходимости (или желательности) анализа локальных минимумов при проектировании сложных технических систем, например, в случае выгодности или предпочтительности данного локального решения с точки зрения других, менее формализованных критериев. Здесь необходимо подчеркнуть ряд принципиальных моментов, важных с точки зрения решения задачи.

Во-первых, очень часто, особенно именно в технических системах, важность рассмотрения точки локального минимума растёт с увеличением расстояния x^0 от точки глобального минимума x^* . Это, возможно, определяется тем, что остальные характеристики решения x^0 , скорее всего, будут сильно отличаться от характеристик точки x^* , что, в свою очередь, повышает важность такого анализа.

Во-вторых, с точки зрения предметной области требуется ввести понятие «мало отличающегося решения». Для этого будем считать, что точка локального минимума x^0 является «удовлетворительной», если со стороны лица, принимающего решение (исследователя) введена полоса эквивалентности на основании параметра $\Delta f > 0$ и точка x^0 удовлетворяет условию:

$$(f(x^0) - f(x^*)) \leq \Delta f. \quad (4)$$

Таким образом, для решения поставленной задачи требуется найти множество

$$D^0 = \left(x^0 \in D \mid \begin{aligned} &f(x^0) \leq f(x), x \in X \cap U_\varepsilon(x^0), U_\varepsilon(x^0) = \{x \in R^n \mid \|x - x^0\| \leq \varepsilon\}; \\ &(f(x^0) - f(x^*)) \leq \Delta f \end{aligned} \right). \quad (5)$$

В дальнейшем будем называть множество D^0 , построенное в соответствии с (5), множеством рациональных локальных минимумов (или множеством рациональных локальных минимумов).

Не ограничивая общности задачи, будем рассматривать данный подход на примере двумерной функции из соображений простоты и наглядности иллюстраций.

3. Анализ и выбор методов решения задачи

Современное состояние вычислительной техники и подходов к формированию моделей и решению задач оптимизации предполагает непосредственное участие человека (лица, принимающего решение, исследователя) в процессе всестороннего анализа возможностей, выбора алгоритмов и параметров, а также использование неформальных и трудно формализуемых знаний и интуиции. Всё это предполагает многошаговый процесс поиска с возможностью наглядного зрительного анализа и изменения параметров соотношений и алгоритмов.

Данной концепции удовлетворяют, в частности, эволюционно-генетические алгоритмы (в дальнейшем – ЭГА), относящиеся к классу эволюционных методов оптимизации. В основе

идеи генетических алгоритмов лежат принципы эволюции в природе. Имитируя процессы, происходящие в биологических системах, этот метод может справиться с теми препятствиями, с которыми не всегда удачно сталкиваются детерминированные методы.

Необходимо отметить важные характеристики и особенности эволюционно-генетических алгоритмов, определивших их выбор для решения данной задачи.

ЭГА являются пошаговыми методами с широкими возможностями вмешательства исследователя в ход решения задачи оптимизации – изменения многочисленных параметров на основе анализа достигнутых на данный момент результатов.

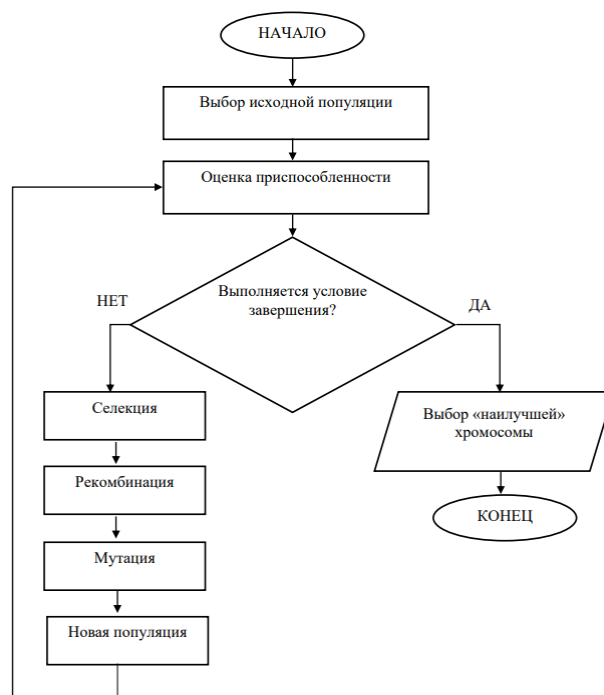


Рис. 1. Схема работы эволюционно-генетического алгоритма

Как известно, на каждом шаге ЭГА генерируются особи (потомки) в количестве, указанном как параметр «количество потомков». Также известно, что во время промежуточных шагов алгоритмов, особи собираются в локальные группы вокруг локальных экстремумов, если они есть. В дальнейшем, на последующих шагах, все потомки будут располагаться в районе глобального минимума, поэтому важно зафиксировать локальные популяции в процессе поиска.

Эволюционно-генетические алгоритмы (как, впрочем, и другие поисковые методы оптимизации) предполагают возможное участие исследователя на каждом шаге работы. Возможное участие может быть как в наглядности представления, так и во вмешательстве в работу метода – например, явном указании возможных точек минимума (исходя из знаний и/или интуиции специалиста).

Известно, что на первых шагах работы эволюционно-генетических алгоритмов, точки на графике, соответствующие особям популяции, группируются около локальных минимумов, образуя локальные популяции. Дальнейшие шаги приводят к перемещению особей в область глобального минимума и к потере локальных популяций. Предлагается подход, который предполагает фиксацию исследователем наличия локальных групп особей и фиксацию наблюдаемые локальных популяций методами кластерного анализа []. Алгоритм кластеризации должен помочь исследователю и самостоятельно выявить локальные популяции, затем принять решение о том, как поделить данные на классы.

Задача кластеризации представляет собой задачу разбиения множества образцов (точек) на классы. Количество классов k может быть задано, а может и быть неизвестно. Поэтому все методы кластеризации могут быть разделены на две большие группы: кластеризация с известным числом классов и кластеризация с неизвестным числом классов.

Для корректной работы алгоритма с известным числом классов разбиения величина k должна быть задана. Проблема заключается в том, что правильное число кластеров на практике чаще всего неизвестно. В данной работе предполагается, что эту задачу возьмёт на себя исследователь, определив классы (в данном случае – локальные популяции) зрительно на основе графического представления.

4. Предлагаемый подход и его реализация

Поисковый алгоритм останавливается исследователем на одном из начальных шагов, когда особи, соответствующие допустимым решениям задачи оптимизации, скапливаются вокруг некоторых точек, и затем область допустимых решений разделяется на кластеры. Затем находится минимум в каждой из образованных областей и полученные значения претендуют быть искомыми локальными минимумами.

Полностью автоматический алгоритм, реализующий данный подход, построить невозможно по нескольким причинам.

Во-первых, неизвестен точный номер шага, на котором необходимо прервать работу алгоритма оптимизации и провести кластеризацию.

Во-вторых, точно неизвестно количество искоемых точек локальных минимумов, а многие алгоритмы кластеризации требуют задать данные о количестве кластеров. При применении методов, не требующих такой априорной информации, велик риск нахождения не всех точек локального минимума или наоборот – нахождения «лишних» точек (то есть, на самом деле на являющихся точками локальных экстремумов).

В третьих, необходимо учитывать, что полученные точки являются предварительными и могут не быть искомыми решениями. проводить анализ полученных подозрительных точек и определить, действительно ли они являются искомыми решениями и как далеко они находятся от глобального оптимума и друг от друга. Для этого применяется подход, при котором алгоритм поисковой оптимизации, остановившись, продолжит поиск глобального минимума. Тогда в итоге полученные точки сравниваются с глобальным оптимумом для принятия решения принадлежности множеству точек локальных минимумов.

Таким образом перед тем, как использовать окончательный ответ, исследователю необходимо сравнить значения функции в найденных точках с глобальным и посмотреть, в какой полосе эквивалентности они лежат, подходит ли ему этот разброс, а также удаленность точек от оптимума.

5. Реализация предлагаемого подхода

Для реализации была использована библиотека распределенных эволюционных вычислений DEAP (Distributed Evolutionary Algorithms in Python), разработанной в канадском университете Лаваль.

В качестве примера рассмотрим задачу поиска глобального минимума многоэкстремальной функции Химмельблау $f(x_1, x_2) = (x_1^2 + x_2 - 11)^2 + (x_1 + x_2^2 - 7)^2$ (рис. 2). Данная функция имеет 4 точки глобального минимума, в которых функция равна нулю.

При запуске алгоритма с параметром максимального количества поколений `Max_Generations = 300` минимальное значение функции было найдено приблизительно на 13 шаге.

В качестве метода кластеризации была выбрана реализация алгоритма DBSCAN (Density-based spatial clustering of applications with noise) из библиотеки Scikit-learn.

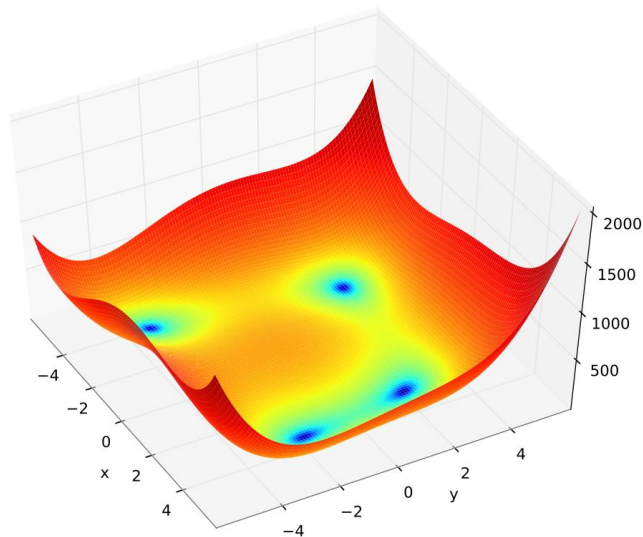


Рис. 2. Функция Химмельблау¹

Его работа основана на понятии плотности для каждой точки x_i выборки - количество других точек, находящихся в ε -окрестности точки x_i . Параметр ε определяет на каком максимальном расстоянии могут располагаться точки, чтобы их можно было отнести к одному кластеру. Также используется значение *min_samples* – минимальное количество точек, образующих кластер. Эти два параметра необходимо задать перед запуском алгоритма.

Алгоритм начинает свою работу по кластеризации с того, что делит точки на три группы:

- *внутренние* – точки, в ε окрестности которых больше, чем *min_samples* точек
- *граничные* – точки, в ε окрестности которых содержатся внутренние, но всего точек меньше, чем *min_samples*
- *шумовые* – точки, в ε окрестности которых меньше, чем *min_samples* точек, и не содержатся основные

Далее основные точки с общей ε окрестностью соединяются ребром и определяются кластеры по выделенным компонентам связности. Граничные точки определяются к кластеру, в котором находится основная ближайшая к ней точка. Шумовые точки не относят ни к одному кластеру.

Метод самостоятельно выявляет количество кластеров и может хорошо распознавать довольно сложные их формы.

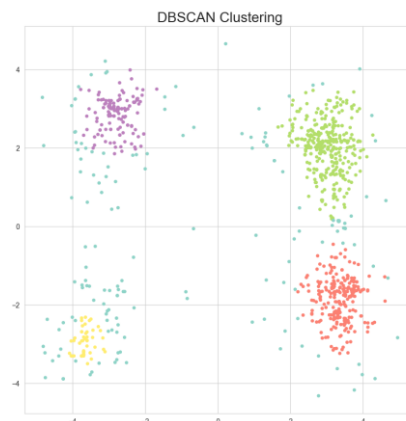


Рис. 3. Результат работы алгоритма кластеризации (найдено 4 кластера)

¹ [Himmelblau's function - Wikipedia](https://en.wikipedia.org/wiki/Himmelblau%27s_function) https://en.wikipedia.org/wiki/Himmelblau%27s_function

DBSCAN дает одинаковый ответ при тех же параметрах и входных данных, если они предоставлены в прежнем порядке.

Параметр ε не должен быть слишком малым, так как в таком случае может быть затруднительно найти количество точек, больше порогового значения, и многие точки будут рассматриваться как шум. Также данный параметр также не должен быть слишком большим – в этом случае точки могут объединиться в единый кластер.

Затем поочередно запускаются локальные алгоритмы для каждой локальной популяции. На рисунке XX представлен пример работы алгоритма (первый и последний шаг) для одного из локальных минимумов.

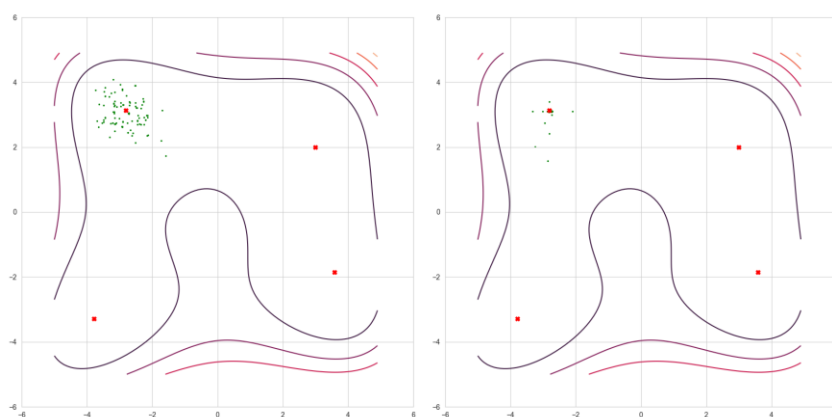


Рис. 4. Результат работы локального алгоритма

На последнем части работы всего алгоритма выполняется поиск глобального минимума и выводятся все найденные глобальные и локальные минимумы.

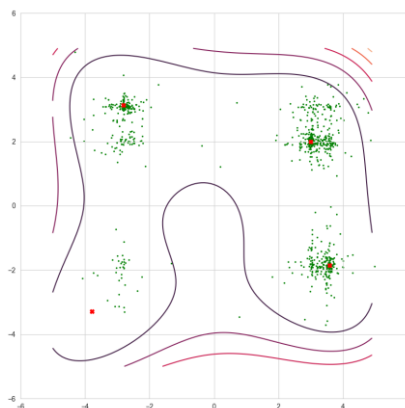


Рис. 5. Общий результат работы алгоритма

Таким образом построена, обоснована и применена на практике интерактивная процедура поиска и фиксации локальных и глобальных минимумов на основе совместного применения эволюционно-генетических алгоритмов и методов кластеризации.

Литература

1. Choromanska, A., Henaff, M., Mathieu, M., Arous, G. B., Lecun, Y. The loss surfaces of multi-layer networks. In Proceedings of the Eighteenth International Conference on Artificial Intelligence and Statistics (ed. G Lebanon & S. V. N Vishwanathan), Proceedings of Machine Learning Research, vol. 38, pp. 192–204. San Diego, California, USA: PMLR, 2015.
2. Сухарев А.Г., Тимохов А.В., Федоров В.В. Курс методов оптимизации: Учеб. пособие. – 2-е изд., – М.: ФИЗМАТЛИТ, 2008. – 368 с.

3. Брантон С. Л., Куц Дж. Н. Анализ данных в науке и технике / пер. с англ. А. А. Слинкина. М.: ДМК Пресс, 2021. – 574 с.
4. Вирсанки Э. Генетические алгоритмы на Python / пер. с англ. А. А. Слинкина. – М.: ДМК Пресс, 2020. – 286 с.

МЕТОДИКА ПОСТРОЕНИЯ ВНУТРЕННИХ ЛИНИЙ ТОКА В ЗАДАЧАХ ТЕОРИИ СТРУЙ

М.А. Шадт

Филиал МГУ им. М.В. Ломоносова в г. Севастополе

Предложенная методика демонстрируется на примере построения внутренних линий тока в классической задаче о натекании плоской струи идеальной жидкости на бесконечную стенку. Методика основана на численном решении уравнения Лапласа в заданной области. Обсуждается вопрос о постановке краевых условий.

Ключевые слова: теория струй, линии тока, уравнение Лапласа.

1. Введение

Задачи теории струй идеальной жидкости относятся к задачам с неизвестной границей.[1] Форма свободной поверхности определяется в процессе решения задачи методом особых точек Чаплыгина.[2] На примере классической задачи о натекании плоской струи идеальной жидкости на плоскую стенку рассматривается метод построения линий тока.

2. Описание методики

Метод построения линий тока заключается в решении уравнения Лапласа для заданной плоской области (рис.1) и построении линий уровня $\psi(x, y) = Const$ Необходимо решить уравнение Лапласа $\frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} = 0$ для функции $\psi(x, y)$ в области Ω . В качестве примера рассматривается задача о натекании струи под углом $\alpha = \frac{\pi}{4}$. На AC: $\psi = \psi_1$, на AB $\psi = \psi_2$ Как показано в [3], расход через сечение CC выражается через угол натекания следующим образом $q_1 = \frac{1+\cos\alpha}{2} = 0.8535$, а через сечение BB расход $q_2 = \frac{1-\cos\alpha}{2} = 0.1465$. На AC поставим условие Дирихле $\psi = 0.8535$. На AB поставим условие Дирихле $\psi = -0.1465$. На BC поставим условие Дирихле $\psi = 0$. На AA выполняются следующие условия: $\frac{\partial \psi}{\partial y} = v_0 \frac{\sqrt{2}}{2}$ и $\frac{\partial \psi}{\partial x} = v_0 \frac{\sqrt{2}}{2}$.

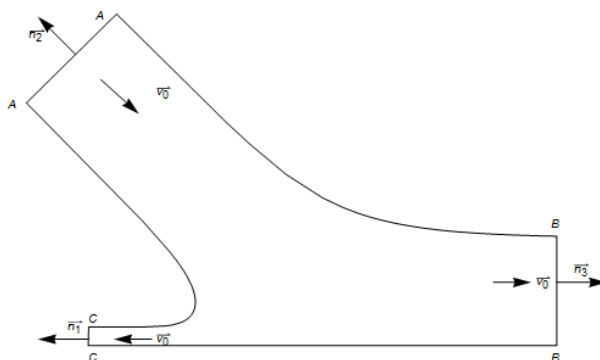


Рис.1. Общая схема течения

Учитывая направление вектора нормали $\vec{n}_2 = \left\{ -\frac{1}{\sqrt{2}}; \frac{1}{\sqrt{2}} \right\}$ приходим к следующему условию Неймана $\frac{\partial \psi}{\partial n_2} = \frac{\partial \psi}{\partial x} \left(-\frac{1}{\sqrt{2}} \right) + \frac{\partial \psi}{\partial y} \frac{1}{\sqrt{2}} = v_0 \frac{\sqrt{2}}{2} \left(-\frac{1}{\sqrt{2}} \right) + v_0 \frac{\sqrt{2}}{2} \frac{1}{\sqrt{2}} = 0$. На BB выполняются следующие условия: $\frac{\partial \psi}{\partial y} = v_0$ и $\frac{\partial \psi}{\partial x} = 0$. Учитывая направление вектора нормали $\vec{n}_2 = \{1; 0\}$

приходим к следующему условию Неймана $\frac{\partial \psi}{\partial n_3} = \frac{\partial \psi}{\partial x} * 1 + \frac{\partial \psi}{\partial y} * 0 = 0$ На СС выполняются следующие условия: $\frac{\partial \psi}{\partial y} = -v_0$ и $\frac{\partial \psi}{\partial x} = 0$. Учитывая направление вектора нормали $\vec{n}_2 = \{-1; 0\}$ приходим к следующему условию Неймана $\frac{\partial \psi}{\partial n_1} = \frac{\partial \psi}{\partial x} * (-1) + \frac{\partial \psi}{\partial y} * 0 = 0$. Поставленная смешанная краевая задача для уравнения Лапласа решалась в пакете Вольфрам Математика методом конечных элементов. Результат построения линий тока представлен на рис.2.

3. Результаты расчетов

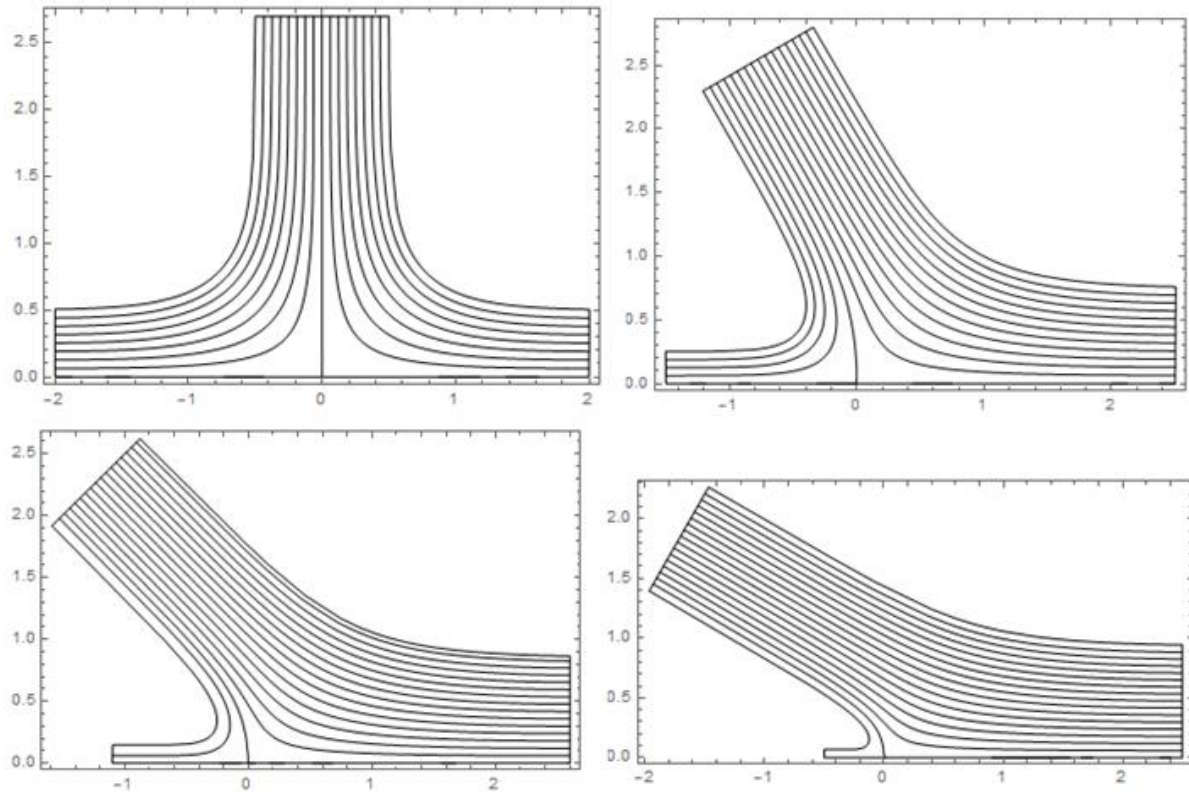


Рис. 2. Линии тока для различных углов натекания струи

Предложенная методика позволяет строить не только линии тока, но и распределение скоростей (рис. 3), а также распределение коэффициента давления (рис. 4), что является ее существенным преимуществом по сравнению с остальными методиками. Во всех расчетах было принято значение $v_0 = 1$. Исходя из уравнения Бернулли, можем записать следующие соотношения: $p_0 + \frac{\rho v_0^2}{2} = p + \frac{\rho v^2}{2}$ или $p - p_0 = \frac{\rho}{2}(v_0^2 - v^2) = \frac{\rho v_0^2}{2} \left(1 - \left(\frac{v}{v_0}\right)^2\right)$.

Тогда коэффициент давления с учетом того, что $v_0 = 1$, выразится следующим простым образом: $C_p = \frac{p - p_0}{\rho v_0^2 / 2} = 1 - \left(\frac{v}{v_0}\right)^2 = 1 - v^2$. К недостаткам данного метода можно отнести относительную сложность постановки краевых условий, индивидуальных для каждой задачи, а также трудности, связанные с написанием программы.

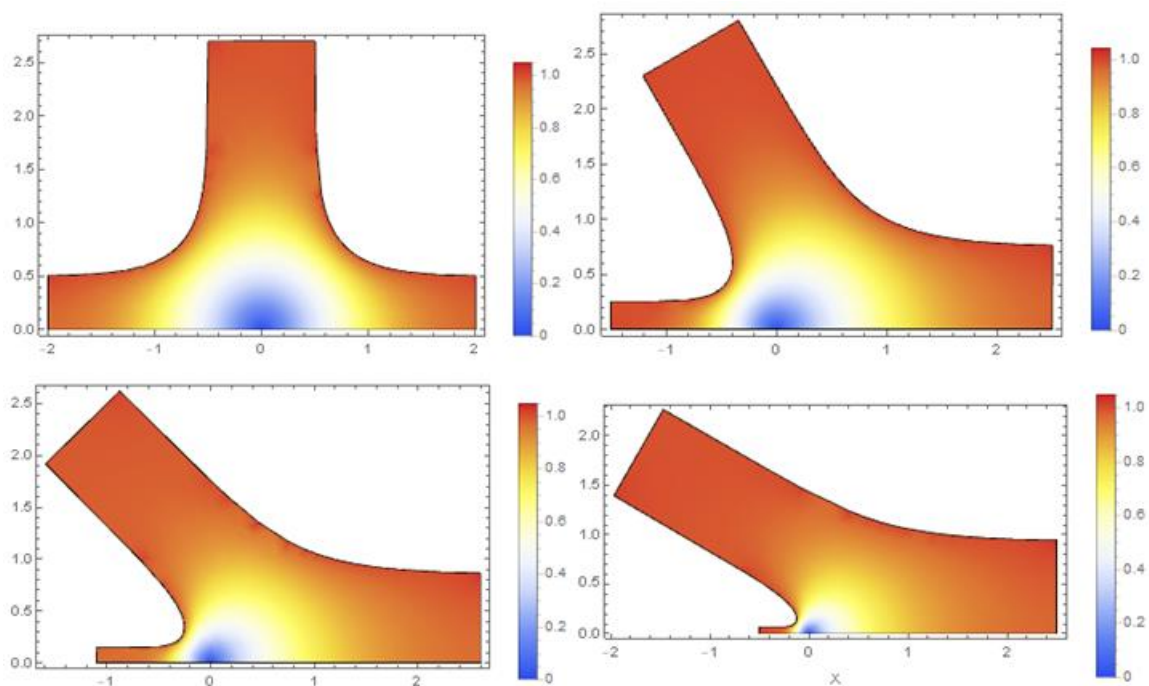


Рис. 3. Распределение модуля скорости

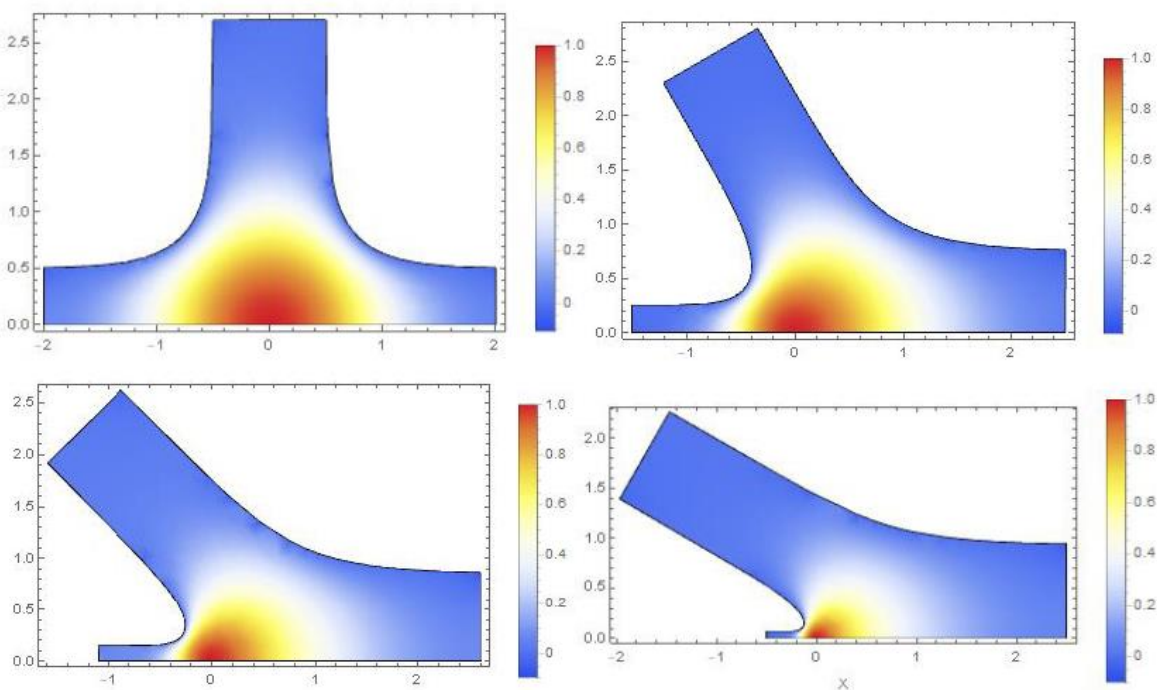


Рис. 4. Распределение коэффициента давления

Литература

1. Гуревич М.И. Теория струй идеальной жидкости. – М.: Наука, 1979. 536 с.
2. Толоконников С.Л. Введение в теорию струйных течений идеальной несжимаемой жидкости (учебное пособие). – М.: Издательство Центра прикладных исследований при механико-математическом факультете МГУ, 2008. 54 с.
3. Кочин Н.Е., Кибель И.А., Розе Н.В. Теоретическая гидромеханика. Часть 1. – М : ГИФМЛ, 1963. 583 с.

СОДЕРЖАНИЕ

<i>А.А. Бусалов, А. В. Калинин, А. А. Тюхтина</i>	
Нелинейная нестационарная система интегро-дифференциальных уравнений переноса излучения и статистического равновесия в диффузионном приближении.....	4
<i>С.И. Веселов, А.С. Поднебеснов</i>	
Задача о рюкзаке с двумя неизвестными	9
<i>А.В. Голубева, А.Л. Сочков</i>	
Применение алгоритмов машинного обучения для прогнозирования типа социального лифта	11
<i>С.В. Гонченко, О.В. Гордеева</i>	
О бифуркациях двумерных диффеоморфизмов с гомоклиническими траекториями к негрубым неподвижным точкам.....	16
<i>Ю.В. Гурьева, Е.П. Васильев, Л.А. Смирнов</i>	
Применение законов сохранения в нейросетевом решении одномерного нелинейного уравнения Шрёдингера.....	22
<i>Н.С. Дюкина</i>	
Применимость упрощенной модели сооружения при расчетах сейсмостойкости подземных коммуникаций.....	28
<i>Е.В. Евстифеев, О.И. Москаленко</i>	
Анализ характеристик перемежающейся обобщенной синхронизации при помощи расчета меры мультистабильности в однонаправленно связанных системах Лоренца	32
<i>В.И. Звонилов</i>	
Почти комплексные многообразия с гомологически нулевой вещественной частью.....	35
<i>А.В. Кондратьева, М.И. Кузнецов</i>	
Неальтернирующие гамильтоновы формы с коэффициентами в алгебре разделенных степеней	37
<i>С.Ю. Коннов, Е.А. Козин</i>	
Реализация библиотеки поддержки распределённого глобального адресного пространства	40
<i>А.Г. Коротченко, В.М. Сморякова</i>	
О некоторых стратегиях численного интегрирования, основанных на использовании конечно-разностной формулы	46
<i>А.А. Коротышева, В.Р. Милов, Ю.С. Егоров, А.Ю. Кербенева, М.С. Дубов</i>	
Применение методов машинного обучения для классификации немаркированных элементов питания	52
<i>Л.Н. Кривдина</i>	
Оптимальное управление максимальным отклонением выхода при ограничениях	57
<i>Т.В. Куприянова, Д.И. Кислицын</i>	
Использование искусственного интеллекта в системе управления освещением умного дома	61
<i>О.В. Любимцев, А.А. Туганбаев</i>	
Об идеалах центрально существенных колец.....	67

Д.О. Любченко, О.Б. Исаева

Коммуникационные схемы на основе синхронизации систем со странными хаотическими и нехаотическими аттракторами..... 69

А.В. Медиевский, А.Г. Зотин, К.В. Симонов, М.А. Курако, А.С. Кругляков

Методика улучшения видимости тканей головного мозга в условиях массивного кровотока по данным NIR-камеры и алгоритмов шиарлет-преобразования изображений 72

И.И. Микулик, Е.А. Благовещенская

Распараллеливание гибридного алгоритма муравьиной колонии с изменяющимися с помощью генетического алгоритма параметрами 78

С.А. Насибли

Обзор методов объясняемого искусственного интеллект в задачах обработки ЭКГ 82

А.А. Оболенский, А.Ю. Нестеров, И.Б. Мееров

Оптимизация слоев нейронных сетей на архитектуре ARM..... 86

Ю.С. Половинкина, И.М. Дорофеев

Применение нейронных сетей для построения классификационной модели обработки медицинских данных 91

Е.В. Пройдакова, З.М. Гитлина, А.Е. Котова

Исследование алгоритма решения оптимизационной задачи распределения коечного фонда в сети медицинских учреждений региона 94

М.М. Пугавко, О.В. Масленников, В.И. Некоркин

Динамика рекуррентной спайковой сети в задачах вычислительной нейронауки 100

Д.С. Пушкарев, С.Ю. Маковкин, И.А. Кастальский, С.Ю. Гордлеева

Исследование синаптических связей в моделях нейронов Ходжкина-Хаксли и Ровата-Сельверстона 106

В.В. Разин

Решение задач дискретной оптимизации при помощи модельных языков программирования 110

Д.А. Рыболовлев, В.Е. Салимоненко

Оценка эффективности модели обнаружения компьютерных атак в отношении атаки нулевого дня 115

В.П. Савельев, Н.И. Сулягина

Математические модели поведения малого бизнеса в условиях выплат по кредиту 119

В.Е. Салимоненко

Построение гибридных систем обнаружения компьютерных атак на основе объединения сигнатурного и эвристического методов 123

Я.А. Середа, И.Д. Никоноров

Исследование реакции глубоких слоев обученных сетей глубокого обучения на семантически незначимые изменения в данных 126

Д.И. Силенко, И.Г. Лебедев

Алгоритм глобальной оптимизации, использующий деревья решений для выявления локальных экстремумов..... 131

М.С. Сорокина

Оптимальное оценивание множеств достижимости в задачах с параметрической неопределенностью..... 137

<i>Е.В. Ушакова</i> Дуополия Хотеллинга на плоскости с информационными рассылками в Евклидовой метрике.....	139
<i>А.М. Федоткин</i> Алгоритмическое управление с дообслуживанием неординарными пуассоновскими потоками малой интенсивности.....	141
<i>А.А. Федюков</i> Стабилизация по измеряемому выходу перевернутого маятника при ограничениях на фазовые переменные.....	145
<i>В.А. Ханадеев, А.Д. Плотникова, О.И. Москаленко</i> Диагностика перемежающейся фазовой синхронизации с учетом мультстабильности в ансамбле однонаправленно связанных систем Ресслера.....	148
<i>Ю.А. Цыбина, И.А. Кастальский, А.А. Заикин, С.Ю. Гордлеева</i> Обработка аналоговой информации спайковой нейрон-астроцитарной сетью	152
<i>В.Г. Чернов</i> Антагонистическая игра с Fuzzy-2 платежной матрицей.....	155
<i>Д.Е. Шапошников, Е.Р. Тренина</i> Поиск локальных экстремумов функций с использованием методов анализа данных.....	158
<i>М.А. Шадт</i> Методика построения внутренних линий тока в задачах теории струй	165

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ И СУПЕРКОМПЬЮТЕРНЫЕ ТЕХНОЛОГИИ

Труды XXII Международной конференции

Нижний Новгород, 14–17 ноября 2022 г.

Под ред. проф. Д.В. Баландина

Отв. за выпуск К.А. Баркалов

Публикуется в авторской редакции

Издательство Нижегородского госуниверситета им. Н.И. Лобачевского.
603022, Н. Новгород, пр. Гагарина, 23.