ISSN 2712-9942

УСПЕХИ КИБЕРНЕТИКИ

RUSSIAN JOURNAL OF CYBERNETICS

2025
Т. 6 № 1

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ УЧРЕЖДЕНИЕ «ФЕДЕРАЛЬНЫЙ НАУЧНЫЙ ЦЕНТР НАУЧНО-ИССЛЕДОВАТЕЛЬСКИЙ ИНСТИТУТ СИСТЕМНЫХ ИССЛЕДОВАНИЙ НАЦИОНАЛЬНОГО ИССЛЕДОВАТЕЛЬСКОГО ЦЕНТРА «КУРЧАТОВСКИЙ ИНСТИТУТ»

Успехи кибернетики

Russian Journal of Cybernetics

Том 6
№ 1

Vol. 6
No. 1

Москва

2025

Учредитель и издатель

Федеральное государственное автономное учреждение «Федеральный научный центр Научно-исследовательский институт системных исследований Национального исследовательского центра «Курчатовский институт»
(НИЦ «Курчатовский институт» — НИИСИ)

«Успехи кибернетики» — это рецензируемый научный журнал, в котором публикуются научные статьи по следующим специальностям (физико-математические, технические науки):

1.2.2. Математическое моделирование, численные методы и комплексы программ;

2.2.2. Электронная компонентная база микро- и наноэлектроники, квантовых устройств;

2.3.1. Системный анализ, управление и обработка информации.

Миссия журнала — развитие научных направлений по заявленной тематике в России и за рубежом, соответствующих приоритетным направлениям развития науки, технологий и техники в Российской Федерации, а также перечню критических технологий РФ.

Журнал ориентирован на пропаганду передовых идей в области физики, математики, технических наук, участие в реализации задач, сформулированных Президентом РФ в Указе от 01.12.2016 № 642, по научно-технологическому развитию РФ, а также импортозамещению по приоритетным направлениям стратегического развития страны, соответствующим тематике журнала, обеспечение печатными площадями высококвалифицированных кадров, повышение качества диссертационных исследований в данных отраслях путем развития механизма профессионального и общественного обсуждения их научных результатов, воспитание молодого поколения ученых.

Журнал принимает к публикации оригинальные статьи; переводы статей, опубликованных в зарубежных журналах (при согласии правообладателя на перевод и публикацию); обзоры; эссе; комментарии; другие информационные материалы.

Издание будет полезно ученым, работающим в соответствующих областях наук, а также аспирантам и студентам.

Журнал зарегистрирован Федеральной службой по надзору в сфере связи, информационных технологий и массовых коммуникаций.
Свидетельство о регистрации средства массовой информации ЭЛ № ФС 77-78857 от 04.08.2020.
Решением Высшей аттестационной комиссии с 27 июня 2022 года журнал включен в «Перечень рецензируемых научных изданий, в которых должны быть опубликованы основные научные результаты диссертаций на соискание ученой степени кандидата наук, на соискание ученой степени доктора наук».

Издается с 2020 года. Выпускается 4 раза в год.

Адрес редакции:
НИЦ «Курчатовский институт» — НИИСИ, 117218, г. Москва, Нахимовский просп., дом 36, корпус 1.
Телефон: +7 9226-54-57-88, факс: +7 (495) 719-76-81, e-mail: taras.gavrilenko@gmail.com. Сайт: jcyb.ru.

Founder and Publisher
Scientific Research Institute for System Analysis of the National Research Centre "Kurchatov Institute"
(NRC "Kurchatov Institute" – SRISA)

Russian Journal of Cybernetics is peer-reviewed.

The journal publishes papers on physics, mathematics, and engineering. The key areas of interest are:
– simulation, numerical methods, and simulation software
– electronic components for micro- and nanoelectronics, and quantum devices
– system analysis, control, and information processing.

Our mission is advancing these research areas in Russia and globally, supporting the national priority fields of research and development, and the critical technologies.

The Journal promotes advanced concepts in physics, mathematics, engineering, and facilitates the national research and technology development under the Russian President Decree No. 642 dated 01.12.2016, and phasing out strategic imports. We offer an opportunity for professional researchers to publish. We improve the doctorate research quality through public peer review and educate the new generation of researchers.

The Journal accepts original papers; translations of papers from international journals (consent of the right holders for the translation and publication is required); reviews; essays; remarks, and event reports.

The Journal is intended for researchers, university, and doctorate students.

The journal is registered with the Federal Service for Supervision of Communications, Information Technology and Mass Media.
Mass media registration certificate EL No. FS 77-78857 dated on 04.08.2020.
On June 27, 2022 the National Board for Academic Degrees and Titles put the Russian Journal of Cybernetics on the list of leading journals for publishing papers by doctorate students.

Published since 2020. 4 issues per year.

Editorial Board Address:
NRC "Kurchatov Institute" – SRISA, Russia 117218, Moscow, Nakhimovsky pr., 36/1.
Tel.: +7(922)654-5788, fax: +7(495)719-76-81, e-mail: taras.gavrilenko@gmail.com. Web: jcyb.ru.

СОДЕРЖАНИЕ

CONTENTS

ПАМЯТИ УЧИТЕЛЯ – К 90-ЛЕТИЮ ПРОФЕССОРА В. А. ТУПЧИЕВА

Профессор, доктор физ.-мат. наук Виль Асадулаевич Тупчиев (26.10.1934–16.09.2013)

Виль Асадулаевич Тупчиев родился 26 октября 1934 г. в г. Махачкале в интеллигентной семье. Окончив школу, поступил на физический факультет МГУ и выбрал себе научное направление, связанное с кафедрой математики, возглавляемой тогда еще членом-корреспондентом АН СССР Андреем Николаевичем Тихоновым. А.Н. Тихонов привлек юного Виля Асадулаевича к своим научным интересам по методам малого параметра для сингулярно возмущенных уравнений, находившихся в области его исследований. Непосредственное руководство исследовательской работой В.А. Тупчиева на кафедре осуществляла А.Б. Васильева, ученица А.Н. Тихонова. После защиты дипломной работы она же стала его руководителем в аспирантуре.

Обучаясь в аспирантуре физфака МГУ, Виль Асадулаевич подружился с Б.Л. Рождественским и его учеником Н.Н. Кузнецовым. Эта творческая дружба сохранилась на всю жизнь. Область научных интересов: теория квазилинейных гиперболических уравнений и приближенные методы их решения, теория квазилинейных систем, связанных с газовой динамикой. В.А. Тупчиевым и Н.Н. Кузнецовым доказана теорема существования для квазилинейных гиперболических систем без условия выпуклости системы. Их обоюдный интерес к системе газовой динамики в немалой степени определялся огромным влиянием на них А.Н. Тихонова и его ученика Б.Л. Рождественского. К началу 60-х годов восходит научная дружба «физфаковца» Виля Асадулаевича с «мехматянином» Станиславом Николаевичем Кружковым, учеником академика Ольги Арсеньевны Олейник, которая, в свою очередь, являлась ученицей академика И.Г. Петровского. Эта дружба заложила основу научного семинара по квазилинейным законам сохранения, которым они руководили около 20 лет совместно с Б.Л. Рождественским на ВМК МГУ вплоть до трагической кончины С.Н. Кружкова 12 июня 1998 г.

В 1964 году Виль Асадулаевич приехал в Обнинск на работу в филиал МИФИ по приглашению первого директора и организатора филиала В.Н. Глазанова. В.А. Тупчиев стал читать курс математической физики и дифференциальных уравнений для будущих физиков-теоретиков и экспериментаторов, перенося физфаковские идеи преподавания математики на обнинскую благодатную почву.

После кончины директора В.Н. Глазанова Виль Асадулаевич перешел в 1966-м году на научную работу в Институт экспериментальной метеорологии, расположенный в Обнинске. Там, в отделе активных воздействий на атмосферу и физики облаков, он занялся по предложению А.Н. Тихонова,

избранного тогда академиком, подготовкой докторской диссертации, которая сочетала методы малого параметра и построение классов корректности автомодельных решений для систем типа газовой динамики в случае знаменитой задачи Римана о «распаде начального разрыва». Здесь получилось мощное объединение теории сингулярно возмущенных систем для обыкновенных дифференциальных уравнений и газовой динамики. Сама по себе эта задача некорректная. Но к этому времени Вилем Асадулаевичем был разработан конструктивный способ отбора во множестве разрывных решений так называемых устойчивых решений, связанный с подсчетом размерностей устойчивых и неустойчивых многообразий на разрыве.

Работа спорилась, к тому же Виль Асадулаевич познакомился в 6-м отделе ИЭМ, который возглавлял его друг д.ф.-м.н. Александр Сергеевич Степанов, с новым для себя классом задач, связанных с явлением коагуляции облачных капель, — кинетическим уравнением М. Смолуховского (являющимся разновидностью кинетического уравнения Л. Больцмана для неупругих столкновений частиц). В этом отделе тогда царила творческая атмосфера, занимались приближенными решениями уравнения Смолуховского, его строгим обоснованием — задача, напрямую связанная с 6-й проблемой Гильберта! Возвратившись в Обнинский филиал МИФИ на кафедру высшей математики, в 1972 г. Виль Асадулаевич привлек к задачам обоснования приближенных методов для уравнения коагуляции Смолуховского своего первого ученика — Валерия Галкина, который в 1978 г. защитил на физфаке МГУ кандидатскую диссертацию «Математическая теория уравнения коагуляции».

Проф. А.Б. Васильева в гостях на кафедре прикладной математики ИАТЭ, Обнинск, 2004 г.

Важнейшей чертой учебной и научной деятельности Виля Асадулаевича всегда была поддержка работы исследовательских семинаров на кафедре для преподавателей и обязательно — для студентов. Через эти семинары в течение сорока лет прошли сотни талантливых молодых людей, уносивших в себе искру научного пламени, зажженного А.Н. Тихоновым.

В 1978 г. после защиты докторской диссертации на ВМК МГУ Виль Асадулаевич возглавил кафедру прикладной математики Обнинского филиала МИФИ. В течение двадцати лет руководства кафедрой профессор В.А. Тупчиев сформировал ее научный и образовательный облик, органично связанный с нуждами атомной отрасли — главной в Обнинске, обеспечил высочайшие стандарты математической культуры преподавателей и студентов.

Кафедра прикладной математики ИАТЭ. В центре — зав. кафедрой профессор В.А. Тупчиев. Обнинск, 1990 г.

В июле 1999 года В.А. Тупчиев передал кафедру прикладной математики своему заместителю по кафедре и первому ученику — профессору В.А. Галкину.

Кафедра ПМ ИАТЭ в начале 2000-х

Профессор В.А. Тупчиев и академик А.А. Самарский. Обнинск, 2000 г.

До конца своих дней Виль Асадулаевич всегда оставался любимейшим преподавателем для выпускников ИАТЭ, развивая новые научные направления, и сохранял высочайший уровень научных исследований, связанных с газовой динамикой, физической кинетикой, задачами атомно-энергетического комплекса.

Профессор В. А. Галкин

О ВЫБОРЕ КЛАССОВ КОРРЕКТНОСТИ

В. А. Галкин
Сургутский филиал федерального государственного автономного учреждения «Федеральный научный центр Научно-исследовательский институт системных исследований Национального исследовательского центра «Курчатовский институт», г. Сургут, Российская Федерация
ORCID: http://orcid.org/0000-0002-9721-4026, ✍ val-gal@yandex.ru

Аннотация: предложен общий топологический подход для анализа широкого класса обратных задач, возникающих при обработке потоков информации. Типичной проблемой, появляющейся при их решении на основе априорного задания топологий в пространстве множества допустимых решений и исходных данных (информации), является некорректность, связанная с отсутствием непрерывности обратного оператора. В теории методов решения некорректных задач широко применяется метод регуляризации А.Н. Тихонова. В работе предлагается подход к выделению классов корректности для решения обратных задач, основанный на согласовании топологий множества исходных данных и обобщения понятия пространства решений, обеспечивающий непрерывную зависимость решения от исходных данных. Рассмотрены классы корректности, основанные на принципе выбора Цермело. Предлагаемый подход может быть использован при построении устойчивых искусственных нейронных сетей (ИНС) для задач распознавания образов. Возможно его распространение на технологии построения систем искусственного интеллекта (ИИ).

Ключевые слова: искусственные нейронные сети, вычислительная неустойчивость, методы регуляризации.

Благодарности: работа выполнена в рамках государственного задания НИЦ «Курчатовский институт» — НИИСИ по теме № FNEF-2024-0001 «Создание и реализация доверенных систем искусственного интеллекта, основанных на новых математических и алгоритмических методах, моделях быстрых вычислений, реализуемых на отечественных вычислительных системах» (1023032100070-3-1.2.1).

Для цитирования: Галкин В. А. О выборе классов корректности. *Успехи кибернетики*. 2025;6(1):12–15.

Поступила в редакцию: 25.02.2025. *В окончательном варианте: 01.03.2025.*

THE CHOICE OF CORRECTNESS CLASSES

V. A. Galkin
Surgut Branch of Scientific Research Institute for System Analysis of the National Research Centre "Kurchatov Institute"
ORCID: http://orcid.org/0000-0002-9721-4026, ✍ val-gal@yandex.ru

Abstract: we present a general topological approach for analyzing a broad class of inverse problems that arise in processing information flows. These problems often involve the lack of continuity in the inverse operator due to the a priori assignment of topologies in the space of admissible solutions and given data. To address such ill-posed problems, researchers widely use A.N. Tikhonov's regularization method. We propose an approach to identifying correctness classes for inverse problems by aligning the topologies of source data and generalizing the concept of the solution space. This ensures the continuous dependence of the solution on the source data. We also examine correctness classes based on the Zermelo selection principle. The proposed approach can support the development of stable artificial neural networks (ANN) for pattern recognition tasks and may extend to artificial intelligence (AI) system technologies.

Keywords: artificial neural networks, numerical instability, regularization methods.

Acknowledgements: this is a part of the FNEF-2024-0001 Development and Deployment of Trusted AI Systems based on New Mathematical and Algorithmic Approaches and Fast Computing Models Compatible with Domestic Computer Hardware government contract (1023032100070-3-1.2.1) granted to the NRC "Kurchatov Institute" – SRISA.

Cite this article: Galkin V. A. The Choice of Correctness Classes. *Russian Journal of Cybernetics*. 2025;6(1):12–15.

Original article submitted: 25.02.2025. *Revision submitted: 01.03.2025.*

Введение

В ряде вычислительных экспериментов по распознаванию образов, относящихся к акустической, видео и текстовой информации, были найдены подходы к созданию программного обеспечения для практического решения задач в соответствующих областях. Эти успехи явились источником интереса к созданию полуэмпирических методов, носящих название искусственные нейронные сети (ИНС). Идейной основой реализации ИНС является принятие гипотезы о возможности создания технического информационного устройства, которое можно обучить на серии заданных примеров распознаванию образа (поиску решения) по информации, вообще говоря, не входящей в обучающий набор.

Указанные замечания относятся к общей фундаментальной проблеме анализа вычислительной устойчивости ИНС при наращивании объемов обучающих наборов и как следствие — границ реализуемости и безопасной применимости систем ИИ, построенных на комплексах ИНС, взаимодействующих между собой путем обмена информацией. Одной из существенных проблем построения технологий ИИ является вычислительная устойчивость ИНС, обеспечивающих надежную интерпретацию входной информации, которая передается между различными ИНС для детализации и уточнения принимаемых решений.

Практическое использование ИНС нацелено на реконструкцию недоступного для непосредственного наблюдения объекта $z \in F$ на основе обработки «сжатой» информации $u \in U$, полученной из входного потока больших объемов данных с целью их классификации и интерпретации.

Процедура сжатия больших объемов данных может быть рассмотрена как действие вполне непрерывного оператора A на заданной паре метрических пространств F, U, т.е. $A : F \to U$ (это отображение трактуется как математическая модель аппаратуры, при помощи которой получают информацию $u \in U$ при наблюдении за объектом $z \in F$).

$$A(z) = u. \tag{1}$$

Соответственно, процедуру «сжатия» входной информации под действием оператора A можно трактовать как то, что ограниченные множества входных данных Big Data zпреобразуются в предкомпактные подмножества выходных данных (почти конечные множества, т.е. хорошо аппроксимируемые конечными ε-сетями). Задачи распознавания образов, задачи их классификации на основе ИНС можно сформулировать как построение обратного оператора $R = A^{-1} \quad : U \to F$

$$R(u) = z \tag{2}$$

на основе полученных данных, лежащих в предкомпактном множестве.

Центральной проблемой для отыскания (1) с компактным непрерывным оператором A является либо отсутствие решения уравнения (1) (т.е. не при всех значениях $u \in U$ существует решение уравнения (1)), либо отсутствие свойства непрерывности у обратного оператора R на множестве значений $A(F) \subset U$ — принципиальная вычислительная неустойчивость ИНС, основанной на формуле (2), см. [1, 2].

Природа отмеченной некорректности задачи (1) связана с произволом в выборе топологий, задаваемых пользователем на множествах F, U. Ниже рассматриваются согласованные топологии на этой паре множеств, позволяющих обеспечить построение классов корректности для широкого класса задач (1).

Таким образом, классы задач, основанные на применении ИНС как средства аппроксимации отображений и построения решений обратных задач с компактным оператором, обнаруживают общее математическое свойство — вычислительную неустойчивость при обработке больших массивов данных. Эти проблемы аналогичны также для задач «восстановления» образов по «цифровым двойникам», поскольку имеют ту же математическую природу.

Вспомогательные построения

Для заданного отображения (1), являющегося математической моделью получения информации $u \in U$, выделим классы корректности задачи об отыскании неизвестной $z \in F$ посредством задания метрики ρ_U на множестве допустимых значений информации U, которую ниже согласованным образом перенесем на множество F.

Определение 1. Назовем классом корректности для уравнения (1) на заданном информационном пространстве (U, ρ_U) такое метрическое подпространство (F_1, ρ_{F_1}), где $F_1 \subset F$, ρ_{F_1} — метрика на множестве F_1, что сужение отображения A на множество F_1

$$A|_{F_1} : (F_1, \rho_{F_1}) \to (U, \rho_U)$$

является гомеоморфизмом.

В общем случае для задания класса корректности считаем заранее определенной метрику ρ_U. Для каждого заданного $u \in A(F) \subset U$ рассмотрим полный прообраз $A^{-1}(u)$, который, очевидно, задает все множество решений уравнения (1) для указанных исходных данных.

Определим на множестве F отношение эквивалентности $\sim$, полагая $z' \sim z''$, если $A(z') = A(z'')$. Обозначим

$$F^* = \{A^{-1}(u), \quad u \in A(F)\}.$$

На множестве классов F^* определим метрику

$$\rho^*(A^{-1}(u'), A^{-1}(u'') \stackrel{def}{=} \rho_U(u', u''). \tag{3}$$

Отметим, что отображение $A : F \to U$ естественным образом продолжается до отображения $A^* : F^* \to U$ по формуле

$$A^*(z^*) = u \qquad \forall z^* \in F^*, \tag{4}$$

так как $\forall z \in z^* : \ A(z) = u$.

Определение 2. Элемент множества $z^* \in F^*$ назовем обобщенным решением уравнения (1), если для него выполнено равенство (4).

Лемма 1. *Отображение $A^* : (F^*, \rho^*) \to (A(F), \rho_U)$ является изометрическим гомеоморфизмом.*

Доказательство. Утверждение леммы 1 является прямым следствием формулы (3). Действительно, в силу (3) для любых элементов $z_1 \in z_1^*$ и $z_2 \in z_2^*$ таких, что $A(z_1) \stackrel{def}{=} u_1$ и $A(z_2) \stackrel{def}{=} u_2$, имеет место равенство

$$\rho^*(z_1^*, z_2^*) \stackrel{def}{=} \rho^*(A^{-1}(u_1), A^{-1}(u_2)) = \rho_U(u_1, u_2).$$

Следовательно, отображение $A^* : (F^*, \rho^*) \leftrightarrow (A(F), \rho_U)$ сохраняет расстояние, т.е. является изоморфизмом. Лемма доказана.

Следствие. Задача (4) является корректной на паре метрических пространств (F^*, ρ^*) $(A(F), \rho_U)$.

В силу множественности элементов, составляющих обобщенное решение $z^* \in F^*$, практическую значимость имеет принцип отбора точек $z \in z^*$. В частности, для этой цели может быть использована аксиома выбора Цермело (Zermelo) [1], постулирующая существование отображения $\tilde{Z} : F^* \to F$ такого, что $\tilde{Z}(z^*) = z(z^*) \in z^*$. Положим $F_1 \stackrel{def}{=} \tilde{Z}(F^*)$. Зададим на этом множестве структуру метрического пространства посредством метрики

$$\rho_{F_1}(z_1, z_2) \stackrel{def}{=} \rho^*(z_1^*, z_2^*). \tag{5}$$

Теорема существования классов корректности

Теорема. *Сужение отображения $A|_{F_1} : (F_1, \rho_{F_1}) \to (A(F_1), \rho_U)$ является изометрическим гомеоморфизмом, а соответствующая пара метрических пространств (F_1, ρ_{F_1}), $(A(F_1), \rho_U)$ задает класс корректности уравнения (1), который назовем $\tilde{Z}$-классом корректности Цермело.*

Замечание. Точность «распознавания образа» z на основе решения задачи (1) в классе корректности Цермело полностью совпадает с точностью получения исходных данных в информационном пространстве (U, ρ_U).

Доказательство теоремы. Действительно, поскольку выполнены равенства

$$A(z_1) = A^*(z_1^*) = u_1,$$
$$A(z_2) = A^*(z_2^*) = u_2,$$

то отображение $A|_{F_1} : F_1 \to A(F_1)$ является взаимно однозначным. Таким образом, справедливы следующие равенства: $\rho_{F_1}(z_1, z_2) = \rho^*(z_1^*, z_2^*) = \rho_U(u_1, u_2)$. Следовательно, $\rho_{F_1}(A^{-1}(u_1), A^{-1}(u_2)) = \rho_U(u_1, u_2)$ и, значит, отображение A является изометрическим гомеоморфизмом на паре пространств $F_1, A(F_1)$.

Теорема доказана.

Практическую значимость имеет вопрос построения отображения выбора Цермело $\tilde{Z}$. Например, задание структуры вполне упорядоченного пространства на множестве F, что гарантируется теоремой Цермело [3] о вполне упорядоченных пространствах, индуцирует соответствующий полный порядок на подмножестве $z^* \subset F$. В этом случае в качестве значения отображения Цермело $\tilde{Z}$ можно взять минимальный элемент на z^*, то есть

$$\tilde{Z}(z^*) = \min\{z \in z^*\}. \tag{6}$$

Таким образом, построение классов корректности, указанных в теореме 1, по формуле (6) порождается всеми возможными способами полного упорядочивания множества неизвестных величин F. Отметим, что для счетных множеств F полный порядок определяется заданием взаимно однозначного его отображения на натуральный числовой ряд.

Замечание

Задание функций выбора Цермело на иерархических обобщенных последовательностях в смысле Мура [3] параметризованных ИНС можно трактовать как технологии построения ИИ, детализирующие подходы из работы [4].

ЛИТЕРАТУРА

1. Тихонов А. Н., Арсенин В. Я. *Методы решения некорректных задач*. М.: Наука; 1986.
2. Бетелин В. Б., Галкин В. А. Математические задачи, связанные с искусственным интеллектом и искусственными нейронными сетями. *Успехи кибернетики*. 2021;2(4):6–14. DOI: 10.51790/2712-9942-2021-2-4-1.
3. Данфорд Н., Шварц Дж. Т. *Линейные операторы. Общая теория*. Т. 1. М: Издательство иностранной литературы; 1962.
4. Гавриленко Т. В., Галкин В. А. Интуитивные логические системы и их приложения в технологиях искусственного интеллекта. *Успехи кибернетики*. 2024;5(1):8–16. DOI: 10.51790/2712-9942-2024-5-1-01.

ВЛИЯНИЕ ПРЕСИНАПТИЧЕСКИХ СИНАПСОВ НА РАСПОЗНАЮЩУЮ СПОСОБНОСТЬ НЕЙРОННОЙ СЕТИ

Б. В. Крыжановский
Федеральное государственное автономное учреждение «Федеральный научный центр Научно-исследовательский институт системных исследований Национального исследовательского центра «Курчатовский институт», Москва, Российская Федерация
ORCID: https://orcid.org/0000-0002-0901-6370, ✍ kryzhanov@mail.ru

Аннотация: исследованы свойства нейронной сети в случае, когда ранг тензора связей выше двух, т.е. помимо матрицы синаптических связей присутствуют пресинаптические синапсы. Такого рода тензоры связи имеют место при аппаратной реализации нейронной сети на основе кроссбаров. Неотъемлемым свойством кроссбарной схемы является наличие паразитных токов: сигнал с одной шины, идущей на некоторый нейрон, перетекает через ячейки памяти (синапсы) на шины всех других нейронов. Тем самым на входной сигнал нейрона накладывается шум — ослабленные сигналы, идущие на все остальные нейроны. При этом проводимость аналоговой кроссбарной ячейки изменяется пропорционально шумовому сигналу и выходной сигнал ячейки приобретает нелинейный по входу характер. Показано, что при определенном виде тензора связей качество нейросетевого алгоритма значительно улучшается. Анализ проведен на примере сети, подобной сети Хопфилда.

Ключевые слова: тензор связи, искусственные нейронные сети, кроссбар, модель Хопфилда.

Благодарности: работа выполнена в рамках госзадания НИЦ «Курчатовский институт» — НИИСИ, проект FNEF-2024-0001.

Для цитирования: Крыжановский Б. В. Влияние пресинаптических синапсов на распознающую способность нейронной сети. *Успехи кибернетики*. 2025;6(1):16–22.

Поступила в редакцию: 20.01.2025. *В окончательном варианте: 27.01.2025.*

EFFECTS OF PRESYNAPTIC SYNAPSES ON THE RECOGNITION RATE OF A NEURAL NETWORK

B. V. Kryzhanovsky
Scientific Research Institute for System Analysis of the National Research Centre "Kurchatov Institute", Moscow, Russian Federation
ORCID: https://orcid.org/0000-0002-0901-6370, ✍ kryzhanov@mail.ru

Abstract: we studied the properties of a neural network where the rank of the coupling tensor is greater than two. In this case, in addition to the matrix of synaptic connections, there are presynaptic synapses. Such coupling tensors occur in hardware implementations of neural networks based on crossbar arrays. An inherent property of the crossbar scheme is the presence of parasitic currents: a signal from one wire going to a specific neuron flows through memory cells (synapses) to the wires of all other neurons. As a result, noise is superimposed on the input signal of the neuron, weakening signals going to all other neurons. In this scenario, the conductivity of the analog crossbar cell changes proportionally to the noise signal, causing the output signal of the cell to acquire a nonlinear relationship with the input. We showed that with a specific type of coupling tensor, the quality of the neural network algorithm improved significantly. We used a network similar to the Hopfield network as an example.

Keywords: coupling tensor, artificial neural networks, crossbar, Hopfield model.

Acknowledgements: this study is a part of the FNEF-2024-0001 government order contracted to the Kurchatov Institute.

Cite this article: Kryzhanovsky B. V. Effects of Presynaptic Synapses on the Recognition Rate of a Neural Network. *Russian Journal of Cybernetics*. 2025;6(1):16–22.

Original article submitted: 20.01.2025. *Revision submitted: 27.01.2025.*

Введение

Наиболее сложной проблемой при аппаратной реализации искусственной нейронной сети является создание связей (синапсов) между нейронами. В большинстве алгоритмов, используемых в различных приложениях, число связей превышает значение 10^4 – 10^5. Соответственно, реализация

такого числа связей стандартными методами электроники представляет собой чрезвычайно сложную задачу.

Одним из подходов к решению означенной проблемы является организация связей в виде кроссбаров [1–8], основанных на аналоговых ячейках памяти. Как правило, рассматриваются энергонезависимые ячейки памяти на сегнетоэлектрических [9, 10], магнитных [11], органических [12, 13] и металлооксидных [14–19] материалах, на плавающем затворе транзистора [20–22] и на фазовых переходах [23–25]. Такие кроссбары являются компонентами смешанных сигнальных схем, реализующих векторно-матричное умножение, которое является основной операцией в любой искусственной нейронной сети. Важно, что такие схемы позволяют выполнять вычисления на физическом уровне, используя фундаментальные законы Ома и Кирхгофа. По сравнению с цифровыми решениями это обеспечивает значительно более высокую эффективность использования энергии и площади.

Существенным недостатком кроссбарных схем является наличие паразитных токов и токов утечек. Если с утечками можно бороться тем или иным способом, то паразитные токи являются неотъемлемой частью кроссбарной схемы: сигнал с одной шины, идущей на некоторый (j-й) нейрон, перетекает через ячейки памяти на шины всех других нейронов. Тем самым, на входной сигнал i-го нейрона накладывается шум - ослабленные сигналы, идущие на все остальные нейроны. При этом проводимость аналоговой ячейки изменяется пропорционально шумовому сигналу и выходной сигнал ячейки приобретает нелинейный по входу характер. Такую ситуацию в каком-то виде описывает приведенное ниже выражение (1). Как правило, наличие шума ухудшает работу нейронной сети. Ниже мы покажем, что ухудшение не является обязательным: при определенных условиях наличие такого рода «шума» может существенно улучшить качество нейросетевого алгоритма. Анализ проведен на примере сети, подобной сети Хопфилда [26]. Этот тип нейронной сети наиболее хорошо исследован [27–40], и результаты аппаратной реализации легко проверить, используя многочисленные аналитические результаты.

Описание модели

Рассмотрим сеть Хопфилдова типа, состоящую из N нейронов, в которой входной сигнал на i-й нейрон задается выражением:

$$S_i = S_{0i} + \sum_{r=1}^{\infty} \sum_{j_1,\ldots,j_r} T^{(r)}_{i,j_1,\ldots,j_r} y_{j_1} y_{j_2} \ldots y_{j_r} \tag{1}$$

где y_j- выходной сигнал от j-го нейрона ($i,j = 1,2,\ldots,N$), S_{0i} — некий параметр, не обязательно равный нулю (как в модели Хопфилда), а $T^{(r)}_{i,j_1,\ldots,j_r}$ — связи между нейронами.

Будем полагать, что ассоциативная память построена на M случайных бинарных паттернах $x_m = \left(x_1^{(m)}, x_2^{(m)}, \ldots, x_N^{(m)}\right)$, $x_i^{(m)} = \pm 1$, $m = 1, 2, \ldots, M$. Связи, приходящие на i-й нейрон задаются подобно правилу Хэбба [41]:

$$T^{(r)}_{i,j_1,\ldots,j_r} = a_r \sum_{m=1}^{M} x_i^{(m)} x_{j_1}^{(m)} \ldots x_{j_r}^{(m)} \tag{2}$$

В стандартной модели Хопфилда отсутствуют нелинейные отклики, т.е. учитывается только линейный член $a_1 = 1$ ($a_r = 0$ при $r \geq 2$). В нейрофизиологических исследованиях известны ситуации [42], когда образуется пресинаптический синапс, т.е. аксон одного нейрона образует синаптическую связь с аксоном другого нейрона. Это соответствует удержанию в сумме (1) только двух первых членов $r = 1,2$ ($a_r = 0$ при $r \geq 3$).

Для простоты выражений не будем требовать, чтобы диагональные элементы тензора связей были строго равны нулю, как в модели Хопфилда. Основные результаты от этого мало изменятся, и все сведется только к переобозначениям некоторых коэффициентов. В этом случае входной сигнал S_i можно записать в виде:

$$S_i = S_{0i} + \sum_{m=1}^{M} \sum_{r=1}^{\infty} a_r x_i^{(m)} (x_m y)^r \tag{3}$$

где $y = (y_1, y_2, \ldots, y_N)$ есть вектор текущего состояния сети.

Энергия такой системы имеет вид:

$$E = S_0 y + \sum_{m=1}^{M} \sum_{r=1}^{\infty} \frac{a_r}{r+1} (x_m y)^{r+1} \tag{4}$$

Распознавание паттернов

В общем случае анализ качества распознавания паттернов не представляется возможным. Ограничимся рассмотрением частного случая, который позволит сделать некоторые общие выводы. В частности, влияние порогового вектора S_{0i} на эффективность распознающей способности сети рассмотрим в последующей работе. Здесь ограничимся случаем $S_{0i} = 0$, $i = 1,...,N$.

Рассмотрим наиболее простой случай, когда связи определяются следующим выражением:

$$a_r = \frac{\alpha^r}{r!} \tag{5}$$

В этом случае выражение (3) преобразуется к виду:

$$S_i = \sum_m x_i^{(m)} \left(e^{\alpha x_m y} - 1\right) \tag{6}$$

С учетом (6) перейдем к вопросу о распознающей способности рассматриваемой нейросети. Пусть на вход сети подается вектор y, представляющий собой искаженный вариант одного из паттернов, например паттерна x_μ: $y_i = -x_i^{(\mu)}$ с вероятностью p, $y_i = x_i^{(\mu)}$ с вероятностью $1-p$, где Np — доля искаженных компонент. Тогда (6) можно преобразовать к виду:

$$S_i = x_i^{(\mu)} \left(e^{\alpha x_\mu y} - 1\right) + \sum_{m \neq \mu} x_i^{(m)} \left(e^{\alpha x_m y} - 1\right) \tag{7}$$

Условие правильного распознавания имеет вид $S_i x_i^{(\mu)} > 0$ для всех $i = 1,...,N$ одновременно. На языке спиновых систем это условие означает, что всякий спин $x_i^{(\mu)}$ направлен вдоль воздействующего на него локального поля S_i. Выполнение условия правильного распознавания рассмотрим в двух предельных случаях.

а) В пределе $\alpha N \ll 1$ мы переходим к стандартной модели Хопфилда. Действительно, оставляя в (7) только члены первого порядка по α, получим, что условие $S_i x_i^{(\mu)} > 0$ преобразуется в известное соотношение:

$$N(1-2p) + \sum_{m \neq \mu} \sum_{j \neq i} x_i^{(\mu)} x_i^{(m)} x_j^{(m)} y_j > 0 \tag{8}$$

Второй из членов в левой части (9) — это внутренний шум: случайная нормально распределенная величина с нулевым средним и дисперсией равной $\sigma_M^2 = NM$. Применяя теоретико-вероятностный подход [29–31], получим: условие $S_i x_i^{(\mu)} > 0$ на одном нейроне выполняется с вероятностью $Q = 1 - 0.5 \cdot \operatorname{erf} z$, где:

$$z = \frac{N(1-2p)}{\sqrt{2}\sigma_M} = \sqrt{\frac{N}{2M}}(1-2p) \tag{9}$$

Соответственно, вероятность правильного распознавания паттерна в целом равна Q^N, а вероятность ошибки распознавания Pr имеет вид:

$$\mathrm{Pr} = 1 - \left(1 - \frac{1}{2}\operatorname{erf} z\right)^N \tag{10}$$

Правильное распознавание имеется в случае $z \to 0$. В этом случае вероятность ошибки с точностью до несущественных коэффициентов принимает вид:

$$\mathrm{Pr} \sim N \exp\left[\frac{N(1-2p)^2}{2M}\right] \tag{11}$$

Отсюда на величину ассоциативной памяти из условия $\Pr \ll 1$ вытекает известное ограничение $M < M_{\max}$, где:

$$M_{\max} = \frac{N}{2\ln N}(1-2p)^2 \tag{12}$$

Отметим, что в пределе $N \to \infty$ более корректное выражение $M_{\max} \simeq 0.138N$ получено в [27] на основе статфизического подхода к описанию модели Хопфилда с хэббовской матрицей связи. В работе [28] проведено обобщение этого результата на случай взвешенной хеббовской матрицы.

b) В противоположном пределе $\alpha N \gg 1$ пренебрежем в (7) единицами по сравнению с экспонентами. Тогда это выражение примет вид:

$$S_i \approx x_i^{(\mu)} e^{\alpha x_\mu y} \left[1 + \sum_{m \neq \mu} x_i^{(\mu)} x_i^{(m)} e^{\alpha (x_m - x_\mu) y} \right] \tag{13}$$

Сумма в квадратных скобках (13) играет роль шума, т.е.

$$Noise \approx \sum_{m \neq \mu} x_i^{(\mu)} x_i^{(m)} e^{\alpha (x_m - x_0) y} \tag{14}$$

Соответственно, условие правильного распознавания $S_i x_i^{(\mu)} > 0$ выполняется при $Noise < 1$. Нетрудно заметить, что величина $Noise$ — это случайная нормально распределенная величина с нулевым средним. Дисперсию этой величины трудно оценить в общем случае, однако этого и не требуется — ниже мы поступим проще.

Для простоты рассмотрим случай распознавания паттерна без искажений, т.е. $y = x_\mu$. Обозначим через γN максимальную меру сходства между различными паттернами. Тогда, с учетом соотношений $x_m x_\mu < \gamma N$ и $y x_\mu \equiv N$ из (14) получим:

$$Noise = \sum_{m \neq \mu} x_i^{(\mu)} x_i^{(m)} e^{\alpha x_m x_\mu - \alpha N} < \sum_{m \neq \mu} x_i^{(\mu)} x_i^{(m)} e^{-\alpha N (1-\gamma)} \tag{15}$$

Отсюда следует:

$$Noise < M e^{-\alpha N (1-\gamma)} \tag{16}$$

Если положить

$$\alpha > \frac{\ln M}{(1-\gamma)N} \tag{17}$$

то для любого $i \in \overline{1,N}$ шум всегда будет меньше сигнала ($Noise < 1$). Соответственно, размер памяти ($M \leq M_{\max}$) будет ограничиваться величиной:

$$M_{\max} = e^{\alpha N (1-\gamma)} \tag{18}$$

Как видим, в рассматриваемом пределе $\alpha N \gg 1$ размер ассоциативной памяти M может быть значительно больше числа нейронов N.

Выражение (18) получено для случая, когда на вход сети поступает неискаженный паттерн ($p = 0$). Это выражение легко обобщается и на случай распознавания искаженного паттерна:

$$M_{\max} = e^{\alpha N (1-2p-\gamma)} \tag{19}$$

Как видим, для эффективного распознавания достаточно условия $1 - 2p > \gamma$, которое означает, что искажения не слишком велики: противоположное условие $\gamma > 1 - 2p$ означает, что паттерн x_μ в результате сильного искажения стал ближе к одному из паттернов $x_m (m \neq \mu)$ памяти. Естественно, что в этом случае паттерн x_μ распознан не будет, а будет распознан паттерн x_m.

Обсуждение результатов

Проведенный выше анализ показал, что наличие в сети нелинейных связей с коэффициентами типа (5) приводит к существенному повышению эффективности распознавания. В частности, при $\alpha N \to \infty$ нейросеть может уверенно распознавать паттерны с искажениями $p \to 0.5$. Причем объем ассоциативной памяти сети возрастает экспоненциально с ростом N в соответствии с (19). На самом деле в этом нет ничего удивительного, поскольку и память нейросети и ее распознающая способность определяются не числом нейронов, а числом связей: в пределе $\alpha N \to \infty$ число связей также стремится к бесконечности.

Модель, описываемая выше выражениями (1)–(4), скорее всего далека от того, чтобы корректно описывать аппаратную реализацию нейронной сети на базе мемристорных кроссбаров. Однако, она дает некоторое представление о влиянии паразитных токов на эффективность распознавания. Скорее всего, для описания этого влияния достаточно учесть в сумме (1) только первые два члена, т.е. представить (1) в виде:

$$S_i = \sum_{j \neq i} T_{ij}^{(1)} y_i y_j + \sum_{j,k \neq i}^{\infty} T_{ijk}^{(2)} y_i y_j y_k \tag{20}$$

где первый член в правой части — описывает стандартную модель Хопфилда, а второй — кубическую нелинейность, обусловленную наличием шумовых сигналов:

$$T_{ij}^{(1)} = \sum_{m=1}^{M} x_i^{(m)} x_j^{(m)}, T_{ijk}^{(2)} = a \sum_{m=1}^{M} x_i^{(m)} x_j^{(m)} x_k^{(m)} \tag{21}$$

Чтобы проверить влияние второго члена на эффективность распознавания мы провели численное моделирование для сети небольшого размера $N = 100$. Для каждого значения параметра a, величины искажений входного паттерна p и параметра загрузки M/N генерировалось по 10^3 хэббовских матриц и вычислялись усредненные величины Pr. Результаты моделирования приведены на рисунках 1, 2.

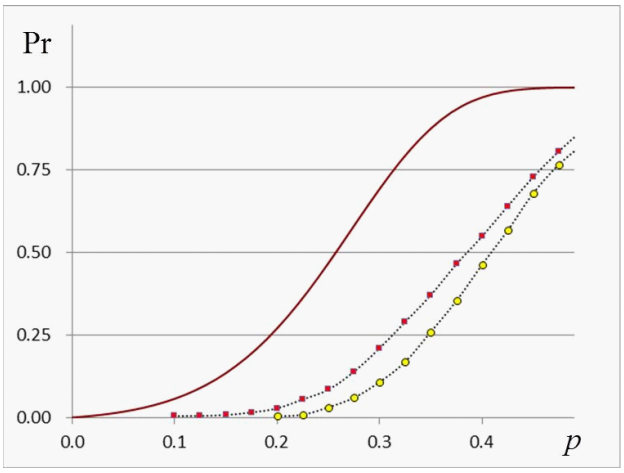


Рис. 1. *Зависимость ошибки распознавания Pr от величины искажений p. Кривые слева направо соответствуют* $a = 0, 0.01, 0.1$. *Загрузка фиксирована* $M = 0.13N, N = 100$

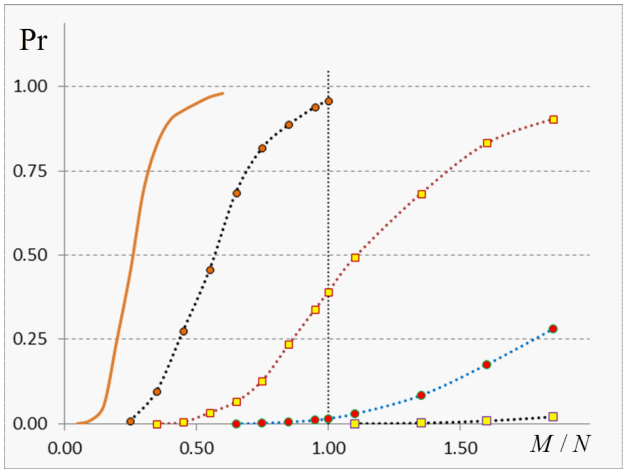


Рис. 2. *Зависимость ошибки распознавания Pr от величины загрузки* M/N. *Кривые слева направо соответствуют* $a = 0, 0.01, 0.02, 0.05, 0.1$

На рисунке 1 показана зависимость ошибки распознавания Pr от величины искажения входного паттерна p. Как видим, при заданной загрузке $M/N = 0.13$ стандартная сеть Хопфилда (сплошная кривая, $a = 0$) начинает выдавать ошибку уже при очень малых искажениях $p \sim 0.05$. В то же время, небольшая примесь нелинейности существенно улучшает картину: в случае $a = 0.01$ сеть начинает ошибаться при $p \geq 0.1$, а в случае $a = 0.1$ отличная от нуля ошибка появляется только при $p \geq 0.2$.

На рисунке 2 показана зависимость ошибки распознавания Pr неискаженных паттернов ($p = 0$) от величины загрузки M/N. Как видим, стандартная сеть Хопфилда (сплошная кривая, $a = 0$)

перестает распознавать неискаженный паттерн уже при малых загрузках ($M/N \geq 0.07$). В то же время, наличие даже небольшой примеси нелинейности $a = 0.01$ гарантирует безошибочное распознавание при объеме памяти $M \leq 0.25N$. С ростом параметра a величина памяти резко возрастает: так при $a = 0.1$ отличная от нуля ошибка появляется только при $M > 1.2N$.

ЛИТЕРАТУРА

1. Pershin Y., Di Ventra M. Experimental Demonstration of Associative Memory with Memristive Neural Networks. *Neural Networks*. 2010;23(7):881–886.
2. Jo S. H., Chang T., Ebong I. et al. Nanoscale Memristor Device as Synapse in Neuromorphic Systems. *Nanoletters*. 2010;10(4):1297–1301.
3. Truong S. N., Ham S. J., Min K. S. Neuromorphic Crossbar Circuit with Nanoscale Filamentary-Switching Binary Memristors for Speech Recognition. *Nanoscale Res. Lett*. 2014;9:629–629.
4. Zhu X., Yang X., Wu C. et al. Hamming Network Circuits Based on CMOS/Memristor Hybrid Design. *IEICE Electronics Express*. 2013;10(12):20130404-20130404.
5. Palagushkin A. N., Roshchupkin D. V., Yudkin F. A. et al. Aspects of the a-TiOx Memristor Active Medium Technology. *J. Appl. Phys*. 2018;124(20):205109.
6. Морозов А. Ю., Ревизников Д. Л., Абгарян К. К. Известия вузов. Вопросы реализации нейросетевых алгоритмов на мемристорных кроссбарах. *Материалы электронной техники*. 2019;22(4):272–278.
7. Kim H., Mahmoodi M. R., Nili H., Strukov D. B. 4K-memristor Analog-Grade Passive Crossbar Circuit. *Nat. Commun.* 2021;12(1):5198.
8. Liu X., Zeng Z. Memristor Crossbar Architectures for Implementing Deep Neural Networks. *Complex & Intelligent System*. 2022;8:787–802.
9. Kaneko Y., Nishitani Y., Ueda M. Ferroelectric Artificial Synapses for Recognition of a Multishaded Image. *IEEE Trans. Electron Devices*. 2014;61:2827–2833.
10. Boyn S. et al. Learning Through Ferroelectric Domain Dynamics in Solid-State Synapses. *Nat. Commun*. 2017;8(1):14736.
11. Romera M. et al. Vowel Recognition with Four Coupled Spin-Torque Nanooscillators. *Nature*. 2018;563:230–234.
12. Fuller E. J. et al. Parallel Programming of an Ionic Floating-Gate Memory Array for Scalable Neuromorphic Computing. *Science*. 2019;364:570–574.
13. Goswami S. et al. Robust Resistive Memory Devices Using Solution-Processable Metal-Coordinated Azo Aromatics. *Nat. Mater*. 2017;16:1216–1224.
14. Indiveri G. et al. Integration of Nanoscale Memristor Synapses in Neuromorphic Computing Architectures. *Nanotechnology*. 2013;24:384010.
15. Prezioso M. et al. Training and Operation of an Integrated Neuromorphic Network Based on Metal-Oxide Memristors. *Nature*. 2015;521:61–64.
16. Ambrogio S. et al. Neuromorphic Learning and Recognition with One-Transistor-One-Resistor Synapses and Bistable Metal Oxide RRAM. *IEEE Trans. Electron Devices*. 2016;63:1508–1515.
17. Adam G. C. et al. 3-D Memristor Crossbars for Analog and Neuromorphic Computing Applications. *IEEE Trans. Electron Devices*. 2017;64:312–318.
18. Hu M. et al. Memristor-Based Analog Computation and Neural Network Classification with a Dot Product Engine. *Adv. Mater*. 2018;30:1705914.
19. Merrikh B. F. et al. Implementation of Multilayer Perceptron Network with Highly Uniform Passive Memristive Crossbar Circuits. *Nat. Commun*. 2018;9(1).
20. Chakrabartty S., Cauwenberghs G. Sub-Microwatt Analog VLSI Trainable Pattern Classifier. *IEEE J. Solid-State Circuits*. 2007;42:1169–1179.
21. Ramakrishnan S., Hasler J. Vector-Matrix Multiply and Winner-Take-All as an Analog Classifier. *IEEE Trans. Very Large Scale Integr. Syst*. 2014;22:353–361.
22. Merrikh B. F. et al. High-Performance Mixed-Signal Neurocomputing with Nanoscale Floating-Gate Memory Cells. *IEEE Trans. Neural Netw. Learn. Syst*. 2018;29:4782–4790.
23. Eryilmaz S. B. et al. Brain-Like Associative Learning Using a Nanoscale Nonvolatile Phase Change Synaptic Device Array. *Front. Neurosci*. 2014;8:205.

24. Burr G. W. et al. Experimental Demonstration and Tolerancing of a Large-Scale Neural Network (165 000 Synapses) Using Phase-Change Memory as the Synaptic Weight Element. *IEEE Trans. Electron Devices*. 2015;62:3498–3507.
25. Boybat I. et al. Neuromorphic Computing with Multi-Memristive Synapses. *Nat. Commun*. 2018;9(1).
26. Hopfield J. J. Neural Networks and Physical Systems with Emergent Collective Computational Abilities. *Proc. Nat. Acad. Sci. USA*. 1982;79:2554-2558.
27. Amit D. J., Gutfreund H., Sompolinsky H. Storing Infinite Numbers of Patterns in a Spin-Glass Model of Neural Networks. *Physical Review Letters*. 1985;55(14):1530–1533.
28. Karandashev I., Kryzhanovsky B., Litinskii L. Weighted Patterns as a Tool to Improve the Hopfield Model. *Physical Review E*. 2012;85:041925.
29. Kryzhanovsky B. V., Kryzhanovsky V. M., Mikaelian A. L. et al. Parametric Dynamic Neural Network Recognition Power. *Optical Memory & Neural Network*. 2001;10(4):211-218.
30. Kryzhanovsky B. V., Mikaelian A. L. On the Recognition Ability of a Neural Network on Neurons with Parametric Transformation of Frequencies. *Doklady Mathematics*. 2002;65(2):286-288.
31. Kryzhanovsky B., Litinskii L., Fonarev A. An Effective Associative Memory for Pattern Recognition. *Lecture Notes in Computer Science*. 2003;2810:179-186.
32. Kryzhanovsky B. V., Kryzhanovsky V. M. Modified q-state Potts Model with Binarized Synaptic Coefficients. Lecture Notes in Computer Science. 2008;5164:72-80.
33. Karandashev I., Kryzhanovsky B., Litinskii L. Properties of the Hopfield Model with Weighted Patterns. *Lecture Notes in Computer Science*. 2012;7552:41-48.
34. Agliari E., Albanese L., Alemanno F. et al. Dense Hebbian Neural Networks: A Replica Symmetric Picture of Supervised Learning. *Physica A: Statistical Mechanics and Its Applications*. 2023;626:129076.
35. Barra A., Bernacchia A., Santucci E., Contucci P. On the Equivalence of Hopfield Networks and Boltzmann Machines. *Neural Networks*. 2012;34:1–9.
36. Bertram R., Rubin J. E. Multi-Timescale Systems and Fast-Slow Analysis. *Mathematical Biosciences*. 2017;287:105–121.
37. Fanaskov V., Oseledets I. *Associative Memory and Dead Neurons*. arXiv:2410.13866.
38. Hinton G. E. A Practical Guide to Training Restricted Boltzmann Machines. *Neural Networks: Tricks of the Trade*. Springer-Verlag Berlin Heidelberg; 2012:599–619.
39. Hoover B., Chau D. H., Strobelt H. et al. A Universal Abstraction for Hierarchical Hopfield Networks. *The Symbiosis of Deep Learning and Differential Equations II*. NeurIPS; 2022.
40. Krotov D., Hopfield J. Dense Associative Memory is Robust to Adversarial Inputs. *Neural Computation*. 2018;30(12):3151–3167.
41. Hebb D. O. *The Organization of Behavior. A Neuropsychological Theory*. NY: Wiley; 1949.
42. Экклс Д. К. *Физиология нервных клеток*. М.: ИЛ; 1959.

МОДЕЛИРОВАНИЕ ЦУНАМИ НА БАЗЕ ПЕРСОНАЛЬНОГО КОМПЬЮТЕРА

М. М. Лаврентьев[1,2,а], **Ан. Г. Марчук**[3,б], **К. К. Облаухов**[1,в] , **М. Ю. Шадрин**[1,г]
[1] *Институт автоматики и электрометрии СО РАН, г. Новосибирск, Российская Федерация*
[2] *Институт математики им. С.Л. Соболева СО РАН, г. Новосибирск, Российская Федерация*
[3] *Институт вычислительной математики и математической геофизики СО РАН, г. Новосибирск, Российская Федерация*
[а] *ORCID: http://orcid.org/0000-0003-1574-0629, ✍ mmlavrentiev@gmail.com*
[б] *ORCID: http://orcid.org/0000-0003-3401-1982, mag@omzg.sscc.ru*
[в] *ORCID: http://orcid.org/0009-0000-8540-7616, oblaukhov@sl.iae.nsk.su*
[г] *ORCID: http://orcid.org/0009-0007-6870-9971, mikesha@sl.iae.nsk.su*

Аннотация: в работе представлены архитектуры аппаратных ускорителей для численного решения нелинейной системы уравнений мелкой воды для работы в составе персонального компьютера или автономно. Достигнутая производительность расчетов позволила провести численные эксперименты по изучению трансокеанских цунами. Предложена концепция корректировки параметров источника после получения данных о профиле волны с глубоководного датчика давления. За счет оптимизации вычислительного конвейера (реализация конечно-разностной схемы МакКормака, имеющей второй порядок аппроксимации) за один такт получаются значения искомых величин в узлах сетки на семи последовательных шагах по времени. При использования микросхемы Xilinx Virtex-7 VC709 в качестве со-процессора персонального компьютера на расчетной сетке из 9601×6781 узлов для выполнения 36 тысяч шагов по времени с интервалом в 3 сек (то есть для моделирования 30 часов распространения волны) требуется всего 1352 сек (то есть 22.5 минуты). Приводятся результаты сравнения получаемого численного решения с известными точными решениями, исследование зависимости расчетов от используемых глобальных батиметрических банков данных и обосновывается возможность применения метода вложенных сеток. Применение предложенной технологии предоставляет исследователям новые возможности в изучении явления цунами.

Ключевые слова: моделирование цунами, аппаратное ускорение, трансокеанское цунами.

Благодарности: работа выполнена при поддержке проектов FWNG-2024-0014 ИАиЭ СО РАН и FWNM-2022-0004 ИВМиМГ СО РАН.

Для цитирования: Лаврентьев М. М., Марчук Ан. Г., Облаухов К. К., Шадрин М. Ю. Моделирование цунами на базе персонального компьютера. *Успехи кибернетики*. 2025;6(1):23–34.

Поступила в редакцию: 07.12.2024. *В окончательном варианте: 28.12.2024.*

TSUNAMI SIMULATION ON A PC

M. M. Lavrentiev[1,2,a], **An. G. Marchuk**[3,b], **K. K. Oblaukhov**[1,c], **M. Yu. Shadrin**[1,d]
[1] *Institute of Automation and Electrometry, Siberian branch of the Russian Academy of Sciences, Novosibirsk, Russian Federation*
[2] *Sobolev Institute of Mathematics, Siberian branch of the Russian Academy of Sciences, Novosibirsk, Russian Federation*
[3] *Institute of Computational Mathematics and Mathematical Geophysics, Siberian branch of the Russian Academy of Sciences, Novosibirsk, Russian Federation*
[a] *ORCID: http://orcid.org/0000-0003-1574-0629, ✍ mmlavrentiev@gmail.com*
[b] *ORCID: http://orcid.org/0000-0003-3401-1982, mag@omzg.sscc.ru*
[c] *ORCID: http://orcid.org/0009-0000-8540-7616, oblaukhov@sl.iae.nsk.su*
[d] *ORCID: http://orcid.org/0009-0007-6870-9971, mikesha@sl.iae.nsk.su*

Abstract: we designed hardware code acceleration architectures for solving the nonlinear system of shallow water equations, enabling to run it on a PC or other device. We achieved sufficient computational performance to conduct numerical experiments on transoceanic tsunamis. We proposed a method to adjust source parameters based on wave profile data from deep-sea pressure sensors. By optimizing the computational pipeline and implementing the McCormack finite-difference scheme with second-order approximation, we computed values at seven consecutive time steps within a single clock cycle. Using a Xilinx Virtex-7 VC709 chip as a co-processor on a computational grid of $9{,}601 \times 6{,}781$ nodes, we performed 36,000 time

steps with a 3-second interval, simulating 30 hours of wave propagation in just 1,352 seconds (22.5 minutes). We validated our numerical solution against known exact solutions, analyzed how global bathymetric data affect calculations, and justified the use of the nested grid method. The proposed technology enhances our ability to study tsunamis.

Keywords: tsunami simulation, hardware acceleration, transoceanic tsunami.

Acknowledgements: this study is supported by the government contract FWNM-2022-0004 with the Institute of Computational Mathematics and Mathematical Geophysics, Siberian branch of the Russian Academy of Sciences, and FWNG-2024-0014 with the Institute of Automation and Electrometry, Siberian branch of the Russian Academy of Sciences.

Cite this article: Lavrentiev M. M., Marchuk An. G., Oblaukhov K. K., Shadrin M. Yu. Tsunami Simulation on a PC. *Russian Journal of Cybernetics*. 2025;6(1):23–34.

Original article submitted: 07.12.2024. *Revision submitted: 28.12.2024.*

Введение

Высокопроизводительные вычисления (High Performance Computing — HPC) часто рассматриваются как способ решения любых технических и даже этических проблем (искусственный интеллект, ChatGPT и так далее). Однако за последние годы энергопотребление таких систем резко возросло. Помимо списков Top 500 (см. http://top500.org) самых мощных суперкомпьютеров, регулярно публикуются списки Green-500 (см. http://green500.org) самых энергоэффективных компьютерных систем. Суперкомпьютер (скажем, HPC-система) — это вычислительная система, производительность которой на порядки выше производительности компьютера общего назначения.

Как отмечается в [1], в начале 2000-ых по параметру производительность графические процессоры (Graphic Processing Unit — GPU) превосходили центральный процессор (Central Processing Unit — CPU) примерно на порядок (естественно, в зависимости от решаемой задачи), а вентильные матрицы, программируемые пользователем (Field Programmable Gates Array — FPGA) — примерно в 30 раз. Примерно также, на порядок, авторы оценивают преимущество GPU по сравнению с CPU по параметру энергоэффективности — отношения производительности к энергопотреблению. В случае же FPGA, эта архитектура превосходит энергоэффективность CPUв 175 раз, а GPU — в 18 раз. Примерно такие же соотношения сохраняются и у современных вычислительных платформ.

Платформа FPGA обладает этими преимуществами поскольку предоставляет возможность (с помощью программных средств) проектировать вычислительные конвейеры, оптимизированные для реализации того или иного алгоритма. Другими словами, это возможность создать «суперкомпьютер на столе» с помощью программных средств. Печатная плата с микрочипом FPGA, работающая в составе персонального компьютера в качестве сопроцессора или самостоятельно, способна решить конкретную вычислительную задачу с очень высокой производительностью. При наличии соответствующей программной библиотеки та же аппаратная конфигурация может стать высокопроизводительной вычислительной системой для решения целого набора задач.

В данной работе возможности этой технологии демонстрируются на примере решения задачи распространения волны цунами.

Катастрофическая волна цунами 11 марта 2011 года у побережья Японии, возникшая после сильнейшего землетрясения (The Great Tohoku Earthquake) с магнитудой более 9, показала необходимость совершенствования методов оценки параметров волны с целью своевременного применения (в случае необходимости) эвакуационных мер и снижения последствий для инженерной инфраструктуры. В работе [2] проводится предварительный анализ возможностей существующей системы предупреждения об опасности цунами, обсуждаются нерешенные проблемы для случаев событий так называемой ближней зоны, когда между сейсмическим событием и приходом волны цунами к ближайшему побережью проходит совсем немного времени, что составляет около 20 минут в случае землетрясения 11 марта 2011 года.

Современные средства компьютерного и математического моделирования позволяют с достаточной точностью рассчитать распространения волны цунами по заданной акватории при доступности корректной информации о начальном возмущении (деформации морского дна или водной поверхности в источнике цунами), цифровой сеточной батиметрии и наличии достаточного времени для проведения расчетов. Для решения задачи об определении параметров начального возмущения в источнике

активно развиваются системы глубоководных донных датчиков давления, позволяющих в режиме реального времени получать данные о профиле волны, проходящей над местом расположения датчика, см., например, [3–7].

Существует целый набор пакетов программ для численного моделирования генерации и распространения волны цунами. Среди наиболее известных из них мы отметим следующие: MOST (Method of Splitting Tsunamis, NOAA Pacific Marine Environmental Laboratory, Seattle, WA, USA) [8, 9]; COMCOT (Cornell University, Ithaca, NY, USA; NS Science, New Zeland) [10]; TUNAMI-N1/TUNAMI-N2 (Tohoku University, Sendai, Japan) [11, 12]; VOLNA and NAMI-DANCE (TSUNAMI Modeling Software by Special Bureau of Sakhalin, Russia and METU Turkey) [13].

Время расчетов зависит как от характеристик используемой аппаратной платформы, так и от размеров шагов расчетной сетки, которые не могут выбираться произвольно. Критерий устойчивости численного решения [14] устанавливает зависимость между шагами по пространственным переменным и времени. Так, в областях с глубинами до 1-3 км использование шагов по пространственным переменным порядка 5-10 м, что необходимо для адекватной оценки зон затопления, требует шага по времени не более 0.05 сек. Это приводит к значительному увеличению времени счета для областей размером 1000×1000 км даже при использовании суперкомпьютерных систем. В случае действительно катастрофических событий масштаба The Great Tohoku Earthquake возможны отключения электроэнергии, что препятствует применению суперкомпьютеров. Поэтому, в частности, развивается концепция использования вложенных сеток, см., например, [15].

В настоящее время разрабатывается технология быстрой оценки максимальных значений амплитуды волны цунами вдоль выделенных участков побережья. Начальным моментом времени считаем приход волны в место расположения одного из глубоководных датчиков давления. Ориентиром является расчет максимальных значений амплитуды волны за несколько минут.

В качестве элементов технологии быстрой (в пределах нескольких минут) оценки параметров волны цунами вдоль защищаемого участка побережья без применения ресурсов суперкомпьютерных центров предложены: метод ортогонального разложения для определения параметров источника цунами по части волнового профиля, записанного в одной точке и метод ускорения численного решения нелинейной системы уравнений мелкой воды за счет аппаратного ускорения (специализированный вычислитель на базе FPGA) [16]. В работе кратко описан метод ортогонального разложения и суммированы результаты исследования качества предложенного метода аппаратного ускорения расчетов.

Математическая постановка задачи

Распространение волны цунами в акваториях с переменной глубиной при отсутствии внешних сил кроме гравитации корректно описывается нелинейной системой уравнений мелкой воды [17, 18]:

$$\begin{aligned} &\eta_t + (u(D+\eta))_x + (v(D+\eta))_y = 0, \\ &u_t + uu_x + vu_y + g\eta_x = 0, \\ &v_t + uv_x + vv_y + g\eta_y = 0. \end{aligned} \tag{1}$$

Здесь $u(x,y,t)$ и $v(x.y,t)$ — компоненты горизонтальной скорости водного потока, постоянные во всем слое от поверхности до дна, $\eta(x,y,t)$ — смещение водной поверхности относительно нулевого уровня (искомая высота волны), g — ускорение свободного падения, $D(x,y)$ — считающийся известным профиль глубин. Обычно для этой системы уравнений ставится задача Коши с начальными условиями, когда в некоторой подобласти в начальный момент времени задается начальное вертикальное смещение водной поверхности при нулевой скорости водного потока. Эта подобласть называется очагом цунами.

Результатом численного решения системы уравнений мелкой воды (хорошо описывающей процесс распространения волны цунами) являются параметры волны (смещение водной поверхности и компоненты скорости водного потока) в узлах расчетной сетки через определенные интервалы времени (шаги по времени). В работе проводится сравнение точности расчетов распространения волны цунами на модельной и реальной батиметрии, полученные с применением трех различных алгоритмов численного решения системы уравнений мелкой воды.

Для численного решения нелинейной системы уравнений мелкой воды (1) использована конечно-разностная схема МакКормака [19], имеющая второй порядок аппроксимации по всем

переменным:

$$\frac{\overset{\frown}{H}{}_{ij}^{n+1} - H_{ij}^n}{\tau} + \frac{H_{ij}^n u_{ij}^n - H_{i-1j}^n u_{i-1j}^n}{\Delta x} + \frac{H_{ij}^n v_{ij}^n - H_{ij-1}^n v_{ij-1}^n}{\Delta y} = 0$$

$$\frac{\overset{\frown}{u}{}_{ij}^{n+1} - u_{ij}^n}{\tau} + u_{ij}^n \frac{u_{ij}^n - u_{i-1j}^n}{\Delta x} + v_{ij}^n \frac{u_{ij}^n - u_{ij-1}^n}{\Delta y} + g\frac{\eta_{ij}^n - \eta_{i-1j}^n}{\Delta x} = 0$$

$$\frac{\overset{\frown}{v}{}_{ij}^{n+1} - v_{ij}^n}{\tau} + u_{ij}^n \frac{v_{ij}^n - v_{i-1j}^n}{\Delta x} + v_{ij}^n \frac{v_{ij}^n - v_{ij-1}^n}{\Delta y} + g\frac{\eta_{ij}^n - \eta_{ij-1}^n}{\Delta y} = 0$$

$$\frac{H_{ij}^{n+1} - \left(\overset{\frown}{H}{}_{ij}^{n+1} + H_{ij}^n\right)/2}{\tau/2} + \frac{\overset{\frown}{H}{}_{i+1j}^{n+1}\overset{\frown}{u}{}_{i+1j}^{n+1} - \overset{\frown}{H}{}_{ij}^{n+1}\overset{\frown}{u}{}_{ij}^{n+1}}{\Delta x} + \frac{\overset{\frown}{H}{}_{ij+1}^{n+1}\overset{\frown}{v}{}_{ij+1}^{n+1} - \overset{\frown}{H}{}_{ij}^{n+1}\overset{\frown}{v}{}_{ij}^{n+1}}{\Delta y} = 0$$

$$\frac{u_{ij}^{n+1} - \left(\overset{\frown}{u}{}_{ij}^{n+1} + u_{ij}^n\right)/2}{\tau/2} + u_{ij}^n \frac{\overset{\frown}{u}{}_{i+1j}^{n+1} - \overset{\frown}{u}{}_{ij}^{n+1}}{\Delta x} + v_{ij}^n \frac{\overset{\frown}{u}{}_{ij+1}^{n+1} - \overset{\frown}{u}{}_{ij}^{n+1}}{\Delta y} + g\frac{\overset{\frown}{\eta}{}_{i+1j}^{n+1} - \overset{\frown}{\eta}{}_{ij}^{n+1}}{\Delta x} = 0$$

$$\frac{v_{ij}^{n+1} - \left(\overset{\frown}{v}{}_{ij}^{n+1} + v_{ij}^n\right)/2}{\tau/2} + u_{ij}^n \frac{\overset{\frown}{v}{}_{i+1j}^{n+1} - \overset{\frown}{v}{}_{ij}^{n+1}}{\Delta x} + v_{ij}^n \frac{\overset{\frown}{v}{}_{ij+1}^{n+1} - \overset{\frown}{v}{}_{ij}^{n+1}}{\Delta y} + g\frac{\overset{\frown}{\eta}{}_{ij+1}^{n+1} - \overset{\frown}{\eta}{}_{ij}^{n+1}}{\Delta y} = 0$$

Здесь H_{ij}^n, u_{ij}^n, и v_{ij}^n — сеточные значения переменных, соответствующие функциям $H(x,y,t)$, $u(x,y,t)$ и $v(x.y,t)$ в дифференциальной системе (1). Параметры τ, Δx и Δy — это шаги расчетной сетки по времени и по пространственным переменным. Для учета сферической формы Земли использовался уменьшающийся шаг сетки по долготе при больших значениях широты. Обозначения F_{ij}^n представляют переменные на временном слое n, знак волны сверху соответствует промежуточным значениям, а F_{ij}^{n+1} соответствует переменным на временном слое n+1.

Удобный вычислительный шаблон позволил реализовать эффективный параллельный вычислительный конвейер, оптимизированный для архитектуры FPGA. При использовании микросхемы XilinxVirtex-7 VC709 в качестве сопроцессора персонального компьютера за один такт вычисляются значения искомых функций на 7 последовательных шагах по времени [20].

Быстрый расчет распространения волны по акватории

Аппаратная архитектура

Для реализации эффективного параллельного конвейера вычислений по схеме МакКормака были предложены архитектуры аппаратного ускорителя расчетов на базе FPGA [21], схемы которых показаны на рис. 1.

Предложенная реализация схемы МакКормака была протестирована. Высокая производительность достигается в случае, когда все данные решаемой задачи находятся во внутренней памяти устройства, и отсутствуют обращения к внешней памяти. Это является ограничением предлагаемой технологии. В случае решения системы уравнений мелкой воды (1) и использования микросхемы Xilinx Virtex-7 VC709 в качестве со-процессора персонального компьютера достигнуты следующие параметры: на расчетной сетке из 9601×6781 узлов для выполнения 36 тысяч шагов по времени с интервалом в 3 сек (то есть для моделирования 30 часов распространения волны) требуется всего 1352 сек (то есть 22.5 минуты).

Тестирование по точности решения

Сравнение получаемого численного решения как с известными точными решениями [22], так и с результатами применения программного пакета MOST [8, 9], показало их хорошее соответствие [23].

Рассмотрим распространение волны цунами от круглого источника в модельной области с наклонным профилем дна. Рассматривалась область размером 1000x1000 км с расчетной сеткой, имеющей пространственный шаг $\Delta x = \Delta y = 1000$ м. Центр кругового источника цунами радиусом 50 км располагался в центре области на расстоянии 300 км от нижней границы, где y = 0, на которой глубина обращается в нуль. В остальной части расчетной области глубина линейно увеличивалась по формуле

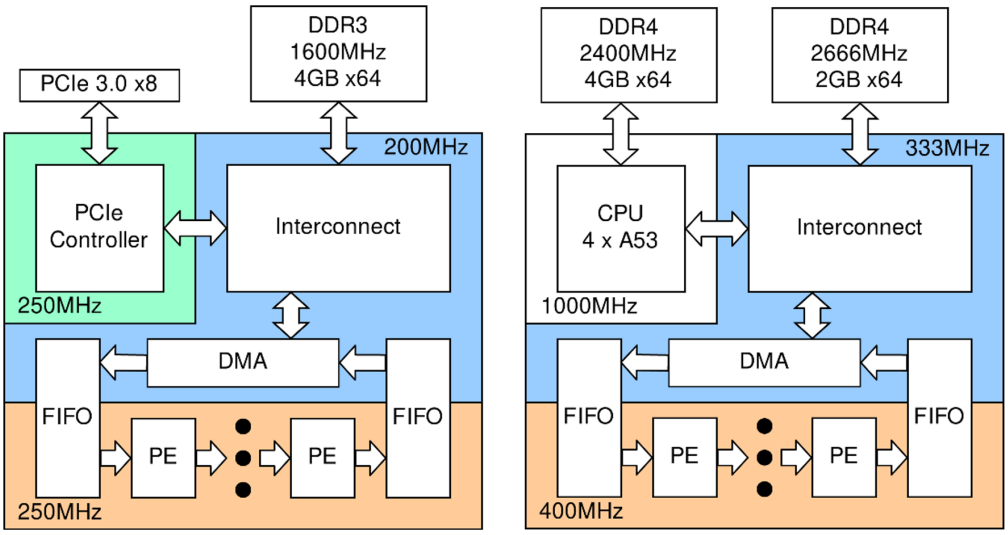


Рис. 1. *Архитектура специализированного вычислителя в составе персонального компьютера (слева) и в качестве независимого устройства (справа)*

D (x, y) = 0.01 y, где y — расстояние до нижней границы расчетной области. Начальное вертикальное смещение h в фокусе определяется по формуле

$$h(r) = 1 + \cos\left(\frac{\pi \cdot r}{r_0}\right), \quad 0 \le r \le r_0. \tag{2}$$

Здесь параметр r представляет собой расстояние до центра источника. Так, в центре очага начальное смещение водной поверхности составило +2 м. Такой очаг породил круговую волну высотой 0,95 м на расстоянии 50 км от центра. Именно такая высота волны на гребне начальной круговой волны радиусом 50 км использовалась при оценке амплитуды волны во всех точках региона по лучевому приближению [22]. Это распределение амплитуды волн цунами (полученное аналитическим методом) сравнивалось с аналогичными распределениями, полученными при численном моделировании распространения цунами с помощью алгоритмов MOST и MacCormack (рис. 2).

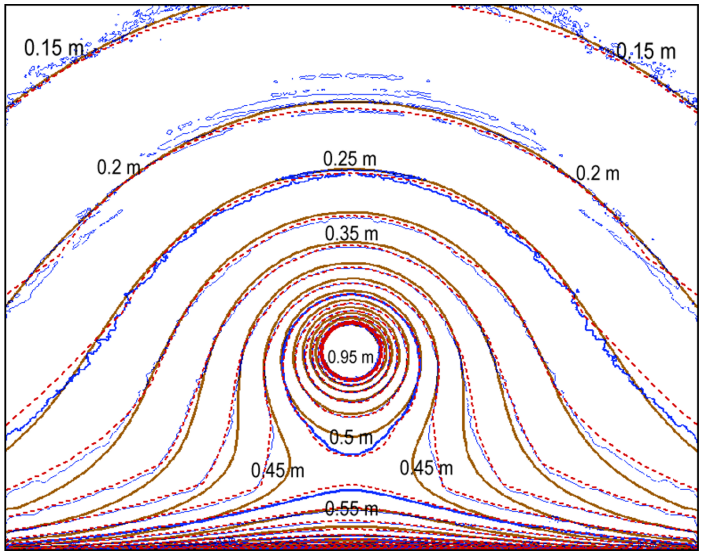


Рис. 2. *Изолинии распределения максимальных высот волн в области с наклонным дном, полученные: точным решением системы уравнений мелкой воды* (1) *(коричневые линии), численным решением на FPGA Calculator (красные пунктирные линии) и с помощью пакета MOST (синие линии). Расстояние до берега измеряется по вертикальной оси относительно нижней границы рисунка*

Исследование зависимости численного решения от цифровой батиметрии

Было исследовано также влияние данных батиметрии на результаты расчетов. Для расчета распространения волн цунами необходимо использовать доступную цифровую батиметрию. В настоящее время существует как минимум 3 глобальных батиметрических банка данных с детальностью 0,5 географических градуса (около 1 км в направлении север-юг). К ним относятся: GEBCO-30 [24], ETOPO-30 [25] и батиметрия Смита и Сандвелла [26, 27]. Кроме того, существуют региональные и местные цифровые наборы батиметрических данных с более высокой детализацией. В частности, Центр океанографических данных Японии (JODC) создал цифровой рельеф дна вокруг Японии [28] с разрешением 500 м на основе данных эхолокации. Эта цифровая батиметрия является наиболее достоверной, так как была разработана японскими специалистами с использованием огромного количества измерений глубин различными приборами. Поэтому батиметрия JODC используется в качестве эталона, с которым сравниваются другие батиметрии и результаты расчетов с их использованием.

Для проведения вычислительных экспериментов, связанных с батиметрией, была выбрана акватория, прилегающая к северо-восточному побережью острова Хонсю (Япония). В расчетах использовались цифровые батиметрические наборы данных детальностью 30 географических секунд из глобальных баз данных, перечисленных во введении. Сетка батиметрии размером 961x841 расчетных узлов охватывала регион, простирающийся от 140° до 148° восточной долготы и от 34° до 41° северной широты. Набор батиметрических данных, разработанный Японским центром океанографических данных (JODC) на основе эхолокационных данных вокруг Японских островов, представляется наиболее надежным для численных расчетов на шельфе Японии, поскольку он основан на расширенных данных зондирования. Поэтому в качестве справочных данных будут рассматриваться как сама цифровая батиметрия, так и результаты численного моделирования с ее использованием. Численные результаты, полученные с использованием альтернативных наборов батиметрических данных, будут сравниваться с этими данными. На рис. 3 представлена визуализация батиметрических данных JODC, дополненных рельефом суши. На средних широтах выбранного района длина пространственных шагов составляет 927 м в направлении юг-север и 740 м в направлении запад-восток.

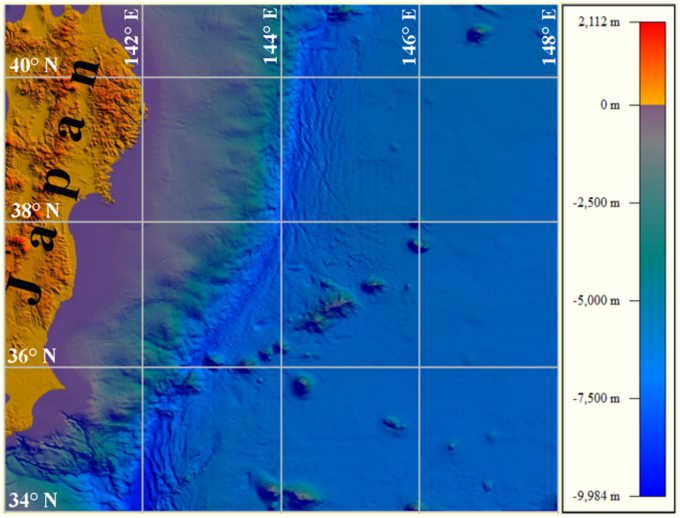


Рис. 3. *Рельеф дна рассматриваемой акватории, визуализированный на основе цифровой батиметрии, созданной JODC. Шкала в правой части показывает цветовое соответствие глубине*

Для численного моделирования возникновения и распространения цунами в выбранном регионе был выбран модельный источник в форме эллипса с длинами осей 180 и 80 км и максимальной высотой смещения водной поверхности в его центре, равной 1 м. На рис. 4 показано распределение максимумов высоты цунами во всем регионе по результатам расчета с использованием цифровой батиметрии JODC с разрешением 0,5 географических градусов.

На рис. 5a — 5b показано распределение разницы глубин (в процентах) между цифровыми батиметриями GEBCO и JODC (оконтурено темно-синим цветом) и относительная разница максимальных высот, рассчитанных по этим батиметриям, во всей расчетной области (оконтурено белым цветом). Изолинии очерчивают только те места, где эта разница превышает 10 %.

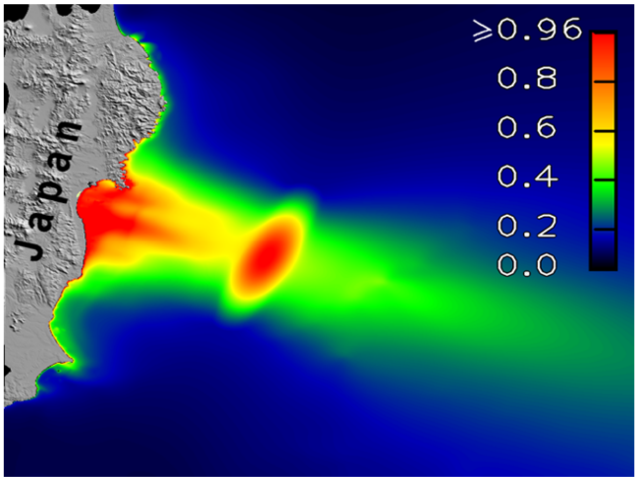


Рис. 4. *Распределение расчетных максимумов волны цунами, полученных с использованием батиметрии JODC*

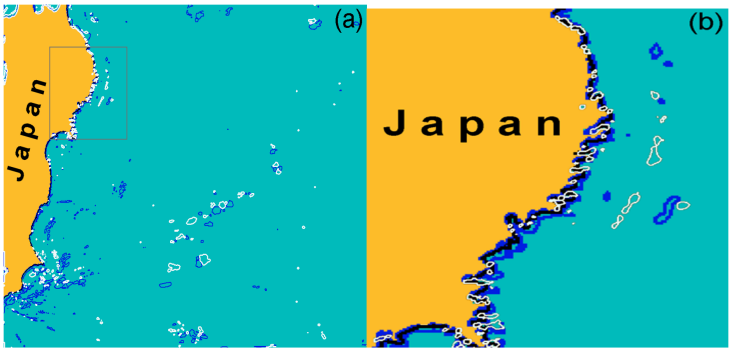


Рис. 5. *Изолинии окрашены в темно-синий цвет и ограничивают области, где разница между глубинами по данным батиметрии GEBCO и JODC превышает 10 %. Белые изолинии ограничивают области, где максимумы рассчитанных волновых форм по этим батиметриям также отличаются на 10 %. На рис. 5(b) более подробно показана подобласть, ограниченная прямоугольником на рис. 5(a)*

Применение метода вложенных сеток

Даже предложенная технология аппаратного ускорения расчетов не позволяет быстро получать результаты при использовании шагов сетки порядка 10 м, что требуется для более корректной оценки ожидаемых зон затопления. Поэтому была проведена проверка возможности применения метода вложенных сеток [29].

Расчетная область, представленная на рис. 6, охватывает географическую область от 137° до 143° восточной долготы и от 34° до 38° северной широты. Размерность расчетной сетки составляет 3000 x 2000 узлов с разрешением 0,002 угловых градуса (180 м в направлении запад-восток и 222,6 м в направлении юг-север). Для реализации метода вложенных сеток в численных экспериментах рассматриваются три уточняющие сетки: B1, B2 и B3. Углы подобласти B2 размером 1481×2401 узлов имеют следующие координаты по отношению к узлам исходной сетки: (1780, 380), (2150, 380), (1780, 980), (2150, 980). Отсчет вертикальных положений узлов сетки начинается с верхней границы B1. Шаг сетки в подобласти B2 в 4 раза меньше по сравнению с B1, что составляет 45 м по горизонтали и 55,6 м по вертикали.

Расчетная область B3 — это область 1302×871 узлов с разрешением сетки 0,000075 угловых градусов в обоих направлениях (примерно 8,35 м в направлении север-юг и 6,68 м в направлении запад-восток). Она охватывает район порта Онахама в префектуре Фукусима. Эта батиметрия была создана путем оцифровки бумажной морской карты Онахамы, выпущенной Японской гидрографической ассоциацией.

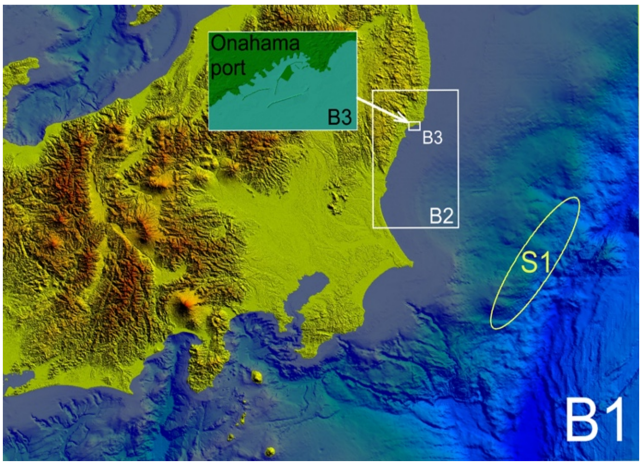


Рис. 6. *Рельеф дна расчетной области B1 к востоку от Японии. Эллипс S1 показывает примерное положение модельного источника цунами, использованного для численных экспериментов. Также указаны вложенные подобласти B2 и B3*

На рис. 7 (левая часть) показана расчетная поверхность океана (волна цунами) во всей расчетной области B1, включая подобласть B2, после 1200 секунд генерации волны.

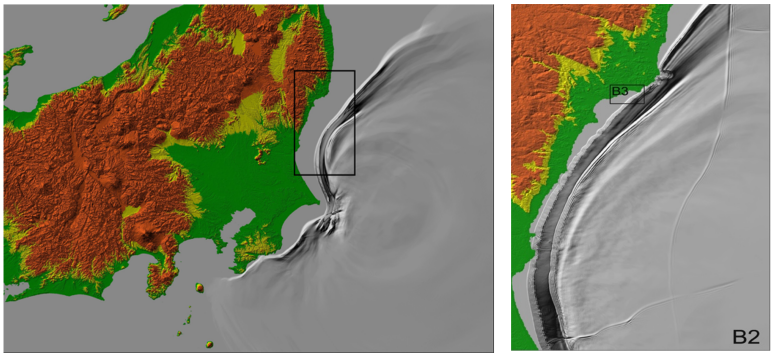


Рис. 7. *Морская поверхность через 1200 с после возникновения цунами как результат моделирования распространения волн в расчетной области B1 (левая часть). Поверхность океана через 2100 с после возникновения цунами как результат второго этапа моделирования распространения волн в расчетном домене B2 (правая часть)*

На рис. 7 (правая часть) представлена поверхность океана через 2100 с после начала процесса распространения. Здесь черный прямоугольник обозначает географическое положение расчетной области B3.

На последнем этапе численного эксперимента расчеты выполняются в районе B3 на оригинальной сетке, привязанной к цифровой батиметрии, разработанной авторами, с наименьшей длиной шага пространственной сетки (0,000075 углового градуса). Это примерно 8,35 м в направлении север-юг и 6,68 м в направлении запад-восток. Конечной целью данного исследования является получение распределения максимумов высоты цунами внутри подрайона B3 и вдоль береговой линии, рис. 8. Таким образом, продемонстрирована возможность вести расчет практически до береговой линии. На настоящий момент не решен вопрос о возможности проведения расчетов по трем сгущающимся сеткам на одной микросхеме, чтобы не тратить время на перезагрузку данных.

Коррекция параметров источника по мере поступления новых данных

Наличие инструмента быстрого расчета распространения волны цунами на базе персонального компьютера позволило предложить концепцию корректировки в режиме реального времени параметров источника по мере поступления новых данных от глубоководных датчиков гидростатического давления [21].

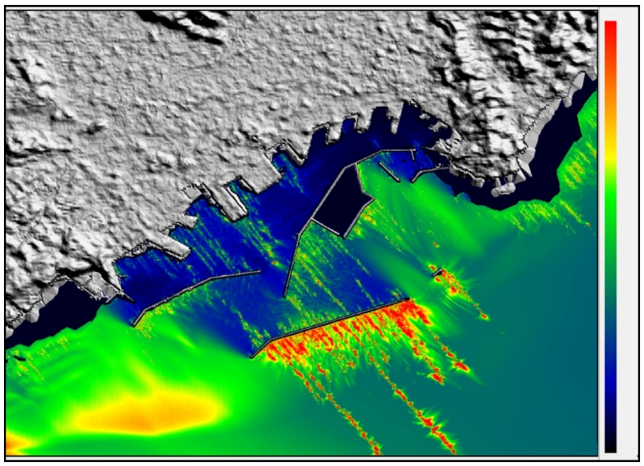


Рис. 8. *Распределение максимумов высоты волн цунами в расчетной области В3*

Рассматривается акватория всего Тихого океана, батиметрия которого представлена в [26]. Она построена с использованием данных спутниковой гравиметрии и имеющихся данных эхолокации (рис. 9).

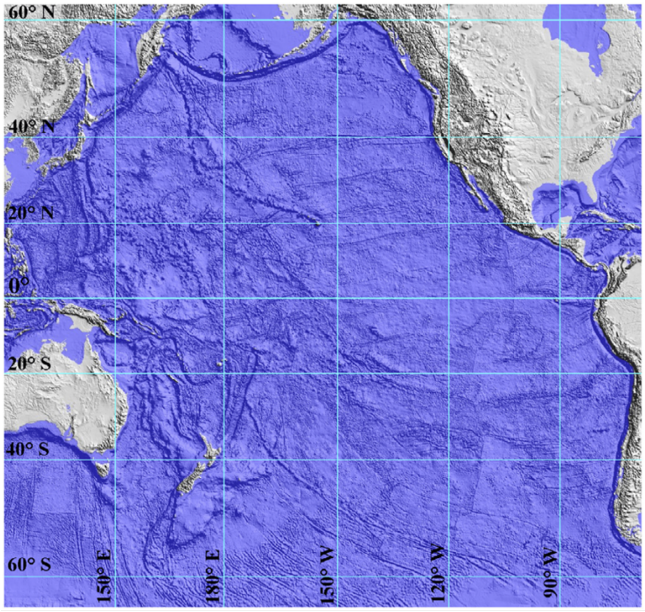


Рис. 9. *Визуализация донного рельефа расчетной области — Тихий океан*

Как показало численное моделирование распространения трансокеанского цунами на сетке размером 2581x2541 расчетных узлов, требуется всего лишь 3 минуты для расчета 36-ти часов распространения волны от очага, расположенного около п-ова Камчатка до берегов Южной Америки. Рис. 10 показывает распределение максимальных высот в акватории Тихого океана в результате землетрясения магнитудой 9.0 у берегов Камчатки. На этом же рисунке черными квадратиками обозначены места установки нескольких регистрирующих станций системы DARTS, через которые проходит волны цунами с наибольшей амплитудой.

При прохождении реальной волны над местами установки регистраторов DART можно видеть высоту цунами в данных точках. Получение результатов моделирования цунами за короткое время позволяет сравнивать вычисленные высоты волны с реально измеренными и корректировать параметры очага, добиваясь лучшего соответствия результатов расчета с натурными данными. Способы корректировки могут быть различными, но это предмет другого большого исследования, и здесь мы не будем этого касаться. Самый простой путь — это изменять высоту начального смещения в очаге.

Рассмотрим предлагаемый порядок действий, имеющих целью оценку высот цунами у удален-

ных побережий при возникновении мега-цунами, имеющим трансокеанский характер. Как уже говорилось, после обработки сейсмических данных можно определить местоположение и оценить размеры, пространственную ориентацию очага цунами и приблизительное максимальное смещение водной поверхности там. Рассмотрим начальное смещение поверхности (рис. 11), моделирующее очаг цунами в результате землетрясения магнитудой 9.0, произошедшее около полуострова Камчатка.

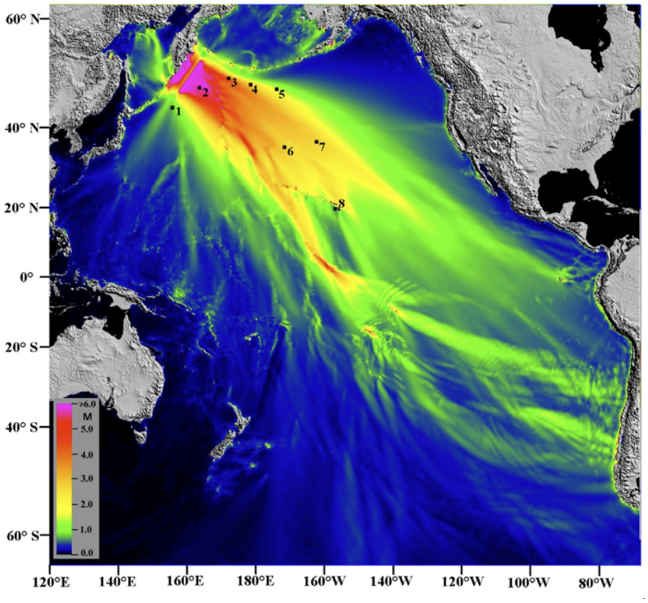


Рис. 10. *Распределение максимумов высот волн во всей расчетной области после 36 часов распространения цунами от очага, расположенного около полуострова Камчатка. Черными квадратиками с номерами отмечено расположение некоторых глубоководных регистрирующих станций системы DARTS*

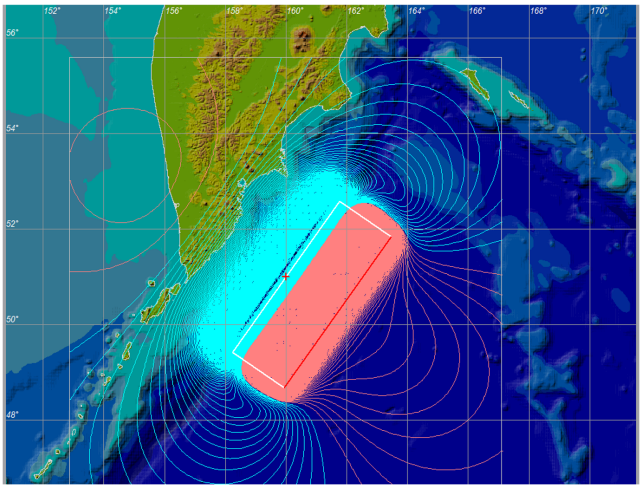


Рис. 11. *Местоположение эпицентра (отмечено маленьким красным крестом) и изолинии вертикального смещения донной поверхности в результате землетрясения магнитудой 9.0 около полуострова Камчатка. Красные изолинии показывают область поднятия, а голубые — область отрицательного смещения дна. Прямоугольником показана проекция плоскости разрыва в Земной коре*

Оно берется в качестве начальных условий для численного моделирования по модели мелкой воды (1). На рис. 12 показано положение волны цунами через 7200 сек после ее генерации, а также местоположение некоторых регистраторов уровня океана системы DARTS.

Из рис. 12 видно, что к этому моменту времени волна прошла над регистраторами 1-4 и приблизилась к DART-станции номер 5. В регистратор #2, расположенный ближе других к очагу цунами,

волна вступила на 240-й секунде, а окончание полного первого периода зафиксировано на 2800-й секунде. Соответственно, в регистратор #3 волна вступила на 1960-й секунде, а ее первый период завершился на 3600-й секунде. Регистратором #1, расположенным сбоку от очага, вступление цунами зарегистрировано на 1088-й секунде, а момент окончания полного первого периода волны наступил на 2624-й секунде после начала моделирования. При этом в отличие от регистраторов 1 и 2 здесь волновой сигнал начинается с отрицательной фазы. После прохода реальной волны цунами над DART-станциями 2, 3 и 1 можно сравнить реальные и вычисленные мареограммы и провести коррекцию положения очага и величины начального смещения в очаге. Можно это делать при проходе цунами над каждым последующим регистратором и затем в течение 1–2 минут повторять численное моделирование с подкорректированным очагом.

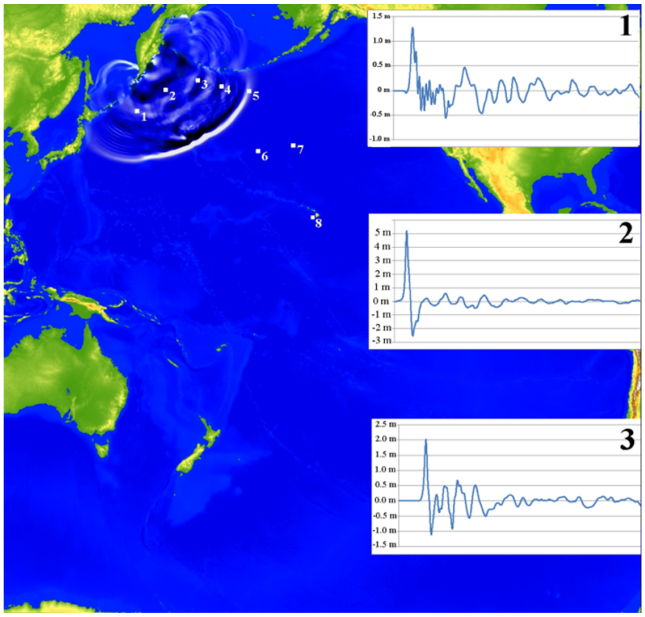


Рис. 12. *Положение волнового фронта в момент времени 7200 сек. В правой части рисунка приведены рассчитанные мареограммы от 0 до 30000 сек в точках установки регистраторов #1–3*

ЛИТЕРАТУРА

1. Thomas D., Howes L., Luk W. A Comparison of CPUs, GPUs, FPGAs, and MPPAs for Random Number Generation. *Proc. of FPGA*. 2009.
2. Wei Y., Chamberlin C., Titov V. V. et al. Modeling of the 2011 Japan Tsunami: Lessons for Near-Field Forecast. *Pure Appl. Geophys*. 2012;170(6–8):1309–1331. DOI: 10.1007/s00024-012-0519-z.
3. *National Data Buoy Center*. Режим доступа: https://www.ndbc.noaa.gov/obs.shtml.
4. *DART (Deep-ocean Assessment and Reporting of Tsunamis).* Режим доступа: http://nctr.pmel.noaa.gov /Dart/.
5. *Research and Development Center for Earthquake and Tsunami*. Режим доступа: https://www.jamstec.go.jp/donet/e/.
6. *S-net Scheme.* Режим доступа: https://www.seafloor.bosai.go.jp/S-net/. (In Japanese)
7. *S-net Coordinates.* Режим доступа: https://www.seafloor.bosai.go.jp/st_info/. (In Japanese)
8. Titov V. V., Gonzalez F. I. Implementation and Testing of the Method of Splitting Tsunami (MOST) Model. *NOAA Technical Memorandum ERL PMEL-112*. NOAA: Washington, DC, USA; 1997.
9. Gica E., Spillane M., Titov V. et al. Development of the Forecast Propagation Database for NOAA's Short-Term Inundation Forecast for Tsunamis (SIFT). *NOAA Technical Memorandum.* NOAA: Washington, DC, USA; 2008. Режим доступа: https://nctr.pmel.noaa.gov/Pdf/brochures/sift_Brochure.pdf.
10. Wang X., Power, W. L. COMCOT: A Tsunami Generation Propagation and Run-Up Model. *GNS Science*. Lower Hutt City, New Zealand; 2011.
11. Shuto N., Goto C., Imamura F. Numerical Simulation as a Means of Warning for Near Field Tsunamis. *Coast. Eng. Jpn*. 1990;33:173–193.

12. Yalciner A. C., Alpar B., Altinok Y. et al. Tsunamis in the Sea of Marmara: Historical Documents for the Past, Models for Future. *Mar. Geol*. 2002;190:445–463.
13. Zaytsev A., Kurkin A., Pelinovsky E. et al. Numerical Tsunami Model NAMI_DANCE. *Sci. Tsunami Hazards*. 2019;38:151–168.
14. Lax P. D., Richtmyer R. D. Survey of the Stability of Linear Finite Difference Equations. *Commun. Pure Appl. Math*. 1956;9:267–293.
15. Liang Q., Hou J., Amouzgar R. Simulation of Tsunami Propagation Using Adaptive Cartesian Grids. *Coastal Engineering J.* 2015;57(4):150729195708001. DOI: 10.1142/S0578563415500163.
16. Lavrentiev M., Lysakov K., Marchuk An. et al. Fundamentals of Fast Tsunami Wave Parameter Determination Technology for Hazard Mitigation. *Sensors*. 2022;22:7630. DOI: https://doi.org/10.3390/s22197630.
17. Stoker J. J. *Water Waves. The Mathematical Theory with Applications.* New York: Interscience Publishers; 1957.
18. Kowalik Z., Murty T. S. *Numerical Modeling of Ocean Dynamics.* World Scientific; 1993.
19. MacCormack R. W., Paullay A. J. Computational Efficiency Achieved by Time Splitting of Finite Difference Operators. *AIAA*. 1972:72–154.
20. Lavrentiev M., Lysakov K., Marchuk An. et al. Hardware/Software Solution for Low Power Evaluation of Tsunami Danger. *J. Low Power Electronics and Appl*. 2022;12(1):6. DOI: 10.3390/jlpea12010006.
21. Lavrentiev M., Marchuk An., Oblaukhov K. Low Power Computation of Transoceanic Wave Propagation for Tsunami Hazard Mitigation. *Ocean Modelling*. 2024;192(102459). DOI: 10.1016/j.ocemod.2024.102459.
22. Marchuk A. G. Estimating Tsunami Wave Height over a Sloping Bottom in the Ray Approximation. *Numerical Analysis and Applications*. 2015;8(4):304–313.
23. Lavrentiev M., Marchuk An., Oblaukhov K. et al. Comparative Testing of MOST and Mac-Cormack Numerical Schemes to Calculate Tsunami Wave Propagation. *Journal of Physics: Conference series.* 2020;1666(1):012028. DOI: 10.1088/1742-6596/1666/1/012028.
24. *GEBCO Homepage*. Режим доступа: https://www.gebco.net/data_and_products/gridded_bathymetry_data/.
25. *NOAA National Centers for Environmental Information.* Режим доступа: https://www.ncei.noaa.gov/products/etopo-global-relief-model#, last accessed 2024/09/15.
26. Smith W. H. F., Sandwell D. T. Conventional Bathymetry, Bathymetry from Space, and Geodetic Altimetry. *Oceanography*. 2004;17(1):8–3. DOI: https://doi.org/10.5670/oceanog.2004.63.
27. *Scripps Institution of Oceanography, University of California San Diego*. Режим доступа: https://topex.ucsd.edu/WWW_html/srtm30_plus.html.
28. *JODC Data On-line Service System*. Режим доступа: https://www.jodc.go.jp/vpage/depth500_file.html.
29. Lavrentiev M., Lysakov K., Marchuk An. et al. FPGA Based Tsunami Wave Propagation Calculator. *Journal of Physics: Conference Series*. 2021;1789(1):012011.

ИНТЕРПОЛЯЦИЯ ОБЛАКА ТОЧЕК (НЕРЕГУЛЯРНОЙ СЕТКИ) С ПОМОЩЬЮ КУБИЧЕСКИХ ТРЕУГОЛЬНИКОВ БЕЗЬЕ (CURVED POINT-NORMAL TRIANGLES)

А. Л. Леготин[а], Г. К. Виноградов[б]

Федеральное государственное автономное учреждение «Федеральный научный центр Научно-исследовательский институт системных исследований Национального исследовательского центра «Курчатовский институт», г. Москва, Российская Федерация

[а] ✍ *andrlegotin@yandex.com* [б] *vinogradovgk@mail.ru*

Аннотация: интерполяция из облака точек (данных, расположенных нерегулярным образом) важна для многих аспектов научной деятельности и практического применения. В статье рассматривается алгоритм интерполяции на нерегулярной сетке с помощью параметризованного кубического треугольника Безье. Уделяется внимание проблеме выбора нормалей вершин поверхности.

Ключевые слова: интерполяция, облако точек, нерегулярная сетка, треугольник Безье, нормаль вершин, криволинейный треугольник, PN треугольники.

Благодарности: авторы выражают благодарность К. А. Медникову за предоставленные данные эллипсометрических измерений распределения толщины оксида кремния.

Для цитирования: Леготин А. Л., Виноградов Г. К. Интерполяция облака точек (нерегулярной сетки) с помощью кубических треугольников Безье (curved point-normal triangles). *Успехи кибернетики*. 2025;6(1):35–46.

Поступила в редакцию: 04.12.2024. *В окончательном варианте: 03.02.2025.*

POINT CLOUD (IRREGULAR GRID) INTERPOLATION WITH THE CURVED POINT-NORMAL TRIANGLES

A. L. Legotin[a], G. K. Vinogradov[b]

Scientific Research Institute for System Analysis of the National Research Centre "Kurchatov Institute", Moscow, Russian Federation

[a] ✍ *andrlegotin@yandex.com*, [b] *vinogradovgk@mail.ru*

Abstract: we examine interpolation from a point cloud, where data is arranged in an irregular grid, as it plays a crucial role in both scientific research and practical applications. We developed an algorithm for interpolating irregular data using a parameterized curved point-normal triangle. We also address the challenge of selecting vertex normals for the surface.

Keywords: interpolation, point cloud, irregular grid, Bezier triangle, vertex normal, curved point-normal triangles, PN triangles.

Acknowledgements: the authors would like to thank K. A. Mednikov for providing the ellipsometric measurements of the silicon oxide thickness distribution.

Cite this article: Legotin A. L., Vinogradov G. K. Point Cloud (Irregular Grid) Interpolation with the Curved Point-Normal Triangles. *Russian Journal of Cybernetics*. 2025;6(1):35–46.

Original article submitted: 04.12.2024. *Revision submitted: 03.02.2025.*

Введение

Визуальное представление необходимо для общей оценки данных. Если имеются искажения, то это в свою очередь отразится на интерпретации результатов.

При измерении толщин тонких пленок на круглой подложке формируется набор точек. Программное обеспечение производителя измерительного оборудования на основе этих данных строит поверхность, 2D вид которой представлен на рис. 1.

Было замечено, что карта высот не совпадает во многих точках. Некоторые точки находятся не в своей цветовой зоне и в целом границы цветовых зон сдвинуты относительно ожидаемых.

Для проверки был произведен расчет по точкам, расположенным вдоль диаметров (в плоскостях 0 и 90 градусов в системе координат производителя).

Результат сравнения приведен на рис. 2, где «Текущий» — градиент взят из 2D карты производителя (стрелками показано место), «Расчетный» — градиент, полученный из точек.

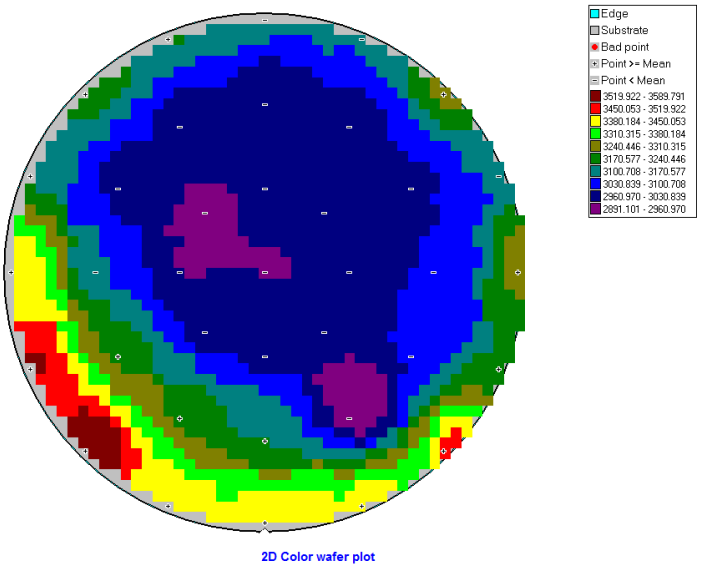


Рис. 1. *Карта высот (ПО производителя)*

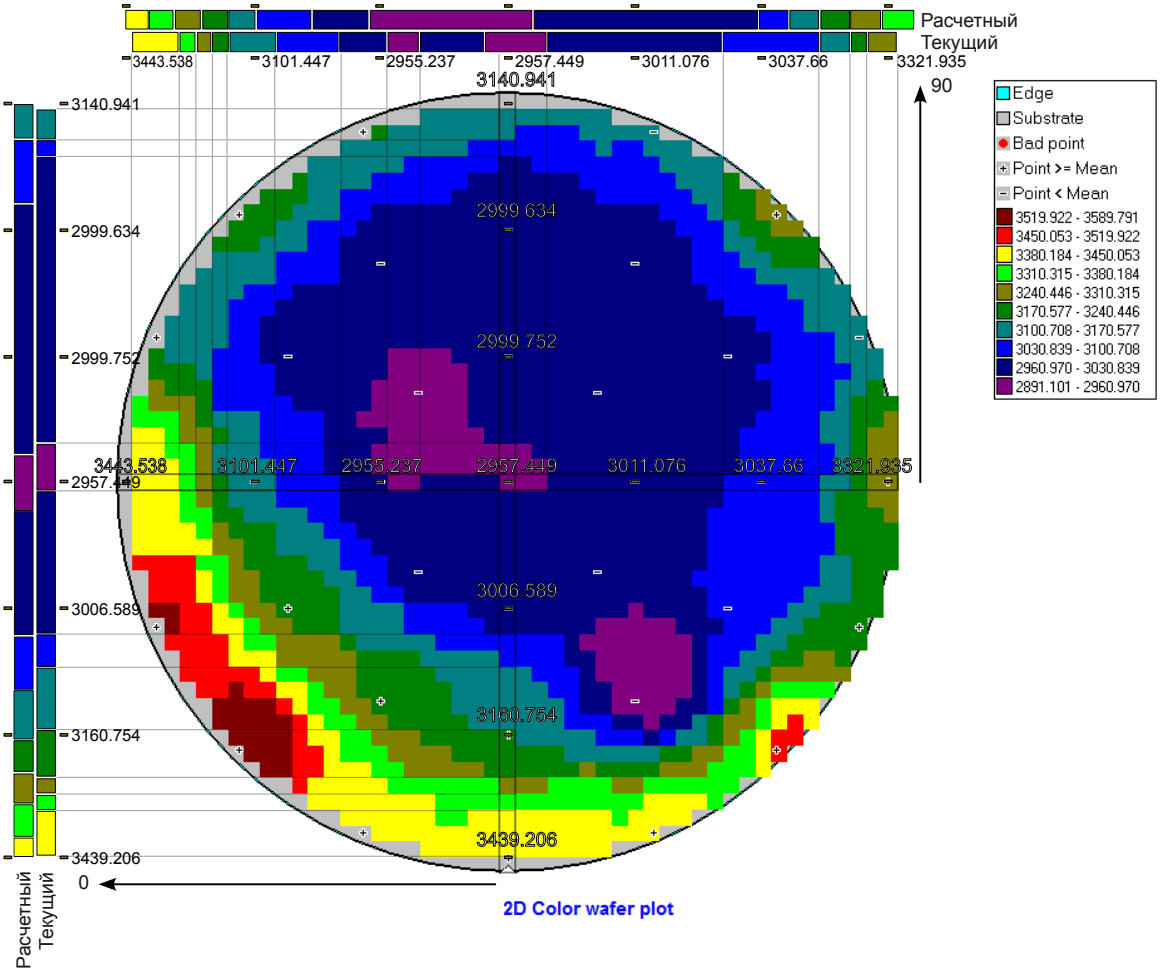


Рис. 2. *Сравнение с расчетными градиентами на карте высот*

Ожидаемые профили поверхности показаны на рис. 3 (по вертикали — высоты, по горизонтали — координаты в мм).

По всей длине градиентов видны расхождения между расчетным и текущим положением цветовых зон. Особо стоит заметить артефакт между точками 2955.237 и 2957.449, где на карте высот наблюдается холм, в то время как по расчету должна быть впадина (на профиле 90 градусов в районе

координат –20 мм) в случае сплайн интерполяции или практически горизонтальная площадка в случае линейной интерполяции.

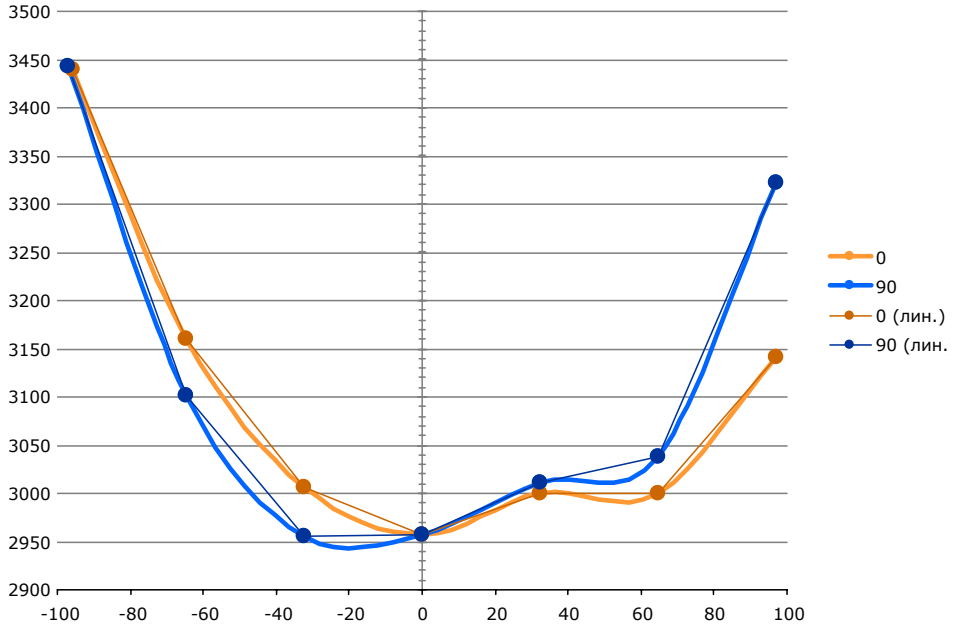


Рис. 3. *Расчетные профили поверхности*

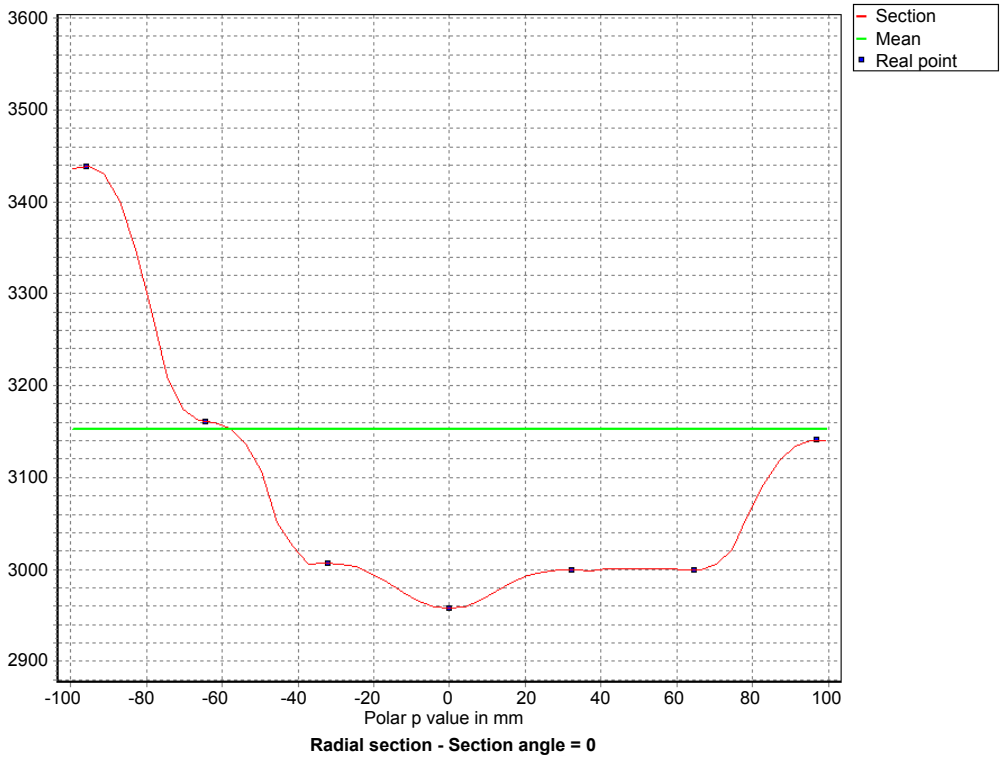


Рис. 4. *Профиль поверхности в сечении 0 градусов (ПО производителя)*

Алгоритм построения поверхности, применяемый производителем, неизвестен, но многое можно понять из дополнительных данных. Таковым может стать профиль поверхности в сечении 0 градусов, который показан на рис. 4 (профиль сгенерирован ПО производителя измерительного оборудования).

В каждой точке наблюдается «ступенька», вход и выход кривой у каждой точки всегда строго горизонтален несмотря на ярко выраженные области наклона.

То есть поверхность формируется из плоских площадок вблизи точек и сглаженных переходов от точки к точке, как следствие, наклон между точками ближе к средней зоне оказывается более крутым. Из-за подобного расчета поверхности происходят сдвиги градиентов. Наблюдаемые ореолы вокруг точек (круговые выступы особенно заметны в верхней половине на границе синей и темно-синей областей, см. рис. 2) появляются в силу этих же причин. Кроме этого, в районе координат –40 (рис. 4) есть излом с небольшим уклоном вниз. Все это говорит о том, что применяемые алгоритмы работают не должным образом.

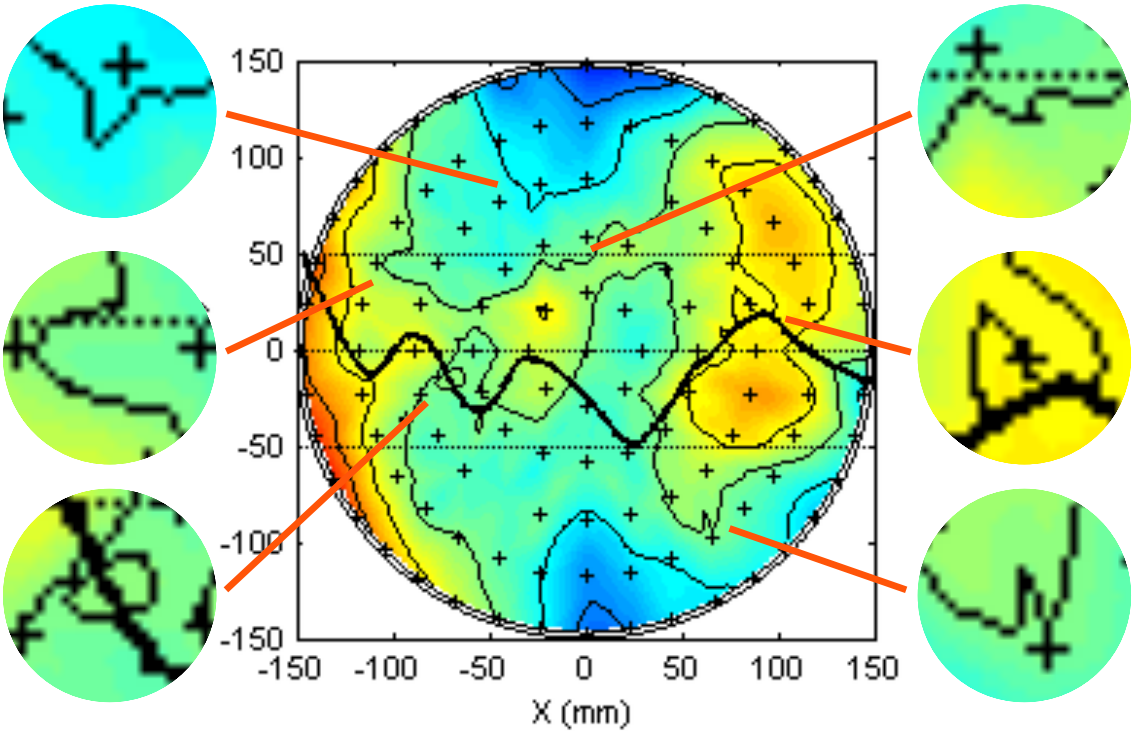


Рис. 5. *Карта высот (сторонняя программа) и проблемы в реализации алгоритмов*

Имеется также сторонняя программа по визуализации подобных распределений. К сожалению, и там присутствуют проблемы в реализации алгоритмов. На рис. 5 показаны увеличенные области с артефактами. Градиент в этих местах слабый, ожидается довольно ровная поверхность, на изолиниях между тем присутствуют острые углы и узкие выступы, которых быть не должно при интерполяции редких точек (в том числе и в случае больших перепадов высот).

Такие результаты визуализации искажают полученные данные, поэтому возникла необходимость улучшить графическое представление.

Набор данных имеет вид вложенных концентрических колец с центральной точкой. Например, рассмотренные выше данные имеют от центральной точки кольцо из 8 точек, затем из 12 и 16 (см. рис. 6). Количество точек может быть разным в зависимости от требуемой плотности измерений.

Из-за того, что точки расположены подобным образом, нельзя воспользоваться методами построения поверхности на основе прямоугольной сетки. Так как кольца имеют разное кол-во точек, затруднительно и применение методов с полярными координатами.

В этой статье представлен способ построения сглаженной поверхности (интерполяции) для более общего случая — для нерегулярно распределенных данных (облака точек), что будет применимо и здесь.

Описанный ниже алгоритм относится к методам локальной интерполяции. Локальные методы интерполяции опираются на ближайшие точки, при этом в зависимости от метода выбор и количество точек разнится. В данном алгоритме используется сетка на основе треугольников, интерполянтом служит параметризованный кубический треугольник Безье.

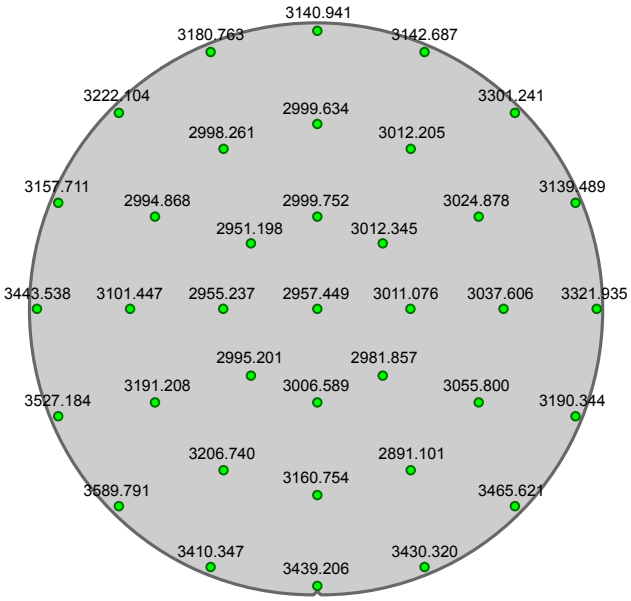


Рис. 6. *Карта точек*

Поверхность Безье

Уравнение поверхности Безье порядка (n,m) в общем виде описывается следующим образом [1, 2]:

$$p(u,v) = \sum_{i=0}^{n} \sum_{j=0}^{m} B_i^n(u) B_j^m(v) P_{i,j} \tag{1}$$

где u,v — параметрические координаты, P — контрольные точки, B — многочлены Бернштейна.

$$B_i^n(u) = \frac{n!}{i!(n-i)!} u^i (1-u)^{n-i} \tag{2}$$

Поверхность Безье порядка (n,m) потребует задания $(n+1)\cdot(m+1)$ контрольных точек.

Частный вид этой поверхности — треугольник Безье [2, 3] порядка n:

$$b(u,v) = \sum_{i+j+k=n} \frac{n!}{i!j!k!} u^i v^j w^k a^i b^j c^k \tag{3}$$

где u, v, w — параметрические (барицентрические) координаты ($u,v,w \geq 0$; $u+v+w=1$); $a^i b^j c^k$ — контрольные точки.

Для треугольника Безье n-го порядка необходимо задание $(n+1)\cdot(n+2)/2$ контрольных точек.

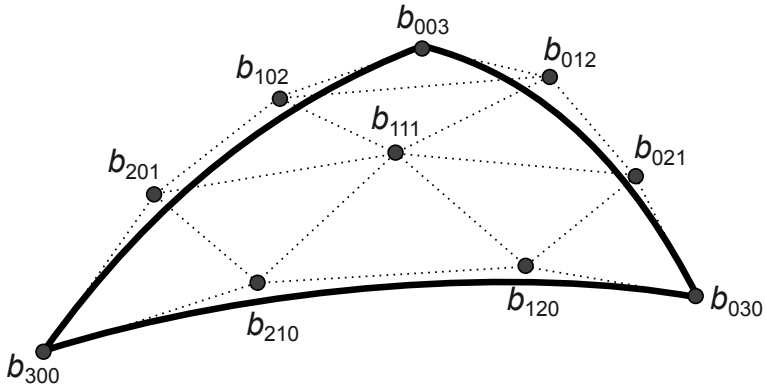


Рис. 7. *Контрольные точки кубического треугольника Безье*

Кубический треугольник Безье

Для кубического треугольника $n = 3$. Уравнение поверхности в этом случае можно представить, как:

$$b(u,v) = \sum_{i+j+k=3} b_{ijk} \frac{3!}{i!j!k!} u^i v^j w^k \tag{4}$$

где b_{ijk} — коэффициенты.

Для построения такого треугольника нужно задать 10 контрольных точек. Из них 3 относятся к вершинам, 6 — к касательным и 1 — к центральной точке (см. рис. 7).

Поверхность из треугольников Безье

С целью обеспечения неразрывности и гладкости поверхности, составленной из треугольников Безье, коэффициенты касательных смежных сторон двух соседних треугольников должны быть равны и все касательные всех треугольников, проведенных от общей вершины, должны лежать в одной плоскости [4].

То есть необходимо:

1) найти для каждой вершины общие треугольники;
2) вычислить касательные и свести их для каждой вершины в одну плоскость.

Если п. 1 алгоритмически легко реализуем, то в п. 2 встает проблема вычисления касательных и выбора их общей плоскости, так как поверхность мы еще строим и как раз хотим получить направление изгиба поверхности в соответствующих точках.

Эта проблема решается следующим образом. Приняв замечание, что нормаль общей плоскости касательных должна совпадать с нормалью искомой поверхности в этой точке, можно прийти к выводу, что п. 2 сводится к нахождению нормалей вершин искомой поверхности и соответственно коэффициенты касательных и центральной точки в конечном итоге должны зависеть от нормалей вершин.

Для наших целей удобна параметризация, приведенная в статье [5]. Данный тип параметризации был разработан для графического применения, стал широко использоваться в этой среде, для обработки информации практически не применялся, между тем подходит для интерполяции и тем важнее опробовать его в этом качестве.

Для построения криволинейного треугольника в этой параметризации достаточно задать расположение точек треугольника в пространстве и векторов нормалей в этих точках.

Уравнение поверхности:

$$\begin{aligned} b(u,v) &= \sum_{i+j+k=3} b_{ijk} \frac{3!}{i!j!k!} u^i v^j w^k = \\ &= b_{300}w^3 + b_{030}u^3 + b_{003}v^3 + b_{210}3w^2u + b_{120}3wu^2 + b_{201}3w^2v + b_{021}3u^2v + b_{102}3wv^2 + \\ &\quad + b_{012}3uv^2 + b_{111}6wuv \end{aligned} \tag{5}$$

где u, v, w — барицентрические координаты ($u,v,w \geq 0; \quad u + v + w = 1$).

Коэффициенты треугольника Безье:

$$\begin{aligned}
&b_{300} = P_1 \\
&b_{030} = P_2 \\
&b_{003} = P_3 \\
&w_{ij} = (P_j - P_i) \cdot N_i \\
&b_{210} = (2P_1 + P_2 - w_{12}N_1)/3 \\
&b_{120} = (2P_2 + P_1 - w_{21}N_2)/3 \\
&b_{021} = (2P_2 + P_3 - w_{23}N_2)/3 \\
&b_{012} = (2P_3 + P_2 - w_{32}N_3)/3 \\
&b_{102} = (2P_3 + P_1 - w_{31}N_3)/3 \\
&b_{201} = (2P_1 + P_3 - w_{13}N_1)/3 \\
&E = (b_{210} + b_{120} + b_{021} + b_{012} + b_{102} + b_{201})/6 \\
&V = (P_1 + P_2 + P_3)/3 \\
&b_{111} = E + (E - V)/2
\end{aligned}$$

где P_1, P_2, P_3 — координаты точек треугольника; N_1, N_2, N_3 — нормализованные векторы нормали (т.е. $\|N_i\| = 1$); $b_{300}, b_{030}, b_{003}$ — коэффициенты вершин; $b_{210}, b_{120}, b_{021}, b_{012}, b_{102}, b_{201}$ — коэффициенты касательных; b_{111} — коэффициент центральной точки.

Алгоритм построения поверхности (интерполяция облака точек)

Входными данными для построения поверхности из треугольников Безье данным способом служат вершины и нормали вершин. Нормали вершин можно вычислять в ходе построения, так и задавать, если, например, известно, как должна пройти поверхность через конкретные вершины, для граничных условий и т.д.

Основные шаги:

1) создание линейно-интерполированной поверхности из облака точек (триангуляция Делоне);
2) вычисление нормалей для каждой вершины поверхности;
3) замена каждого плоского треугольника криволинейным треугольником Безье.

Построение линейной интерполяции из облака точек, а также определение нормалей вершин зависит от характера данных. Остановимся на некоторых аспектах подробнее.

1. Триангуляция Делоне может быть проведена различными методами. Достаточно большое количество алгоритмов и видов триангуляций описано в [6]. В зависимости от структуры данных может стать более предпочтителен тот или иной алгоритм.

Далее в качестве примера мы будем рассматривать нерегулярную сетку, которая не имеет элементов поверхности с отрицательным наклоном (Z координата нормали каждой точки поверхности положительна). В этом случае можно применить триангуляцию Делоне для двумерного случая (временно игнорируется координата Z) и получить трехмерную поверхность, состоящую из плоских треугольников (для иллюстрации см. рис. 9, слева — режим каркаса, справа — визуализация).

2. Нормаль вершины можно вычислить как среднее взвешенное одной из характеристик близлежащих элементов. Например, сумма нормалей соседних плоских треугольников с весами в виде углов [7], длин ребер, их комбинаций, например, площадь и т.п. Нормаль вершины можно так же найти по методу наименьших квадратов из ближайших соседних вершин, построить в локальной области полином и т.д.

В зависимости от задач тот или иной алгоритм может стать более или менее предпочтительным.

Сравнение некоторых алгоритмов вычисления нормалей вершин проведено в [8]. К сожалению, не все результаты сравнительного анализа алгоритмов авторами включены в суммарную таблицу, есть вопросы и по критериям выводов, как проявили себя те или иные алгоритмы. Например, если на Sphere TP и Torus TP алгоритм MWA явно показывает себя с не лучшей стороны, то на других графиках разница не значительна (особенно Perlin MT и turbulence MT, где кривые практически совпадают), но в таблице некоторые алгоритмы все равно получили оценки «хорошо» и «плохо».

Стоит отметить некоторые особенности подобных алгоритмов.

Если за весовой коэффициент взят угол, то треугольник, имеющий больший угол раствора ребер у данной вершины, будет иметь соответственно и больший вес вне зависимости от его размеров. То есть, если к вершине примыкает подобный треугольник с короткими ребрами (компактный), нормаль его будет иметь более сильное влияние по сравнению с нормалями других возможно более крупных треугольников с меньшими углами раствора у данной вершины. Не всегда это может быть приемлемо.

В общем виде весовой коэффициент должен зависеть и от угла, и от длин ребер в некой их комбинации. Коэффициент, зависящий от площади, этому условию удовлетворяет, но существует геометрия, когда и такой весовой коэффициент будет не лучшим выбором.

Рассмотрим на примере куба, грани которого разбиты на плоские треугольники как на рис. 8. К центральной вершине (по рисунку) от левой грани прилегает два треугольника, поэтому нормаль левой стороны будет иметь вдвое больший вес, чем нормаль правой (или верхней). Алгоритм с весовым коэффициентом, зависящим только от угла, в этом случае будет точным.

Куб для иллюстрации, где это проявляется максимально. Такая ситуация вполне возможна на практике. Триангуляция квадрата (прямоугольника) неопределенна — есть два равнозначных варианта, грани могут быть разбиты на треугольники различным образом. В общем случае, если некоторые точки данных составляют четырехугольник и чем ближе их расположение к прямоугольнику, тем сильнее особенности подобных алгоритмов.

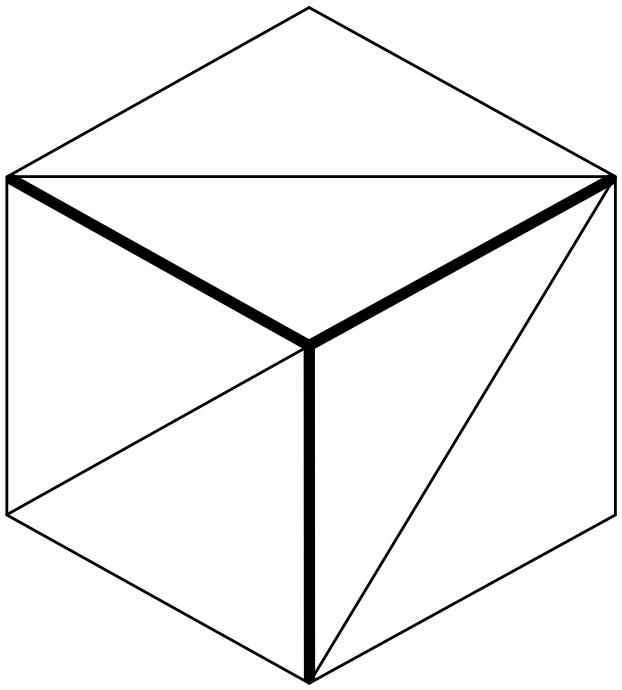


Рис. 8. *Пример разбиения граней на треугольники*

В других алгоритмах, основанных на весовых коэффициентах, те или иные эффекты так же могут проявляться в зависимости от метода (комбинации характеристик) и геометрии (исходных данных).

На случайном наборе лучшие результаты, если судить по графикам в [8], показывает алгоритм MWAAT, где весовой коэффициент — площадь. В качестве примера рассмотрим его.

Направление нормали треугольника ABC можно вычислить векторным произведением двух его ребер (например, $N = AB \times AC$).

Замечание: векторное произведение некоммутативно, необходимо следить за порядком выбора ребер, чтобы не получить в каких-то случаях инвертированную нормаль.

Принимая во внимание, что модуль векторного произведения ребер треугольника пропорционален площади, можно увидеть, что (ненормализованный, вычисленный через векторное произведение) вектор нормали уже включает в себя нужный весовой коэффициент. В силу этого, суммируя такие вектора нормалей всех прилежащих к вершине треугольников, получим средневзвешенную нормаль вершины (которую затем необходимо нормализовать).

Кратко данный способ вычисления нормали вершин выглядит так:

1) посчитать нормали всех треугольников;
2) для каждой вершины вычислить средневзвешенную нормаль от всех прилежащих к вершине треугольников.

3. На этом шаге каждый плоский треугольник заменяется на криволинейный треугольник Безье и с учетом уже известных нормалей, используя (5), производится вычисление значений в нужных точках.

Результат интерполяции по вышеприведенному алгоритму см. на рис. 10.

В данном алгоритме удобно производить оптимизацию вычислений. На основе нормалей и вершин считаются коэффициенты и далее для нужной точки внутри треугольника все необходимые точки можно вычислить быстрее. Часть коэффициентов будут общими для смежных треугольников, в оптимизации можно учесть и это.

Визуальное представление

Требуемого уровня детализации можно достичь разбиением поверхности Безье и вычислением каждой точки разбиения.

Разбиение криволинейного треугольника на простейшие плоские элементы (тесселяция) можно

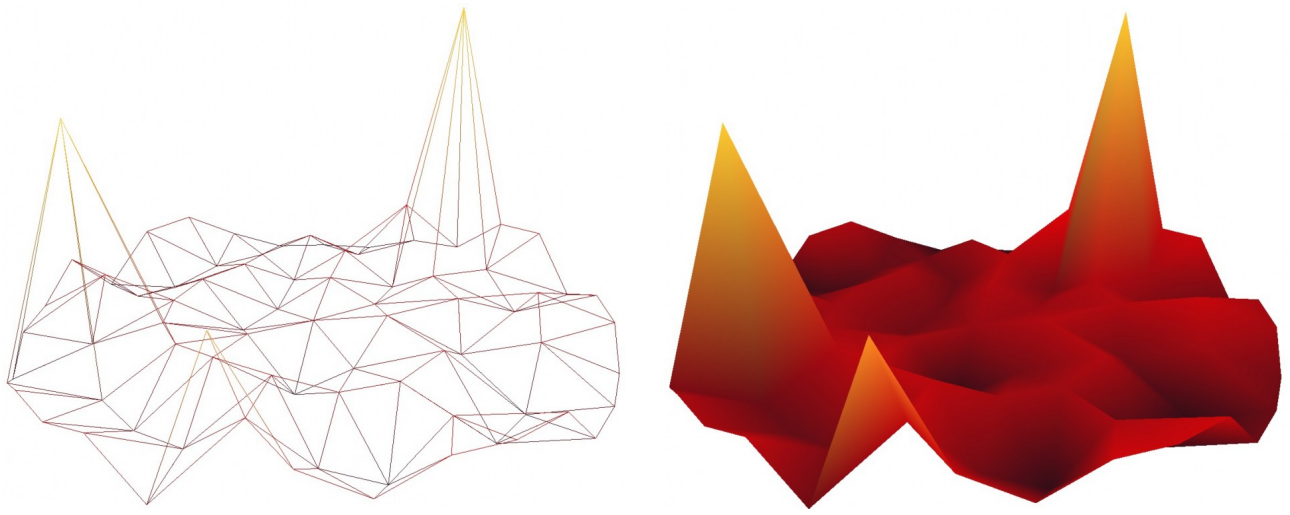


Рис. 9. *Линейная интерполяция поверхности*

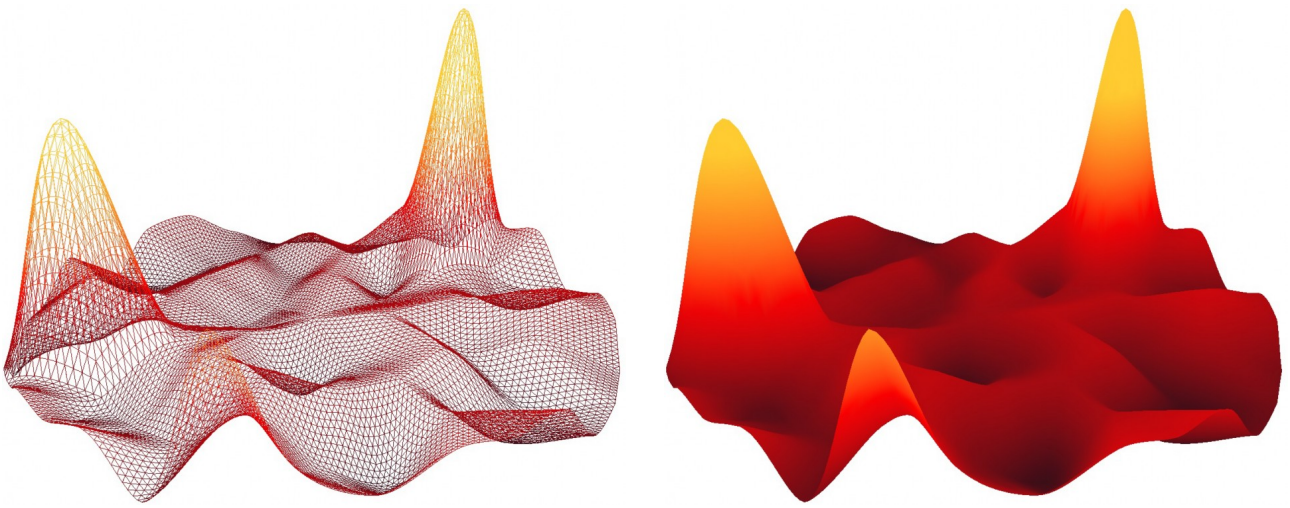


Рис. 10. *Интерполяция треугольниками Безье*

производить разными методами. Чаще применяют сетку из треугольников, возможно разбиение на четырехугольники и т.д.

В соответствии с выбранным методом формируется набор точек на поверхности криволинейного треугольника и вычисляются координаты каждого из новых элементов поверхности.

Так как уравнение использует параметрические координаты, поверхность Безье легко разделяется на нужную сетку (например, заданием шага u/k,v/k, где $k = 2,3,4$ и т. д.). Сделав так для каждого треугольника Безье, в результате получится состоящая из плоских фигур поверхность с необходимым уровнем детализации.

Результаты

На рис. 11 показаны сечения интерполяционной поверхности в тех же плоскостях, что и ранее (0 и 90 градусов).

Идеального совпадения профилей сечения с расчетными кривыми не ожидалось, так как на профиль поверхности оказывают влияние другие ближайшие точки в пространстве, а расчетные кривые строились только по находящимся в плоскости сечения точкам. Вдобавок к этому механизмы сглаживания разные, тем не менее профили сечения довольно точно совпали с расчетными кривыми.

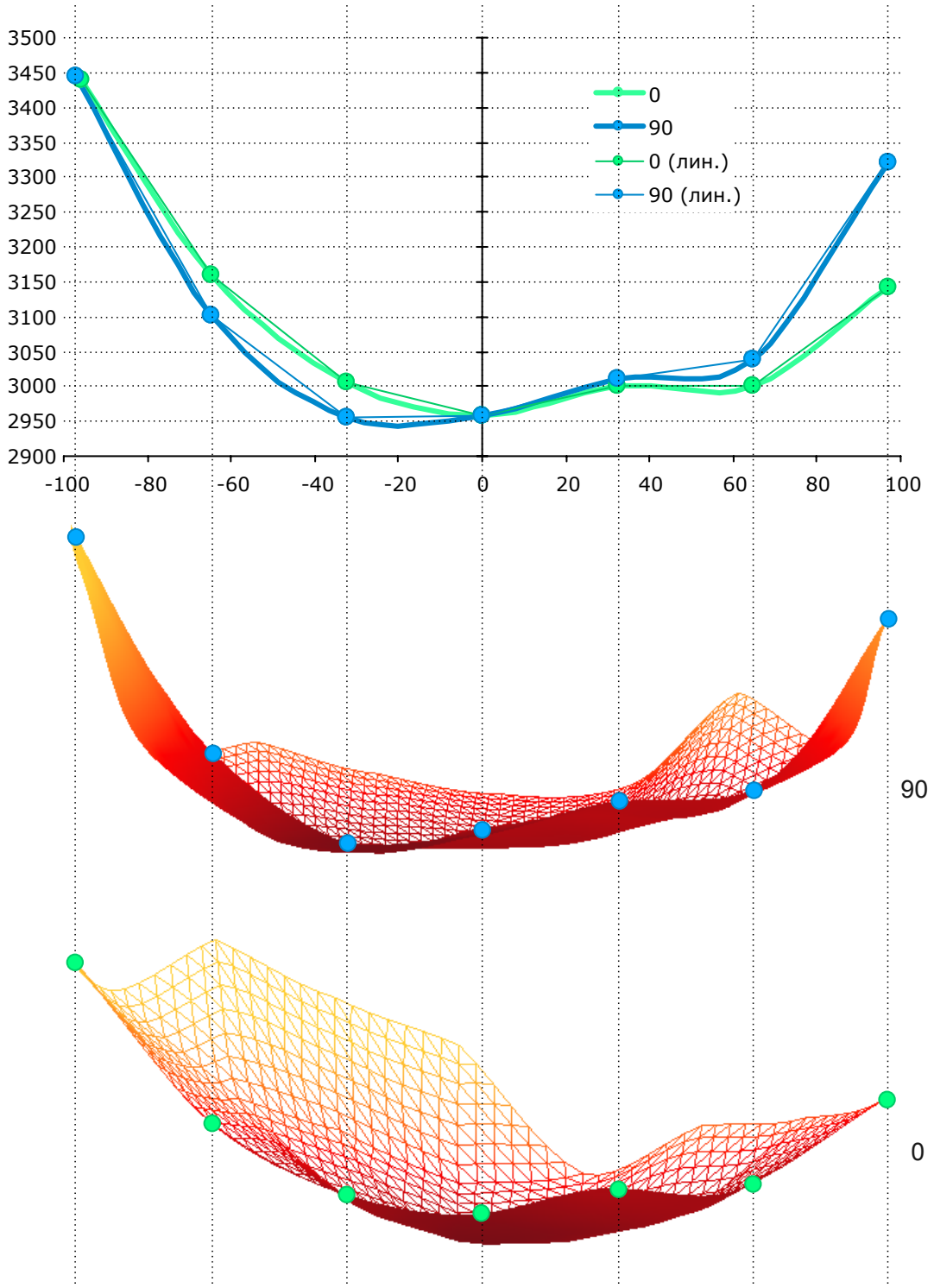


Рис. 11. *Профили и сечения поверхности в плоскостях 0 и 90 градусов*

Алгоритм был протестирован и на других наборах данных.

При более плотных измерениях можно получить довольно детальную поверхность (см. рис. 12).

На сильно зашумленных данных, где в том числе присутствуют одиночные выбросы, алгоритм интерполяции на основе криволинейных треугольников Безье тоже показал себя хорошо (рис. 9, 10, 13).

В некоторых случаях (в дополнение к рассмотренным ранее на шаге 2 алгоритма) может стать особо важен выбор методики расчета нормалей. Например, если имеются высокие выбросы и если нормаль вычисляется по средневзвешенной площади соседних элементов, то весовой коэффициент одной

Рис. 12. *«Лунный рельеф» (данные измерений)*

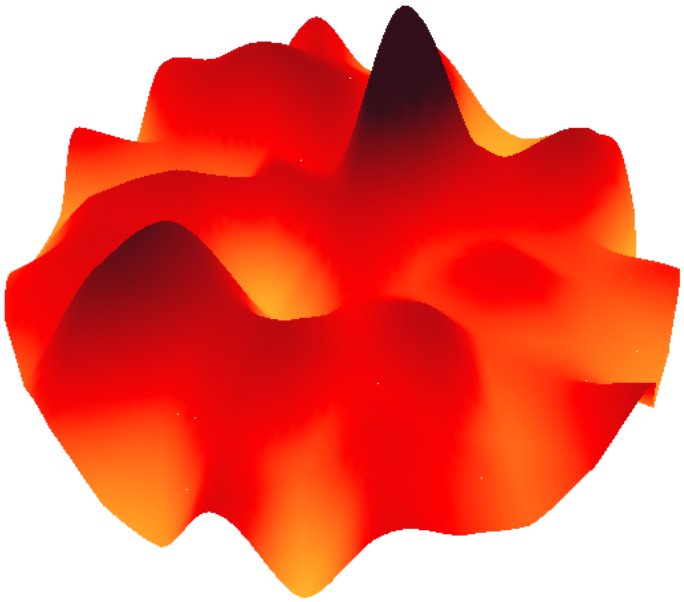


Рис. 13. *Интерполяция зашумленных данных*

из граней может чрезмерно возрасти и возникнет перекос нормали в эту сторону. Неустойчивое поведение алгоритма нормалей с коэффициентом по площади может произойти, когда перепад высот между точками становится больше расстояния между ними (в горизонтальной плоскости). Для таких наборов данных нужно сменить алгоритм вычисления нормалей или как вариант задать статичную вертикальную нормаль. Возможен и следующий прием: перед вычислением нормалей можно отмасштабировать данные по оси Z (уменьшить высоты), рассчитать нормали, затем произвести обратное масштабирование, включая и сами нормали. Таким образом временно нивелируется негативное отношение высоты и размера сетки.

Заключение

Описанный выше алгоритм обеспечивает автоматическую сшивку ребер криволинейных треугольников (касательные смежных треугольников согласованы), не требует итеративных операций по увеличению детализации (она задается сразу), возможность задания нормали для каждой точки исходных данных обеспечивает контроль формирования рельефа, применение разных методов определения нормали для тех или иных случаев увеличивает гибкость алгоритма.

Нормали использовать и проще, и нагляднее (что повышает предсказуемость результата), чем напрямую контрольные точки, которые необходимо связывать между собой как внутри треугольника, так и между соседними треугольниками.

При этом стоит заметить, что можно и не вычислять нормали, а задать вертикальную нормаль для каждой вершины, например, для первичного просмотра данных, при оценке поверхности для выбора алгоритма вычисления нормалей, или использовать такой вид как вариант сглаженной пространственной гистограммы. В том числе этот вариант может подойти для сильно зашумленных данных или с определенной структурой, чтобы не вносить дополнительные искажения, когда проявляются особенности того или иного алгоритма.

ЛИТЕРАТУРА

1. Роджерс Д., Адамс Дж. *Математические основы машинной графики*: Пер. с англ. М.: Мир; 2001. 604 с., ил. ISBN: 5-03-002143-4.
2. Farin G. *Curves and Surfaces for CAGD:* A Practical Guide (5th Edition). Morgan-Kaufmann; 2002. ISBN: 1-55860-737-4.
3. Farin G. Triangular Bernstein-Bézier Patches. *Computer Aided Geometric Design.* 1986;3(2):83–127. DOI: 10.1016/0167-8396(86)90016-6.
4. Liu Y. *Triangular Bezier Surfaces with Approximate Continuity*. University of Waterloo Computer Science Dept. University Avenue Waterloo; Canada. ISBN: 978-0-494-43305-8.
5. Vlachos A., Peters J., Boyd C., Mitchell J. L. Curved PN Triangles. *Proceedings of the 2001 Symposium on Interactive 3D Graphics (I3D '01)*. Association for Computing Machinery, New York, NY, USA; 2001:159–166. DOI: 10.1145/364338.364387.
6. Скворцов А. В., Мирза Н. С. *Алгоритмы построения и анализа триангуляции.* Томск: Изд-во Том. ун-та; 2006. ISBN: 5-7511-2028-5.
7. Leth N. W., Petersen H. G. A New Method for Generating Work Piece Surface Representations for Robotic Machining. *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. Prague, Czech Republic; 2021:7607–7614. DOI: 10.1109/IROS51168.2021.9636774.
8. Jin S., Lewis R., West D. A Comparison of Algorithms for Vertex Normal Computation. *Vis Comput.* 2005;21(1-2):71–82. DOI: 10.1007/s00371-004-0271-1.

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ И ПЕРСПЕКТИВЫ ЕГО РЕАЛИЗАЦИИ НА В-КОМПЬЮТЕРАХ

Г. Е. Деев[а], С. В. Ермаков[б]
Обнинский институт атомной энергетики, Национальный исследовательский ядерный университет «МИФИ», г. Обнинск, Российская Федерация
[а] ✍ *georgdeo@mail.ru*, [б] *ermakov@iate.obninsk.ru*

Аннотация: в статье рассматривается перспектива развития вычислительной техники и искусственного интеллекта (ИИ), связанных с В-технологией, которая использует иной принцип работы, чем А-технология, применяемая в современной компьютерной технике. В-технология позволяет создавать вычислительные устройства для любой системы счисления, обеспечивает максимальную надежность и скорость вычислений, а также имеет упрощенную элементную базу. В-компьютеры обладают замечательными свойствами, такими как насыщенность устройствами для вычисления большого набора функций и наличие устройства с варьируемой структурой (УВС) для создания новых вычислительных устройств внутри компьютера.

В-компьютеры, как и нейронные сети, способны синтезировать вычислительные устройства по фрагментам отображений, причем, в отличие от нейронных сетей, синтезированное устройство оказывается полностью известным, а, значит, известным является и отображение, связанное с ним.

В-компьютеры способны к саморазвитию, к неограниченному увеличению своей вычислительной мощи, причем, не только за счет улучшения технических параметров, но и за счет способности внутри себя создавать вычислительные устройства.

Ключевые слова: искусственный интеллект, устройства с варьируемой структурой, В-компьютеры, В-схемы.

Для цитирования: Деев Г. Е., Ермаков С. В. Искусственный интеллект и перспективы его реализации на В-компьютерах. *Успехи кибернетики*. 2025;6(1):47–54.

Поступила в редакцию: 25.02.2025. *В окончательном варианте: 09.03.2025.*

POSSIBLE B-COMPUTER IMPLEMENTATIONS OF ARTIFICIAL INTELLIGENCE

G. E. Deev[a], S. V. Ermakov[b]
Obninsk Institute for Nuclear Power Engineering, National Research Nuclear University MEPhI, Obninsk, Russian Federation
[a] ✍ *georgdeo@mail.ru*, [b] *ermakov@iate.obninsk.ru*

Abstract: we explore the prospects for computing and artificial intelligence (AI) development based on B-technology, which operates on a different principle than the A-technology used in modern computing. B-technology enables us to create computing devices for any number system, ensuring maximum reliability and speed while using a simplified element base. B-computers offer remarkable capabilities, including a high density of function-computing devices and a variable structure device (VSD) that allows us to generate new computing devices within the computer.

Like neural networks, B-computers synthesize computational devices from fragments of mappings. However, unlike neural networks, the synthesized device is fully known, making its associated mapping completely transparent. B-computers self-develop and infinitely expand their computing power not only by improving technical parameters but also by generating new computing devices within themselves.

Keywords: artificial intelligence, variable structure devices, B-computers, B-circuits.

Cite this article: Deev G. E., Ermakov S. V. Possible B-Computer Implementations of Artificial Intelligence. *Russian Journal of Cybernetics*. 2025;6(1):47–54.

Original article submitted: 25.02.2025. *Revision submitted: 09.03.2025.*

Общие положения

Скажем несколько слов об искусственном интеллекте (ИИ). Это одно из тех понятий, в названии которого достаточно хорошо отражена его суть. Это, во-первых, искусственно созданный объект и, во-вторых, обладающий интеллектом.

Огромное число объектов окружающего мира созданы искусственно. Но все они интеллектом не обладают. Таким образом, к настоящему времени таких объектов, как ИИ в полном смысле этого слова не существует. Объекты, которым приписывается название ИИ, являются, по существу, запрограммированными роботами, моделирующими разумное поведение. Сюда относятся роботы, функционирующие на производствах, нейронные сети и GPT [1–7].

Устройства с варьируемой структурой

Рассмотрим перспективы развития вычислительной техники и перспективы развития ИИ, связанного с этой техникой. Отметим, что речь идет о весьма отдаленной перспективе. Та технология, о которой мы будем сейчас говорить, будет проникать в жизнь постепенно, ей предстоит пройти путь, который прошла современная компьютерная техника. Но, поскольку мы уже имеем опыт развития, связанный с современной вычислительной техникой, можно надеяться, что путь этот будет значительно короче.

Речь идет о так называемой В-технологии (В — латинское) [8, 9]. Современную компьютерную технологию мы называем А-технологий, поскольку ей принадлежит очевидный исторический приоритет. Главный принцип А-технологии состоит в том, что по информационным линиям могут передаваться несколько информационных сигналов. На заре развития вычислительной техники, когда по линиям пытались передавать несколько информационных сигналов, возникали трудности, связанные с надежностью вычислений. Этому делу положил конец Дж. фон Нейман, когда он сформулировал тезис о том, что для создания вычислительной техники наилучшей является двоичная система счисления. В этом случае по линиям передается не более двух сигналов и вычисление становится максимально надежным. С тех пор все компьютеры двоичные.

В В-технологии используется другой принцип: по каждой информационной линии может проходить только один информационный сигнал. Таким образом, между линией и сигналом устанавливается взаимно однозначное соответствие. Казалось бы, обеднение, но в действительности использование этого принципа приводит к массе полезных вещей. Этот принцип мы называем В-принципом. Как уже было сказано, В-принцип в настоящее время в вычислительной технике не используется. Он используется физиками в лабораторных работах, когда они по линии передают только один сигнал. Но им нет необходимости придавать этому принципиальное значение, и они на этом не акцентируют своего внимания.

Нами разработано большое количество вычислительных устройств на базе В-принципа. Оказалось, что можно создавать вычислительные устройства для любой системы счисления (а не только для двоичной). Это важно, так как скорость вычислений растет с ростом основания системы счисления. Кроме того, надежность вычислений становится максимальной, скорость вычислений максимальна и много других параметров, причем, всегда на уровне экстремальных значений. В-компьютеры обладают минимальной элементной базой: она состоит только из элемента &. Для сравнения, элементная база А-компьютеров весьма разнообразна: это элементы «И», «ИЛИ», «ИНЕ» и пр. Из сравнения элементных баз можно сделать вывод, что схемы В-устройств должны быть значительно проще аналогичных А-схем тех же устройств.

Компьютеры на В-схемах обладают массой замечательных свойств, одним из которых является насыщенность компьютера устройствами для вычислений большого набора функций. Это своеобразная библиотека функций компьютера, вычисляемых аппаратно. Наличие такой библиотеки приводит к ускорению вычислений, так как хорошо известно, что вычисление функции устройством происходит быстрее, чем при вычислении ее программой. Заметим, что в А-компьютерах аналогичная функция вычислялась бы программно.

Еще одним замечательным свойством В-компьютера является наличие в нем устройства с варьируемой структурой (УВС), являющегося, по существу, создателем внутри компьютера новых устройств, встраиваемых в архитектуру компьютера. Эти вновь созданные устройства усиливают вычислительную мощь компьютера.

Таким образом, В-компьютеры допускают внутри себя нечто наподобие производства по созданию вычислительных устройств — вещь, совершенно немыслимую в А-компьютерах. Процесс создания нового вычислительного устройства внутри компьютера напоминает процесс настройки нейронной сети, но обладает некоторыми отличительными особенностями, причем, эти особенности в пользу В-компьютера. Опишем далее работу УВС.

Пример синтеза а.в.у. по фрагменту отображения

Всякое отображение f есть набор соответствий вида $f : X \ni \bar{x} \to \bar{y} \in Y$, где X — область определения отображения, $\bar{x}$ — элемент из области определения отображения, Y — область значений отображения, $\bar{y}$ — элемент из области значений отображения. Заметим, что символами с чертой, наподобие, $\bar{x}$ обозначаются числа и числоподобные объекты: $\bar{x} = \ldots x_n \ldots x_2 x_1 x_0$, в то время как без черты, x — цифры, иногда с индексами.

Устройство с варьируемой структурой (УВС) схематически представлено на рис. 1.

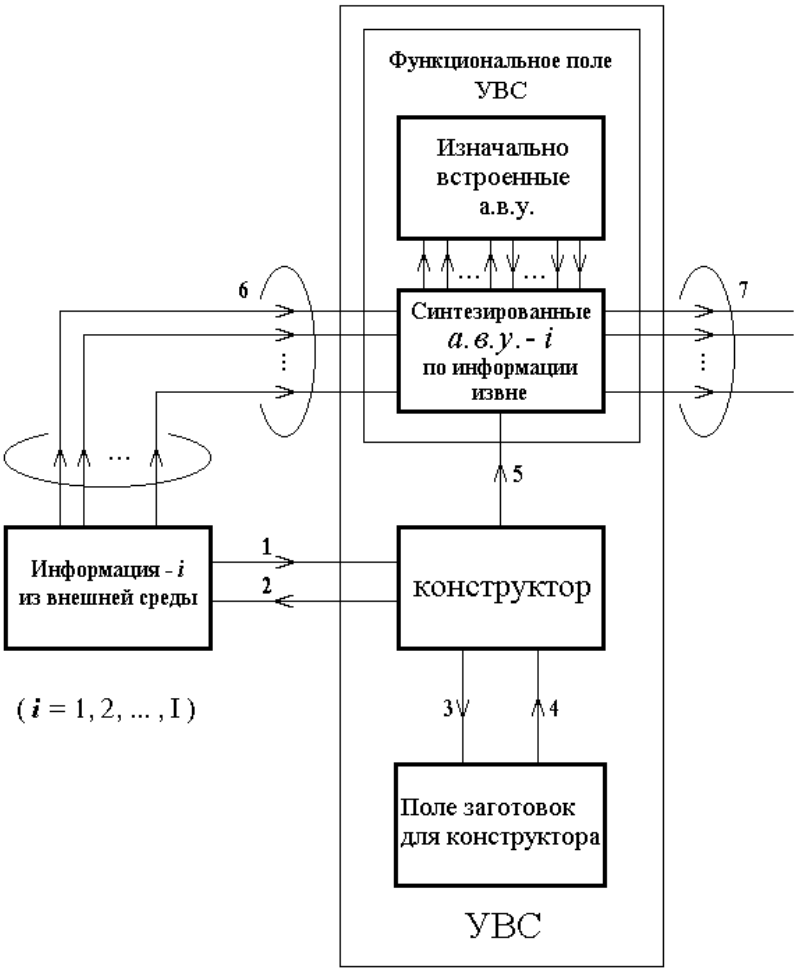


Рис. 1. *Схема УВС*

Конструктор на этой схеме — это и есть упомянутый выше создатель устройств. Обращает на себя внимание поле заготовок конструктора. Оно состоит из «сырых» В-схем устройств, т.е. еще не готовых устройств. После простановки на них соединений они становятся полноценными устройствами и интегрируются в компьютер. Поле заготовок потенциально может состоять из любого количества заготовок. С развитием миниатюризации компьютерной технологии количество их может исчисляться миллионами и когда количество таких структур в компьютере начинает исчисляться миллионами, вот тогда можно говорить, что мы приближаемся к ИИ.

На рис. 2 слева, приведен пример пустой заготовки, и справа, она же, но с уже выбранным вариантом связей. Как видим, превращение пустой заготовки в реально действующий прибор сводится в В-схемах к простому физическому действию — простановке точки на перекрестиях линии. Такая команда исходит от конструктора из УВС. Количество вариантов таких простановок для данной заготовки огромно и равно 7^{84}. Столько вычислительных устройств можно создать на такой заготовке. Если учесть, что заготовки могут быть разными, более простыми и более сложными, а также, что количество их может быть чрезвычайно велико, то легко себе представить, какой огромный вычислительный потенциал скрывается за В-компьютерами.

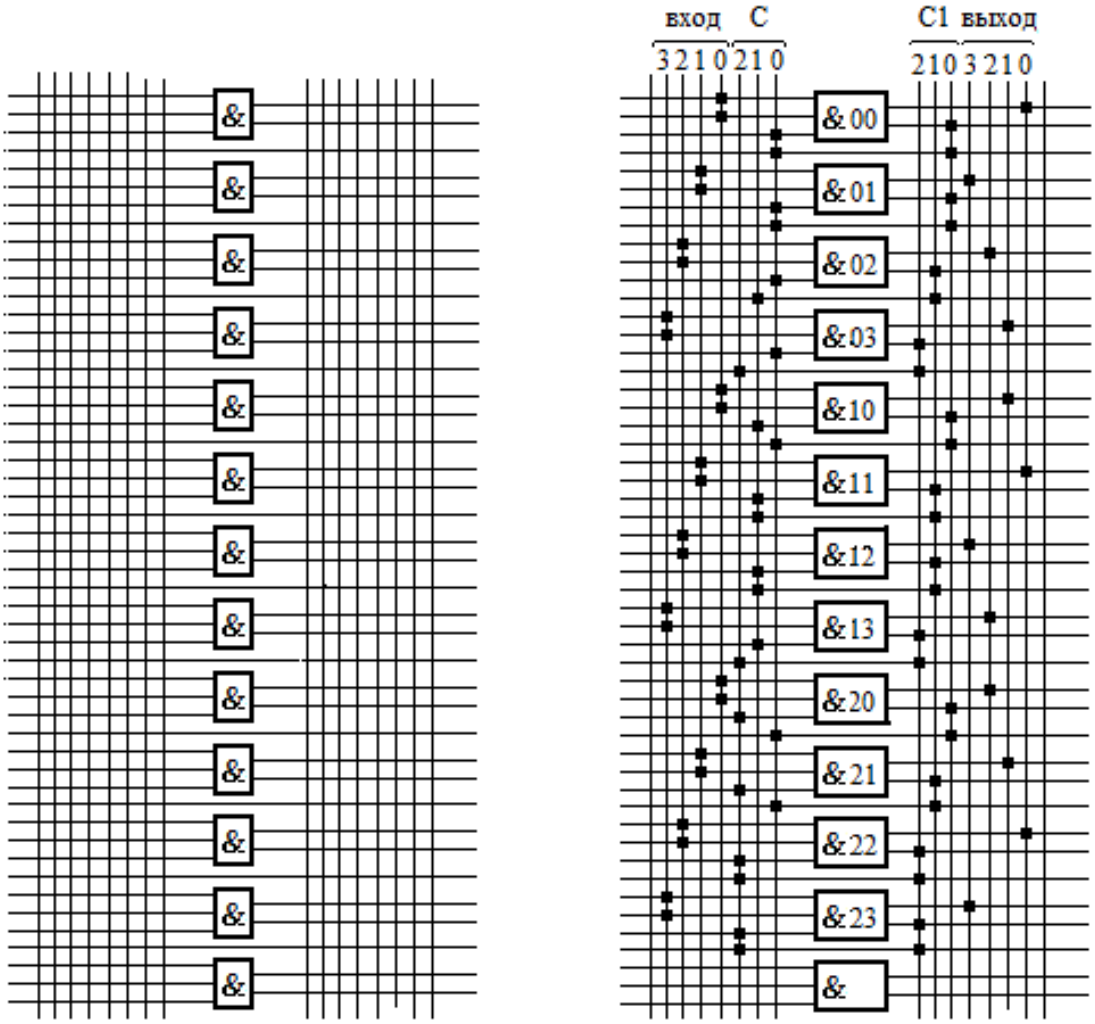


Рис. 2. *Заготовка без соединений (слева) и заготовка с соединениями (справа)*

Способность В-компьютеров к созданию внутри себя других вычислительных устройств и тем самым к самостоятельному увеличению своего вычислительного интеллектуала можно назвать способностью к саморазвитию. Итак, В-компьютеры способны к саморазвитию.

Понятно, что на вход компьютера поступает ограниченная информация из окружающего мира. В частности, ограниченная информация поступает о том отображении, для которого мы хотим построить устройство, вычисляющее любое конкретное соответствие этого отображения. И это принципиально в нашем реальном мире. Поэтому в некоторых случаях нельзя быть уверенным в том, что синтезированное устройство реализует изучаемое отображение в полном объеме.

Перейдем к примеру. Как уже было сказано, любое отображение есть множество соответствий типа $f : X \ni \bar{x} \to \bar{y} \in Y$, причем, в полном объеме это множество обычно не бывает известным. По этой причине мы вынуждены говорить, что то или иное высказывание имеет место «в предположении». Так это или на самом деле нет, выясняется в процессе использования синтезированных устройств. «Практика — критерий истины». Конечный набор соответствий $M = \{\bar{x}_i \to \bar{y}_i\}_{i=1}^{i=m}$ мы называем репрезентативным для отображения *f*, если синтезированное на его основе вычислительное устройство позволяет вычислить любое соответствие $f : X \ni \bar{x} \to \bar{y} \in Y$. Но в репрезентативности любого конкретного набора соответствий $M = \{\bar{x}_i \to \bar{y}_i\}_{i=1}^{i=m}$ мы уверены быть не можем. Но зато мы можем по результатам синтеза сразу определить, что оно не репрезентативно. Это мы можем сделать в том случае, если в результате синтеза получилось два или более устройств.

В самом деле, каждое устройство реализует свое «родное» отображение, и если у нас получилось несколько устройств, то это означает, что набор соответствий M является общим для этих нескольких отображений. Для дальнейшего решения задачи о нахождении устройства именно для нашего отображения надо дополнить набор M несколькими соответствиями, и так поступать до тех пор,

пока не останется только одного синтезированного устройства. Вот оно-то и будет главным претендентом на роль устройства, представляющего отображение f.

Мы не будем описывать алгоритм синтеза, а лишь приведем на примере результат его работы.

Пусть числовой объект $\overleftarrow{x} = 1300122203\underset{0}{1}$ записан в четверичной системе счисления. Это значит, что все его цифры принадлежат множеству цифр $Z_{(4)} = \{0,1,2,3\}$. Стрелка $\leftarrow$ над объектом x указывает на направление роста разрядов, место нулевого разряда отмечено нулем снизу. Пусть, аналогично, объект $\overleftarrow{y} = 4403212103\underset{0}{4}$ записан в пятеричной системе счисления, $Z_{(5)} = \{0,1,2,3,4\}$, и пусть в некотором отображении между объектами $\overleftarrow{x}$ и $\overleftarrow{y}$ есть соответствие, которое мы запишем в разрядной сетке:

	t	0	1	2	3	4	5	6	7	8	9	10
$\overleftarrow{x}$	x_t	1	3	0	2	2	2	1	0	0	3	1
$\downarrow$:	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$
$\overleftarrow{y}$	y_t	4	3	0	1	2	1	2	3	0	4	4

(1)

$$\underset{t=0}{\overset{t=10}{\forall t}} \left[x_t \in Z_{(4)}\right], \qquad \underset{t=0}{\overset{t=10}{\forall t}} \left[y_t \in Z_{(5)}\right].$$

Спрашивается, существует ли а.в.у., реализующее соответствие числовых объектов $\overleftarrow{x} \to \overleftarrow{y}$?

Алгоритм начинает поиск, т.е. вначале ставится вопрос, существует ли абстрактный автомат, реализующий указанное соответствие (отметим, что класс А-0 заполняют абстрактные автоматы).

В результате применения <u>алгоритма синтеза а.в.у. по фрагментам отображения</u> (в дальнейшем будем говорить просто — алгоритма) выясняется, что существуют два минимальных автомата A_1 и A_2 с двумя состояниями, для которых соответствие (1) имеет место. Эти автоматы задаются таблицами:

A_1 :

x \ q	0	1
0	0, 0	0, 3
1	1, 4	1, 2
2	1, 1	0, 2
3	0, 4	0, 3

A_2 :

x \ q	0	1
0	1, 3	0, 0
1	1, 4	0, 2
2	1, 1	0, 2
3	0, 4	1, 3

(2)

В таблицах $x \in Z_{(4)}$, q — *состояние* автомата, $q \in Q = \{0,1\}$, Q — множество состояний автомата. На пересечении строки-x и столбца-q стоит пара (θ,σ), где $\theta = \theta(q,x)$ — значение функции переходов на паре (q, x), а $\sigma = \sigma(q,x)$ — значение функции выходов на той же паре.

Таким образом, согласно алгоритму существует два автоматных отображения f_1 и f_2, реализуемых соответственно, автоматами A_1 и A_2, для каждого из которых имеет место, в частности, данное соответствие $\overleftarrow{x} \to \overleftarrow{y}$.

Проверим, что это действительно так. Для этого пропустим объект $\overleftarrow{x}$ через автоматы A_1 и A_2 и убедимся, что в обоих случаях на выходе будет $\overleftarrow{y}$. Изобразим развертки во времени работы автоматов при прохождении через них объекта $\overleftarrow{x}$.

Через A_1:

t	0	1	2	3	4	5	6	7	8	9	10	11
x_t	1	3	0	2	2	2	1	0	0	3	1	
	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	
q_t	$0\to$	$1\to$	$0\to$	$0\to$	$1\to$	$0\to$	$1\to$	$1\to$	$0\to$	$0\to$	$0\to$	1
	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	
y_{t-1}	4	3	0	1	2	1	2	3	0	4	4	

(3)

Через A_2:

t	0	1	2	3	4	5	6	7	8	9	10	11
x_t	1	3	0	2	2	2	1	0	0	3	1	
	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	
q_t	$0\rightarrow$	$1\rightarrow$	$1\rightarrow$	$0\rightarrow$	$1\rightarrow$	$0\rightarrow$	$1\rightarrow$	$0\rightarrow$	$1\rightarrow$	$0\rightarrow$	$0\rightarrow$	1
	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	
y_{t-1}	4	3	0	1	2	1	2	3	0	4	4	

(4)

Символ t обозначает автоматное время и одновременно, для строки x_t, — номер разряда; начальное состояние автоматов в обоих случаях $q = 0$. Результаты работы автоматов прочитываются в нижних строках разверток, в обоих случаях совпадают и оказываются равными объекту $\overleftarrow{y}$, как и должно быть. Тот факт, что алгоритм выделил в классе А-0 два автомата A_1 и A_2, означает, что информации, заложенной во фрагменте соответствия $\overleftarrow{x}\rightarrow\overleftarrow{y}$, недостаточно для <u>однозначного</u> выделения автомата. Чтобы однозначно найти автомат в классе А-0, дополним соответствие $\overleftarrow{x}\rightarrow\overleftarrow{y}$ в 11-ом разряде и перейдем к новому соответствию $\overleftarrow{x}_1\rightarrow\overleftarrow{y}_1$:

	t	0	1	2	3	4	5	6	7	8	9	10	11
$\overleftarrow{x}_1$	x_t	1	3	0	2	2	2	1	0	0	3	1	0
$\downarrow$:	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$
$\overleftarrow{y}_1$	x_t	4	3	0	1	2	1	2	3	0	4	4	3

(5)

Пропустим объект $\overleftarrow{x}_1$ через автоматы A_1 и A_2 и посмотрим, получится ли на выходе объект $\overleftarrow{y}_1$. При прохождении $\overleftarrow{x}_1$ через A_1 развертка (3) пополнится лишь 12-м тактом:

t	0	1	...		10	11	12
x_t	1	3	...		1	0	
	$\downarrow$	$\downarrow$			$\downarrow$	$\downarrow$	
q_t	$0\rightarrow$	$1\rightarrow$	...	$\rightarrow$	$0\rightarrow$	$1\rightarrow$	0
	$\downarrow$	$\downarrow$			$\downarrow$	$\downarrow$	
y_{t-1}	4	4	...		4	3	

(6)

Как видно, в 12-ом такте, в полном соответствии с (5), получился результат $y_{11} = 3$, т.е. автомат A_1 реализует соответствие $\overleftarrow{x}_1\rightarrow\overleftarrow{y}_1$. При прохождении $\overleftarrow{x}_1$ через A_2 будем иметь:

t	0		1		...		10		11		12
x_t	1		3		...		1		0		
	$\downarrow$		$\downarrow$				$\downarrow$		$\downarrow$		
q_t	0	$\rightarrow$	1	$\rightarrow$	...	$\rightarrow$	0	$\rightarrow$	1	$\rightarrow$	0
	$\downarrow$		$\downarrow$				$\downarrow$		$\downarrow$		
y_{t-1}	4		4		...		4		0		

(7)

Но на этот раз при $t = 12$ автомат A_2 вырабатывает сигнал $y'_{11} = 0$, отличный от требуемой цифры 3:

$$\overleftarrow{x}_1 \xrightarrow{A_2} \overleftarrow{y}'_1 \neq \overleftarrow{y}_1 . \tag{8}$$

Следовательно, автомат A_2 не реализует соответствие $\overleftarrow{x}_1\rightarrow\overleftarrow{y}_1$ и, таким образом, существует лишь один автомат, A_1 (рис. 3), реализующий соответствие $\overleftarrow{x}_1\rightarrow\overleftarrow{y}_1$, а, значит, в предположении, что соответствие $\overleftarrow{x}_1\rightarrow\overleftarrow{y}_1$ — репрезентативное, существует лишь одно автоматное отображение f_1, для которого $\overleftarrow{x}_1\rightarrow\overleftarrow{y}_1$. Так по одному соответствию можно найти все автоматное отображение.

Факт построения автомата A_1 по соответствию $\overleftarrow{x}_1\rightarrow\overleftarrow{y}_1$ можно интерпретировать как процесс обучения, с которым описанный синтез имеет большое сходство. Действительно, в мозгу школьника при обучении его, скажем, операции сложения после выполнения им <u>конечного</u> числа примеров формируется нечто, позволяющее ему в дальнейшем решить <u>любой</u> пример на выполнение операции

сложения. Здесь мы имеем аналогию. Через УВС, в котором содержится алгоритм синтеза, пропускается конечное число соответствий (в рассмотренном примере одно: $\overleftarrow{x}_1 \to \overleftarrow{y}_1$) и в результате в УВС формируется структура (автомат A_1), позволяющая найти любое другое соответствие автоматного отображения f_1. Конечно, мы не утверждаем, что в мозгу происходит буквально то же самое, мы лишь подчеркиваем аналогию. Более того, мы уже ранее говорили, что не будем стремиться моделировать природу, не только по незнанию того, как она что-то делает, но и потому, что количественного эффекта при достижении одной и той же цели мы чаще всего добиваемся на путях, отличных от тех, которым следует природа.

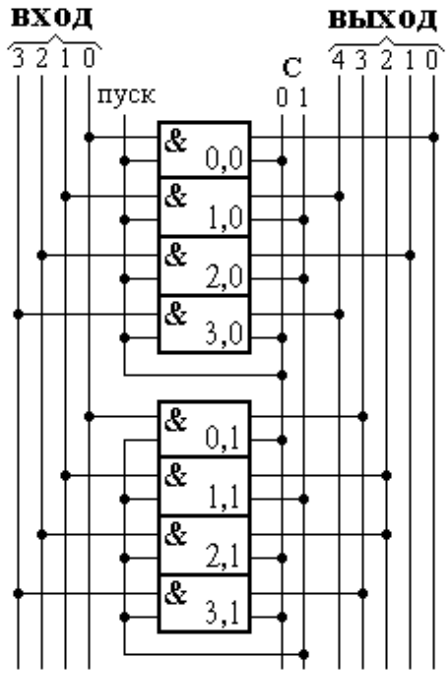


Рис. 3. *В-схема автомата* A_1*, синтезированного с помощью алгоритма синтеза а.в.у. по фрагментам отображения*

Между данным процессом синтеза и синтезом нейронных сетей наблюдается сходство, состоящее в том, что и там, и там синтез происходит на парах «вход-выход» типа $f : X \ni \bar{x} \to \bar{y} \in Y$. И это единственное сходство. В описанном методе, в отличие от нейронных сетей, нам известно синтезированное отображение, ибо известно табличное задание устройства, реализующего отображение, а, следовательно, и В-схема этого устройства. А все свойства любого отображения извлекаются из его табличного задания и В-схемы. Таким образом, результат описанного синтеза — не черный ящик, как жалуются «нейронщики». Нейронную сеть невозможно интегрировать в компьютер в виде устройства, а В-схемы — можно.

Выводы

В-компьютеры, как и нейронные сети, способны синтезировать вычислительные устройства по фрагментам отображений, причем, в отличие от нейронных сетей, синтезированное устройство оказывается полностью известным, а, значит, известным отображение, связанное с ним.

В-компьютеры способны к саморазвитию, к неограниченному увеличению своей вычислительной мощи, причем, не только за счет улучшения технических параметров, но и за счет способности внутри себя создавать вычислительные устройства.

ЛИТЕРАТУРА

1. Спиноза Б. *Избранные произведения*. Ростов-на-Дону: Феникс; 1998. 604 с.
2. Лейбниц Г. В. *Сочинения*. В 4 томах. Т. 2. М.: Мысль; 1983. 686 с.
3. Russell B. *The Analysis of Mind*. Routledge; 2022. 268 с.

4. Рассел Б. *Введение в математическую философию. Избранные работы* / пер. с англ. В. В. Целищева, В. А. Суровцева. Новосибирск: Сиб. унив. изд-во; 2007. 264 с.
5. Локк Дж. *Опыт о человеческом разумении. Сочинения*: В 3 т. Т. 1 / под ред. И. С. Нарского. М.: Мысль; 1985. 623 с.
6. Локк Дж. *Опыт о человеческом разумении. Сочинения*: В 3 т. Т. 2 / под ред. И. С. Нарского. М.: Мысль; 1985. 560 с.
7. Локк Дж. *Избранные философские произведения*. М.: Государственное социально-экономическое издательство; 1960. 1268 с.
8. Деев Г. Е., Ермаков С. В. *В-компьютеры: генезис, развитие, приложения.* Успехи кибернетики. 2022;3(2):74–85. DOI: 10.51790/2712-9942-2022-3-2-9.
9. Деев Г. Е., Ермаков С. В. *В-компьютеры: саморазвитие*. Успехи кибернетики. 2020;1(2):51–60. DOI: 10.51790/2712-9942-2020-1-2-6.

НЕЯВНЫЕ СООБЩЕСТВА В СЕТЯХ И ИНФОРМАЦИОННОЕ ВОЗДЕЙСТВИЕ

А. А. Чеповский[1,а], **А. М. Чеповский**[1,2,б]

[1] *Национальный исследовательский университет «Высшая школа экономики», г. Москва, Российская Федерация*

[2] *Российский экономический университет им. Г. В. Плеханова, г. Москва, Российская Федерация*

[а] *ORCID: https://orcid.org/0000-0002-0938-3517, ✍ aachepovsky@hse.ru*

[б] *ORCID: https://orcid.org/0000-0001-8959-6119, CHepovskiy.AM@rea.ru*,

Аннотация: в настоящей работе авторы изучают проблему как оценки качества выделения неявных сообществ на графе, полученном при импорте данных из социальных сетей и мессенджеров, так и методов анализа таких сетей на предмет выявления информационного воздействия на их акторов. Рассматриваются два подхода к оценке корректности разбиения графа на сообщества. Первый способ оценки основан на методах теории информации и заключается в подсчете нормализованной взаимной информации (NMI) и асимметрично-нормализованной взаимной информации (ANRMI). В рамках второго подхода рассмотрены три метода оценки качества выделения неявных сообществ, основанные на анализе текстовых массивов сопоставленных сообществ. Определяются и сравниваются коэффициенты попарной ранговой корреляции для словарей разных текстовых массивов. Применяется анализ соответствий для исследования корпусов текстов сообществ. Третий метод анализа заключается в подсчете психолингвистических характеристик текстовых массивов неявных сообществ. На примерах реальных данных показана применимость методов исследования для оценки качества разбиений графа взаимодействующих объектов социальных сетей и мессенджеров на сообщества и анализа информационного воздействия в такой сети.

Ключевые слова: выделение неявных сообществ, взаимная информация, ранговая корреляция, психолингвистические характеристики, информационное воздействие..

Для цитирования: Чеповский А. А., Чеповский А. М. Неявные сообщества в сетях и информационное воздействие. *Успехи кибернетики*. 2025;6(1):55–61.

Поступила в редакцию: 13.01.2025. *В окончательном варианте: 26.01.2025.*

IMPLICIT COMMUNITIES IN SOCIAL NETWORKS AND INDOCTRINATION

A. A. Chepovsky[1,a], **A. M. Chepovsky**[1,2,b]

[1] *National Research University Higher School of Economics, Moscow, Russian Federation*

[2] *Plekhanov Russian University of Economics, Moscow, Russian Federation*

[a] *ORCID: https://orcid.org/0000-0002-0938-3517, ✍ aachepovsky@hse.ru*

[b] *ORCID: https://orcid.org/0000-0001-8959-6119, CHepovskiy.AM@rea.ru*

Abstract: we studied the problem of assessing the quality of implicit community detection in a graph constructed from data imported from social networks and instant messengers. Additionally, we analyzed methods for identifying indoctrination within such networks. We considered two approaches to evaluating the accuracy of community detection. The first approach, based on information theory, involved calculating normalized mutual information (NMI) and asymmetrically normalized mutual information (ANRMI). The second approach examined three methods for assessing the quality of implicit community detection using text analysis. We determined and compared pairwise rank correlation coefficients for dictionaries derived from different text arrays. We also applied correspondence analysis to study corpora of community texts. The third method involved calculating the psycholinguistic characteristics of text arrays associated with implicit communities. Using real data, we demonstrated the applicability of these methods for evaluating the partitioning of a social network graph and analyzing information influence within the network.

Keywords: implicit community detection, mutual information, rank correlation, psycholinguistic characteristics, indoctrination..

Cite this article: Chepovsky A. A., Chepovsky A. M. Implicit Communities in Social Networks and Indoctrination. *Russian Journal of Cybernetics*. 2025;6(1):55–61.

Original article submitted: 13.01.2025. *Revision submitted: 26.01.2025.*

Введение

Описание информационного воздействия на пользователей социальных сетей и мессенджеров связано с проблемой выделения неявных сообществ на графе взаимодействия акторов, широко описанной в научной литературе, например, в работах [1–3]. Задача выделения неявных сообществ на графе заключается в разбиении графа на подграфы, такие, что плотность связей внутри этих подграфов намного выше плотности связей между ними. Представляет интерес нахождение и пересекающихся сообществ, подразумевающих наличие общих вершин, принадлежащих сразу нескольким сообществам. Особенность рассматриваемой проблематики заключается в том, что задача выделения сообществ не имеет точного решения: нет однозначного математического определения понятия «сообщество».

Актуальной задачей при анализе структуры графов является оценка качества выделения сообществ [1, 2, 4]. Под качеством выделения сообществ на графе социальной сети можно понимать соответствие полученных неявных сообществ задачам выявления информационного взаимодействия между акторами сети. Один из классических методов сравнения результатов выделения сообществ на графах произвольной природы основан на применении методов теории информации и заключается в подсчете нормализованной взаимной информации (NMI) для двух (и более) различающихся разбиений одного и того же графа [5, 6]. Разными авторами рассматривались и ее модификации, в том числе и ассиметричная версия (ANRMI) [7–9]. В целом, такой подход описывает результаты, основанные на структуре сети, без учета дополнительных метаданных, которые для случая графов социальных сетей представляют себой важный фактор.

В предыдущих работах [3, 10–12] нами были предложены иные методики для оценки качества выделения неявных сообществ на графах социальных сетей и мессенджеров. Эти подходы основаны на анализе текстовых массивов, поставленных в соответствие выявленным сообществам и составленных из всех сообщений входящих в них вершин графа. В работах [3, 10] представлен метод корреляционного анализа применительно к задаче сравнения неявных сообществ. Работа [11] посвящена применению метода анализа соотвествий к корпусам текстов неявных сообществ. В работах [3, 12] показано, как использовать вычисление психолингвистических характеристик обобщенных текстов сообществ, выделенных на графе сети Telegram-каналов, для качественного анализа расмматриваемого выделения.

Отметим вышедшие недавно работы [13, 14], в которых методы компьютерной лингвистики успешно примененяются для поиска скрытых сообществ в корпусах социальных медиа. Эти работы подтверждают перспективность предложенных нами в работах [3, 10–12] подходов использования методов компьютерной лингвистики для исследования неявных сообществ, выделенных алгоритмами анализа топологии графов.

В даной работе мы обобщаем предложенное ранее [3, 10–12] использование методов компьютерной лингвистики для оценки качества выделения неявных сообществ на графах взаимодействия в социальных сетях и их применение для анализа информационного воздействия. Дополнительно показано использование в контексте этой задачи стандартого инструментария по оценке разбиения на сообщества с использованием показателей NMI.

Анализируемые сообщества

Массив данных, на которых будем применять различные методики, получен при импорте из мессенджера Telegram, если быть точнее, то строятся графы его публичных каналов, как это описано в работах [3, 12]. Рассмотрим четыре графа, полученных таким образом и соответствующих различным тематикам обсуждения: развлекательные, образовательные, общественно-политические. Данные были получены в конце 2022 года. Размеры исходных графов: G_1 состоит из 619 вершин и 2973 ребер; G_2 состоит из 458 вершин и 2698 ребер; G_3 состоит из 89 вершин и 207 ребер; G_4 состоит из 472 вершин и 16249 ребер; G_5 состоит из 288 вершин и 2167 ребер. Далее к этим графам был применен метод Галактик [3, 12] для выделения неявных пересекающихся сообществ. На графах G_1, G_2 и G_3 было выделено по 8 пересекающихся сообществ, на графе G_4 выделено 3, а на графе G_5 — 18 пересекающихся сообществ.

Для методик, основанных на процедурах компьютерной лингвистики, формировались массивы текстов выделенных сообществ. Эксперименты с корпусами текстов проводились для графа G_1, подробно описанного в работе [12]. Для каждого из выделенных на этом графе сообществ S_i $(i = 0, \ldots, 7)$ брались текстовые сообщения всех пользователей, представляющих вошедшие в S_i вершины. Эти тексты объединялись по сообществам в единый для каждого из них текстовый массив на русском языке.

Размеры полученных корпусов, очищенных от специальных символов, для этих восьми сообществ составили (в Кб): {335; 39208; 13571; 3485; 1633; 37897; 2925; 10794}. Из текстов выделялись различные лингвистические характеристики процедурами, описанными в [3, 15], и составлялись их частотные словари: буквосочетаний различной длины; начальных форм слов различных частей речи; именных и глагольных групп; псевдооснов словоупотреблений (часть слова без некоторых аффиксов).

Метод взаимной информации

Для начала посмотрим на классический подход оценки качества разбиения, который можно применить на рассматриваемом наборе данных. Подход основан на подсчете взаимной информации и на предположении, что для близких разбиений достаточно сообщить небольшое количество информации, чтобы получить из одного разбиения другое. Применение методов теории информации к задаче обнаружения сообществ помогает понять через энтропию, сколько информации требуется для перемещения объектов из одного сообщества в другое [5, 6]. Более подробно энтропия для разбиения C графа на сообщества определяется как:

$$H(C) = -\sum_{i=1}^{k(C)} P(C_i) \log_2 P\ (C_i), \tag{1}$$

где $P(C_i)$ — вероятность того, что случайная вершина принадлежит сообществу C_i, $k(C)$ — количество сообществ в разбиении C.

Условная энтропия $H(S \mid C)$:

$$H(S \mid C) = -\sum_{i=1}^{k(C)} \sum_{j=1}^{k(S)} P\left(C_i, S_j\right) \log_2 P\ \left(C_i, S_j\right), \tag{2}$$

где $P\left(C_i, S_j\right)$ — вероятность того, что случайная вершина принадлежит сообществу C_i в разбиении C и сообществу S_j в разбиении S.

Взаимная информация между двумя разбиениями C и S измеряет, насколько знание одного разбиения уменьшает неопределенность относительно другого. Нормализованная взаимная информация (NMI) используется для улучшения интерпретации взаимной информации, нормируя ее относительно энтропий двух разбиений:

$$NMI\ (C;S) = \frac{(H(S) - H(G|C))}{\frac{1}{2}(H(C) + H(S))}, \tag{3}$$

получаем [0; 1] как диапазон принимаемых значений, где $NMI = 1$ означает полное соответствие между разбиениями.

Мера 3 симметрична, что означает: значение не зависит от порядка разбиений C и S. Для устранения искажений и более точного оценивания качества выделения сообществ разными авторами в работах [7–9] была введена асимметрично-нормализованная преобразованная взаимная информация (ANRMI), нормализующая взаимную информацию с использованием энтропии только одного из разбиений для сравнения разбиений на сообщества в результате применения алгоритмов к графам с заранее известной первичной структурой:

$$ANRMI\ (C;S) = \frac{(H(S) - H(S|C))}{H(S)}, \tag{4}$$

где $H(S)$ — энтропия первичного разделения S на сообщества графа.

В качестве вычислительного эксперимента для пяти описанных выше графов по 3 и 4 были подсчитаны нормализованная взаимная информация (NMI) и асимметрично-нормализованная преобразованная взаимная информация (ANRMI). Подсчеты производились для исходного выделения сообществ методом Галактик и результатов работы на этих же графах одного из классических алгоритмов *Louvain*, основанного на максимизации модулярности [16]. Усредненные значения подсчитанных показателей взаимной (ассиметричной) информации по этим двум методам выделения сообществ представлены в таблице 1.

Таблица 1

Подсчет показателей NMI и ANRMI для разбиений G_1–G_5 методом Галактик и алгоритмом Louvain

Граф	NMI	ANRMI
G_1	0,25	0,28
G_2	0,47	0,48
G_3	0,65	0,64
G_4	0,07	0,035
G_5	0,21	0,01

Из приведенных данных видно, что в большинстве случаев значения нормализованной взаимной информации и асимметрично-нормализованной преобразованной взаимной информации не отличаются существенно. Наибольшее различие наблюдается для случая большого числа сообществ (граф G_5), тут подсчет *ANRMI* указывает на слабое совпадение разбиений графа, полученных двумя методами.

В любом случае подобные подсчеты могут лишь оценить различие между двумя разбиениями, что дает возможность оценки результатов работы алгоритмов для случаев, когда граф сгенерирован искусственным образом и заранее известна «заложенная» структура сообществ. Для случая реальных данных и полученных при импорте графов эти оценки явно не указывают на качество выделения сообществ с точки зрения «смыслового» содержания. И тем более не могут помочь с оценкой информационного воздействия на акторов сети.

Перейдем теперь к тем методикам оценки качества разбиения графа на неявные сообщества, которые могут помочь с этой задачей и применимы при работе с данными, полученными из реальных сетей.

Корреляционный анализ корпусов текстов сообществ

Суть данного метода заключается в том, что корпуса текстов неявных сообществ сопоставляются попарным сравнением частотных словарей различных лингвистических характеристик (лексических единиц отдельных частей речи, словосочетаний, псевдооснов слов), составленных для каждого из анализируемых наборов текстов. Данная методика подробно описана в [3, 10].

В частотных словарях отсортированные записи рассматриваются как случайные величины с заданными рангами записей. Для каждой пары словарей подсчитываются коэффициенты попарной ранговой корреляции, которые являются оценками наличия связи между случайными величинами [17, 18]. Сравниваются коэффициенты попарной ранговой корреляции для словарей разных текстовых массивов одних и тех же лингвистических характеристик.

В расчетах каждая пара словарей приводится к одному и тому же размеру отбрасыванием характеристик с низкими частотами использования. Но анализируемые словари имеют для реальных текстовых массивов достаточно большие размеры. Поэтому учитываются только первые 10000 записей по убыванию частоты элементов каждого из словарей.

Коэффициент попарной ранговой корреляции принимает значения на $[-1;\ 1]$. Значения, близкие к 1, указывают на согласованность словарей с точки зрения частот распределения лингвистических характеристик в сравниваемых текстах. Обратная ситуация –при значениях коэффициента попарной ранговой корреляции, близких к -1. Если значения коэффициента попарной ранговой корреляции находятся в окрестности 0, то словари несогласованы, что указывает на различия текстовых массивов.

К приведенным в работе [10] результатам добавим значения коэффициентов попарной ранговой корреляции для словарей глагольных групп (которые считаются одинаковыми, если совпадают псевдоосновы словоупотреблений их составляющих) текстовых массивов сообществ для графа G_1, полученных при разбиении графа, описанного в [12]. Результаты представлены в таблице 2.

Представленные в таблице 2 результаты показывают, что словари словосочетаний глагольных групп текстовых массивов практически всех сообществ при их попарном сравнении «обратны» по частотам использования словосочетаний в текстах разных сообществ. Это указывает на несовпадение текстов различных сообществ по их тематической и психологической направленности. Такое несовпадение можно трактовать как показатель корректности анализируемого разбиения графа на сообщества.

Таблица 2

Коэффициенты корреляции словарей глагольных групп (в псевдоосновах) текстов сообществ S_i графа G_1

	S_0	S_1	S_2	S_3	S_4		S_5	S_6	S_7
S_0	1								
S_1	-0.7	1							
S_2	-0.79	-0.93	1						
S_3	0.84	-0.91	-1	1					
S_4	-0.52	-0.14	-0.92	-0.91	1				
S_5	-0.7	0.69	-0.93	-0.9	-0.08		1		
S_6	-0.59	-0.12	-0.92	-0.89	-0.25		-0.09	1	
S_7	-0.66	0.2	-0.93	-0.89	-0.045		0.31	-0.15	1

На следующем этапе анализа текстов выделенных сообществ описанным методом подсчета коэффициентов корреляции словарей текстов сообществ со словарями текстов эталонного набора, например, текстов «общественного», «политического, «экстремистского» характера, можно решать для реальных данных задачу выявления информационного воздействия.

Анализ соответствий корпусов текстов сообществ

Анализ соответствия был предложен в качестве метода лингвистического анализа текстовых данных Ж.-П. Бензекри в [19] с целью выявления факторов, группирующих характеристики текстов, и определения взаимной «близости» или «удаленности» подкорпусов на основе анализа совместного появления значений переменных. Анализ соответствия дает геометрическое представление связей между двумя наборами характеристик, определяющими каждый из подкорпусов текстов с точки зрения их смыслового содержания (тематики, психологической направленности, степени и направленности информационного воздействия текстов).

Корпуса текстов неявных сообществ исследовались методом анализа соответствий в [3, 11]. Для исследований корпусов текстов сообществ средствами анализа соответствий использовалась корпусная платформа ТХМ [20], в которой библиотека для анализа соответствий включена как один из инструментов платформы. Каждый подкорпус суть массив текстов одного из неявных сообществ. Близость между значениями характеристик подкорпусов интерпретируется как оценка, указывающая на сходство или различие между подкорпусами.

На рис. 1 представлена диаграмма расположения подкорпусов по результатам анализа соответствий для частотных таблиц слов для корпуса текстов сообществ графа G_1. На диаграмме подкорпуса разнесены в презентационном пространстве результаты анализа соответствий. В названии осей указывается процент вхождения в выделенный фактор характеристик относительно всего корпуса. По осям откладывается отклонение набора признаков от указанного в названии оси процента вариации для данного подкорпуса. Удаление от пересечения осей координат в расположении подкорпуса по осям координат указывает на отличие набора признаков (частота встречаемости) от признаков по всему корпусу.

Результаты анализа соответствий явно показывают разделение подкорпусов текстов выделенных неявных пересекающихся сообществ по лингвистическим характеристикам. По нашему мнению, это можно рассматривать как свойство выделенных на графе сообществ, которое может подтверждать корректность выделения.

В частности, результаты применения анализа соответствий и корреляционного анализа показывают возможность выявлять наборы лингвистических характеристик в качестве дифференцирующих признаков, характеризующих различия в массивах текстов неявных сообществ. Этот же инструментарий позволяет оценить содержание текстов выделенных сообществ и с точки зрения задачи выявления информационного воздействия на акторов сети.

Психолингвистические характеристики текстов сообществ

При анализе текстовых массивов неявных сообществ вычисляются статистические лингвистические характеристики, которые можно трактовать как психолингвистические характеристики. Данные

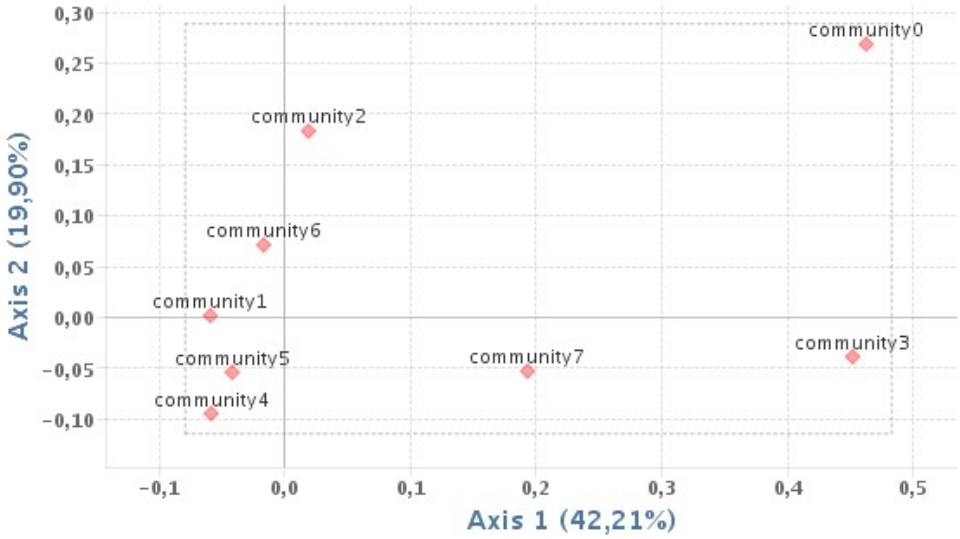


Рис. 1. *Анализ соответствий для слов корпуса текстов сообществ графа* G_1

процедуры описаны в [3, 12]. Исследуется 23 различных фактора, из которых выбираются маркеры, которые могут характеризовать психологическую направленность массивов текстов различных сообществ.

Рассмотим, в частности, два таких фактора — два коэффициента действия, вычисляемых как отношение количества глаголов к количеству прилагательных. Отношение берется в первом случае без учета причастий и деепричастных оборотов, а во втором случае — с учетом причастий и деепричастий. Для исследуемых нами текстов сообществ значения этих коэффициентов изменяются от 0,6 до 1,5. Различные значения для текстов сообществ указывают на различные установки на действия в разных сообществах. Это показывает, что коэффициенты действия являются маркерами различных сообществ.

Для текстов сообществ с близкими значениями коэффициентов действия наблюдаются различия в значениях другого фактора — коэффициента лексического разнообразия, который вычисляется как отношение числа уникальных псевдооснов к числу словоупотреблений в тексте. Для таких текстовых массивов коэффициент лексического разнообразия изменяется от 0,05 до 0,14. Такие различия связаны с различной тематикой, обсуждаемой в сообществах (например, новостные политические каналы и каналы, посвященные искусству, имеют различающиеся значения этого фактора). Поэтому коэффициент лексического разнообразия является дополнительным маркером, указывающим на принадлежность текстов разным сообществам.

Другими дополнительными маркерами текстов сообществ являются, как следует из наших экспериментов, такие характеристики синтаксических конструкций, как среднее количество числа «подгрупп» в одной именной группе и среднее количество числа «подгрупп» в одной глагольной группе. Они дают различия в значениях (от 2,7 до 4,0) для текстов сообществ с близкими значениями других перечисленных характеристик.

Таким образом, на реальных данных показана возможность использовать психолингвистические характеристики для анализа выделенных сообществ на графе взаимодействующих объектов социальных сетей и мессенджеров. С учетом возможности оценки по психолингвистическим характеристикам направленность членов сообществ данных храктеристик позволяет выделять группы, в которых ведется активное информационное воздействие.

Заключение

В данной работе сформулирован в полном объеме принципиально новый подход к оценке корректности выделения неявных сообществ на графах взаимодействующих объектов, основанный на лингвистическом анализе текстов, построенных на основе текстовых атрибутов вершин соответствующих сообществ. На основе исследования реальных данных показана возможность оценки корректности выделения неявных сообществ на графе, импортированном из социальных сетей и мессенджеров.

В работе показано, что классические методы оценки качества разбиения графа на сообщества,

такие как вычисление характеристик взаимной информации, не позволяют оценивать содержательно полученное разбиение. Это ограничивает их применение при решении проблем, связанных с анализом информационного воздействия на акторов сети.

Представленные в работе методы компьютерной лингвистики: корреляционный анализ, анализ соответствий, анализ психолингвистических характеристик, — применимы для решения как задачи оценки корректности выделения сообществ на графе социальной сети, так и задачи выявления информационного воздействия в такой сети.

ЛИТЕРАТУРА

1. Fortunato S., Newman M. E. J. 20 Years of Network Community Detection. *Nat. Phys*. 2022;18:848–850.
2. Чеповский А. А. О неявных сообществах на графе взаимодействующих объектов. Успехи кибернетики. 2023;4:56-64. DOI: 10.51790/2712-9942-2023-4-1-08.
3. Чеповский А. А. *Анализ графов взаимодействующих объектов*. М.: Национальный открытый университет «ИНТУИТ»; 2022. 270 с.
4. Fortunato S. Community Detection in Graphs. *Physics Reports*. 2010;486:75-174.
5. Danon L., Díaz-Guilera A., Duch J., Arenas A. Comparing Community Structure Identification. *J. Stat. Mech*. 2005;P09008.
6. Lancichinetti A., Fortunato S., Radicchi F. Benchmark Graphs for Testing Community Detection Algorithms. *Physical Review*. 2008;E78:046110. 6 p.
7. Newman M. E. J., Cantwell G. T., Young J.-G. Improved Mutual Information Measure for Clustering, Classification, and Community Detection. *Physical Review E*. 2020;101:042304. DOI: 10.1103/PhysRevE.101.042304.
8. Jérôme A., Kirkley A. Asymmetrically Normalized Mutual Information: Unbiased Measure for Comparing Clusterings with Different Numbers of Clusters. *Physical Review E*. 2020;101:042301.
9. Jerdee M., Kirkley A., Newman M. E. J. Normalized Mutual Information is a Biased Measure for Classification and Community Detection. arXiv:2307.01282. DOI: 10.48550/arXiv.2307.01282.
10. Аванесян Н. Л., Зенькова В. В., Чеповский А. А., Чеповский А. М. Анализ текстов сообществ социальных сетей. *Успехи кибернетики*. 2023;4:33-39. DOI: 10.51790/2712-9942-2023-4-2-05.
11. Фокина А. И., Чеповский А. А., Чеповский А. М. Использование платформы TXM корпусного анализа для анализа текстов сообществ социальных сетей. *Вестник НГУ. Серия: Информационные технологии*. 2023;21:29–38. DOI: 10.25205/1818-7900-2023-21-2-29-38.
12. Чеповский А. А. Об особенностях построения и анализа графов взаимодействующих объектов в сети Telegram-каналов. *Вопросы кибербезопасности*. 2023;1:75-81. DOI: 10.21681/2311-3456-2023-1-75-81.
13. Мамаев И. Д., Митрофанова О. А. Лингвистические параметры для идентификации скрытых сетевых сообществ. *Terra Linguistica*. 2024;15:102–115. DOI: 10.18721/JHSS.15108.
14. Мамаев И. Д. Лингвистические профили скрытых сообществ: морфосинтаксический аспект. *Филологические науки. Вопросы теории и практики*. 2024;17:1155-1162. DOI: 10.30853/phil20240168.
15. Соловьев Ф. Н. Автоматическая обработка текстов на основе платформы TXM с учетом анализа структурных единиц текста. *Вестник НГУ. Серия: Информационные технологии*. 2020;18:74–82. DOI: 10.25205/1818-7900-2020-18-1-74-82.
16. Blondel V. D., Guillaume J.-L., Lambiotte R., Lefebvre E. Fast Unfolding of Communities in Large Networks. *Journal of Statistical Mechanics: Theory and Experiment*. 2008;10:P10008.
17. Бендат Дж., Пирсол А. *Прикладной анализ случайных данных*. М.: Мир; 1989. 540 с.
18. Деза Е. И., Деза М. М. *Энциклопедический словарь расстояний*. М.: Наука; 2008. 444 с.
19. Benzécri J.-P., Bellier L. *L'analyse des données. V. 2: L'analyse des Correspondances*. Paris: Dunod; 1976. 616 p.
20. Heiden S. The TXM Platform: Building Open-Source Textual Analysis Software Compatible with the TEI Encoding Scheme. *Proceedings of the 24th Pacific Asia Conference on Language, Information and Computation*. Sendai, Japan. 2010:389–398.

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ КОЛЛЕКТИВНОГО ПОВЕДЕНИЯ РОЯ МИНИАТЮРНЫХ РОБОТОВ

К. В. Костарев[а], Д. А. Брацун[б]
Пермский национальный исследовательский политехнический университет,
г. Пермь, Российская Федерация
[а] ✍ *kost.kreal@gmail.com*
[б] *ORCID: https://orcid.org/0000-0002-3229-2330, DABracun@pstu.ru*

Аннотация: в данной работе представлена двумерная математическая модель коллективного поведения роя миниатюрных роботов. Габариты, масса и ряд необходимых характеристик были установлены согласно экспериментальным данным. Роботы представлены в виде совокупности дискретных тел, взаимодействующих друг с другом в соответствии с эффективным потенциалом, вид которого зависит от взаимодействия сенсоров и излучателей с микроконтроллером. Каждый робот является источником тепла и при этом следует в направлении градиента глобального температурного поля. Для расчета температуры используется двумерное уравнение теплопроводности. Представлены результаты численного моделирования взаимодействия от нескольких штук до 500 роботов. Разработанная программа позволяет воспроизводить различные механизмы самоорганизации и коллективного поведения, например, самосборку роя роботов, находящихся на расстоянии. Для оптимизации времени выполнения применяется программно-аппаратная архитектура параллельных вычислений Nvidia CUDA. Для визуализации полученных данных использована программа gnuplot.

Ключевые слова: самоорганизация, роботы, коллективное поведение, групповая робототехника, активная материя, эмерджентность, математическое моделирование.

Благодарности: работа выполнена при финансовой поддержке Министерства науки и высшего образования Российской Федерации (грант № FSNM-2023-0003).

Для цитирования: Костарев К. В., Брацун Д. А. Математическое моделирование коллективного поведения роя миниатюрных роботов. *Успехи кибернетики*. 2025;6(1):62–69.

Поступила в редакцию: 07.10.2024. *В окончательном варианте: 10.10.2024.*

SIMULATION OF THE COLLECTIVE BEHAVIOR OF A MINIATURE ROBOT SWARM

K. V. Kostarev[a], D. A. Bratsun[b]
Perm National Research Polytechnic University, Perm, Russian Federation
[a] ✍ *kost.kreal@gmail.com*
[b] *ORCID: https://orcid.org/0000-0002-3229-2330, DABracun@pstu.ru*

Abstract: we developed a two-dimensional model to simulate the collective behavior of a miniature robot swarm. The model incorporates size, mass, and other experimental characteristics. It represents the robots as discrete bodies that interact through an effective potential as the interaction of sensors and emitters with a microcontroller. Each robot generates heat and moves along the gradient of the global temperature field, calculated using the 2D heat conduction equation. We present numerical simulations for interactions ranging from a few robots to swarms of up to 500 robots. The software simulates various self-organization and collective behavior mechanisms, including the self-assembly of robot swarms. To optimize runtime, we implemented the Nvidia CUDA parallel computing architecture. We used gnuplot for visualization of the results.

Keywords: self-organization, robots, swarm intelligence, collective behavior, group robotics, active matter, emergence, simulation.

Acknowledgements: this study is supported by the Ministry of Science and Higher Education of the Russian Federation (grant No. FSNM-2023-0003).

Cite this article: Kostarev K. V., Bratsun D. A. Simulation of the Collective Behavior of a Miniature Robot Swarm. *Russian Journal of Cybernetics*. 2025;6(1):62–69.

Original article submitted: 07.10.2024. *Revision submitted: 10.10.2024.*

Введение

Явление самоорганизации наблюдается как в природных, так и в синтетических системах. В результате самоорганизации система приобретает новые свойства, изменяя свои качества [1]. Самоорганизация на нашей планете наблюдается на различных масштабных уровнях от молекулярного до

экосистемного. Примеры включают кристаллизацию, тепловую конвекцию жидкостей [2], колебания в химических реакциях [3], коллективное поведение живых организмов [4].

Аналог хорошо изученной тепловой конвекции жидкости неоднократно обсуждался учеными в контексте живой природы. В частности, авторы [5] утверждают, что тепловая конвекция существенно влияет на распределение температуры в пчелином улье. Аналог тепловой конвекции наблюдается также при коллективном поведении императорских пингвинов во время антарктической зимовки [6]. Обнаруженные в живой природе механизмы могут быть применены для построения синтетических систем, состоящих из большого числа роботов.

На сегодняшний день на производствах и в научных исследованиях широко распространены одиночные роботизированные системы. Однако существует множество задач, при решении которых рой относительно простых роботов справится лучше, чем один сложный [7], подобно колониям муравьев, пчел или термитов в природе. Исследования в этой области направлены на разработку алгоритмов и технологий, которые позволят роботам работать максимально слаженно, открывая новые горизонты для применения в различных сферах от промышленности до медицины.

Исследования в области коллективной робототехники требуют постройки значительного количества единиц. По причинам стоимости и сложности редко в проектах используется более 100 роботов. Алгоритмы коллективного поведения роя из 1000 и более единиц проверяются только с помощью математического моделирования.

Чтобы обеспечить возможность создания большого количества роботов, учеными из Гарвардского университета были разработаны миниатюрные роботы под названием Kilobot [8]. Данный робот имеет круглую форму диаметром 33 миллиметра. Передвигается в пространстве благодаря трем ножкам и двум вибромоторам. Управляя величиной вибрации для двух двигателей независимо в дифференциальном приводе, робот может двигаться в непрерывном диапазоне от вращения по часовой стрелке, прямо вперед, до вращения против часовой стрелки. Это позволяет килоботам двигаться примерно на 1 см/сек и вращаться примерно на 45 градусов/сек. Роботы способны обмениваться сигналами и измерять расстояние до ближайших соседей. В зависимости от модификации, роботы могут иметь дополнительные функции, например, измерение температуры или уровня освещения.

Математическая модель

Представим, что рой состоит из роботов, которые имеют одинаковые характеристики и работают по одному и тому же алгоритму. Каждый бот выполняет следующую последовательность действий: сначала он генерирует тепло с фиксированной мощностью, затем измеряет температуру в нескольких точках на своей периферии и определяет направление глобального теплового градиента. После этого бот перемещается по плоскости в направлении этого градиента. В случае нулевого градиента температуры робот не будет двигаться.

Совокупность роботов в модели рассматривается как сложная система локально взаимодействующих между собой микроскопических элементов на основе Аристотелевой механики. Это обусловлено диссипативным характером перемещения роботов:

$$m_i \frac{d\overrightarrow{r_i}}{dt} = \sum \overrightarrow{F_{ij}}(\overrightarrow{r}). \tag{1}$$

Для прогнозирования поведения i-го робота в определенный промежуток времени необходимо учесть все силы, действующие на него в момент времени t. Также важно задать начальные условия для координат каждого робота.

В рамках модели мы принимаем каждого робота как круг с определенным радиусом и массой. Физические аспекты взаимодействия особей в группе определяются с помощью потенциальных сил, которые имеют специальный вид.

Таким образом, у роботов в совокупности существует общее тепловое поле. Зона максимальной температуры будет находиться в центре коллективной структуры. Каждый робот будет стремиться продвигаться в этом направлении. Из этого следует:

$$\overrightarrow{F} = \mathrm{k}\nabla T, \tag{2}$$

где k — размерный коэффициент пропорциональности между градиентом температуры и силой. Сила $\overrightarrow{F}$ является основным фактором, который определяет самоорганизацию в системе.

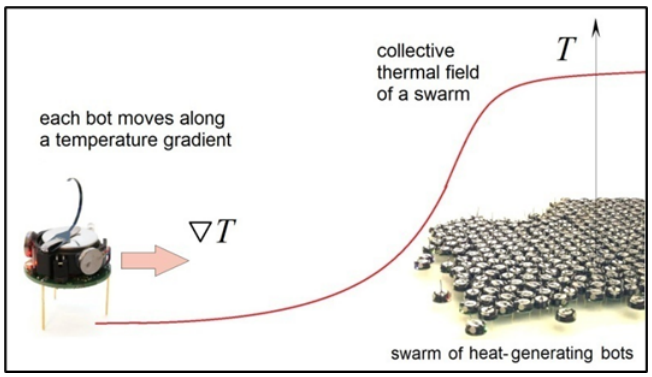


Рис. 1. *Самосборка роя роботов происходит благодаря алгоритму, который предписывает им следовать градиенту коллективного теплового поля*

Чтобы учесть столкновения роботов, необходимо ввести механизм отталкивания. На i-го робота со стороны j-го будет действовать отталкивающая сила:

$$\vec{\Phi} = \vec{n}_{ij} A e^{\frac{r_{ij} - d_{ij}}{B}}, \tag{3}$$

где A и B — некоторые постоянные, от которых зависит величина силы отталкивания, r_{ij} — сумма радиусов r_i и r_j.

Обозначим расстояние между центрами масс через:

$$d_{ij} = \left\| r_{ij} - d_{ij} \right\| \tag{4}$$

и вектор нормали, направленный от робота i к j:

$$n_{ij} = \frac{(\vec{r}_i - \vec{r}_j)}{d_{ij}}. \tag{5}$$

Мы рассматриваем особи как круги одинакового радиуса $R = 0.2$ (м), следовательно, формулу (3) можно упростить:

$$\vec{\Phi} = \vec{n}_{ij} A e^{\frac{2R - d_{ij}}{B}}. \tag{6}$$

С учетом всех описанных выше сил, запишем систему обыкновенных дифференциальных уравнений:

$$\begin{cases} m\dfrac{dx_i}{dt} = k\dfrac{\partial T}{\partial x} + \displaystyle\sum_{\substack{j=1 \\ j\neq i}}^{M} A e^{\frac{2R - \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}}{B}} \cdot \dfrac{(x_i - x_j)}{\sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}} \\ m\dfrac{dy_i}{dt} = k\dfrac{\partial T}{\partial y} + \displaystyle\sum_{\substack{j=1 \\ j\neq i}}^{M} A e^{\frac{2R - \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}}{B}} \cdot \dfrac{(y_i - y_j)}{\sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}} \end{cases} \tag{7}$$

Во второй части модели мы строим общее тепловое поле. Каждый робот является источником тепла с определенной мощностью. Уравнение, которое описывает это тепловое поле, представляет собой двумерное уравнение теплопроводности, учитывающее наличие источников тепла:

$$\frac{\partial T}{\partial t} = \lambda \left(\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} \right) + \sum_{\mu=1}^{M} P_\mu \left(x_\mu, y_\mu \right), \tag{8}$$

где $\sum_{\mu=1}^{M} P_\mu \left(x_\mu, y_\mu \right)$ — это сумма мощностей источников тепла, λ — коэффициент теплопроводности.

Двумерное уравнение теплопроводности — это дифференциальное уравнение в частных производных второго порядка, описывающее распределение температуры в заданной области пространства и ее изменение во времени. Для его численного решения мы используем метод конечных разностей,

который позволяет заменить дифференциальное уравнение системой алгебраических уравнений на конечной разностной сетке.

$$\begin{gathered} T\,(x,y) \Longrightarrow T\,(i\Delta x, j\Delta y) \\ T_{ij} = T\,(x_0, y_0) \\ T_{i+1,j} = T(x_0 + \Delta x,\, y) \\ T_{i-1,j} = T(x_0 - \Delta x,\, y) \\ T_{i,j+1} = T(x_0,\, y + \Delta y) \\ T_{i,j-1} = T(x_0,\, y - \Delta y). \end{gathered} \tag{9}$$

В результате первоначальная задача редуцируется до задачи вычисления значений переменных в дискретных узлах сетки. Такая процедура приводит к формированию явной разностной схемы, которая позволяет найти приближенное решение исходного дифференциального уравнения.

$$\begin{gathered} \frac{T_{ij}^{n+1} - T_{ij}^{n}}{\Delta t} = \lambda \left(\frac{T_{i-1,j}^{n} - 2T_{ij}^{n} + T_{i+1,j}^{n}}{(\Delta x)^2} + \frac{T_{i,j-1}^{n} - 2T_{ij}^{n} + T_{i,j+1}^{n}}{(\Delta y)^2} \right) \\ - V_x \frac{T_{i+1,j}^{n} - T_{i-1,j}^{n}}{2\Delta x} - V_y \frac{T_{i,j+1}^{n} - T_{i,j-1}^{n}}{2\Delta y} \quad - DT_{i,j}^{n} + \sum_{\mu=1}^{M} P_\mu(x_\mu, y_\mu) \end{gathered} \tag{10}$$

$$\begin{gathered} T\,(x,y,0) = T_0(x,y) \\ T\,(x,y,t)\,|_\Gamma = T_0(x,y,t). \end{gathered}$$

В результате модель состоит из двух разных частей, связанных нелинейным образом. Расчет теплового поля использует данные о местоположении каждой особи, так как именно они являются источниками тепла. С другой стороны, общее тепловое поле определяет динамику самих особей.

$$\frac{d\overrightarrow{r_i}}{dt} = k\frac{\partial T}{\partial \overrightarrow{r_i}} + \sum_{j=1, i\neq j}^{M} \overrightarrow{n}_{ij} A e^{\frac{2R - d_{ij}}{B}}, \tag{11}$$

$$\frac{T_{ij}^{n+1} - T_{ij}^{n}}{\Delta t} = \lambda \left(\frac{T_{i-1,j}^{n} - 2T_{ij}^{n} + T_{i+1,j}^{n}}{(\Delta x)^2} + \frac{T_{i,j-1}^{n} - 2T_{ij}^{n} + T_{i,j+1}^{n}}{(\Delta y)^2} \right) + \sum_{\mu=1}^{M} P_\mu\left(x_\mu, y_\mu\right). \tag{12}$$

Следует заметить, что при расчете обеих частей модели шаг по времени остается постоянным. Для обеспечения устойчивости явной схемы он определяется условием Куранта.

Программная реализация

Расчеты, лежащие в основе данной модели, возможно выполнить на одном процессорном ядре. Однако такой подход потребует значительных временных затрат. Для достижения оптимальной производительности целесообразно использовать возможности параллельных вычислений на графическом процессоре. Реализация модели выполнена на языке Cuda C, который позволяет создавать программы, выполняемые на GPU.

При распараллеливании вычислений на GPU важно обеспечить синхронизацию доступа к разделяемым данным между потоками выполнения. В противном случае возможны ошибки в результатах расчетов. В данной модели для синхронизации потоков используются различные механизмы, такие как атомарные операции и барьеры синхронизации.

Для визуализации полученных данных используется кроссплатформенная программа gnuplot. Она может работать одновременно с выполнением вычислений и позволяет генерировать выходные изображения с помощью встроенной системы команд.

Код программы можно представить в виде блок схемы.

В таблице приведены значения основных переменных, использованных в нашей модели, где k, A, B — коэффициенты притяжения и отталкивания между роботами, R, m — радиус и масса робота, hx — шаг по сетке, dif — коэффициент тепловой диффузии. После инициализации начальных усло-

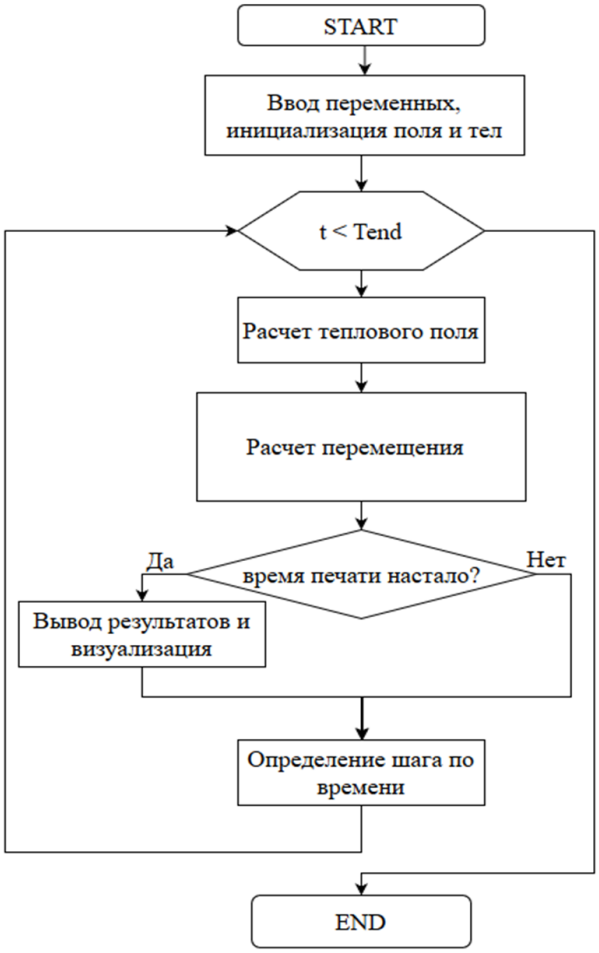


Рис. 2. *Блок-схема реализованной программы*

Таблица

Значения основных переменных

Переменная	Значение
m	0.02 кг
k	0.0006 Н
A	0.0008 Н
B	0.015
hx	0.01 м
R	0.02 м
dif	0.02

вий программа входит в цикл по времени, в котором происходит взаимодействие двух пространств: непрерывного и дискретного. Сначала на основе данных поиска источников тепла происходит расчет значений теплового поля. После расчета теплового поля программа начинает определение новых координат роботов, которые изменяются благодаря изменению направления и величин векторов движения. Сначала рассчитывается сила отталкивания. Следующим шагом является вычисление направления вектора линейной силы для каждого робота. Алгоритм вычисления можно описать следующим образом: мы выделяем многоугольную область вокруг каждого робота, а далее проходим по периметру, находя точку с максимальным значением теплового поля. Вектор силы будет направлен в сторону области с

максимальным значением температуры. Новые координаты особей меняют местоположение источников тепла в пространстве, что приводит к изменению теплового поля в пространстве. Этот процесс повторяется, пока время не достигнет заданного значения.

Результаты численных экспериментов

Первоначально работы находятся в беспорядочном движении. Это обусловлено отсутствием выраженного центра тепла. В зонах с наибольшей плотностью температура начинает подниматься быстрее, и там скапливается все большее количество роботов. Далее происходит плавное поглощение маленьких групп более крупными.

Диффузия выступает в роли механизма передачи тепловой информации между роботами. Каждый робот оценивает интенсивность теплового поля вокруг себя, определяет свои дальнейшие действия и движется в направлении градиента. Это приводит к образованию единой группы, как показано на рисунках 3 и 4. В некоторых случаях образуются несколько групп, которые могут существовать долго независимо друг от друга.

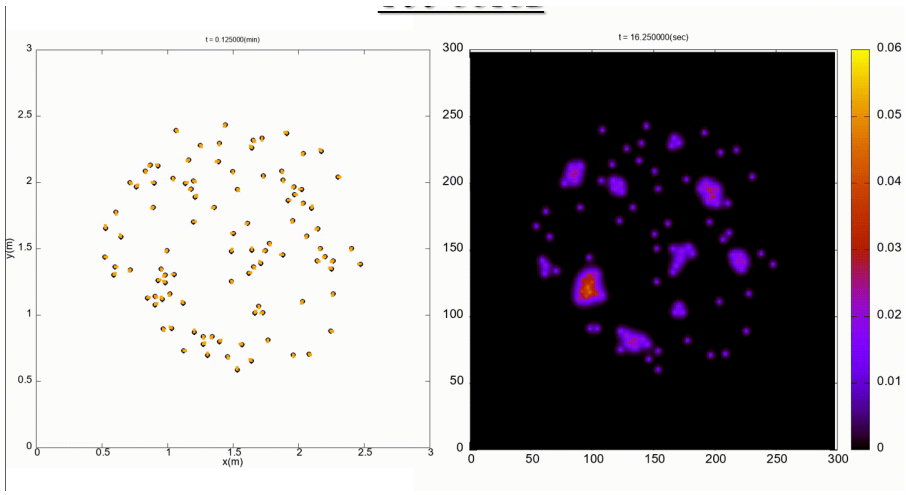


Рис. 3. *В начале численного эксперимента роботы распределены по пространству случайным образом*

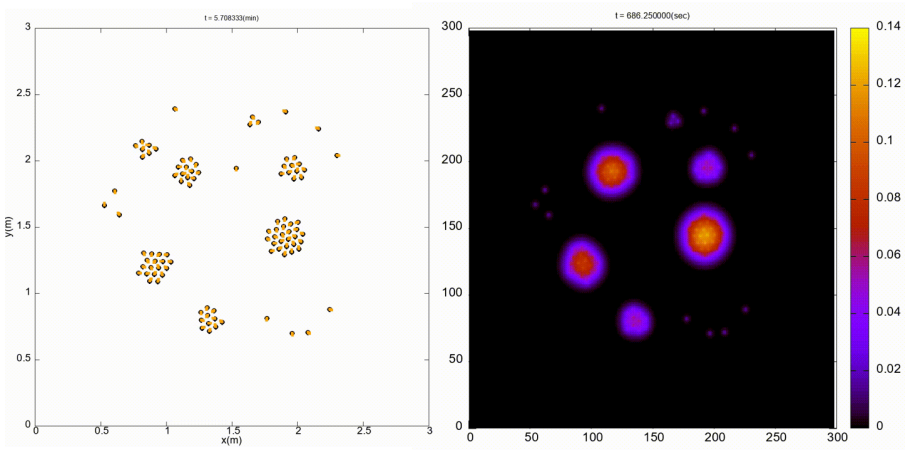


Рис. 4. *Образование нескольких скученных групп роботов*

На этом заканчивается первый этап самоорганизации в системе. Проводя аналогию между коллективной динамикой роботов и агрегатными состояниями вещества, можно сделать вывод, что рой ботов переходит в квазитвердую фазу. С течением времени группы роботов приобретают круглую форму, что с точки зрения термодинамики является оптимальным состоянием для минимизации тепловых потерь.

С увеличением количества роботов усложняется поведение роя в целом. После сборки в одну крупную структуру группа из 500 роботов демонстрирует образование вихревых потоков. Роботы продвигаются от края к центру структуры, после чего снова выталкиваются наружу. Данные движения имеют схожую природу с тепловой конвекцией в сплошных средах. Далее движение роботов в группе синхронизируется и вихревые движения прекращаются. В конечном счете вся группа приобретает одинаковое направление движения по часовой или против часовой стрелки в зависимости от начальных положений роботов на плоскости. Эволюцию системы можно наблюдать на изображениях 5 и 6.

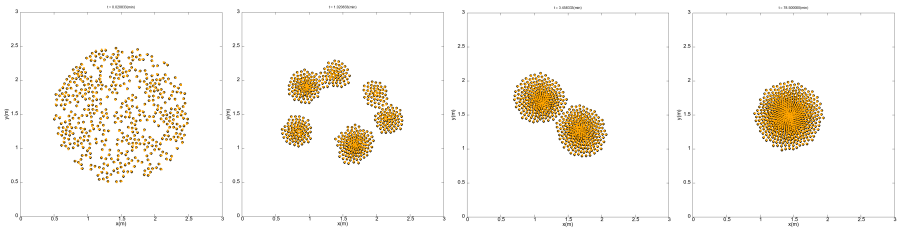


Рис. 5. *Образование крупной квазитвердой группы из 500 роботов*

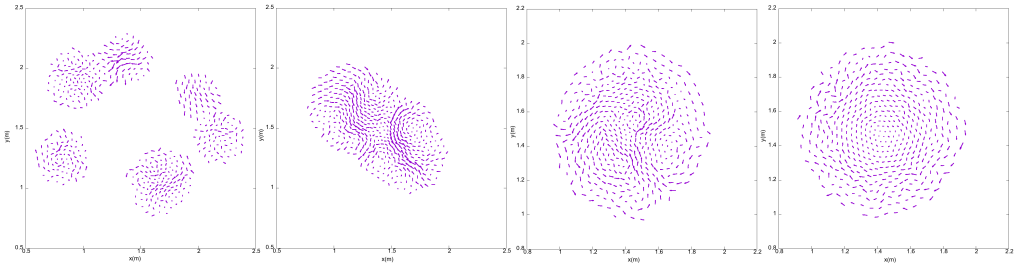


Рис. 6. *Векторные поля, демонстрирующие перемещения роботов в разные моменты времени*

Выводы

Данная модель адекватно отображает групповую динамику роя роботов, включая процессы, аналогичные фазовому переходу первого порядка, происходящему при снижении температуры окружающей среды, движению в направлении температурного градиента и формированию нескольких тепловых центров в крупных группах.

С увеличением количества роботов система в целом показывает все более сложное поведение. Модель демонстрирует, что в группе возможно появление вихревых движений, сходных по механизму с тепловой конвекцией. Анализ эволюции вихревых структур показывает, что со временем данные макроскопические движения затухают. Дальнейшие исследования системы будут направлены на подбор параметров, которые позволят получить устойчивые вихревые движения.

ЛИТЕРАТУРА

1. Хакен Г. *Синергетика. Иерархии неустойчивостей в самоорганизующихся системах и устройствах.* М.: Мир; 1985. 424 с. Режим доступа: https://spkurdyumov.ru/uploads/2018/07/sinergetika-ierarxii-neustoiychivostey.pdf.
2. Гершуни Г. З., Жуховицкий Е. М. *Конвективная устойчивость несжимаемой жидкости.* М.: Наука; 1972. 392 с.
3. Жаботинский А. М. Периодический ход окисления малоновой кислоты в растворе (исследование кинетики реакции Белоусова). *Биофизика*. 1964;9(3):306–311.
4. Trenchard H. The Peloton Superorganism and Protocooperative Behavior. *Applied Mathematics and Computation.* 2015;270:179–192. Режим доступа: https://arxiv.org/pdf/1509.05965.
5. Гакашев А. И., Тарунин Е. Л. Интенсивность тепловой конвекции в ульях. *Вычислительная механика сплошных сред.* 2008;1(2):16–26. Режим доступа: https://journal.permsc.ru/index.php/ccm/article/download/CCMv1n2a2/345.

6. Брацун Д. А., Костарев К. В. Сплошносредная модель биоконвекции с центростремительной силой. *Вестник Пермского университета*. 2022;2:36–46. Режим доступа: https://cyberleninka.ru/article/n/sploshnosrednaya-model-biokonvektsii-s-tsentrostremitelnoy-siloy.
7. Xie H., Sun M., Fan X., Lin Z., Chen W., Wang L., He Q. Reconfigurable Magnetic Microrobot Swarm: Multimode Transformation, Locomotion, and Manipulation. *Science Robotics*. 2019;4:28. Режим доступа: 10.1126/scirobotics.aav8006.
8. Rubenstein M., Ahler C., Hoff N., Cabrera A., Nagpal R. Kilobot: A Low Cost Robot with Scalable Operations Designed for Collective Behaviors. *Robotics and Autonomous Systems*. 2014;62(7):966–975. Режим доступа: 10.1016/j.robot.2013.08.006.

КОМПЬЮТЕРНОЕ МОДЕЛИРОВАНИЕ ПОЛНОЙ ЭНЕРГИИ ДВУХАТОМНОЙ МОЛЕКУЛЫ КРЕМНИЯ В ПЕРВОМ ПОРЯДКЕ ТЕОРИИ ВОЗМУЩЕНИЙ

Ю. Н. Штанов[1,а], **В. П. Кощеев**[2,б]

[1] *Тюменский индустриальный университет, филиал «Сургутский институт нефти и газа», г. Сургут, Российская Федерация*

[2] *Московский авиационный институт (национальный исследовательский университет), филиал «Стрела», г. Жуковский, Московская область, Российская Федерация*

[а] *ORCID: https://orcid.org/0000-0002-5822-3368,* ✍ *yuran1987@mail.ru*

[б] *ORCID: https://orcid.org/0000-0002-0724-9760, koshcheev1@yandex.ru*

Аннотация: разработана компьютерная программа, которая предназначена для моделирования полной энергии двухатомной молекулы с помощью волновых функций, которые аппроксимируют решения уравнения Хартри–Фока для изолированных атомов. В работе выполнено моделирование полной энергии двухатомной молекулы кремния с помощью метода компьютерного моделирования в первом порядке теории возмущений, параметры которого найдены с помощью численного решения самосогласованной системы уравнений.

Ключевые слова: теория возмущений, атомный форм-фактор, волновые функции, уравнение Хартри–Фока, самосогласованная система уравнений, атом кремния.

Для цитирования: Штанов Ю. Н., Кощеев В. П. Компьютерное моделирование полной энергии двухатомной молекулы кремния в первом порядке теории возмущений. *Успехи кибернетики*. 2025;6(1):70–75.

Поступила в редакцию: 16.11.2024. *В окончательном варианте: 30.11.2024.*

COMPUTER SIMULATION OF THE TOTAL ENERGY OF A DIATOMIC SILICON MOLECULE USING THE FIRST-ORDER PERTURBATION THEORY

Yu. N. Shtanov[1,a], **V. P. Koshcheev**[2,b]

[1] *Industrial University of Tyumen, Surgut Branch, Surgut, Russian Federation*

[2] *Moscow Aviation Institute (National Research University), Strela Branch, Zhukovsky, Moscow Region, Russian Federation*

[a] *ORCID: https://orcid.org/0000-0002-5822-3368,* ✍ *yuran1987@mail.ru*

[b] *ORCID: https://orcid.org/0000-0002-0724-9760, koshcheev1@yandex.ru*

Abstract: we developed an application to simulate the total energy of a diatomic molecule using wave functions that approximate the solutions of the Hartree-Fock equation for isolated atoms. The paper presents a simulation of the total energy of a diatomic silicon molecule using computer simulation based on the first-order perturbation theory. We found the parameters through a numerical solution of a self-consistent system of equations.

Keywords: perturbation theory, atomic form factor, wave functions, Hartree–Fock equation, self-consistent system of equations, silicon atom.

Cite this article: Shtanov Yu. N., Koshcheev V. P. Computer Simulation of the Total Energy of a Diatomic Silicon Molecule Using the First-Order Perturbation Theory. *Russian Journal of Cybernetics*. 2025;6(1):70–75.

Original article submitted: 16.11.2024. *Revision submitted: 30.11.2024.*

Введение

С помощью компьютерного эксперимента в [1, 2] исследуются новые типы упорядоченных структур, состоящих из атомарного кремния. Полная энергия взаимодействия между атомами вычисляется с помощью уравнения теории функционала плотности, решение которого для основного состояния совпадает с решением уравнения Хартри–Фока в первом порядке теории возмущений [3]. С другой стороны, достоверность вычисления с помощью многочисленных версий теории функционала плотности достигается тогда, когда результаты совпадают между собой [4]. В настоящей публикации вычислена полная энергия двухатомной молекулы кремния с помощью метода компьютерного моделирования в первом порядке теории возмущений [5].

Двухатомную молекулу будем описывать с помощью стационарного уравнения Шредингера:

$$H\psi = E\psi. \tag{1}$$

Гамильтониан уравнения (1) представим в виде

$$H = H^0 + U, \tag{2}$$

$$U = \frac{Z_1 Z_2 e^2}{|\vec{r}_1 - \vec{r}_2|} + \sum_{j_1=1}^{Z_1}\sum_{j_2=1}^{Z_2} \frac{e^2}{\left|\vec{r}_1 + \vec{r}_{1j_1} - \vec{r}_2 - \vec{r}_{2j_2}\right|} - \sum_{j_1=1}^{Z_1} \frac{Z_2 e^2}{\left|\vec{r}_1 + \vec{r}_{1j_1} - \vec{r}_2\right|} - \sum_{j_2=1}^{Z_2} \frac{Z_1 e^2}{\left|\vec{r}_1 - \vec{r}_2 - \vec{r}_{2j_2}\right|}, \tag{3}$$

где U — потенциальная энергия взаимодействия двух атомов; $\vec{r}_1$ и $\vec{r}_2$ — координаты первого и второго атомного ядра; $\vec{r}_1 + \vec{r}_{1j_1}$ и $\vec{r}_2 + \vec{r}_{2j_2}$ — координаты j_1-го и j_2-го электронов первого и второго атома, соответственно; $\vec{r}_1 - \vec{r}_2 = \vec{r} + \delta\vec{r}$; коллективные колебания атомных электронов и ядер описываются вектором $\delta\vec{r} = \delta\vec{r}_{nucl.} + \delta\vec{r}_{el.}$; $r = |\vec{r}|$ — расстояние между атомами в молекуле.

Решение уравнения (1) с гамильтонианом (2) будем искать с помощью теории возмущений $\psi = \psi^0 + \psi^1 + \ldots$ и $E = E^0 + E^1 + \ldots$. Потенциальную энергию (электронные термы) двухатомной молекулы будем искать в первом порядке теории возмущений

$$E^1 = \left\langle \psi^0 \middle| U \middle| \psi^0 \right\rangle, \tag{4}$$

где угловые скобки $\langle \ldots \rangle$ были введены Дираком [6].

Гамильтониан H^0 представим в виде $H^0 = H_1^0 + H_2^0$, где H_i^0 — гамильтониан i-го атома; $i = 1{,}2$.

Решение уравнения Шредингера $H^0\psi^0 = E^0\psi^0$, будем искать в виде $\psi^0 = \psi_1^0\psi_2^0$ и $E^0 = E_1^0 + E_2^0$, где уравнение Шредингера для i-го изолированного атома имеет вид

$$H_i^0\psi_i^0 = E_i^0\psi_i^0, \tag{5}$$

где $\psi_i^0 = \psi_i^0\left(\vec{r}_{i1}, \vec{r}_{i2}, \ldots, \vec{r}_{iZ_i}\right)$.

Усреднение по квантовым флуктуациям местоположения атомных электронов будем осуществлять с помощью метода [7], который Бете использовал для вычисления атомного форм-фактора, а усреднение по коллективным колебаниям атомных электронов и ядер выполним по квадрату модуля волновой функции гармонического осциллятора в основном состоянии. Соответствующие средние будем обозначать $\left\langle \psi_1^0 \middle| U \middle| \psi_1^0 \right\rangle = \langle \ldots \rangle_{e1}$, $\left\langle \psi_2^0 \middle| U \middle| \psi_2^0 \right\rangle = \langle \ldots \rangle_{e2}$ и $\left\langle \psi^0 \middle| U \middle| \psi^0 \right\rangle_{pl} = \langle \ldots \rangle_{pl}$.

Разложим потенциальную энергию взаимодействия (3) в интеграл Фурье:

$$U = \int \frac{d^3\vec{k}}{(2\pi)^3} \frac{4\pi e^2}{k^2} \exp\left(i\vec{k}\left(\vec{r} + \delta\vec{r}\right)\right) \left[Z_1 Z_2 + \sum_{j_1=1}^{Z_1}\sum_{j_2=1}^{Z_2} \exp\left(i\vec{k}\left(\vec{r}_{1j_1} - \vec{r}_{2j_2}\right)\right) - Z_2 \sum_{j_1=1}^{Z_1} \exp\left(i\vec{k}\vec{r}_{1j_1}\right) - Z_1 \sum_{j_2=1}^{Z_2} \exp\left(i\vec{k}\vec{r}_{2j_2}\right)\right]. \tag{6}$$

Усредним потенциальную энергию взаимодействия двух атомов (6)

$$U(r,\Delta) = \int \frac{d^3\vec{k}}{(2\pi)^3} \frac{4\pi e^2}{k^2} \left[Z_1 - F_1(k)\right]\left[Z_2 - F_2(k)\right] \exp\left[-k^2\Delta\right] \exp\left(i\vec{k}\vec{r}\right), \tag{7}$$

где $\left\langle \sum_{j_1=1}^{Z_1} \exp\left(i\vec{k}\vec{r}_{1j_1}\right)\right\rangle_{e1} = F_1(k)$; $F_1(0) = Z_1$ и $\left\langle \sum_{j_2=1}^{Z_2} \exp\left(i\vec{k}\vec{r}_{2j_2}\right)\right\rangle_{e2} = F_2(k)$; $F_2(0) = Z_2$ — атомные форм-факторы первого и второго атомов; $\left\langle \exp\left(i\vec{k}\delta\vec{r}\right)\right\rangle_{pl} = \exp\left[-k^2\Delta\right]$; $2\Delta = \sigma_{nucl.}^2 + \sigma_{el.}^2$ — сумма средних квадратов амплитуды коллективных атомных ядерных $\sigma_{nucl.}^2 = \frac{\hbar}{2\omega_{nucl.}\mu_{nucl.}}$ и электронных $\sigma_{el.}^2 = \frac{\hbar}{2\omega_{el.}\mu_{el.}}$ колебаний в расчете на одну степень свободы; упругая постоянная $\omega_{el.}^2\mu_{el.} = \omega_{nucl.}^2\mu_{nucl.} = U''(r_{\min})$ есть значение второй производной в минимуме потенциальной энергии молекулы.

Известно, что волновая функция двухатомной молекулы должна быть антисимметричной относительно перестановки координат, определяющих местоположения электронов. Если волновая функция двухатомной молекулы выбрана в виде произведения волновых функций изолированных атомов

$\psi^0 = \psi_1^0\psi_2^0$, то обменные силы будут учитываться отдельно для электронов изолированных атомов и отдельно для электронов, которые принадлежат первому и второму атомам.

Выражение (7) для потенциальной энергии взаимодействия без учета обменных сил между электронами изолированных атомов запишем в виде

$$U\left(r,\Delta\right) = \frac{Z_1 Z_2 e^2}{r}\Phi\left(r,\Delta\right), \tag{8}$$

где $\Phi\left(r,\Delta\right) = \frac{2}{\pi}\int_0^\infty \frac{1}{k}\left[1-\frac{F_1(k)}{Z_1}\right]\left[1-\frac{F_2(k)}{Z_2}\right]\sin\left(kr\right)\exp\left[-k^2\Delta\right]dk$ — функция экранирования потенциальной энергии двухатомной молекулы.

Непосредственной подстановкой можно показать, что функция экранирования потенциальной энергии является решением уравнения диффузионного типа

$$\frac{\partial\Phi(r,\Delta)}{\partial\Delta} = \frac{\partial^2\Phi(r,\Delta)}{\partial r^2}. \tag{9}$$

В [8] было показано, что учет коллективных колебаний атомных электронов приводит к самосогласованной системе уравнений

$$\begin{cases} \dfrac{\partial U(r,\Delta)}{\partial r} = 0, \\ \dfrac{\partial^2 U(r,\Delta)}{\partial r^2} = \dfrac{const.}{\Delta^2}, \end{cases} \tag{10}$$

где $U(r,\Delta)$ зависит от Δ согласно (8).

С помощью уравнения (9) перепишем (10) в виде

$$\begin{cases} \dfrac{\partial U\left(r,\Delta\right)}{\partial r} = 0, \\ \dfrac{\partial}{\partial\Delta}\left(U\left(r,\Delta\right) + \Delta\dfrac{\partial^2 U\left(r,\Delta\right)}{\partial r^2}\right) = 0. \end{cases} \tag{11}$$

В результате решения системы нелинейных уравнений (11) получим величины Δ и $r_{\min}$.

Известно [3], что вместо многочастичного уравнения Шредингера (5) строятся одночастичные приближения на основе уравнений Хартри–Фока или формализма теории функционала плотности. Атомный форм-фактор является Фурье-компонентой плотности распределения атомных электронов

$$F_{1,2}\left(k\right) = \int n_{1,2}\left(\vec{r}\right)\exp\left(-i\vec{k}\vec{r}\right)d^3\vec{r}, \tag{12}$$

где $n_{1,2}\left(\vec{r}\right) = 2\left|\psi_{1\mathrm{s}}\left(\vec{r}\right)\right|^2 + 2\left|\psi_{2\mathrm{s}}\left(\vec{r}\right)\right|^2 + 6\left|\psi_{2\mathrm{p}}\left(\vec{r}\right)\right|^2 + 2\left|\psi_{3\mathrm{s}}\left(\vec{r}\right)\right|^2 + 2\left|\psi_{3\mathrm{p}}\left(\vec{r}\right)\right|^2$ — электронная плотность для атома кремния Si.

Волновые функции, которые аппроксимируют решение уравнения Хартри–Фока для изолированных атомов кремния и энергии E_i^0, представлены в [9].

Условие применимости поправки в первом порядке теории возмущений к энергии системы в невозмущенном состоянии имеет вид

$$\left|E^1\right| = \left|U\left(r\right)\right| \ll \left|E^0\right|. \tag{13}$$

Известно [10], что поправка второго порядка малости к полной энергии строится с помощью волновых функций, вычисленных с точностью до первого порядка теории возмущений, то есть учитывается взаимная поляризация изолированных атомов. Поправка второго порядка малости к полной энергии должна быть много меньше расстояния между уровнями энергии (электронными термами) двухатомной молекулы, которые вычислены в первом порядке теории возмущений [10], что требует дополнительного исследования.

Программная реализация

Для решения задачи компьютерного моделирования полной энергии взаимодействия двух атомов была разработана программа TotalEnergy [11, 12]. В программе атомный форм-фактор строится

с помощью волновых функций [9], которые аппроксимируют решения уравнения Хартри–Фока для изолированных атомов (для краткости «*приближение Хартри–Фока*»). Программа позволяет выполнять расчеты для атомов с числом $Z = 2\ldots18$, с учетом электронной конфигурации атома. Программа написана на языке программирования C++, а графический интерфейс строится с использованием библиотеки Qt [13, 14].

Для численного вычисления выражения (8) используется численное интегрирование с использованием квадратур Кленшоу–Кертиса (*Clenshaw–Curtis*) и алгоритма CQUAD, который реализован в библиотеке GNU Scientific Library (GSL) [15]:

$$\int_0^{+\infty} f(x)dx = \int_0^1 \frac{1}{t^2} f\left(\frac{1-t}{t}\right) dt. \tag{14}$$

Для построения графиков высокого качества библиотека MathGL [13] позволяет экспортировать результаты в форматы postscript, svg, png, jpeg и другие. Результаты расчетов можно сохранять в формат: txt. Графический интерфейс программы представлен на рис. 1.

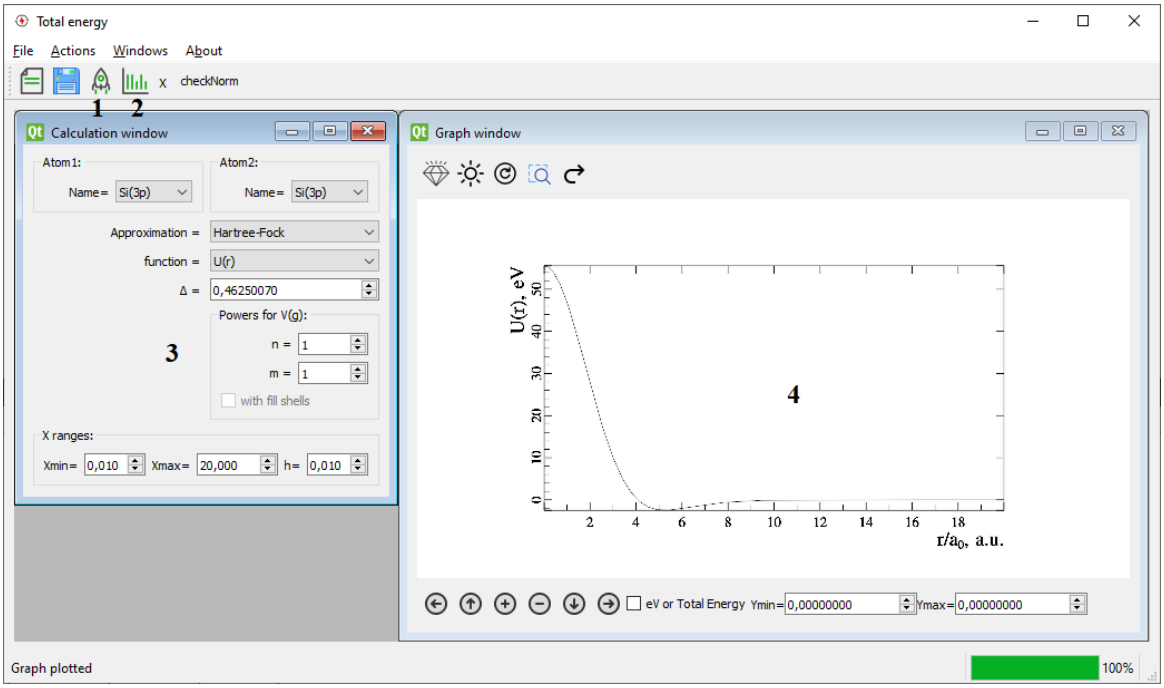


Рис. 1. *Графический интерфейс программы TotalEnergy. Обозначения: 1 — запуск расчетов, 2 — построение графиков, 3 — ввод постоянных, 4 — результаты расчетов*

Численное решение системы уравнений (11) выполнено с использованием алгоритма Ричарда Брента [16], результат с использованием приближения Хартри–Фока представлен на рис. 2 для различных состояний молекулы кремния. Результаты вычисления полной энергии для различных состояний молекулы кремния представлены на рис. 3. Видно, что глубина потенциальной ямы для молекулы кремния в основном состоянии $\mathrm{Si}\,(3p) + \mathrm{Si}\,(3p)$ в первом порядке теории возмущений наименьшая.

Заключение

Видно, что программа TotalEnergy позволяет строить графики потенциальной и полной энергий для молекул нейтральных атомов, положительных и отрицательных ионов с $Z = 2-18$. Программа является кроссплатформенной и имеет лицензию GPL. Достоинством построенного метода в сравнении с методом функционала плотности является самосогласованная система уравнений (11), которая имеет единственное решение Δ и $r_{\min}$.

Дополнительные материалы к статье размещены в [11].

ЛИТЕРАТУРА

1. Lima K. A. L. et al. Computational Insights into Popsilicene as a New Planar Silicon Allotrope Composed of 5–8–5 Rings. *Scientific Reports.* 2024;14(1):18884. DOI: 10.1038/s41598-024-69788-4.

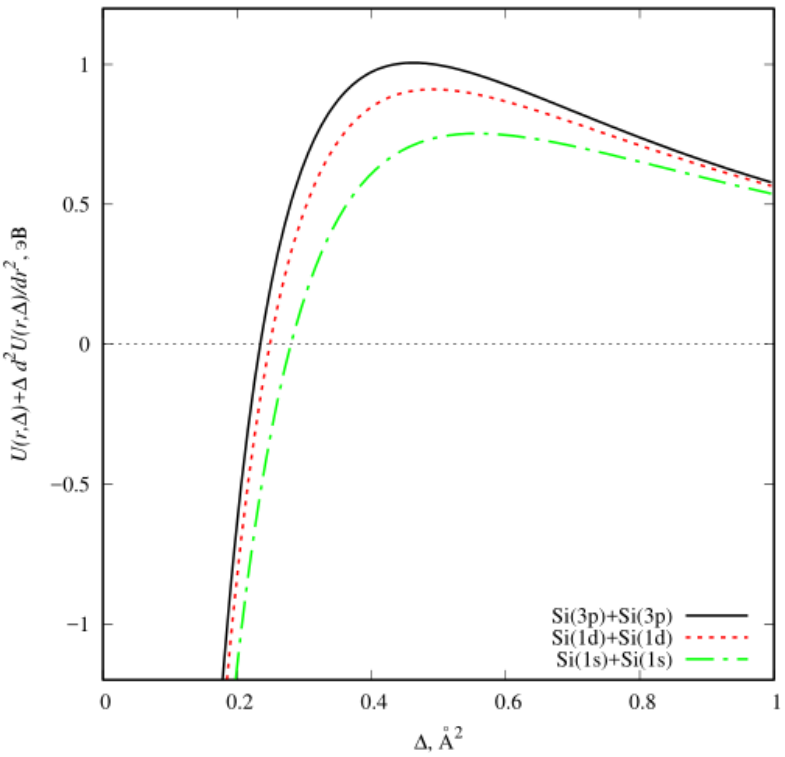


Рис. 2. *График функции* $U(r,\Delta) + \Delta\partial^2 U(r,\Delta)/\partial r^2$ *при* $\partial U(r,\Delta)/\partial r = 0$ *для* $\mathrm{Si}(3p) + \mathrm{Si}(3p)$*(черная линия),* $\mathrm{Si}(1d) + \mathrm{Si}(1d)$*(красная линия) и* $\mathrm{Si}(1s) + \mathrm{Si}(1s)$ *(зеленая линия)*

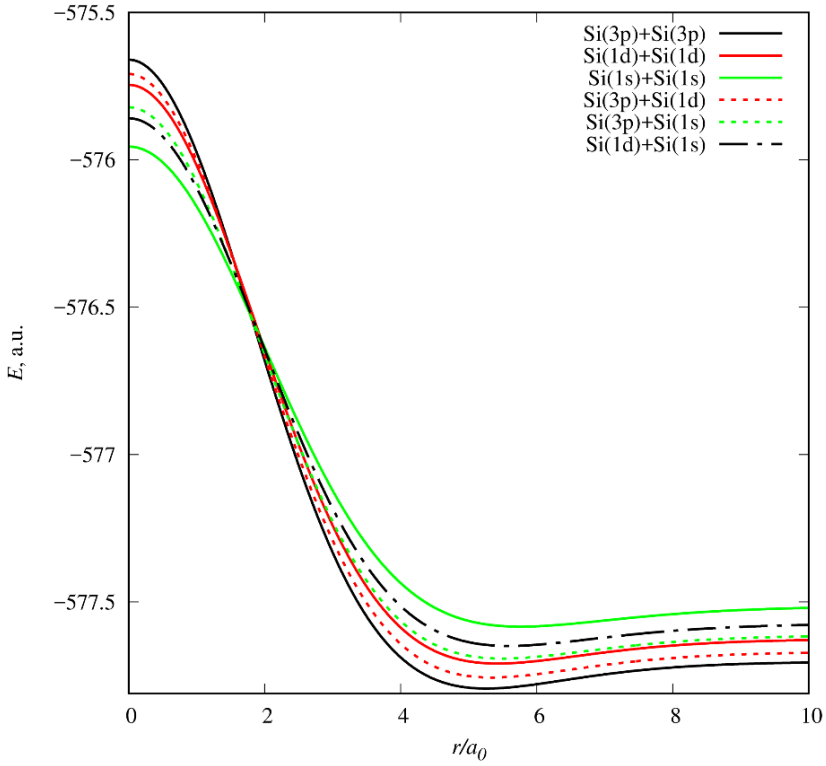


Рис. 3. *Результаты вычисления полной энергии молекулы кремния для состояний: черная сплошная линия —* $\mathrm{Si}(3p) + \mathrm{Si}(3p)$*; красная сплошная линия —* $\mathrm{Si}(1d) + \mathrm{Si}(1d)$*; зеленая сплошная линия —* $\mathrm{Si}(1s) + \mathrm{Si}(1s)$*, красная пунктирная линия —* $\mathrm{Si}(3p) + \mathrm{Si}(1d)$*; зеленая пунктирная линия —* $\mathrm{Si}(3p) + \mathrm{Si}(1s)$*, черная штрихпунктирная линия —* $\mathrm{Si}(1d) + \mathrm{Si}(1s)$

2. Wang J. et al. Computational Prediction of a New Metallic Silicon Allotrope. *Computational Materials Science.* 2025;246:113380. DOI: 10.1016/j.commatsci.2024.113380.
3. Сарры А. М., Сарры М. Ф. К теории функционала плотности. *Физика твердого тела*. 2012;54(6):1237–1243. Режим доступа: https://journals.ioffe.ru/articles/642.
4. Teale A. M. et al. DFT Exchange: Sharing Perspectives on the Workhorse of Quantum Chemistry and Materials Science. *Physical Chemistry Chemical Physics*. 2022;24(47):28700–28781. DOI: 10.1039/d2cp02827a.
5. Koshcheev V. P., Shtanov Y. N. Computer Simulation of the Total Energy and Shielding Function of a Carbon Molecule Using the First-Order Perturbation Theory. *Bulletin of the Russian Academy of Sciences: Physics*. 2024;88(4):441–446. DOI: 10.1134/S1062873823706049.
6. Дирак П. А. М. *Принципы квантовой механики*. М.: Наука; 1979. 479 с.
7. Бете Г. Квантовая механика. М.: Мир; 1965. 333 с.
8. Koshcheev V. P., Shtanov Y. N. Modeling of the Potential Energy of Interaction of Two Atoms by Solving a System of Nonlinear Equations. *Journal of Surface Investigation: X-ray, Synchrotron and Neutron Techniques*. 2023;17(1):265–269. DOI: 10.1134/S1027451023010317.
9. Clementi E., Roetti C. Roothaan-Hartree-Fock Atomic Wavefunctions: Basis Functions and Their Coefficients for Ground and Certain Excited States of Neutral and Ionized Atoms, Z≤54. *Atomic Data and Nuclear Data Tables*. 1974;14(3):177–478. DOI: 10.1016/S0092-640X(74)80016-1.
10. Ландау Л. Д., Лифшиц Е. М. *Курс теоретической физики. Том III. Квантовая механика (нерелятивистская теория).* М.: Физматлит; 2004. 800 с.
11. Штанов Ю. Н., Кощеев В. П. *Свидетельство о государственной регистрации программы для ЭВМ №2024660504.* М.: Роспатент; 2024.
12. Штанов Ю. Н., Кощеев В. П., Моргун Д. А. Библиотека программ «JINRLIB». Режим доступа: http://wwwinfo.jinr.ru/programs/jinrlib/tropics/index.html.
13. Алексеев Е. Р., Злобин Г. Г., Костюк Д. А. и др. Программирование на языке C++ в среде Qt Creator. М.: ALT Linux; 2015. 448 с. Режим доступа: https://www.altlinux.org/Books:Qt-C++.
14. Прохоренок Н. А. Qt 6. Разработка оконных приложений на C++. СПб.: БХВ-Петебург; 2022. 512 с.
15. GNU Project. GNU Scientific Library. Режим доступа: https://www.gnu.org/software/gsl/.
16. Brent R. P. Algorithms for Minimization without Derivatives. Prentice-Hall. Englewood Cliffs, New Jersey; 1973. 195 p. ISBN 0-13-022335-2.

ЭФФЕКТ УСИЛЕНИЯ НАЧАЛЬНОГО МАГНИТНОГО ПОЛЯ В МОДЕЛИ ГЕОДИНАМО

И. В. Бычин[1,2,а], **А. В. Гореликов**[1,2,б]

[1] *Сургутский филиал федерального государственного автономного учреждения «Федеральный научный центр Научно-исследовательский институт системных исследований Национального исследовательского центра «Курчатовский институт», г. Сургут, Российская Федерация*

[2] *Сургутский государственный университет, г. Сургут, Российская Федерация*

[а] ✍ *bychin_iv@surgu.ru*, [б]*gorelikov_av@surgu.ru*

Аннотация: в гидромагнитном динамо магнитное поле генерируется механизмом самовозбуждения, который заключается в том, что определенные конфигурации течений проводящей жидкости могут усиливать слабое начальное магнитное поле, а затем поддерживать его в стационарном или квазистационарном состоянии, препятствуя затуханию. Представляется важным продемострировать эффект усиления магнитного поля методами вычислительной магнитной гидродинамики. В данной работе рассматривается модель геодинамо с вакуумными граничными условиями для магнитного поля на внешней границе сферического слоя. Определяющими задачу безразмерными параметрами являются: Pr — число Прандтля, Pm — магнитное число Прандтля, E — число Экмана и Ra^* — модифицированное число Релея. Начальный уровень магнитного поля определяется параметром Λ, числом Эльзассера. Вычислительные эксперименты проводились с помощью разработанного авторами МГД-кода, адаптированного для работы на гибридных вычислительных системах с графическими процессорами. В коде реализован конечно-объемный метод решения задач резистивной магнитной гидродинамики вязкой несжимаемой жидкости. В качестве начального магнитного поля рассматривалось слабое ($\Lambda = 10^{-2}$) однородное магнитное поле, направленное вдоль оси вращения сферического слоя. При проведении вычислительных экспериментов для значений, определяющих задачу параметров $Pr = 1$, $Pm = 5$, $E = 5 \cdot 10^{-4}$, $Ra^* = 200$, получено квазистационарное решение, в котором начальный уровень энергии магнитного поля в процессе генерации возрастает на три порядка. В полученном решении поля температуры и скорости симметричны относительно экваториальной плоскости, а во внешнем магнитном поле преобладает дипольная составляющая.

Ключевые слова: вычислительная магнитная гидродинамика, геодинамо, слабое однородное начальное магнитное поле.

Благодарности: работа выполнена в рамках государственного задания НИЦ «Курчатовский институт» — НИИСИ по теме № FNEF-2024-0001 «Создание и реализация доверенных систем искусственного интеллекта, основанных на новых математических и алгоритмических методах, моделях быстрых вычислений, реализуемых на отечественных вычислительных системах» (1023032100070-3-1.2.1).

Для цитирования: Бычин И. В., Гореликов А. В. Эффект усиления начального магнитного поля в модели геодинамо. *Успехи кибернетики*. 2025;6(1):76–83.

Поступила в редакцию: 01.12.2024. *В окончательном варианте: 16.12.2024.*

INITIAL MAGNETIC FIELD AMPLIFICATION IN THE GEODYNAMO MODEL

I. V. Bychin[1,2,a], **A. V. Gorelikov**[1,2,b]

[1] *Surgut Branch of Scientific Research Institute for System Analysis of the National Research Centre "Kurchatov Institute", Surgut, Russian Federation*

[2] *Surgut State University, Surgut, Russian Federation*

[a] ✍ *bychin_iv@surgu.ru*, [b]*gorelikov_av@surgu.ru*

Abstract: in a hydromagnetic dynamo, the self-excitation mechanism generates the magnetic field by amplifying a weak initial field through specific configurations of conductive fluid flows. These flows sustain the field in a stationary or quasi-stationary state, preventing its attenuation. Demonstrating this amplification effect requires computational magnetohydrodynamics methods. In this paper, we analyze a geodynamo model with vacuum boundary conditions for the magnetic field at the outer boundary of a spherical layer. The dimensionless parameters that determine the problem are: Pr is the Prandtl number, Pm is the magnetic Prandtl number, E is the Ekman number, and Ra^* is the modified Rayleigh number. The initial level of the magnetic field depends on the parameter Λ, the Elsasser number. We conducted computational experiments using our MHD code, which we adapted for hybrid computing systems with graphics processors. Our

code employs a finite-volume method to solve resistive magnetohydrodynamics problems for a viscous incompressible fluid. A weak ($\Lambda = 10^{-2}$) homogeneous magnetic field directed along the axis of rotation of the spherical layer was considered as the initial magnetic field. When conducting computational experiments, for the values of the parameters defining the problem $Pr = 1$, $Pm = 5$, $E = 5 \cdot 10^{-4}$, $Ra^* = 200$, we obtained a quasi-stationary solution in which the initial level of the magnetic field energy increases by three orders of magnitude during the generation process. In the obtained solution, the temperature and velocity fields are symmetrical relative to the equatorial plane, and the dipole component predominates in the external magnetic field.

Keywords: computational magnetohydrodynamics, geodynamo, weak uniform initial magnetic field.

Acknowledgements: this study is a part of the government contract No. FNEF-2024-0001 Development and Implementation of Trusted AI Systems Using New Mathematical and Algorithmic Methods; Fast Computing Models on Domestic Hardware with the Kurchatov Institute (1023032100070-3-1.2.1).

Cite this article: Bychin I. V., Gorelikov A. V. Initial Magnetic Field Amplification in the Geodynamo Model. *Russian Journal of Cybernetics*. 2025;6(1):76–83.

Original article submitted: 01.12.2024. *Revision submitted: 16.12.2024.*

Введение

В теории геодинамо предполагается, что при достаточно развитой конвекции начальное (затравочное) магнитное поле может быть произвольным и сколь угодно слабым, и оно должно существенно усиливаться в процессе генерации механизмом магнитного гидродинамо. Однако, стоит заметить, что во многих работах по численному исследованию геодинамо начальное магнитное поле имеет специальный вид, который предполагает наличие тороидальной составляющей [1–4], или вопрос о начальном магнитном поле вообще не рассматривается. Вопросы влияния фонового магнитного поля на движение проводящей жидкости в задаче геодинамо исследовались [5, 6], но количество таких работ невелико.

В связи с этим, представляется важным продемонстрировать, что эффект усиления слабого начального магнитного поля воспроизводится методами вычислительной магнитной гидродинамики с использованием разработанных программных комплексов для моделирования геодинамо.

Постановка задачи и математическая модель

Рассматривается естественная конвекция в центральном поле тяжести электропроводной вязкой несжимаемой жидкости, заполняющей сферический слой. Величина силы тяжести линейно зависит от радиуса. Полагается, что твердые границы сферического слоя равномерно вращаются вокруг оси O_z и являются изотермическими. Естественная конвекция возникает за счет положительной разности температур на внутренней и внешней границах слоя. Предполагается, что внутри жидкого сферического слоя находится твердое изотермическое электропроводящее шаровое ядро, скорость вращения которого совпадает со скоростью границ слоя. Электрические проводимости жидкости, заполняющей сферический слой и внутреннего твердого ядра считаются одинаковыми.

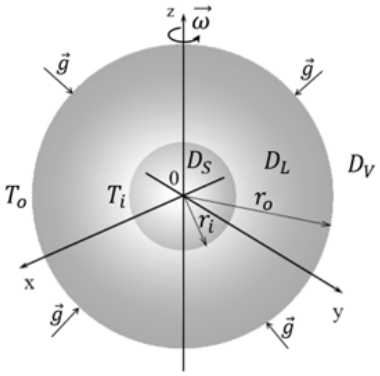


Рис. 1. *Иллюстрация к постановке задачи*

В безразмерной форме система уравнений, описывающая термогравитационную магнитогидродинамическую конвекцию в равномерно вращающейся системе отсчета, имеет вид [1]:

$$r_i < r < r_o :$$

$$E\left(\frac{\partial \boldsymbol{u}}{\partial \tau}+(\boldsymbol{u}\nabla)\,\boldsymbol{u}-\nabla^2\boldsymbol{u}\right)+2\boldsymbol{e}_z\times\boldsymbol{u}+\nabla P=Ra^*\frac{\boldsymbol{r}}{r_o}\Theta+\frac{1}{Pm}\mathrm{rot}\boldsymbol{B}\times\boldsymbol{B}, \tag{1}$$

$$\mathrm{div}\boldsymbol{u}=0, \tag{2}$$

$$\frac{\partial\Theta}{\partial\tau}+(\mathrm{u}\nabla\Theta)=\frac{1}{Pr}\nabla^2\Theta, \tag{3}$$

$$r<r_o:$$

$$\frac{\partial\boldsymbol{B}}{\partial\tau}-\mathrm{rot}\,(\boldsymbol{u}\times\boldsymbol{B})-\frac{1}{Pm}\nabla^2\boldsymbol{B}=0, \tag{4}$$

$$\mathrm{div}\boldsymbol{B}=0. \tag{5}$$

Безразмерные переменные в системе (1)–(5): τ — время, r — расстояние от центра сферического слоя, $\boldsymbol{u}$ — скорость жидкости (в уравнении индукции (4) для $r<r_i$, т.е. во внутреннем твердом ядре $\boldsymbol{u}\equiv 0$), P — отклонение от гидростатического давления вызванное конвекцией, Θ — температура, $\boldsymbol{B}$ — вектор индукции магнитного поля соответственно. Значения безразмерных внешнего r_o и внутреннего r_i радиусов принимались равными: $r_o=20/13$ и $r_i=7/13$, что примерно соответствует соотношению между радиусами жидкого и твердого ядра Земли. Безразмерные параметры задачи: $E=\frac{\nu}{\Omega H^2}$ — число Экмана, $Pr=\frac{\nu}{k}$ — число Прандтля, $Pm=\frac{\nu}{\eta}$ — магнитное число Прандтля, $Ra^*=\frac{\alpha g_0\Delta TH}{\nu\Omega}$ — модифицированное число Рэлея, где ν — кинематическая вязкость, η — коэффициент магнитной диффузии, H – толщина сферического слоя, k — коэффициент температуропроводности, g_0 — ускорение свободного падения на внешнем радиусе, α — коэффициент теплового расширения, Ω — угловая скорость вращения границ сферического слоя, ΔT — разность температур на внутренней и внешней границах слоя. Уравнение движения (1) записано в приближении Буссинеска [7] в равномерно вращающейся с угловой скоростью $\boldsymbol{\omega}=\Omega\boldsymbol{e}_z$ системе отсчета, в нем учитываются: сила Кориолиса ($-2\boldsymbol{e}_z\times\boldsymbol{u}$), сила Лоренца ($\frac{1}{Pm}\mathrm{rot}\boldsymbol{B}\times\boldsymbol{B}$) и сила Архимеда ($Ra^*\frac{\boldsymbol{r}}{r_o}\Theta$), влияние центробежной силы считается пренебрежимо малым.

Граничные (6), (7) и начальные (8) условия для температуры и скорости:

$$r=r_i:\Theta=1,\,\boldsymbol{u}=0, \tag{6}$$

$$r=r_o:\Theta=0,\,\boldsymbol{u}=0, \tag{7}$$

$$r_i<r<r_o,\,\tau=0:$$

$$\boldsymbol{u}=0,\Theta=\frac{r_or_i}{r}-r_i+\frac{21}{\sqrt{17920\pi}}\left(1-3x^2+3x^4-x^6\right)\sin^4\theta\cos(m\varphi)\ , \tag{8}$$

где $x=2r-r_i-r_o$, m — волновое число, параметр, определяющий азимутальную симметрию начального поля температуры [1].

При постановке задачи предполагается, что за внешней границей сферического слоя ($r>r_o$) находится неограниченная диэлектрическая среда, в которой магнитное поле потенциально $\boldsymbol{B}=\nabla\psi$. На внешней границе сферического слоя магнитное поле полагается непрерывным и его тангенциальные составляющие определяются через соответствующие компоненты $\nabla\psi$ в сферических координатах:

$$r=r_o:B_\theta=\frac{1}{r_o}\frac{\partial\psi}{\partial\theta},\qquad B_\varphi=\frac{1}{r_o\sin\theta}\frac{\partial\psi}{\partial\varphi}. \tag{9}$$

Скалярный потенциал ψ магнитного поля является регулярным на бесконечности решением внешней задачи Неймана для уравнения Лапласа:

$$r>r_o:\Delta\psi=0, \tag{10}$$

$$r=r_o:\frac{\partial\psi}{\partial r}=B_r. \tag{11}$$

Радиальная компонента магнитного поля B_r на внешней границе сферического слоя в (11) находится из решения уравнения индукции (4). Граничные условия (9), в которых потенциал находится, как решение задачи (10)-(11), называются вакуумными граничными условиями.

В данной работе предполагалось, что до определенного момента времени магнитное поле отсутствует, а затем уже после того, как в сферическом слое устанавливается режим развитой конвекции, в проводящей среде в момент времени $\tau^* > 0$ возникает флуктуация магнитного поля, которая при определенных условиях может быть усилена механизмом магнитного гидродинамо. В качестве начального возмущения магнитного поля рассматривалось однородное поле, направленное вдоль оси вращения слоя:

$$r \leq r_o, \tau = \tau^* : \boldsymbol{B} = -\sqrt{\Lambda}\boldsymbol{e_z}, \tag{12}$$

где $\Lambda = \sigma\mathcal{B}^2/\rho\Omega$ — число Эльзассера, которое представляет отношение силы Лоренца к силе Кориолиса, действующим на поток проводящей жидкости (σ — проводимость, $\mathcal{B}$ — характерное значение индукции магнитного поля, ρ — плотность жидкости).

В ходе численного решения вычислялись средние плотности кинетической и магнитной энергии:

$$E_{kin} = \frac{1}{2V_L}\int_{D_L} \boldsymbol{u}^2 dV, \qquad E_{mag} = \frac{1}{2V_L EPm}\int_{D_L} \boldsymbol{B}^2 dV, \tag{13}$$

где V_L — объем жидкого сферического слоя.

Численное решение

Система уравнений МГД-конвекции (1)–(5) с условиями (6)–(12) решалась численно в сферических координатах на смещенных структурированных сетках конечно-объемным методом решения задач резистивной магнитной гидродинамики [8], в котором для получения соленоидального магнитного поля используется алгоритм переноса ограничения (CTA — constrained transport algorithm) [9]. Использованный метод позволяет решать начально-краевые задачи для уравнения индукции магнитного поля (4) с вакуумными граничными условиями на сферических поверхностях (9)–(11) [10]. Вычислительные эксперименты проводились с помощью разработанного авторами МГД-кода CVMHD (Control Volume Magneto-HydroDynamics), адаптированного для работы на гибридных вычислительных системах с графическими процессорами [8, 10, 19–21]. Программный комплекс CVMHD предназначен для математического моделирования МГД-течений и гидромагнитного динамо. На рис. 2 продемонстрирована модульная структура комплекса. Комплекс программ разработан с использованием технологий параллельного программирования CUDA и OpenMP [11, 12].

Рис. 2. *Структура комплекса CVMHD*

В комплексе CVMHD реализованы следующие основные численные методы и алгоритмы: метод контрольного объема [13, 14] для сферических координат на структурированных разнесенных сетках; схема дискретизации уравнения индукции на смещенных сетках в сферических координатах [8], схема QUICK [15] для аппроксимации индуктивной части электрического поля уравнения индукции совместно с методом отложенной коррекции [16]; алгоритм переноса ограничения [9]; алгоритмы SIMPLER [13, 14] и PISO [17, 18] для расчета поля течения жидкости, а также реализованы вакуумные [10] и псевдовакуумные [19] граничные условия для компонентов магнитного поля. На рис. 3

представлена блок-схема, которая концептуально описывает вычислительный процесс на временном слое при решении типовой задачи МГД-конвекции с заданными вакуумными граничными условиями. Комплекс протестирован на задачах с точным аналитическим решением и эталонных задачах конвекции и геодинамо [10, 19–21].

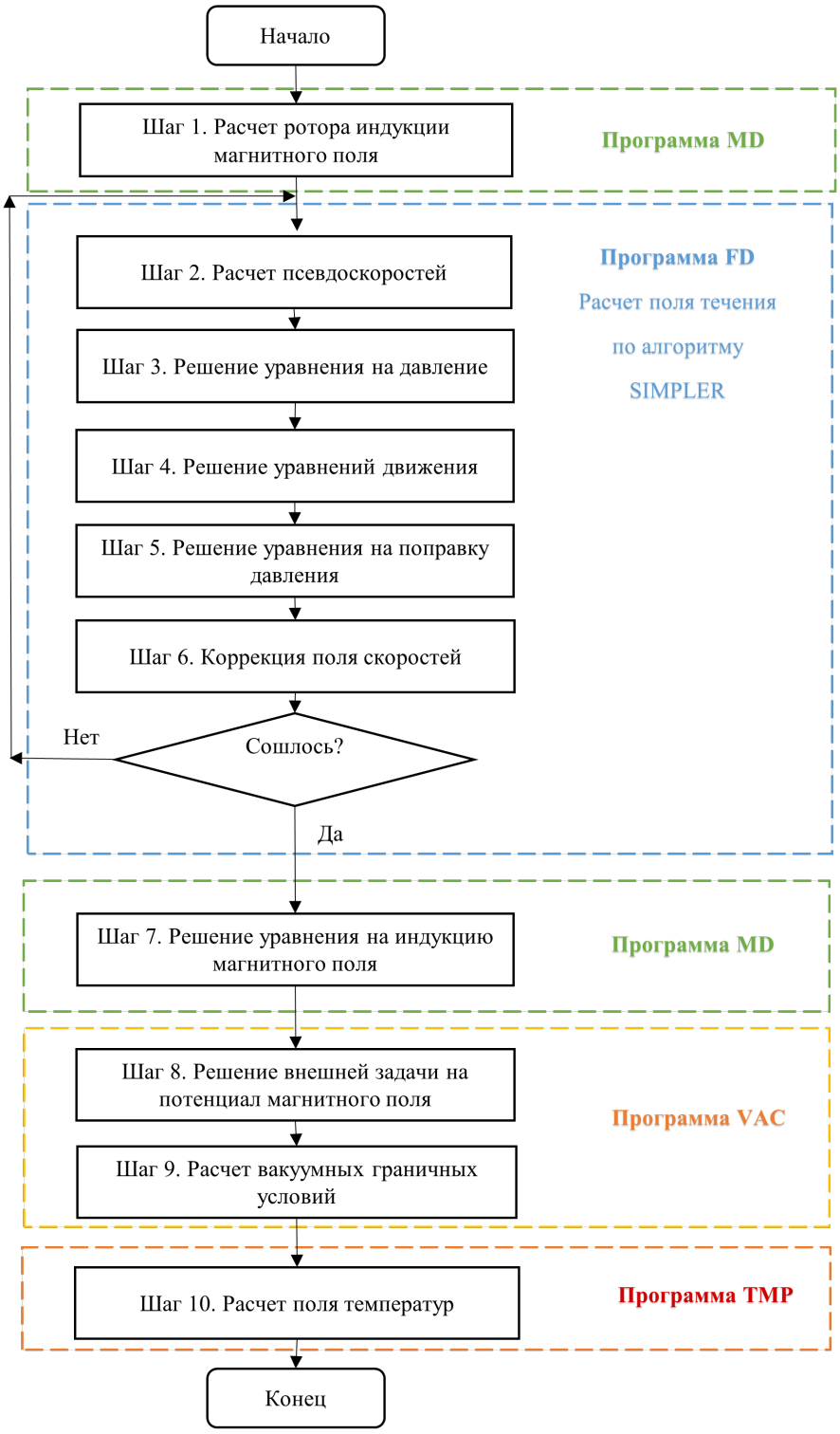
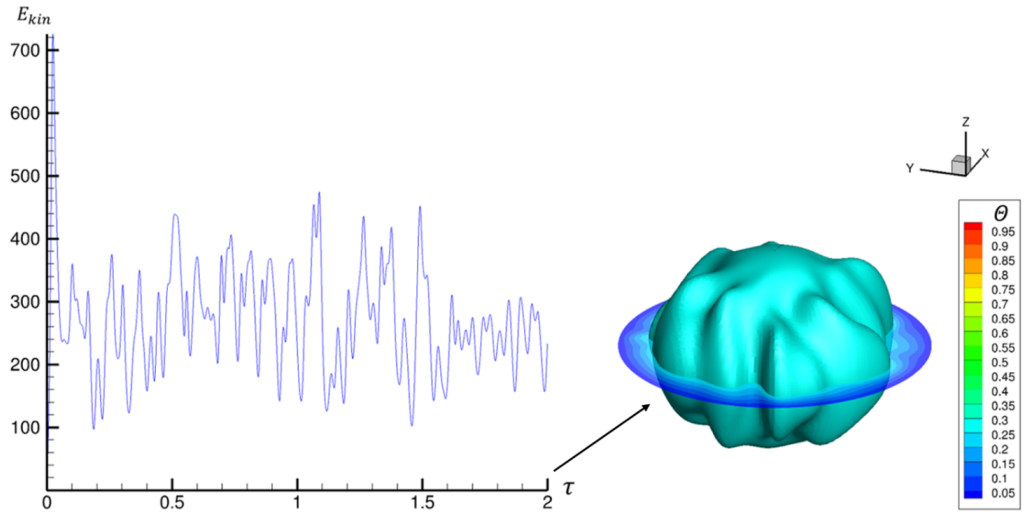


Рис. 3. *Блок-схема вычислительного процесса в рамках 1-го временного слоя*

Результаты расчетов

Расчет проводился на сетке $n_r = 70$, $n_\theta = 140$, $n_\varphi = 160$, в сферических координатах ($r \in [0,r_o]$, $\theta \in [0,\pi]$, $\varphi \in [0,2\pi]$) с шагом по времени $\delta\tau = 10^{-4}$. Сначала на момент времени $\tau^* = 2$ было по-



а) б)

Рис. 4. *а) График средней плотности кинетической энергии в режиме естественной конвекции ($\boldsymbol{B} \equiv 0$); б) Изоповерхность температуры ($\Theta = 0.5$) на момент времени $\tau^* = 2$*

лучено квазистационарное решение задачи о естественной конвекции во вращающемся сферическом слое (1)–(3) с граничными (6)-(7) и начальными (8) условиями при $\boldsymbol{B} \equiv 0$, и значениях управляющих параметров $m = 5$, $E = 5 \cdot 10^{-4}$, $Pr = 1$, $Ra^* = 200$. Для данного решения на рис. 4 представлены график средней плотности кинетической энергии (рис. 4а) и типичная изоповерхность температуры (рис. 4б), которая демонстрирует структуру течения, состоящую из десяти конвективных колонн, вытянутых вдоль оси вращения сферического слоя. Затем полученные поля температуры, давления, скорости рассматривались, как начальные условия (на момент времени $\tau^* = 2$) для задачи о генерации магнитного поля (1)–(5) с граничными условиями (6), вакуумными граничными условиями (9)–(11) и начальными условиями (12), магнитное число Прандтля принималось равным 5 ($Pm = 5$). В качестве начального магнитного поля рассматривалось слабое однородное поле, направленное вдоль оси вращения слоя (12): $\Lambda = 0{,}01$. Соответствующее начальное значение средней плотности магнитной энергии $E_{mag} \approx 2$.

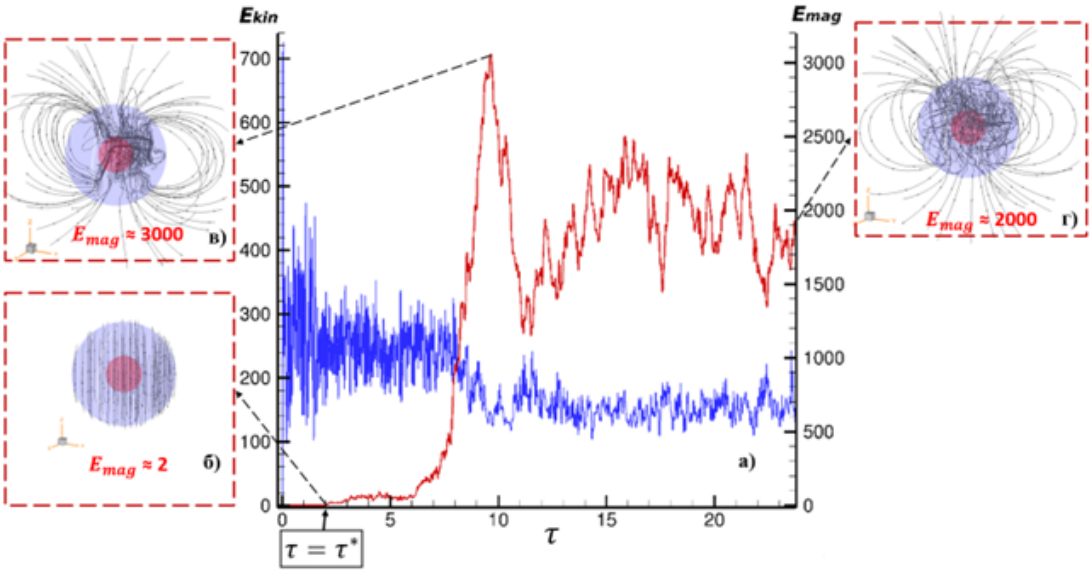


Рис. 5. *(а) Графики средних плотностей кинетической E_{kin} (синяя линия) и магнитной E_{mag} (красная линия) энергий; б), в), г) Силовые линии магнитного поля в различные моменты времени*

Результаты расчета, представленные на рис. 5а, демонстрируют эволюцию средних плотностей кинетической и магнитной энергий в процессе генерации и явный (на три порядка) эффект усиления энергии начального магнитного поля. Необходимо отметить, что при этом уровень кинетической энер-

гии падает приблизительно в полтора раза. На рис. 5б–г показана эволюция магнитного поля, которая демонстрирует, что преобладание дипольной составляющей сохраняется, но при этом наблюдаются изменения угла наклона оси симметрии магнитного поля относительно оси вращения слоя, проходящей через географический полюс.

Заключение

Получено решение задачи геодинамо, в котором в процессе генерации магнитного поля значение начальной магнитной энергии возрастает на три порядка, а во внешнем магнитном поле преобладает дипольная составляющая. В полученном решении поля температуры и скорости симметричны относительно экваториальной плоскости, а во внешнем магнитном поле преобладает дипольная составляющая.

ЛИТЕРАТУРА

1. Christensen U. R. et al. A Numerical Dynamo Benchmark. *Physics of the Earth and Planetary Interiors*. 2001;128(1–4):25–34. DOI: 10.1016/S0031-920100275-8.
2. Matsui H. et al. Performance Benchmarks for a Next Generation Numerical Dynamo Model. *Geochemistry, Geophysics, Geosystems*. 2016;17:1586–1607. DOI: 10.1002/2015GC006159.
3. Marti P. et al. Full Sphere Hydrodynamic and Dynamo Benchmarks. *Geophysical Journal International*. 2014;197:119–134. DOI: 10.1093/gji/ggt518.
4. Jackson A. et al. A Spherical Shell Numerical Dynamo Benchmark with Pseudo-Vacuum Magnetic Boundary Conditions. *Geophysical Journal International*. 2014;196:712–723. DOI: 10.1093/gji/ggt425.
5. Sarson G. R., Jones C. A., Zhang K. Dynamo Action in a Uniform Ambient Field. *Physics of the Earth and Planetary Interiors*. 1999;111(1–2):47–68. DOI: 10.1016/S0031-920100145-9.
6. Sakuraba A., Kono M. Effect of a Uniform Magnetic Field on Nonlinear Magnetocenvection in a Rotating Fluid Spherical Shell. *Geophysical and Astrophysical Fluid Dynamics*. 2000;92(3–4):255–287. DOI: 10.1080/03091920008203718.
7. Ландау Л. Д., Лифшиц Е. М. *Гидродинамика*. М.: Физматлит; 2001. 731 с.
8. Бычин И. В., Гореликов А. В., Ряховский А. В. Схема дискретизации уравнения индукции на смещенных сетках в ортогональных криволинейных координатах. *Успехи кибернетики*. 2022;3:60–73. DOI: 10.51790/2712-9942-2022-3-2-8.
9. Iskakov A. B., Descombes S., Dormy E. An Integro-Differential Formulation for Magnetic Induction in Bounded Domains: Boundary Element–Finite Volume Method. *Journal of Computational Physics*. 2004;197:540–554. DOI: 10.1016/j.jcp.2003.12.008.
10. Бычин И. В., Гореликов А. В., Ряховский А. В. Численное решение начально-краевой задачи с вакуумными граничными условиями для уравнения индукции магнитного поля в шаре. *Вестник Томского государственного университета. Математика и механика*. 2020;64:15–30. DOI: 10.17223/19988621/64/2.
11. Ansorge R. *Programming in Parallel with CUDA: A Practical Guide*. Cambridge: Cambridge University Press; 2022. 395 p.
12. Mattson T. G., He Y., Koniges A. E. *The OpenMP Common Core: Making OpenMP Simple Again*. Cambridge: The MIT Press; 2019. 295 p.
13. Patankar S. V. *Numerical Heat Transfer and Fluid Flow*. New York: Hemisphere Publishing Corporation; 1980. 214 p.
14. Versteeg H. K., Malalasekera W. *An Introduction to Computational Fluid Dynamics.* Harlow: Pearson Education Limited; 2007. 503 p.
15. Leonard B. P. A Stable and Accurate Convective Modelling Procedure Based on Quadratic Upstream Interpolation. *Computer Methods in Applied Mechanics and Engineering*. 1979;19:59–98. DOI: 10.1016/0045-782590034-3.
16. Hayase T., Humphrey J. A. C., Greif R. A Consistently Formulated QUICK Scheme for Fast and Stable Convergence Using Finite-Volume Iterative Calculation Procedures. *Journal of Computational Physics*. 1992;98:108–118. DOI: 10.1016/0021-999190177-Z.
17. Issa R. I. Solution of the Implicitly Discretised Fluid Flow Equations by Operator-Splitting. *Journal of Computational Physics*. 1986;62:40–65. DOI: 10.1016/0021-999190099-9.

18. Issa R. I., Gosman A. D., Watkins A. P. The Computation of Compressible and Incompressible Recirculating Flows by a Non-Iterative Implicit Scheme. *Journal of Computational Physics*. 1986;62:66–82. DOI: 10.1016/0021-999190100-2.
19. Бычин И. В. Тестирование магнитогидродинамического кода на задачах естественной конвекции и геодинамо. *Успехи кибернетики*. 2021;2:6–13. DOI: 10.51790/2712-9942-2021-2-1-1.
20. Бычин И. В., Гореликов А. В., Ряховский А. В. Тестирование программного комплекса для численного моделирования теплообмена и течения жидкости в сферических слоях. *Вестник кибернетики*. 2013;12:81–88.
21. Галкин В. А., Гореликов А. В., Бычин И. В., Дубовик А. О., Ряховский А. В. Тестирование алгоритмов вычислительной магнитной гидродинамики на задаче с точным решением. *Успехи кибернетики*. 2020;1:29–37. DOI: 10.51790/2712-9942-2020-1-4-4.

КВАНТОВОЕ СИНГУЛЯРНОЕ РАЗЛОЖЕНИЕ

И. А. Суров
Университет ИТМО, г. Санкт-Петербург, Российская Федерация
ORCID: https://orcid.org/0000-0001-5690-7507, ✍ ilya.a.surov@itmo.ru

Аннотация: используемое в искусственных нейронных сетях вещественное исчисление описывает амплитуды нервных сигналов, игнорируя их фазы — важнейшую группу параметров, управляющую композицией когнитивных волн биологического мозга. Для снятия связанных с этим ограничений в статье представлена комплекснозначная модификация сингулярного разложения, являющегося прообразом тензорной алгебры современных нейросетей. При этом диагональная матрица сингулярных значений обобщается до самосопряженной матрицы, аналогичной матрице плотности квантовой теории. Добавляемые таким образом недиагональные элементы учитывают нестационарную алгоритмику породившей моделируемые данные поведенческой системы в малоразмерном «семантическом» пространстве. Как и в обычном сингулярном разложении, эта «скрытая» алгоритмика разворачивается в пространство наблюдаемых событий домножением матрицы плотности на пару ортогональных действительных матриц. Полученная матрица комплекснозначных амплитуд порождает наблюдаемые действительные данные посредством квадратного модуля аналогично квантовому правилу Борна. По сравнению с обычным сингулярным разложением, на небольших случайно сгенерированных матрицах разработанный метод существенно повышает точность аппроксимации ценой небольшого увеличения числа параметров. Это преимущество обусловлено эффективностью квантово-волновых принципов обработки информации в естественном мышлении, выраженных в представленной алгебре. Разработанный метод открывает новые возможности для семантического анализа данных, а также для совершенствования современных нейросетевых архитектур.

Ключевые слова: матричное разложение, матрица плотности, амплитуда, интерференция, фаза, комплексные нейронные сети.

Благодарности: работа выполнена за счет гранта РНФ № 23-71-01046.

Для цитирования: Суров И. А. Квантовое сингулярное разложение. *Успехи кибернетики*. 2025;6(1):84–93.

Поступила в редакцию: 18.12.2024. *В окончательном варианте: 22.01.2025.*

QUANTUM SINGULAR-VALUE DECOMPOSITION

I. A. Surov
ITMO University, Saint-Petersburg, Russian Federation
ORCID: https://orcid.org/0000-0001-5690-7507, ✍ ilya.a.surov@itmo.ru

Abstract: real-valued calculations in artificial neural networks captures the amplitudes of neural signals but neglects their phases, which are critical parameters controlling the composition of cognitive waves in natural brains. To address this limitation, we present a complex-valued modification of singular matrix decomposition, a conceptual precursor to the tensor algebra underlying modern neural networks. In this approach, we generalize the diagonal matrix of singular values to a self-adjoint complex matrix, analogous to the density matrix in quantum theory. Within low-dimensional “semantic” spaces, the additional non-diagonal elements of the complex matrix account for the non-stationary logic of the cognitive systems that generate the data. As in the standard formulation, the density matrix is sandwiched between two orthogonal real-valued matrices, unfolding semantic regularities into the space of observable events. The squared modulus of the resulting set of complex-valued amplitudes then produces observable real-valued data, following the quantum-mechanical Born rule. By introducing a minor increase in the number of parameters, our method significantly enhances the precision of classical singular value decomposition. This improvement highlights the efficiency of wave-like and quantum-inspired principles in natural cognition, as expressed in the proposed algebra. The method provides new opportunities for semantic data analysis and offers pathways to advance modern neural network architectures.

Keywords: matrix decomposition, density matrix, amplitude, interference, phase, complex neural networks.

Acknowledgements: the work is supported by an RNF grant, project 23-71-01046.

Cite this article: Surov I. A. Quantum Singular-Value Decomposition. *Russian Journal of Cybernetics*. 2025;6(1):84–93.

Original article submitted: 18.12.2024. *Revision submitted: 22.01.2025.*

Введение

Современные алгоритмы искусственного интеллекта и машинного обучения реализованы посредством тензорной алгебры, в которой матрицы описывают переходы между соседними слоями искусственной нейросети. При этом элементами матриц обычно являются действительные числа, кодирующие «веса связей» между парами вершин. С физической точки зрения, однако, целесообразность этого обычного выбора вызывает вопросы; действительные числа удобны для формализации механических и термодинамических процессов, однако взаимодействие нервных клеток к таковым не относится. Реальные нейронные сигналы передаются электромагнитными волнами, идущими по дендритам и аксонам — фактически, по межклеточным волноводам [1]. Возникающие при этом интерференционные эффекты следуют не логике частиц, а логике «когнитивных волн» [2, 3]; при моделировании этой волновой активности удобным оказывается комплекснозначный формализм волновых и квантовых состояний [4–6].

Комплексные числа отличаются от действительных наличием дополнительной степени свободы, определяющей фазу нейронной волны или колебательного процесса. Эта фаза также кодируется действительным числом, однако не с линейной, а с круговой топологией. Важность этой разницы хорошо известна в физике и ряде задач обработки сигналов [7]; формализации волновой и квантовой физики в рамках действительного исчисления возможны, но более громоздки и сложны. Сходным образом обычные нейронные сети способны имитировать некоторые свойства нервно-волновых процессов [8–10], однако ряд ограничений современных вычислительных систем [11, 12], по видимому, обусловлен их несоответствием истинной физике «нервной деятельности».

В силу накопленного опыта, теоретических и прикладных результатов полный переход искусственных нейросетей на новые логические принципы и математические формализмы вряд ли оправдан. Целесообразно дополнение существующих моделей новыми модулями: отдельными переходами или группами слоев, встраиваемых в существующие архитектуры нейронных сетей и позволяющих снять отмеченные ограничения. В качестве такого модуля может быть использован, в частности, матричный вид квантовой модели контекстно-зависимых решений [13], а также двухматричное разложение на принципах квантовой логики [14]. При этом комплекснозначной становится одна из двух матриц разложения, что позволяет реализовать в этом модуле логику интерференции нервно-волновых амплитуд. Число возникающих при этом фазовых параметров, однако, линейно по размерам исходной матрицы, что затрудняет масштабирование этих подходов.

В данной статье представлено комплекснозначное матричное разложение, в котором число добавочных параметров не связано с размером анализируемой матрицы. Это достигается переходом от двухматричного к трехматричному разложению, прототипом которого является классическое сингулярное разложение.

Классическое сингулярное разложение

Отправной точкой представляемого метода является усеченное сингулярное матричное разложение [15, 16], объектом которого обычно является набор M численных признаков, измеренный в N экспериментах — отсчетах или пробах. Массив этих данных в виде матрицы $\boldsymbol{D}$ размером $M \times N$ аппроксимируется произведением двух ортогональных и одной диагональной матрицы:

$$\boldsymbol{D}_{M\times N} \approx \boldsymbol{U}_{M\times K} * \boldsymbol{\Lambda}_{K\times K} * \boldsymbol{V}_{K\times N}. \tag{1}$$

При этом число строк M и число столбцов N матрицы $\boldsymbol{D}$ совпадают с числами строк и столбцов матриц $\boldsymbol{U}$ и $\boldsymbol{V}$. Вторые размеры этих матриц совпадают с размером K диагональной матрицы $\boldsymbol{\Lambda}$, элементами которой являются сингулярные значения $\lambda_1 \ldots \lambda_K \geq 0$, как показано на рис. 1 для $K = 2$. Этот размер K является свободным параметром разложения и не превышает наименьшего из размеров M и N

$$1 \leq K \leq K_{\max} = \min(M,N),$$

в каком случае равенство в формуле (1) становится точным. Усечение до $K < K_{\max}$ уменьшает соответствующие размеры матриц разложения ценой ошибки, по норме равной сумме отброшенных сингулярных значений.

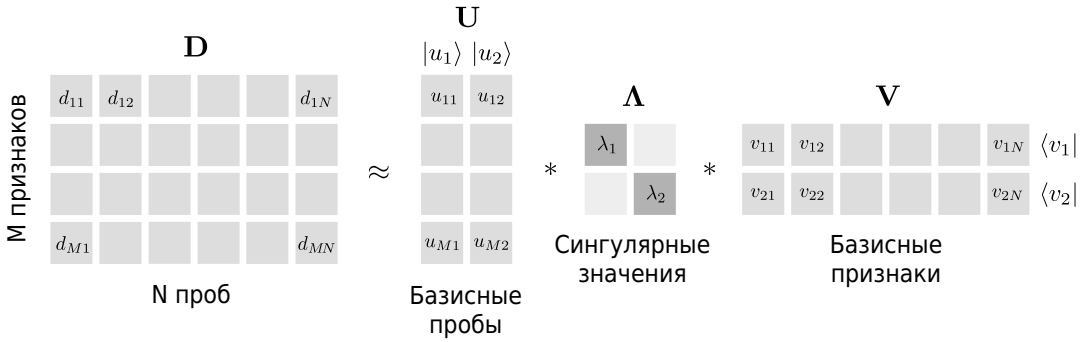


Рис. 1. *Сингулярное матричное разложение* (1)–(2) *матрицы* $\boldsymbol{D}$ *из* M *строк-признаков и* N *столбцов-проб, усеченное до* $K = 2$ *сингулярных значений*

Разложение (1) удобно записать в тензорном виде

$$\boldsymbol{D} \approx \sum_{k=1}^{K} |u_k\rangle \lambda_k \langle v_k| = \sum_{k=1}^{K} \lambda_k |u_k\rangle\langle v_k|, \tag{2}$$

где $|u_k\rangle$ и $\langle v_k|$ есть вектор-столбцы матрицы $\boldsymbol{U}$ и вектор-строки матрицы $\boldsymbol{V}$, соответствующие сингулярным значениям λ_k. Ортогональность матриц $\boldsymbol{U}$ и $\boldsymbol{V}$ означает, что K вектор-столбцов $|u_k\rangle$ длиной M образуют неполный ортонормированный базис в «пространстве проб» размерности M; аналогичным образом вектор-строки $\langle v_k|$ длиной N образуют неполный ортонормированный базис «пространства признаков» размерности N:

$$\langle v_k|v_l\rangle = \langle u_k|u_l\rangle = \delta_{kl}. \tag{3}$$

Из записи (2) видно, что сингулярное разложение представляет исходную матрицу $\boldsymbol{D}$ в виде суммы K «главных компонент» $|u_k\rangle\langle v_k|$, взятых с весами λ_k.

Операторный вид

Матрицу исходных данных $\boldsymbol{D}$ можно рассматривать как оператор, отображающий вектор-признаки длиной N в вектор-пробы длиной M:

$$\boldsymbol{D}|x_N\rangle = |x_M\rangle. \tag{4}$$

Столбцы матрицы $\boldsymbol{V}$ и строки матрицы $\boldsymbol{U}$ тогда представляют собой ортонормированные базисы пространств признаков и проб, в которых оператор (4) принимает диагональный вид $\boldsymbol{\Lambda}$. При этом благодаря ортогональности (3) каждый базисный вектор $|v_k\rangle$ отображается в базисную пробу $|u_k\rangle$, умноженную на соответствующее сингулярное значение (2):

$$\boldsymbol{D}|v_k\rangle = \sum_{j=1}^{K_{\max}} \lambda_j |u_j\rangle\langle v_j|v_k\rangle = \lambda_k |u_k\rangle, \qquad \langle u_k|\boldsymbol{D} = \lambda_k \langle v_k|. \tag{5}$$

Соответственно, сингулярное разложение аппроксимирует полное отображение (4) последовательностью трех этапов, показанной на рис. 2 в сетевой форме:

- Умножение на матрицу $\boldsymbol{V}$ раскладывает произвольный вектор-признак $|x_N\rangle$ по ортонормированному базису $|v_k\rangle$, проектируя его в сжатое пространство размерности K:

$$\boldsymbol{V}|x_N\rangle = \begin{bmatrix} \langle v_1|x_N\rangle \\ \vdots \\ \langle v_K|x_N\rangle \end{bmatrix} = |x_K\rangle.$$

- Умножение полученной проекции $|x_K\rangle$ на диагональную матрицу $\boldsymbol{\Lambda}$ взвешивает ее компоненты с сингулярными значениями λ_k:

$$\boldsymbol{\Lambda}|x_K\rangle = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_K \end{bmatrix} \begin{bmatrix} \langle v_1|x_N\rangle \\ \vdots \\ \langle v_K|x_N\rangle \end{bmatrix} = |x_{K\Lambda}\rangle. \tag{6}$$

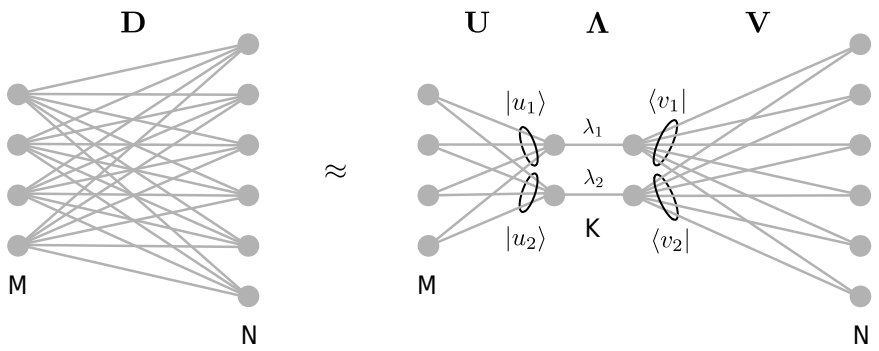


Рис. 2. *Сетевая форма классического сингулярного разложения (рис. 1), в которой полносвязный переход между слоями признаков и проб посредством матрицы* $\boldsymbol{D}$ *аппроксимируется тремя последовательными переходами* $\boldsymbol{U}$, $\boldsymbol{\Lambda}$, $\boldsymbol{V}$ *с двумя внутренними слоями из* K *вершин каждый*

- Умножение на ортогональную матрицу $\boldsymbol{U}$ разворачивает результат в пространство проб размерности M, аппроксимируя полное отображение (4):

$$\boldsymbol{U}|x_{K\Lambda}\rangle \approx |x_M\rangle = \boldsymbol{D}|x_N\rangle.$$

Таким образом информация x переводится из пространства признаков в пространство проб и обратно не напрямую (как при умножении на исходную матрицу $\boldsymbol{D}$), а через промежуточное пространство малой размерности K. Это малоразмерное представление позволяет, в частности, вычислять похожесть векторов из разномерных пространств признаков и проб, которыми могут быть, например, слова и тексты, в которых эти слова встречаются [17]. Кроме того, в ряде задач информативной оказывается сама сжатая форма информации. Представление слов естественного языка 100–1000 мерными векторами [18], например, получено из внутреннего, малоразмерного слоя искусственной нейросети, подобного средним слоям на рис. 2. Алгоритмы такого типа позволяют моделировать неявные закономерности речи [18], в силу чего отображения из пространств исходных данных N,M в пространство (много) меньшей размерности K рассматриваются как переход от синтаксической информации в исходной матрице $\boldsymbol{D}$ к ее неявному смыслу, или «скрытой семантике» [19].

Смысловая (не)стационарность

Согласно диагональному виду матрицы $\boldsymbol{\Lambda}$ каждая малоразмерная «смысловая» компонента $|k\rangle$ размерности K переходит сама в себя (5), то есть

$$\boldsymbol{\Lambda}|k\rangle = \lambda_k|k\rangle, \quad 1 \le k \le K, \tag{7}$$

как показано в середине правой части рис. 2. Соответствующее действие оператора $\boldsymbol{D}$ в сингулярном представлении сводится к преобразованию координат смыслового состояния $|k\rangle$ между пространствами признаков и проб посредством ортогональных матриц $\boldsymbol{U}$ и $\boldsymbol{V}$. Таким образом, классическое сингулярное представление любого (вещественный или комплекснозначный) оператора $\boldsymbol{D}$ является стационарным, когда его «ответ» на любой базисный «запрос» семантически совпадает с этим «запросом» с точностью до положительного множителя.

Реальные поведенческие системы, однако, редко бывают стационарными. На практике смысловые состояния обладают свойством более или менее спонтанного перехода в другие смысловые состояния, что является свойством творческого поведения живых организмов. Большинство поведенческих данных, включая таблицы «признак-проба», порождается именно такими, нестационарными и недетерминированными системами [20].

Оператор плотности

Ключевым элементом нестационарного и недетерминированного поведения является свободный выбор одной из W взаимоисключающих альтернатив. В квантовой когнитивистике эти альтернативы образуют базис $|1\rangle, |2\rangle, \ldots, |W\rangle$ гильбертова пространства, в котором действует моделируемая

поведенческая система [24]. Вероятностные закономерности такого поведения кодируются так называемым оператором плотности — самосопряженной (эрмитовой) матрицей размером $W \times W$, которая в простейшем случае $W = 3$ имеет вид

$$\mathcal{R} = \begin{bmatrix} \rho_{11} & \rho_{21} & \rho_{31} \\ \rho_{12} & \rho_{22} & \rho_{32} \\ \rho_{13} & \rho_{23} & \rho_{33} \end{bmatrix} = \sum_{n,m}^{3} \rho_{nm} |m\rangle\langle n| \tag{8}$$

и работает следующим образом.

Пусть на вход поведенческой системы подан вектор-запрос $|z\rangle = |2\rangle$. Оператор (8) обрабатывает его по правилам матричного умножения и возвращает вектор

$$\mathcal{R}|z\rangle = \begin{bmatrix} \rho_{11} & \rho_{21} & \rho_{31} \\ \rho_{12} & \rho_{22} & \rho_{32} \\ \rho_{13} & \rho_{23} & \rho_{33} \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} \rho_{21} \\ \rho_{22} \\ \rho_{23} \end{bmatrix} = \rho_{21}|1\rangle + \rho_{22}|2\rangle + \rho_{23}|3\rangle. \tag{9}$$

Этот вектор кодирует возможные варианты ответа моделируемой системы на запрос $|2\rangle$, каждый из которых характеризуется комплексным числом ρ_{2k}. В частности, числа ρ_{21} и ρ_{23} соответствуют переходам от вопросного состояния $|2\rangle$ к ответам $|1\rangle$ и $|3\rangle$, тогда как число ρ_{22} соответствует «утвердительному» ответу $|2\rangle$, совпадающему с вопросом. Таким образом, элементы ρ_{nm} матрицы плотности соответствуют переходам поведенческой системы от вопроса n к ответу m, как видно из правой части формулы (8).

Наряду с базисными состояниями $|1\rangle \ldots |W\rangle$, вопросом $|z\rangle$ к поведенческой системе может быть любой нормированный вектор в рассматриваемом гильбертовом пространстве, например $(|1\rangle + e^{i\phi}|3\rangle)/\sqrt{2}$. В этом случае

$$\mathcal{R}|z\rangle = \begin{bmatrix} \rho_{11} & \rho_{21} & \rho_{31} \\ \rho_{12} & \rho_{22} & \rho_{32} \\ \rho_{13} & \rho_{23} & \rho_{33} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ e^{i\phi} \end{bmatrix} / \sqrt{2} = \frac{1}{\sqrt{2}} \begin{bmatrix} \rho_{11} + \rho_{31} e^{i\phi} \\ \rho_{12} + \rho_{32} e^{i\phi} \\ \rho_{13} + \rho_{33} e^{i\phi} \end{bmatrix}, \tag{10}$$

то есть числовая мера каждого из возможных ответов складывается из двух компонент: переход из вопросной компоненты $|1\rangle$ (первые слагаемые) и переход из вопросной компоненты $|3\rangle$ (вторые слагаемые). Результат этого сложения зависит от фазы ϕ третьей компоненты вектор-запроса $|z\rangle$. Благодаря этому модель может учитывать переменные внешние условия и внутренние обстоятельства поведенческой системы — так называемый контекст принятия решений, фиксируемый в теории вероятности Колмогорова [21]. Таким образом, формализм оператора плотности позволяет моделировать контекстно-чувствительные закономерности нестационарного поведения, не находящие отражения в классической теории вероятности [22, 24].

Будучи оператором в гильбертовом пространстве, матрица плотности (8) также является комплекснозначной. Модуль элемента ρ_{nm} есть вероятность P перехода $n \to m$, тогда как его фаза есть физическое действие S на этом переходе в единицах постоянной Планка для рассматриваемой системы [25, 26]:

$$\rho_{nm} = \langle m|\mathcal{R}|n\rangle = P_{nm} * \exp\left(i\frac{S_{nm}}{\hbar}\right). \tag{11}$$

Фазы элементов ρ_{nm} определяют отклик поведенческой системы на небазисные запросы $|z\rangle$ аналогично величине ϕ в формуле (10). Поскольку при обращении перехода $nm \to mn$ действие S меняет знак, а его вероятность P не меняется, каждый недиагональный элемент матрицы (8) является комплексным сопряжением от своей зеркальной пары: $\rho_{mn} = \rho^*_{nm}$.

На переходах из каждого базисного состояния $|1\rangle \ldots |W\rangle$ в само себя действие в формуле (11) $S_{nn} = 0$, в силу чего диагональные элементы матрицы (8) являются вещественными:

$$\rho_{nn} = \langle n|\mathcal{R}|n\rangle = P_n \geq 0, \qquad \sum_n P_n = 1. \tag{12}$$

Как видно из формулы (9), каждый такой элемент представляет собой вероятность утвердительного ответа поведенческой системы на базисный вопрос $|n\rangle$. Благодаря нормировке на единицу (12) эта диагональ соответствует классическому распределению вероятности P_n множества событий $|1\rangle \ldots |W\rangle$ [21].

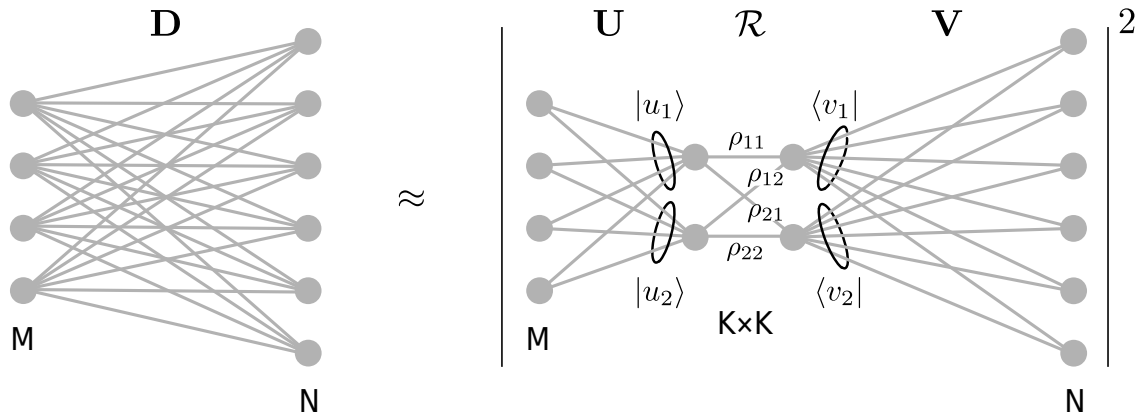


Рис. 3. *Комплекснозначное разложение* (15) *в сетевой форме при* $K = 2$. *Сингулярный переход* $\boldsymbol{\Lambda}$ *на рис. 2 заменен на полносвязную эрмитову матрицу плотности* $\mathcal{R}$ (8)

Комплекснозначное сингулярное разложение

Таким образом, диагональный оператор плотности (8) с элементами (12) описывает логику поведенческих систем, не способных к переходу в состояния, отсутствующие в вектор-запросе. Это и есть свойство (смысловой) стационарности, отмеченное в конце раздела о классическом сингулярном разложении. Снять связанные с этим ограничения можно путем замены диагональной матрицы сингулярных значений $\boldsymbol{\Lambda}$ (6), (7) размером $K \times K$ на матрицу плотности $\mathcal{R}$ (8) такого же размера. При такой замене, однако, произведение

$$\boldsymbol{U}\mathcal{R}\boldsymbol{V} = \boldsymbol{Q} \tag{13}$$

становится комплекснозначным, что не соответствует вещественному типу данных матрицы $\boldsymbol{D}$. Мнимая часть этого произведения, а значит и матрицы плотности $\mathcal{R}$, могла бы лишь увеличить норму Фробениуса $\|\boldsymbol{D} - \boldsymbol{Q}\|$, повышая ошибку аппроксимации; известно, что по такой ошибке сингулярное разложение (1) матрицы $\boldsymbol{D}$ оптимально среди всех ее линейных аппроксимаций фиксированного ранга K [16].

Решение этой проблемы подсказывается квантовоподобными принципами когнитивно-поведенческого моделирования [22, 23], согласно которым наблюдаемые вещественные величины порождаются комплексными амплитудами квантово-волновых состояний через операцию квадратного модуля. Элементы произведения (13) следует считать именно такими амплитудами. Соответственно, аппроксимацией исходных данных должна быть не сама матрица $\boldsymbol{Q}$, а поэлементный квадрат ее модуля с общей мерой ошибки

$$loss = \|\boldsymbol{D} - |\boldsymbol{Q}|^2\|. \tag{14}$$

Внесенная таким образом нелинейность выводит эту аппроксимацию за рамки теоремы об оптимальности классического сингулярного разложения [15], в принципе позволяя превзойти его по точности.

Как и в классическом разложении (1), матрицы $\boldsymbol{U}$ и $\boldsymbol{V}$ в формуле (13) остаются вещественными и ортогональными (3). Следовательно, аппроксимация (14) представима в виде

$$\boldsymbol{D} \approx |\boldsymbol{U}\mathcal{R}\boldsymbol{V}|^2 = \left|\sum_{n,m}^{K} \rho_{nm} |u_m\rangle\langle v_n|\right|^2, \tag{15}$$

отличающемся от классического разложения (5) наличием перекрестных членов, вносимых недиагональными элементами матрицы $\mathcal{R}$ (8), как показано на рис. 3. При этом из-за отсутствия нормировки матрицы $\boldsymbol{D}$ требование единичного следа (12) с диагональных элементов ρ_{nn} целесообразно снять. Элементы ρ_{nm} в (15) образуют эрмитову матрицу $\mathcal{R}$ с неотрицательными собственными значениями, приводимую к обычной матрице плотности делением на след.

Аппроксимация (14), (15) также предполагает, что матрица $\boldsymbol{D}$ ограничивается неотрицательными значениями, интерпретируемыми как интенсивности амплитуд $\boldsymbol{Q}$. Произвольные действительные матрицы приводятся к такому виду посредством экспоненциирования с последующим логарифмированием (преобразование Больцмана, softmax).

Число параметров и оптимизация

В отличие от сингулярной матрицы $\boldsymbol{\Lambda}$ с K действительными значениями (6), матрица $\mathcal{R}$ в разложении (13), (15) имеет K^2 независимых параметров (включая модули и фазы комплексных элементов). Число свободных параметров матриц $\boldsymbol{U}$ и $\boldsymbol{V}$

$$npU = MK - \frac{K(K+1)}{2}, \qquad npV = NK - \frac{K(K+1)}{2},$$

напротив, остается без изменений. Таким образом, переход от классического к комплекснозначному разложению увеличивает полное число независимых параметров на $K^2 - K$ единиц. Для рассмотренного далее случая квадратных матриц $\boldsymbol{D}$ размером полное число параметров обоих разложений для $N = M \leq 11$ показано на рис. 4. В частности, при $N = M = 10$ матрица $\boldsymbol{D}$ имеет $NM = 100$ свободных параметров, тогда как классическое сингулярное разложение с $K = 2{,}3{,}4$ аппроксимирует ее, используя 36, 51 и 64 параметра. Комплекснозначное разложение (15) увеличивает эти значения на 2, 6 и 12 единиц соответственно.

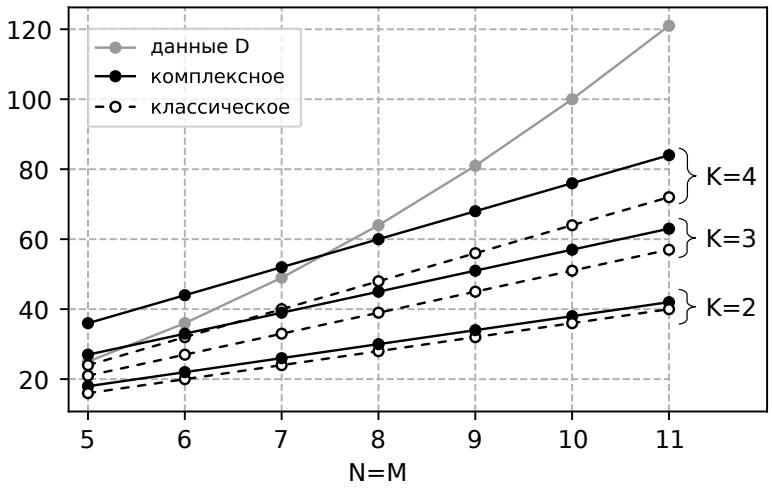


Рис. 4. *Число параметров классического (пунктир, белые точки) и комплекснозначного (сплошная линия, черные точки) разложения квадратной матрицы* $\boldsymbol{D}$ *в зависимости от ее размера* $N = M$ *для различных* K*. Серым показано число параметров* N^2 *исходных данных*

Параметры матриц $\boldsymbol{U}, \mathcal{R}, \boldsymbol{V}$ разложения (15) определялись с помощью алгоритма многомерной оптимизации Бройдена — Флетчера — Гольдфарба — Шенно (BFGS) [27, 28] с функцией ошибки (14). Этот алгоритм обеспечивает сравнительно быструю сходимость к локальному минимуму, выбор которого зависит от начальной точки. В этой связи для каждой исходной матрицы $\boldsymbol{D}$ проводилось от 20 до 100 оптимизаций со случайным набором начальных значений, из которых бралась наилучшая.

Результаты испытаний

Разложение (15) испытано на квадратных матрицах $\boldsymbol{D}$ размером $5 \leq N = M \leq 11$, составленных из N^2 случайных чисел от 0 до 1. Для каждого размера таким образом готовилось от 20 до 100 матриц, каждая из которых аппроксимировалась для нескольких начальных точек, как отмечено выше. Среднее значение ошибки (14) из полученных таким образом наилучших комплекснозначных аппроксимаций нормировалось на среднюю ошибку классического сингулярного разложения того же ранга K для тех же исходных матриц $\boldsymbol{D}$.

Зависимости этой нормированной ошибки от N показаны красным на рис. 5, где графики соответствуют размеру матрицы плотности $K = 2$, 3 и 4. Во всех случаях ошибка комплекснозначного разложения ниже ошибки классического аналога, и эта разность убывает с ростом N. В частности, для $N = 7$ переход от сингулярного к комплекснозначному разложению при $K = 2$ снижает среднюю ошибку аппроксимации на 21% ценой 2 новых параметров, добавляемых к 24 параметрам классического разложения (рис. 4). При $K = 3$ снижение ошибки составляет 61%, тогда как число параметров увеличивается с 33 до 39; при $K = 4$ ошибка снижается в 18 раз при росте числа параметров с 40 до 52.

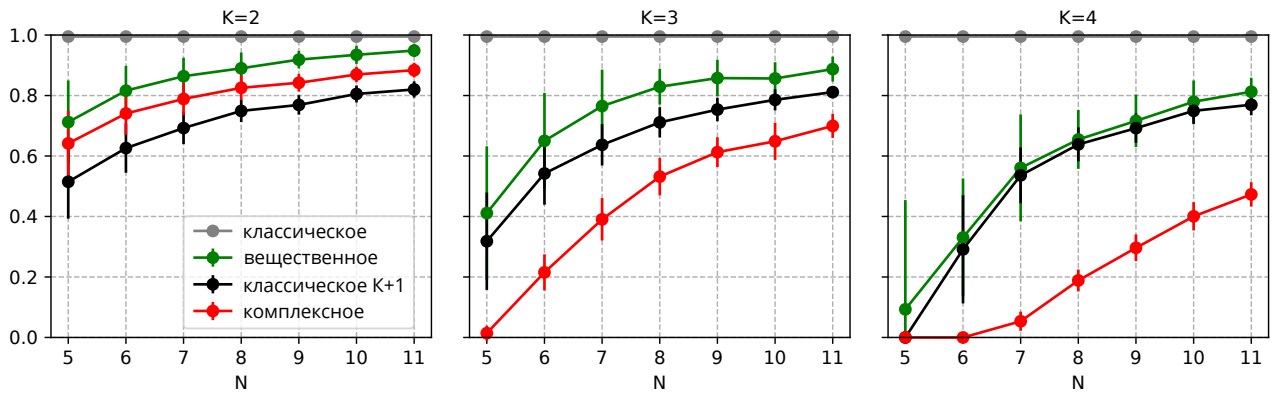


Рис. 5. *Красный: средняя ошибка комплекснозначного разложения* (15) *случайной квадратной матрицы* $\boldsymbol{D}$ *в зависимости от ее размера* $N = M$ *для различных* K*, нормированная на среднюю ошибку обычного сингулярного разложения* (1) *(серый). Черный: обычное сингулярное разложение размерности, повышенной на единицу. Зеленый: вещественная модификация комплексного разложения, при которой фазы недиагональных элементов матрицы* $\mathcal{R}$ (11) *положены равными 0*

В связи с этим целесообразно сравнить полученные ошибки с более низкими ошибками сингулярных разложений, ранг которых на единицу выше при соответственно увеличенном числе параметров (рис. 4). При $K = 3$ эти (также нормированные) ошибки повышенного ранга 4, показанные черным, превосходят ошибки комплекснозначного разложения ранга 3 на рассмотренном диапазоне N. При этом начиная с $N = 7$ параметров у комплекснозначного разложения меньше, чем у сингулярного разложения повышенного ранга: 39 против 40 для $N = 7$, 45 против 48 для $N = 8$, 51 против 56 для $N = 9$, 57 против 64 для $N = 10$ и 63 против 72 для $N = 11$. Аналогичный результат имеет место для $K = 4$, но не для $K = 2$, при котором классическое разложение повышенного ранга 3 выигрывает благодаря существенно большему числу параметров (рис. 4).

Зеленые графики на рис. 5 показывают среднюю ошибку аппроксимации (15), в которой элементы матрицы $\mathcal{R}$ ограничены действительными значениями, что снижает число добавочных параметров вдвое за счет отсутствия фаз. В рассмотренном диапазоне N и K точность такой аппроксимации превосходит классическое разложение того же ранга (серый), но не дотягивает до повышенного ранга (черный). Как видно из графиков, большая часть преимущества комплекснозначного разложения над классическим возникает именно вследствие перехода $\mathcal{R}$ от вещественных к комплексным числам.

Заключение

Существенное повышение точности аппроксимации при незначительном росте числа параметров (рис. 4, 5) означает, что представленный метод повышает удельную информативность параметров модели. Это достигается благодаря переходу от линейной ошибки $\|D - \boldsymbol{U\Lambda V}\|$ классического разложения (1) к квадратичной ошибке (14), в результате чего элементы матриц $\boldsymbol{U}$, $\mathcal{R}$, $\boldsymbol{V}$ начинают работать не как линейные компоненты данных $\boldsymbol{D}$, а как порождающие эти данные амплитуды вероятности рассматриваемых событий. Для поведенческих данных эти амплитуды описывают «когнитивные волны» поведенческой системы [3], возможно, распределенной по множеству удаленных организмов. При этом недиагональные элементы матрицы $\mathcal{R}$ (8) отвечают за интерференцию этих амплитуд — ключевое отличие нелинейной «логики волн» от колмогоровской логики множеств. Как видно из представленных результатов, возможность такой интерференции повышает точность аппроксимации без увеличения числа компонент разложения.

Переход от комплекснозначных амплитуд к наблюдаемым вещественным данным посредством квадратичной нормы (12) ставит задачу нелинейной оптимизации. Представленные в предыдущем разделе результаты получены с помощью общего алгоритма BFGS, ресурсоемкость которого быстро растет при повышении размера раскладываемой матрицы $\boldsymbol{D}$. По этой причине опробовать представленное разложение на больших объемах данных типа [17] пока не удалось. Такая апробация, однако, интересна в связи с возможностью процессно-смысловой интерпретации комплексных элементов матрицы $\mathcal{R}$ [29], что может принципиально расширить задачи семантического анализа текстовых [19] и других поведенческих данных [30–32]. Подобные исследования требуют эффективной минимизации

ошибки (12), в том числе с помощью алгоритмов квантовой и волновой информатики [33–37].

Решение этой оптимизационной задачи также важно для использования представленного метода в искусственных нейросетях. При этом комплекснозначный модуль на рис. 3 мог бы встраиваться в сверточные нейросети, машины Больцмана, автокодировщики, трансформеры и другие классические, квантовые и гибридные архитектуры [38–40], расширяя их возможности благодаря вышеотмеченному переходу от логики множеств к логике волн. Для такого встраивания необходим алгоритм обратного распространения ошибки, проходящий через барьер квадратного модуля — границу между «классической» и «квантовой» частями нейросети.

ЛИТЕРАТУРА

1. Bressloff P. C. *Waves in Neural Media: From Single Neurons to Neural Fields*. Springer; New York; 2014. DOI: 10.1007/978-1-4614-8866-8.
2. Barrett T. W. The Cortex as Interferometer: The Transmission of Amplitude, Frequency and Phase in Cortical Structures. *Neuropsychologia*. 1969;7(2):135–48.
3. Суров И. А. Логика множеств и логика волн в когнитивно-поведенческом моделировании. *Информационные и математические технологии в науке и управлении*. 2023;32(4):51–66. DOI: 10.25729/ESI.2023.32.4.005.
4. Khrennikov A. Yu. Quantum-Like Model of Processing of Information in the Brain Based on Classical Electromagnetic Field. *BioSystems*. 2011;105(3):250–62. DOI: 10.1016/j.biosystems.2011.05.014.
5. Khrennikov A., Basieva I., Pothos E. M., Yamato I. Quantum Probability in Decision Making from Quantum Information Representation of Neuronal States. *Scientific Reports*. 2018;8(1):16225. DOI: 10.1038/s41598-018-34531-3.
6. Surov I. A., Semenenko E., Platonov A. V., Bessmertny I. A., Galofaro F., Toffano Z., Khrennikov A. Yu., Alodjants A. P. Quantum Semantics of Text Perception. *Scientific Reports*. 2021;11(1):4193. DOI: 10.1038/s41598-021-83490-9.
7. Краснов А. Е., Головкин М. Е., Герасимова В. И. Применение причинных преобразований в распознавании сигналов и изображений. *XIV Всероссийское совещание по проблемам управления (ВСПУ-2024)* : сборник научных трудов, 17–20 июня 2024 г., М.: Ин-т проблем упр. им. В.А. Трапезникова Рос. акад. наук; 2024:3186–3190.
8. Busemeyer J. R., Fakhari P., Kvam P. Neural Implementation of Operations Used in Quantum Cognition. *Progress in Biophysics and Molecular Biology*. 2017;130:53–60.
9. Scholes G. D. Quantum-Like States on Complex Synchronized Networks. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*. 2024;480(2295):20240209.
10. Khrennikov A. Yu., Ozawa M., Benninger F., Shor O. Coupling Quantum-Like Cognition with the Neuronal Networks within Generalized Probability Theory. 2024. preprint arXiv:2411.00036v1. DOI: 10.48550/arXiv.2411.00036.
11. Кузнецов О. П. Неклассические парадигмы в искусственном интеллекте. *Теория и системы управления*. 1995;5:3–23.
12. Roli A., Jaeger J., Kauffman S. A. How Organisms Come to Know the World: Fundamental Limits on Artificial General Intelligence. *Frontiers in Ecology and Evolution*. 2022;9:806283. DOI: 10.3389/fevo.2021.806283.
13. Суров И. А. Матрично-кубитный алгоритм семантического анализа вероятностных данных. *Моделирование и анализ информационных систем*. 2024;31(3):280–93. DOI: 10.18255/1818-1015-2024-3-280-293.
14. Kozhisseri S., Surov I. A. Quantum-Probabilistic SVD: Complex-Valued Factorization of Matrix Data. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*. 2022;22(3):567–73. DOI: 10.17586/2226-1494-2022-22-3-567-573.
15. Eckart C., Young G. The Approximation of One Matrix by Another of Lower Rank. *Psychometrica*. 1936;1(3):211–218. DOI: 10.1007/2FBF02288367.
16. Stewart G. W. On the Early History of the Singular Value Decomposition. *SIAM Review*. 1993;35(4):551–566. DOI: 10.1137/1035134.
17. Landauer T. K., Dumais S. T. A Solution to Plato's Problem: The Latent Semantic Analysis Theory of Acquisition, Induction, and Representation of Knowledge. *Psychological Review*. 1997;104(2):211–240.

18. Mikolov T., Sutskever I., Chen K., Corrado G., Dean J. Distributed Representations of Words and Phrases and Their Compositionality. *NIPS'13. Proceedings of the 26th International Conference on Neural Information Processing Systems*. 2013.
19. Landauer T. K., McNamara D. S., Dennis S., Kintsch W., editors. *Handbook of Latent Semantic Analysis*. Routledge. 2011.
20. Газя Г. В., Газя Н. Ф., Еськов В. В., Манина Е. А. Непредсказуемость и неопределенность создают реальную Complexity. *Успехи кибернетики*. 2024;5(2):97–102. DOI: 10.51790/2712-9942-2024-5-2-11.
21. Khrennikov A. Yu. *Interpretations of Probability*. Berlin, New York: De Gruyter; 2009. 219 p. DOI: 10.1515/9783110213195.
22. Khrennikov A. Yu. *Ubiquitous Quantum Structure: From Psychology to Finance*. Berlin, Heidelberg: Springer Berlin Heidelberg; 2010. DOI: 10.1007/978-3-642-05101-2.
23. Haven E., Khrennikov A. *The Palgrave Handbook of Quantum Models in Social Science*. London: Macmillan Publishers Ltd. 2017. 365 p. DOI: 10.1057/978-1-137-49276-0.
24. Суров И. А. Квантовая теория: методология и математика управления. *Труды XIII ВСПУ*. 2019:1589–1593. DOI: 10.25728/vspu.2019.1589.
25. Nottale L. Scale Relativity: A Fractal Matrix for Organization in Nature. Part 2. *Electronic Journal of Theoretical Physics*. 2007;4(16):187–274.
26. Nottale L., Auffray C. Scale Relativity Theory and Integrative Systems Biology: 2. Macroscopic Quantum-Type Mechanics. *Progress in Biophysics and Molecular Biology*. 2008;97(1):115–157.
27. Head J. D., Zerner M. C. A Broyden—Fletcher—Goldfarb—Shanno Optimization Procedure for Molecular Geometries. *Chemical Physics Letters*. 1985;122(3):264–270.
28. Gommers R., et al. *SciPy 1.9.0 Release Notes*. 2022. DOI: 10.5281/zenodo.6940349.
29. Суров И. А. Процессная семантика комплексных чисел. *Математические структуры и моделирование*. 2023;4:71–84. DOI: 10.24147/2222-8772.2023.4.71-84.
30. Кедрин В. С., Сальникова М. К. Прогнозирование нестационарных макроэкономических процессов с помощью методик сингулярного разложения и искусственного интеллекта. *Труды Братского государственного университета. Серия: Естественные и инженерные науки — развитию регионов*. 2007;2:38–41. Режим доступа: https://www.elibrary.ru/item.asp?id=13362489.
31. Ильин П. Л., Самойлова Т. А. Сингулярное разложение пространственных матриц. *Современные информационные технологии и ИТ-образование*. 2022;18(3):578–88. DOI: 10.25559/SITITO.18.202203.578-588.
32. Wang X., Gu L., Lee H., Zhang G. Quantum Context-Aware Recommendation Systems Based on Tensor Singular Value Decomposition. *Quantum Information Processing*. 2021;20(5):190. DOI: 10.1007/s11128-021-03131-y.
33. Lloyd S., Mohseni M., Rebentrost P. Quantum Principal Component Analysis. *Nature Physics*. 2014;10(9):631–633. DOI: 10.1038/NPHYS3029.
34. Rebentrost P. et al. Quantum Singular-Value Decomposition of Nonsparse Low-Rank Matrices. *Phys. Rev. A*. 2018;97(1):012327. DOI: 10.1103/PhysRevA.97.012327.
35. Gilyén A., Su Y., Low G. H., Wiebe N. Quantum Singular Value Transformation and Beyond: Exponential Improvements for Quantum Matrix Arithmetics. *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*. 2019:193–204. DOI: 10.1145/3313276.3316366.
36. Spall J., Guo X., Barrett T. D., Lvovsky A. I. Fully Reconfigurable Coherent Optical Vector-Matrix Multiplication. *Optics Letters*. 2020;45(20):5752. DOI: 10.1364/OL.401675.
37. Wang X., Song Z., Wang Y. Variational Quantum Singular Value Decomposition. *Quantum*. 2021;5:483. DOI: 10.22331/q-2021-06-29-483.
38. Jia Z., Yi B., Zhai R., Wu Y., Guo G., Guo G. Quantum Neural Network States: A Brief Review of Methods and Applications. *Adv Quantum Tech*. 2019;2(7–8):1800077. DOI: 10.1002/qute.201800077.
39. Ваулин Н. В. Регуляризация сверточной нейронной сети сингулярным разложением для обучения на малых выборках. *Интеллектуальные системы: Теория и приложения*. 2022;26(4):20–36.
40. Melnikov A., Kordzanganeh M., Alodjants A. P., Lee R. K. Quantum Machine Learning: from Physics to Software Engineering. *Advances in Physics X*. 2023;8(1). DOI: 10.1080/23746149.2023.2165452.

ОЦЕНКА ПРИМЕНИМОСТИ МЕТОДА РАЗРЫВНЫХ СМЕЩЕНИЙ В ЗАДАЧАХ РОСТА И ВЗАИМОДЕЙСТВИЯ КРИВОЛИНЕЙНЫХ ТРЕЩИН

Д. А. Пестов
Федеральное государственное автономное учреждение «Федеральный научный центр Научно-исследовательский институт системных исследований Национального исследовательского центра «Курчатовский институт», г. Москва, Российская Федерация
ORCID: https://orcid.org/0000-0002-4687-1800, ✍ dmitr-ey94@mail.ru

Аннотация: в работе исследована применимость численных методов на основе метода разрывных смещений к моделированию роста криволинейных трещин в условиях сложного нагружения или взаимодействия трещин. Исследовано влияние разрешения численного метода и точности аппроксимации коэффициентов интенсивности напряжений на точность прогнозирования траектории роста трещины. Рассмотрен метод, имеющий малую вычислительную сложность в предположениях линейной упругости среды, плоского деформированного состояния, а также при условиях квазистатического роста трещин. В качестве численного метода выбран метод разрывных смещений нулевого порядка точности и аппроксимация коэффициентов интенсивности напряжений через значения разрывов смещений в элементах, ближайших к кончику. Показана высокая точность определения коэффициентов интенсивности напряжений для случаев одиночной трещины. Для случая нескольких трещин получены оценки применимости метода: высокая точность аппроксимации коэффициентов интенсивности напряжений достигается при расстоянии между трещинами, превосходящем размер граничного элемента. Показано совпадение численно рассчитанной траектории роста с экспериментальной для образца из оргстекла. Показана слабая зависимость траектории трещины от размера граничного элемента, а также от величины приращения ее длины на каждом шаге при квазистатическом росте. Показана устойчивость траектории трещины к малым отклонениям при наличии перепада напряжений. На основе полученных результатов сделан вывод о границах применимости методов нулевого порядка к задачам о росте криволинейных трещин.

Ключевые слова: рост трещин, метод разрывных смещений, криволинейные трещины.

Благодарности: работа выполнена за счет субсидии, выделенной НИЦ «Курчатовский институт» — НИИСИ на выполнение государственного задания по теме № FNEF-2024-0002 «Математическое моделирование многомасштабных динамических процессов и системы виртуального окружения».

Для цитирования: Пестов Д. А. Оценка применимости метода разрывных смещений в задачах роста и взаимодействия криволинейных трещин. *Успехи кибернетики*. 2025;6(1):94–103.

Поступила в редакцию: 18.02.2025. *В окончательном варианте: 19.03.2025.*

APPLICABILITY OF THE DISPLACEMENT DISCONTINUITY METHOD TO CURVILINEAR CRACK GROWTH AND INTERACTION PROBLEMS

D. A. Pestov
Scientific Research Institute for System Analysis of the National Research Centre "Kurchatov Institute", Moscow, Russian Federation
ORCID: https://orcid.org/0000-0002-4687-1800, ✍ dmitr-ey94@mail.ru

Abstract: we studied the applicability of numerical methods based on the discontinuous displacement method to the simulation of curvilinear crack growth and interaction under complex loads. We examined how the resolution of the numerical method and the accuracy of stress intensity factor approximation affect the accuracy of a predicted crack growth path. We considered a low computational complexity method under the assumptions of linear elasticity, plane-strain conditions, and quasi-static crack growth. We selected the zero-order accuracy discontinuous displacement method and approximated stress intensity factors using displacement discontinuity values in elements closest to the crack tip. We demonstrated high accuracy in determining stress intensity factors for single-crack cases. For multiple cracks, we obtained criteria of the method applicability: high accuracy in approximating stress intensity factors occurs when the distance between cracks exceeds the size of the boundary element. The numerically calculated growth trajectory coincided with the experimental one for a plexiglass sample under shear load. We observed a

weak dependence of the crack trajectory on the boundary element size and the crack length increment at each step under quasi-static growth. We also showed that the crack trajectory remains stable under small deviations in the presence of stress gradients. Based on these results, we concluded the applicability limits of zero-order methods for modeling the growth and interaction of curvilinear cracks.

Keywords: displacement discontinuity method, crack growth, curvilinear cracks.

Acknowledgements: this is a part of the FNEF-2024-0002 Simulation of Multiscale Dynamic Processes and Virtual Environments government contract granted to the Scientific Research Institute for System Analysis of the National Research Centre “Kurchatov Institute”.

Cite this article: Pestov D. A. Applicability of the Displacement Discontinuity Method to Curvilinear Crack Growth and Interaction Problems. *Russian Journal of Cybernetics*. 2025;6(1):94–103.

Original article submitted: 18.02.2025. *Revision submitted: 19.03.2025.*

Введение

Метод разрывных смещений представляет собой граничноэлементный метод для решения задач о телах, ослабленных наличием трещин. Данный метод широко применяется для решения статических задач с трещинами, а также для задач роста прямолинейных трещин, в частности, трещин гидроразрыва [1, 2]. При моделировании криволинейных растущих трещин использование метода разрывных смещений связано с несколькими проблемами, способными существенно влиять на точность решения. К таким проблемам относятся неточное определение коэффициентов интенсивности напряжений в окрестности кончика трещины, неточное определение поля напряжений вблизи кончиков элемента, а также влияние величины приращения длины трещины (совпадающей в большинстве случаев с размером численного элемента) на итоговую траекторию роста. Для решения данных проблем требуется увеличение порядка точности используемых методов [3] и, соответственно, усложнение расчетов. Оценим влияние данных факторов на различные случаи роста и взаимодействия трещин.

Математическая постановка задачи

Рассмотрим задачу о росте трещин в линейно упругой однородной среде в условиях плоской деформации. Будем предполагать, что раскрытие трещин мало по сравнению с их длинами, то есть трещины можно считать отрезками кривых, на которых поле перемещений терпит разрыв. В таком случае поле напряжений и перемещений можно описать с помощью уравнений равновесия и закона Гука, где G — модуль сдвига, а ν — коэффициент Пуассона:

$$\begin{aligned} \frac{\partial \sigma_{xx}}{\partial x} + \frac{\partial \sigma_{xy}}{\partial y} = 0 \\ \frac{\partial \sigma_{xy}}{\partial x} + \frac{\partial \sigma_{yy}}{\partial y} = 0 \end{aligned}, \tag{1}$$

$$\begin{aligned} \frac{\partial u_x}{\partial x} &= \frac{1}{2G}\left[(1-\nu)\sigma_{xx} - \nu\sigma_{yy}\right] \\ \frac{\partial u_y}{\partial y} &= \frac{1}{2G}\left[(1-\nu)\sigma_{yy} - \nu\sigma_{xx}\right]. \\ \frac{1}{2}\left(\frac{\partial u_x}{\partial y} + \frac{\partial u_y}{\partial x}\right) &= \frac{1}{2G}\sigma_{xy} \end{aligned} \tag{2}$$

В общем случае граничные условия ставятся на берегах трещины и на бесконечности. Граничные условия на «бесконечности» или, согласно принципу Сен-Венана, на достаточном удалении от трещины представляют собой внешние напряжения, действующие на трещину. Это могут быть сжимающие напряжения в горной породе, растягивающие напряжения в образце под нагрузкой или иные нагрузки, приложенные на достаточном расстоянии от трещины. Далее предполагаем случай изначально однородного поля напряжений, определяемого следующими условиями, принимая сжимающие напряжения за положительные:

$$\begin{aligned} \sigma_{xx}|_\infty &= \sigma^0_{xx} \\ \sigma_{yy}|_\infty &= \sigma^0_{yy}. \\ \sigma_{xy}|_\infty &= \sigma^0_{xy} \end{aligned} \tag{3}$$

На берегах трещины ставится условие свободной поверхности или заданного давления, в случае если это, например, трещина гидроразрыва. Таким образом, для точек кривой C, представляющей собой трещину, ставятся условия на напряжения, нормальные и касательные к этой кривой в данной точке, где $p(s)$ — заданное распределение внутреннего давления вдоль трещины.

$$\begin{aligned} \sigma_{nn}|_{(x,y)\in C} &= p(s) \\ \sigma_{ns}|_{(x,y)\in C} &= 0 \end{aligned} \tag{4}$$

Ввиду линейности задачи ее решение можно представить в виде суперпозиции решений двух задач: первой, с граничными условиями на бесконечности при условии отсутствия трещины, и второй, где напряжения на бесконечности равны нулю, а условия (3), (4) принимают следующий вид:

$$\begin{aligned} \sigma_{nn}|_{(x,y)\in C} &= p(s) - \sigma^0_{nn}(s) \\ \sigma_{ns}|_{(x,y)\in C} &= 0 - \sigma^0_{ns}(s) \\ \sigma|_{\infty} &= 0 \end{aligned} \quad . \tag{5}$$

Здесь σ^0_{nn} и σ^0_{ns} получаются из $\sigma^0_{xx}, \sigma^0_{xy}, \sigma^0_{yy}$ путем поворота системы координат на угол $\theta(s)$. Такой подход обусловлен тем, что метод разрывных смещений предполагает напряжения на бесконечности равными нулю. Будем предполагать, что решение первой задачи известно и сконцентрируемся на второй. Далее мы будем рассматривать случай одноосного или двухосного растяжения на бесконечности, но основные рассуждения, изложенные в этой работе, могут быть применены и для случаев более сложных внешних напряжений, в том числе неоднородных.

Метод решения Для решения полученной задачи будем использовать численный метод разрывных смещений [4–8]. В данном методе трещина представляется в виде множества прямолинейных отрезков, на каждом из которых раскрытие и сдвиг берегов трещины предполагаются постоянными и обозначаются $D^i_n D^i_s$ соответственно. На рис. 1 изображена схема разбиения трещины на элементы.

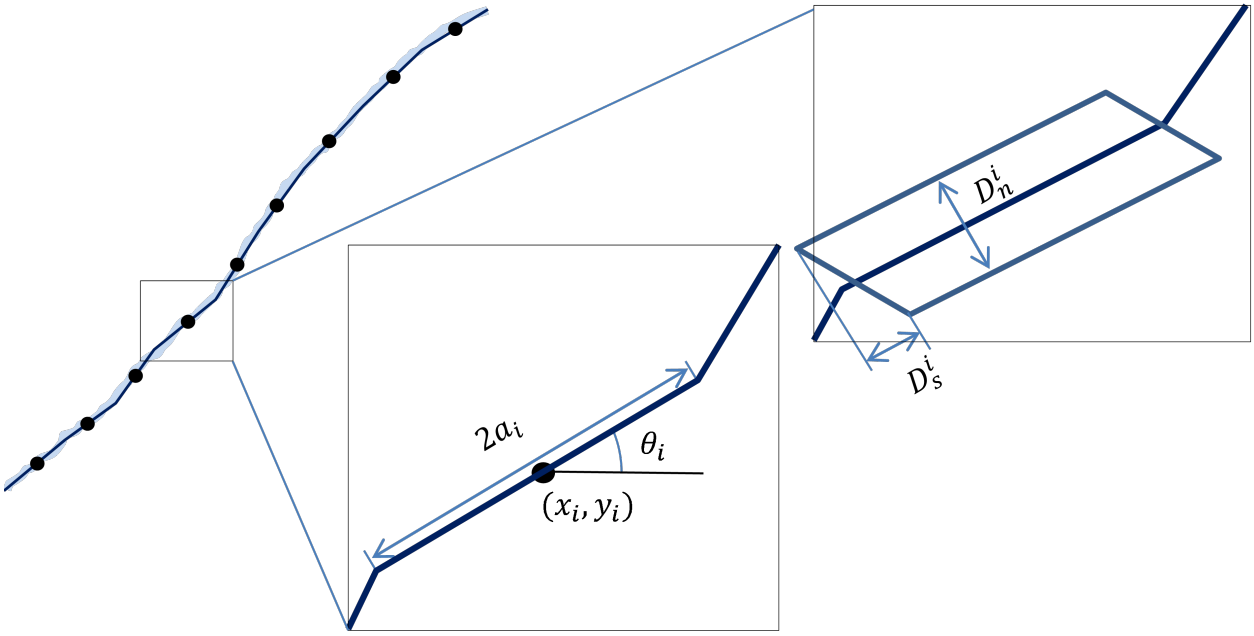


Рис. 1. *Схематичное разбиение трещины на элементы.* $2a_i$ *— длина элемента,* θ_i *— угол наклона от оси x,* (x_i, y_i) *— координаты центра элемента.* D^i_n, D^i_s *— нормальный и касательный разрывы смещений*

В таком случае поле напряжений, вызываемое отдельным элементом, описывается уравнениями (6), где A — известные функции, зависящие от размера элемента, его ориентации и расстояния до

точки (x,y), являющимися точными решениями уравнений (1)–(2) для задачи об одиночном прямолинейном разрыве смещения.

$$\begin{aligned}
\sigma_{xx}(x,y) &= \sum_i A^n_{xx}(x,y,x_i,y_i,\theta_i,a_i)D^i_n + A^s_{xx}(x,y,x_i,y_i,\theta_i,a_i)D^i_s \\
\sigma_{xy}(x,y) &= \sum_i A^n_{xy}(x,y,x_i,y_i,\theta_i,a_i)D^i_n + A^s_{xy}(x,y,x_i,y_i,\theta_i,a_i)D^i_s \\
\sigma_{yy}(x,y) &= \sum_i A^n_{yy}(x,y,x_i,y_i,\theta_i,a_i)D^i_n + A^s_{yy}(x,y,x_i,y_i,\theta_i,a_i)D^i_s
\end{aligned} \tag{6}$$

Граничное условие (5) при использовании этого метода выполняется только в точках, соответствующих центрам элементов.

$$\begin{aligned}
&\sigma_{nn}(x_i,y_i) = p_i - \sigma^0_{i\,nn} = \sigma^i_n \\
&\sigma_{ns}(x_i,y_i) = -\sigma^0_{i\,nn} = \sigma^i_s \quad . \\
&\sigma|_\infty = 0
\end{aligned} \tag{7}$$

Подстановка уравнений (6) в условия (7) дает нам систему линейных алгебраических уравнений относительно переменных D_n и D_s, где A^{ij}_{nn}, A^{ij}_{ns}, A^{ij}_{ss} и A^{ij}_{sn} — это постоянные коэффициенты, зависящие от взаимного расположения элементов i,j.

$$\begin{aligned}
\sigma^j_n &= \sum_{i=1}^N A^{ij}_{nn}D^i_n + A^{ij}_{sn}D^i_s \\
\sigma^j_s &= \sum_{i=1}^N A^{ij}_{ns}D^i_n + A^{ij}_{ss}D^i_s
\end{aligned} \quad , \qquad j = 1,\ldots,N. \tag{8}$$

Решение системы (8) позволяет получить раскрытия трещин и сдвиги берегов D_n и D_s. Использование формул (6) позволяет при известных D_n и D_s определять напряжения в любой точке среды.

Недостатком данного метода является наличие сингулярных точек на концах каждого граничного элемента. В этих точках напряжения принимают бесконечные значения, а вблизи от этих точек напряжения вычисляются с погрешностью. С другой стороны, вне окрестности особых точек метод позволяет определять напряжения с высокой точностью даже при небольшом количестве граничных элементов.

Для решения задачи о росте трещин, помимо уравнений (8) и условий (7), нам потребуются также законы, описывающие рост трещин: его направление и скорость.

Для определения направления роста трещины применим метод наибольших растягивающих напряжений [9]. В нем, с использованием асимптотического разложения поля напряжений в окрестности кончика трещины [10] и предположения, что трещина распространяется в направлении, для которого кольцевые растягивающие напряжения будут наибольшими, получено явное выражение зависимости угла поворота трещины от коэффициентов интенсивности напряжений.

$$\theta = 2\arctan\left(\frac{K_I - \sqrt{K_I^2 + 8K_{II}^2}}{4K_{II}}\right) \tag{9}$$

Распространение трещины будем предполагать квазистатическим, то есть на каждом шаге, в случае выполнения критерия разрушения, трещина увеличивается вдоль полученного направления на фиксированную длину Δs, равную размеру граничного элемента. Метод максимальных растягивающих напряжений предполагает силовой критерий распространения трещин для случая сложных нагружений, где K_I^C — трещиностойкость материала.

$$K_e q = \cos\frac{\theta}{2}\left[K_I\cos^2\frac{\theta}{2} - \frac{3}{2}K_{II}\sin\theta\right] \geq K_{IC} \tag{10}$$

Следует отметить, что определение коэффициентов интенсивности напряжений само по себе является отдельной задачей, которая имеет аналитическое решение для отдельных конфигураций трещин, но в случае криволинейной трещины должна решаться численно. В данной работе воспользуемся приближенным решением [11].

$$K_I = 0.806\frac{G}{2(1-\nu)}\sqrt{\frac{\pi}{\Delta s}}D_n$$
$$K_{II} = 0.806\frac{G}{2(1-\nu)}\sqrt{\frac{\pi}{\Delta s}}D_s \qquad (11)$$

Здесь D_n,D_s — это нормальный и касательный разрывы смещения ближайшего к кончику трещины элемента, Δs — его длина.

Уравнения (8)–(11) представляют собой метод моделирования растущих криволинейных трещин. Основными упрощениями, способными негативно влиять на точность расчета, являются:

- метод разрывных смещений нулевого порядка (8), где разрывы смещений предполагаются кусочно-постоянными функциями вдоль длины трещины, что дает погрешность измерения напряжений вблизи трещины;
- асимптотическое приближение коэффициентов интенсивности напряжений (11), дающее погрешность при уменьшении размера граничного элемента;
- предположение квазистатического роста, ввиду которого траектория трещины может зависеть от величины Δs — приращения длины на каждом шаге.

Для оценки точности рассматриваемого метода было проведено сравнение полученных с его помощью результатов с аналитическими и экспериментальными:

1. Плоская трещина под наклонной нагрузкой

На рис. 2 показана постановка задачи: растяжение образца с одиночной прямолинейной трещиной, ориентированной под углом к растягивающим нагрузкам [12].

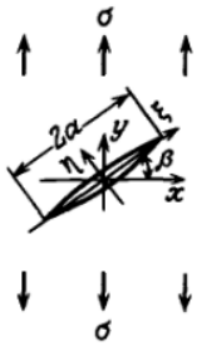


Рис. 2. *Растяжение образца с трещиной длины $2a$, находящейся под углом β к приложенным нагрузкам*

Данная задача имеет аналитическое решение для коэффициентов интенсивности напряжений:

$$K_I^t = \sigma\sqrt{\pi a}\cos^2\beta, K_{II}^t = \sigma\sqrt{\pi a}\cos\beta\sin\beta.$$

В таблице 1 представлены значения относительной погрешности $\Delta K_I = (K_I - K_I^t)/K_I^t$ и $\Delta K_{II} = (K_{II} - K_{II}^t)/K_{II}^t$ коэффициентов интенсивности напряжений K_I,K_{II}, полученных численно, к теоретическим значениям K_I^t и K_{II}^t. Для данной постановки значения ΔK_I оказались равны ΔK_{II}. Кроме того, эти значения оказались независимыми от угла β, поэтому в таблице приведена только зависимость от количества N элементов в разбиении трещины. Видим, что для одиночной прямолинейной трещины даже десяти элементов достаточно для обеспечения высокой точности определения коэффициентов интенсивности напряжений.

Следует также отметить, что повышение количества элементов свыше некоторого предела не увеличивает точность определения КИН. Это обусловлено тем, что повышение точности определения раскрытия трещины при уменьшении разрешения численного метода нивелируется уменьшением точности асимптотического приближения (11) при уменьшении размера граничного элемента.

Таблица 1

Значения относительной погрешности ΔK_I,ΔK_{II} для различных значений количества элементов N

N	10	30	100	300	1000
$\Delta K_I, \Delta K_{II}$	0.23%	0.5%	0.89%	0.9%	1%

2. Две параллельные трещины, расположенные перпендикулярно растягивающим нагрузкам

На рис. 3 представлена постановка задачи [13]. В таблице 2 представлены отношения численного и теоретического значения коэффициентов интенсивности для различных значений отношения расстояния между трещинами к их длине.

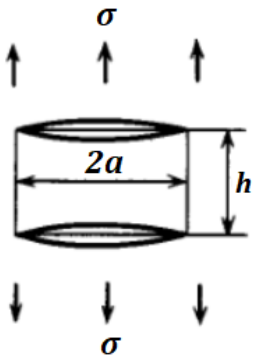


Рис. 3. *Две параллельные трещины длины $2a$ на расстоянии h друг от друга, под действием одноосного растяжения*

Таблица 2

Значения ΔK_I при различных N и $2a/h$

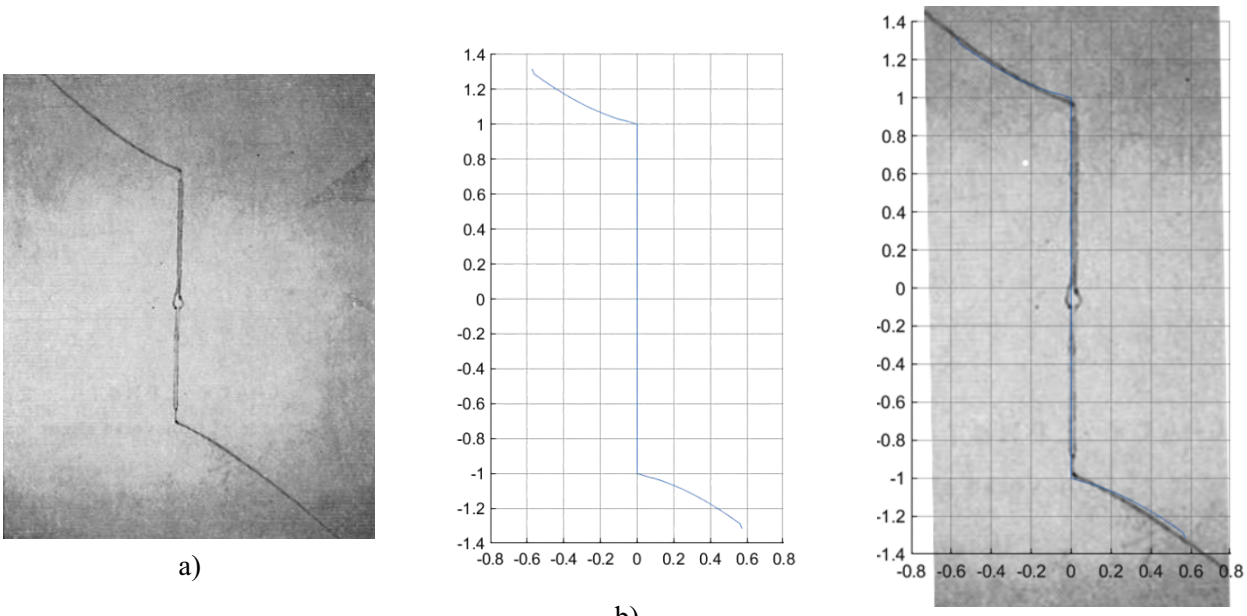
ΔK_I				
2a/h	N=10	N=30	N=100	N=300
0.2	0.35%	0.57%	0.9%	0. 99%
0.4	0.78%	0.39%	0.95%	0. 92%
0.6	1.8%	0.28%	0.82%	0. 97%
0.8	1.8%	0.02%	0.67%	0. 81%
1.0	0.9%	1.2%	2.03%	2.2%
1.25	1.3%	1.3%	2.2%	2.5%
2	2.6%	1.2%	2.6%	3.3%
5	8.1%	0.11%	3%	4.2%
10	16%	2.2%	3.6%	5.3%
100	25%	41%	5%	8.1%

Результаты показывают, что в случае, если расстояние между трещинами превосходит размер элемента, то ошибка между численными и теоретическими значениями коэффициентов интенсивности напряжений не превосходит нескольких процентов. Существенное повышение ошибки происходит только для больших значений $2a/h$, что соответствует случаю очень близкого расположения трещин, где используемый метод неприменим. Также имеет место понижение точности определения КИН при увеличении N, как и для случая одиночной трещины.

3. Эксперимент: трещина, растущая под распределенными сдвиговыми нагрузками

На рис. 4(a) показана экспериментально полученная траектория роста изначально прямолинейной трещины в образце из оргстекла, к которой были приложены сдвиговые нагрузки [9]. На рис. 4(b)

показана траектория роста трещины в той же постановке, полученная численно. Совмещение двух траекторий на рис. 4(c) показывает хорошее совпадение траектории трещины не только в начальный момент, что является показателем применимости критерия (9), но и при дальнейшем росте.



a) b) c)

Рис. 4. *Траектория трещины: a – экспериментальная, b – численная, c – наложение численной на экспериментальную*

Полученные результаты показывают, что данный метод дает хорошую точность для задач с одиночными, а также взаимодействующими и растущими трещинами. Ограничением применимости данного метода являются случаи, когда трещины находятся по отношению друг к другу на расстоянии меньшем, чем размер граничного элемента. В таком случае можно предполагать неточное определение как полей напряжений у кончиков, так и последующего направления роста.

Зависимость траектории трещины от размера граничных элементов

Помимо сравнения численно полученной траектории с экспериментальной, рассмотрим также зависимость траектории от размера элементов. Для этого рассмотрим рост горизонтально расположенной трещины начальной длины $l = 1$ под действием комбинации сдвиговых и растягивающих напряжений: $\sigma_{xx} = \sigma_{yy} = \sigma_{xy}$. На рис. 5 показаны траектории роста трещины для различного размера граничных элементов. Точками на графике отмечены центры элементов. Значения N_0 обозначают начальное количество элементов, на которые разбита трещина.

Полученные результаты показывают отсутствие существенной зависимости траектории трещины от размера используемых элементов. Таким образом, мы можем считать, что в нашей постановке величина приращения длины на каждом шаге не оказывает существенного влияния на траекторию трещины.

Исследование траектории поворота трещины

Отсутствие зависимости траектории роста трещины от размера элемента может обозначать устойчивость траектории трещины к малым отклонениям. Помимо численных ошибок, такие отклонения могут обозначать дефекты в реальных материалах ввиду неоднородности среды или наличия микротрещин. Для проверки данного предположения проведем численный эксперимент.

Рассмотрим трещину, состоящую из трех прямолинейных участков. Первый участок длины l_0 расположен вдоль оси x с центром в начале координат. Два других длиной $l_1 = l_0/20$ расположены в концах первого участка под углом α к нему (рис. 6).

На рис. 7 представлены траектории роста трещин при различных значениях α для одноосного растяжения вдоль оси y. Отметим, что для углов $\alpha \leq 40°$ (сплошные линии на рис. 7) отклонение трещины от оси не увеличивается с первого же шага. Для углов $\alpha \leq 80°$ (пунктирные линии на рис. 7) отклонение трещины от оси по мере роста трещины становится меньше начального. Для

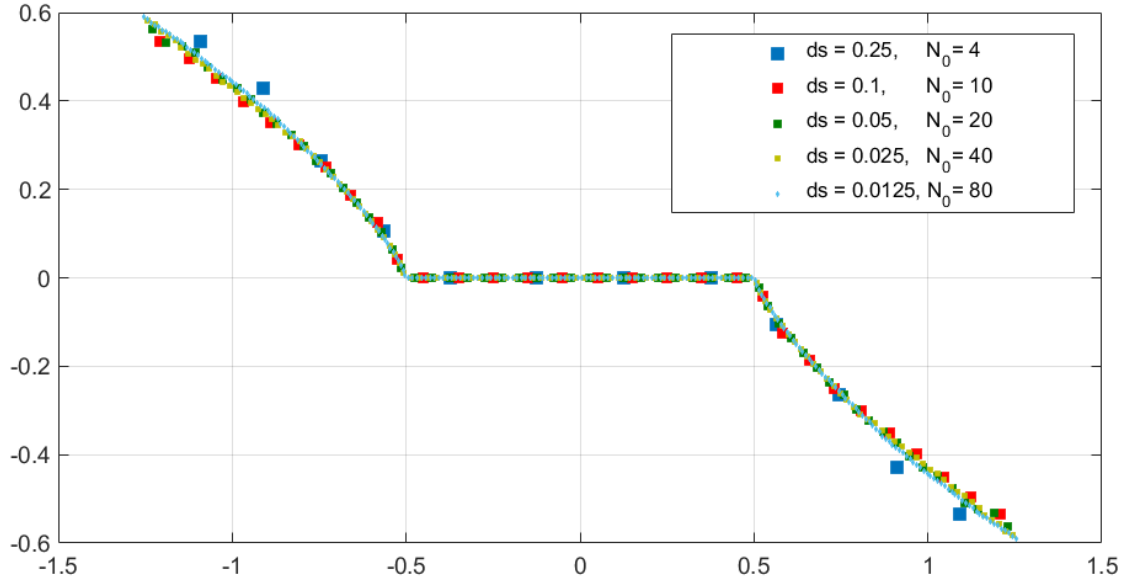


Рис. 5. *Траектория трещины при различных значениях длины ds граничного элемента. N_0 — начальное количество элементов в трещине*

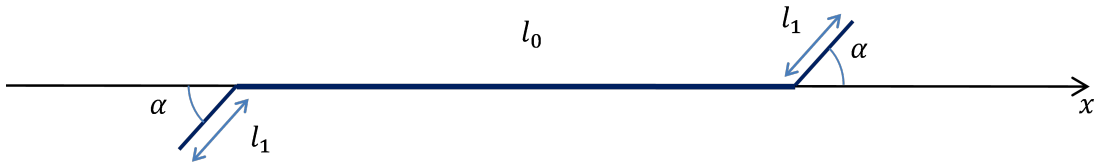


Рис. 6. *Начальная форма трещины с изломом*

$\alpha = 90°$ (штрихпунктирная линия) итоговое отклонение оказывается больше начального, но меньше максимального, достигаемого спустя несколько шагов роста.

На рис. 8 показаны траектории роста трещин в той же постановке для случая двухосного растяжения. В таких условиях направление роста трещины определяется только полем напряжений, вызванным самой трещиной. Полученные результаты показывают, что в этом случае отклонение траектории трещины от начального направления не ограничено. Это обусловлено тем, что в отсутствие перепада внешних напряжений трещина нормального отрыва стремится продолжать рост вдоль своей протяженности. В рассматриваемой постановке это направление можно получить, проведя прямую, соединяющую кончики трещины (штрихованная линия на рис. 8). Ввиду симметрии задачи данная прямая будет проходить через точку (0,0). В таком случае траекторию трещины нельзя считать устойчивой к отклонениям. Соответственно, результат вычисления может отличаться в зависимости от точности используемых методов.

Выводы

Рассмотрен метод на основе метода разрывных смещений с кусочно-постоянными разрывами смещений и коэффициентами интенсивности напряжений, вычисляемыми по значениям разрывов смещений в ближайшем к кончику элементе. Данный метод обеспечивает достаточно высокую точность для решения задач криволинейного роста трещин в случае, если расстояние между трещинами превосходит размер используемых разрывных элементов. Также данный метод обеспечивает точное определение траектории роста трещины, не зависящее существенно от размера используемых в разбиении трещины элементов, при условии, что рост трещины обусловлен не только формой трещины, но и внешними нагрузками. Показана устойчивость траектории трещины к малым отклонениям при наличии перепада напряжений, обеспечивающего наличие предпочитаемого направления роста. Данная устойчивость обеспечивает применимость методов невысокого порядка точности для определения траектории роста криволинейной трещины. С учетом данных факторов можно считать подобные ме-

тоды применимыми в задачах, где рост трещины обусловлен внешними нагрузками, а также в задачах взаимодействия трещин, при условии, что расстояние между этими трещинами достаточно велико.

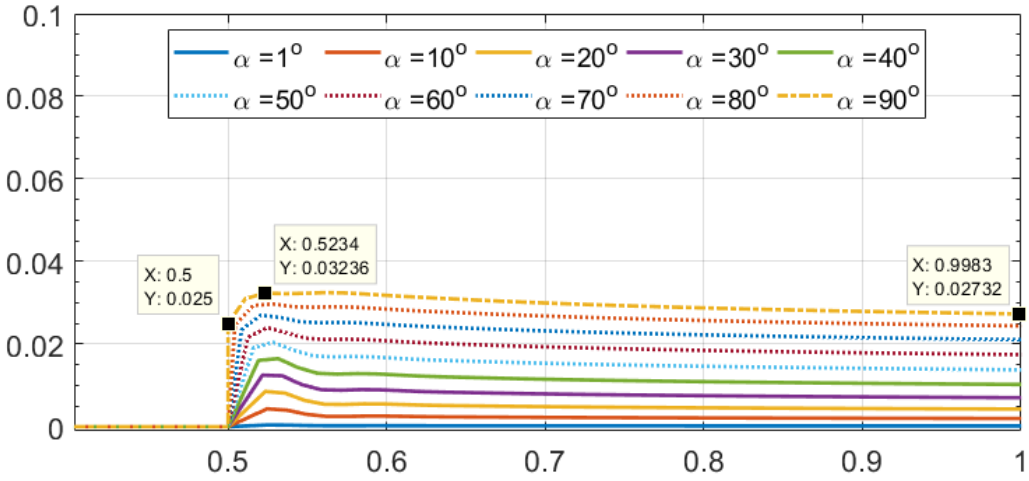


Рис. 7. *Траектории роста трещины при различных углах начального излома в случае одноосного растяжения*

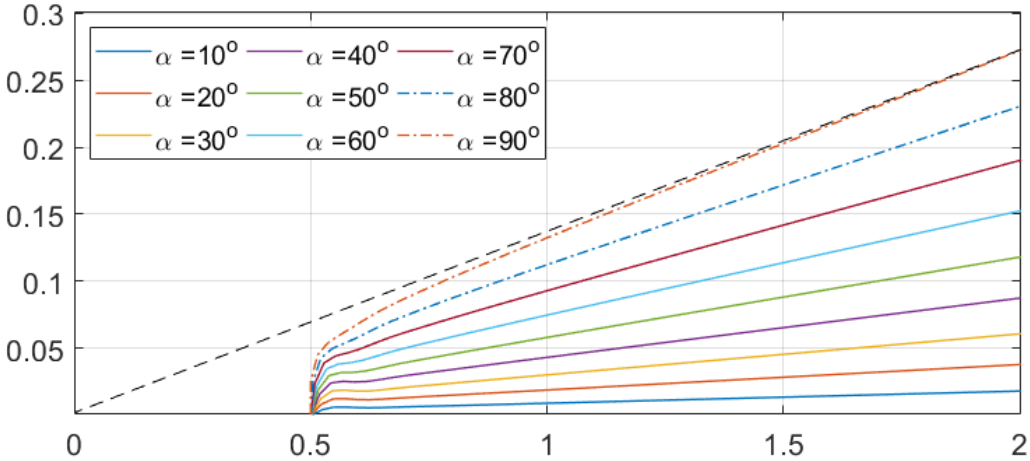


Рис. 8. *Траектории роста трещины при различных углах начального излома в случае двухосного растяжения*

ЛИТЕРАТУРА

1. Черный С. Г., Лапин В. Н., Есипов Д. В., Куранаков Д. С. *Методы моделирования зарождения и распространения трещин*. Новосибирск: Изд-во СО РАН; 2016. 312 с.
2. Akulich A. V., Zvyagin A. V., Pestov D. A., Tyurenkova V. V., Li K. Interaction of a Static Hydraulic Fracture under the Constant Pressure of a Fluid with a Natural Fault. *Mathematical Models and Computer Simulations*. 2019;11(2):209–218. DOI: 10.1134/S2070048219020029.
3. Zvyagin A. V., Udalov A. S. A Displacement Discontinuity Method of High-Order Accuracy in Fracture Mechanics. *Moscow University Mechanics Bulletin.* 2020;75(6):153–159. DOI: 10.3103/s0027133020060060.
4. Crouch S. L. Solution of Plane Elasticity Problems by the Displacement Discontinuity Method. I. Infinite Body Solution. *Int. J. for Numerical Methods in Engineering.* 1976;10(2):301–343. DOI: 10.1002/nme.1620100206.
5. Zvyagin A. V., Udalov A. S., Shamina A. A. Boundary Element Method for Investigating Large Systems of Cracks Using the Williams Asymptotic Series. *Acta Astronautica*. 2022;194:480–487. DOI: 10.1016/j.actaastro.2021.11.024.

6. Shamina A. A., Zvyagin A. V., Smirnov N. N. et al. Computational Modeling of Cracks Different Forms in Three-Dimensional Space. *Acta Astronautica*. 2021;186(7):289–302. DOI: 10.1016/j.actaastro.2021.05.041.
7. Li K., Smirnov N. N., Qi C., Kiselev A. B., Pestov D. A. The Numerical Asymptotic Solution to Initial Condition Problem of Preexisting Plane-Strain Hydraulic Fracture with Fluid Lag. *Engineering Fracture Mechanics*. 2020;239:107296. DOI: 10.1016/j.engfracmech.2020.107296.
8. Smirnov N. N., Li K., Skryleva E. I., Pestov D. A., Shamina A. A., Qi C., Kiselev A. B. Mathematical Modeling of Hydraulic Fracture Formation and Cleaning Processes. *Energies, MDPI*. 2022;15(6):1967. DOI: 10.3390/en15061967.
9. Erdogan F., Sih G. C. On the Crack Extension in Plates under Plane Loading and Transverse Shear. *Journal of Basic Engineering*. 1963;85(4):519–525. DOI: 10.1115/1.3656899.
10. Williams M. On the Stress Distribution at the Base of a Stationary Crack. *Journal of Applied Mechanics*. 1957;24:109–111.
11. Olson J. Fracture Aperture, Length and Pattern Geometry Development under Biaxial: A Numerical Study with Applications to Natural, Cross-Jointed Systems. *Geological Society, London, Special Publications*. 2007;289:123–142. DOI: 10.1144/SP289.8.
12. Westergaard H. Bearing Pressures and Cracks. *Journal of Applied Mechanics*. 1939;6:49-53.
13. Kamei A., Yokobori T. Some Results on the Stress Intensity Factors of Cracks and/or Slip Bands System. *Reports of the Research Institute for Strength and Frac. Mater*. 1974;10(2):29–93.

АНАЛИЗ ОТЕЧЕСТВЕННЫХ ПРОГРАММНЫХ СРЕДСТВ ПО ДВУМЕРНОМУ И ТРЕХМЕРНОМУ ПРОЕКТИРОВАНИЮ В УЧЕБНОМ ПРОЦЕССЕ

Т. Г. Бурдыко[а], К. И. Бушмелева[б]
Сургутский государственный университет, г. Сургут, Российская Федерация
[а] ✍ *burdyko_tg@surgu.ru*, [б] *bushmeleva_ki@surgu.ru*,

Аннотация: в статье рассматривается использование отечественного программного обеспечения для создания чертежей на компьютере, что облегчает усвоение учащимися основных положений чтения и выполнения моделей и сборочных единиц, а также умение применять знания при решении задач с творческим содержанием.

Ключевые слова: черчение, 2D-моделирование, 3D-моделирование, программное обеспечение, импортозамещение, российские аналоги, проектирование, САПР.

Для цитирования: Бурдыко Т. Г., Бушмелева К. И. Анализ отечественных программных средств по двумерному и трехмерному проектированию в учебном процессе. *Успехи кибернетики*. 2025;6(1):104–107.

Поступила в редакцию: 24.10.2024. *В окончательном варианте: 05.12.2024.*

DOMESTIC 2D AND 3D CAD SYSTEMS FOR STUDENTS

T. G. Burdyko[a], K. I. Bushmeleva[b]
Surgut State University, Surgut, Russian Federation
[a] ✍ *burdyko_tg@surgu.ru*, [b] *bushmeleva_ki@surgu.ru*,

Abstract: we explored the use of domestic drafting software. This approach simplifies teaching students the fundamental principles of understanding and making part and assembly drawings. It also improves their ability to solve problems creatively.

Keywords: drafting, 2D modeling, 3D modeling software, import substitution, domestic software, design, CAD.

Cite this article: Burdyko T. G., Bushmeleva K. I. Domestic 2D and 3D CAD Systems for Students. *Russian Journal of Cybernetics*. 2025;6(1):104–107.

Original article submitted: 24.10.2024. *Revision submitted: 05.12.2024.*

Системы автоматизированного проектирования (САПР) сейчас остаются основными инструментами для создания двумерных и трехмерных графических моделей и чертежей. Они значительно ускоряют процесс проектирования и изготовления, повышая точность расчетов и качество конечного продукта.

Autodesk заслужила признание как один из ведущих разработчиков программного обеспечения, в частности, в области САПР. Программный продукт AutoCAD, созданный компанией Autodesk, получил широкое распространение в различных отраслях проектирования. В Российской Федерации данное программное обеспечение пользовалось популярностью среди специалистов, в частности, архитекторов, инженеров-механиков, ландшафтных дизайнеров, подрядчиков, а также студентов. В своей работе эти профессионалы активно применяли AutoCAD для создания как двумерных, так и трехмерных моделей.

Однако компания Autodesk приняла решение о прекращении деятельности на территории Российской Федерации и ограничении доступа к своему программному обеспечению. Данный запрет распространяется на российские организации, но не затрагивает частных лиц.

В настоящей статье рассматриваются отечественные альтернативные программы, которые могут быть использованы для создания и оформления чертежей и графических моделей.

Российские специалисты разработали альтернативные решения в сфере инженерной графики, которые могут полностью заменить, а в некоторых случаях и дополнить функциональность формата DWG, который применяется в AutoCAD. По большей части это коммерческие пакеты, однако их стоимость существенно ниже, чем у международного стандарта.

Использование современного отечественного программного обеспечения в сфере графики представляет собой эффективную стратегию для разработки и работы с чертежами различного типа. Данный подход способствует получению новых знаний студентами в ходе образовательного процесса.

Эти программы [1–4] могут быть успешно использованы в следующих учебных программах высшего образования:

- **архитектурные специальности:** студенты-архитекторы могут применить программное обеспечение для проектирования зданий и сооружений, разработки их интерьеров. Использование такого программного обеспечения позволяет им экспериментировать с различными материалами, текстурами и источниками освещения, что дает возможность получить реалистичное представление о будущей постройке. Трехмерное моделирование также предоставляет возможность проведения виртуальных экскурсий по проекту, изучения его пространственной организации и оценки эргономических характеристик;
- **строительные специальности:** студенты строительных специальностей могут использовать программное обеспечение для проектирования и разработки проектов строительных конструкций, таких как фундаменты, стены и другие элементы. Данный софт позволяет моделировать нагрузки и анализировать устойчивость сооружения на различных этапах строительства. Это способствует глубокому пониманию студентами механических свойств материалов и процессов возведения зданий;
- **машиностроительные специальности:** в области машиностроения данное программное обеспечение выступает в качестве основного инструмента для проектирования деталей и механизмов. С его помощью студенты имеют возможность моделировать разные виды двигателей, шасси, трансмиссий и других составных частей машин. Программное обеспечение предоставляет функциональность для проведения виртуальной сборки и анализа прочности конструкций, что оптимизирует процесс проектирования и сокращает вероятность возникновения ошибок;
- **энергетические специальности:** данное программное обеспечение предоставляет студентам-энергетикам мощный инструмент для проектирования объектов энергетической инфраструктуры, таких как электростанции и линии электропередач. Используя это ПО, студенты могут создавать виртуальные модели сложного оборудования, включая турбины, генераторы и трансформаторы. Возможность проведения симуляций работы энергосистем позволит им глубже понять принципы функционирования энергетической инфраструктуры;
- **транспортные специальности:** студенты указанных специальностей обладают возможностью применения программного обеспечения для проектирования транспортных средств, разработки схем их движения и анализа безопасности дорожного движения. Они могут создавать виртуальные модели автомобилей, поездов, самолетов и иных видов транспорта, а также проводить симуляции движения на дорогах и железнодорожных путях;
- **геодезические специальности:** студенты-геодезисты могут использовать ПО для создания топографических карт, планирования земельных участков и анализа рельефа местности. ПО позволяет создавать виртуальные модели рельефа и окружающей среды, а также проводить виртуальные измерения и анализировать геодезические данные.

На сегодняшний день, после ухода Autodesk из России, сравнивать САПР-пакеты следует, в первую очередь, по их функциональности, гибкости, стабильности, и здесь выясняется, что подобрать замену легендарному стандарту особого труда не составляет. Главное — четко понимать, для каких именно задач нужна цифровая проектировочная платформа: просто ли для отрисовки чертежей, или для создания 3D-образов сложных механических конструкций, или же для процессного моделирования и BIM. Под любую задачу найдется достойный аналог AutoCAD, в том числе и российской разработки, во многих случаях еще и превосходящие оригинал [7].

Программы для создания 2D- и 3D-моделей являются мощным инструментом для обучения студентов технических специальностей [5]. Эта технология превращает образование в более практичный, визуальный и интерактивный процесс, что позволяет студентам получить более глубокие знания и подготовиться к работе в современных инженерных компаниях (таблица 1).

В России существует несколько программ по графическому черчению, которые используются в различных отраслях. В этом анализе мы сравним некоторые из самых популярных российских ПО, подчеркнув их функции, преимущества и недостатки (таблица 2) [1–3, 6].

Каждая из российских САПР имеет свои сильные и слабые стороны. *КОМПАС-3D* — это мощ-

Таблица 1

Преимущества использования графического софта в техническом образовании

Характеристика	Преимущество
практическое обучение	студенты могут применять теоретические знания на практике, работая с как с 2D-, так и с 3D-моделями
визуализация	3D-моделирование позволяет студентам видеть проекты в пространстве, что делает их более понятными и доступными для восприятия
интерактивность	студенты могут взаимодействовать с созданными моделями, изменять их и проводить различные симуляции
эффективность	за счет понятного интерфейса и большого количества инструментов облегчается процесс проектирования, что приводит к снижению количества ошибок

Таблица 2

Анализ отечественного графического софта

Программа	**Отрасль**	**Функции**	**Преимущества**	**Недостатки**
КОМПАС-3D	Машиностроение, архитектура, энергетика, строительство и транспорт	2D- и 3D-моделирование, параметрическое проектирование, инструменты для совместной работы, симуляция	Интуитивно понятный интерфейс, мощные инструменты моделирования, доступная цена	Ограниченная совместимость с международными стандартами САПР, ограниченные онлайн-ресурсы
NanoCAD	Архитектура, энергетика, транспорт, инженерия и дизайн	2D- и 3D-моделирование, параметрическое проектирование, инструменты для совместной работы	Бесплатная версия, совместимость с файлами AutoCAD, интуитивно понятный интерфейс, что облегчает новичкам освоение программы. Он напоминает привычные CAD-системы, что позволяет быстро адаптироваться	Ограниченные онлайн-ресурсы
Astra R-Nesting	Проектирование мебели, деревообработка и металлообработка	2D- и 3D-моделирование, вложение и оптимизация	Специализированное программное обеспечение для вложения и оптимизации, удобный интерфейс, доступная цена	Ограниченные возможности, по сравнению с программным обеспечением САПР общего назначения, ограниченные онлайн-ресурсы
QGIS	Геоинформационные системы	Гибкая платформа для работы с геоданными, поддерживающая создание и анализ трехмерных моделей	Специализированное программное обеспечение, удобный интерфейс, открытый код	Из-за открытого кода нет классической защиты данных, прикладные разработки предлагаются как коммерческие продукты, большая часть технической документации доступна только на английском языке

ная САПР-программа с широким набором функций, но она может иметь ограниченную совместимость с международными стандартами САПР. *NanoCAD* — это бесплатный и удобный вариант, но с ограниченными возможностями, по сравнению с коммерческим программным обеспечением. Подходит для тех, кто только собирается освоить подобные программы. Совместим с файлами AutoCAD, что позволяет быстро адаптироваться [1, 6]. *Astra R-Nesting* — это специализированное программное обеспечение для вложения и оптимизации, но оно может не подходить для задач САПР общего назначения. *QGIS* — это специализированное программное обеспечение, имеющее удобный интерфейс, также может не подходить для задач общего назначения в области САПР.

Использование программ для черчения в высшем образовании имеет множество положительных аспектов. Во-первых, использование такой программы, как NanoCAD, значительно повышает качество проектирования и визуализации, позволяя студентам создавать точные и профессиональные чертежи. Во-вторых, это улучшает процесс обучения, так как студенты могут быстрее усваивать материал, благодаря интерактивным и наглядным инструментам.

Кроме того, навыки работы с современными CAD-программами являются важными конкурентными преимуществами на рынке труда, что способствует повышению уровня трудоустройства выпускников. Тем не менее, следует учитывать необходимость подготовки преподавателей и обеспечение доступа к необходимым программам и оборудованию. В целом, интеграция данных технологий в учебный процесс способствует более глубокому пониманию дисциплин, связанных с инженерией и дизайном.

ЛИТЕРАТУРА

1. Зимина О. *Лучшие программы для черчения на компьютере.* Режим доступа: https://amssoft.ru/repair/programmy-dlya-chertezhej.php?ysclid=lvulpt6wgf970800073.
2. *ТОП-10 российских аналогов AutoCAD.* Режим доступа: https://otzyvmarketing.ru/articles/rossijskie-analogi-autocad/?ysclid=lvujx300d9534494965.
3. Антипина Н. А., Буркова С. П., Долотова Р. Г., Нехорошева Ю. Г. *Инженерная графика*: уч. пособие. Томск: Томский политехнический университет; 2013. 168 с.
4. Берлинер Э. *Актуальность применения САПР в машиностроении.* Режим доступа: https://sapr.ru/article/7837?ysclid=lvulpkui43642026369.
5. Саидова С. А. Роль инженерной и компьютерной графики в учебном процессе. *Наука и образование в современном мире: вопросы теории и практики*: материалы Международ. (заочной) науч.-практ. конф. Нефтекамск; 2023. С. 32–40.
6. Баденко В. Л., Волкова Ю. В. *Основы создания трехмерных моделей в среде NanoCAD*: уч. пособие. СПб.: Санкт-Петербургский политехнический университет Петра Великого; 2024. 59 с.
7. *САПР-замещение: что использовать вместо AutoCAD?* Режим доступа: https://www.it-world.ru/tech/gfnkf071n34gso4gso4wksocks4ss4o.html.

СОВРЕМЕННЫЕ ПРОБЛЕМЫ БИОКИБЕРНЕТИКИ

В. М. Еськов[1,а], **Ю. М. Попов**[2,б], **О. А. Ведясова**[3,в], **Т. Я. Корчина**[4,г]

[1] *Сургутский филиал федерального государственного автономного учреждения «Федеральный научный центр Научно-исследовательский институт системных исследований Национального исследовательского центра «Курчатовский институт», г. Сургут, Российская Федерация*

[2] *Самарский государственный социально-педагогический университет, г. Самара, Российская Федерация*

[3] *Самарский национальный исследовательский университет имени академика С.П. Королева, г. Самара, Российская Федерация*

[4] *Ханты-Мансийская государственная медицинская академия, г. Ханты-Мансийск, Российская Федерация*

[а] ✍ *firing.squad@mail.ru*, [б] *kafedra_afgch@mail.ru*, [в] *o.a.vedyasova@gmail.com*, [г] *t.korchina@mail.ru*

Аннотация: в 1999 году В. Л. Гинзбург представил фундаментальную статью о проблемах физики и астрофизики. В ней он особым образом поднимает проблему тоталитаризма в науке. При этом Гинзбург говорит о лысенковщине, которая не просто представила интенсивное направление в науке, но и вызвала ярую негативную реакцию на ряд других (новых) наук. Речь идет в первую очередь о генетике и кибернетике, которые тогда объявили «лженауками» и, политически, преследовали ученых. Удивительно, но история с кибернетикой, тоже с биокибернетикой, повторилась. Сейчас активно отрицаются (или замалчиваются) работы Н. А. Бернштейна, W. Weaver и эффект Еськова-Зинченко. Фактически игнорируются научные факты, которые могут дать для биомедицины, психологии, экологии даже больше, чем дали генетика и кибернетика. Что это за факты, и о какой новой науке говорил Weaver, и говорится в нашем сообщении. Ответы на эти вопросы мы и представляем.

Ключевые слова: стохастика, хаос, тоталитаризм, эффект Еськова–Зинченко.

Благодарности: работа выполнена в рамках государственного задания НИЦ «Курчатовский институт» — НИИСИ по теме № FNEF-2024-0001 «Создание и реализация доверенных систем искусственного интеллекта, основанных на новых математических и алгоритмических методах, моделях быстрых вычислений, реализуемых на отечественных вычислительных системах» (1023032100070-3-1.2.1).

Для цитирования: Еськов В. М., Попов Ю. М., Ведясова О. А., Корчина Т. Я. Современные проблемы биокибернетики. *Успехи кибернетики*. 2025;6(1):108–115.

Поступила в редакцию: 13.11.2024. *В окончательном варианте: 15.12.2024.*

MODERN PROBLEMS IN BIOCYBERNETICS

V. M. Eskov[1,a], **Yu. M. Popov**[2,b], **O. A. Vedyasova**[3,c], **T. Ya. Korchina**[4,d]

[1] *Surgut Branch of Scientific Research Institute for System Analysis of the National Research Centre "Kurchatov Institute", Surgut, Russian Federation*

[2] *Samara State University of Social Sciences and Education, Samara, Russian Federation*

[3] *Samara National Research University, Samara, Russian Federation*

[4] *Khanty-Mansiysk State Medical Academy, Khanty-Mansiysk, Russian Federation*

[a] ✍ *firing.squad@mail.ru*, [b] *kafedra_afgch@mail.ru*, [c] *o.a.vedyasova@gmail.com*, [d] *t.korchina@mail.ru*

Abstract: in 1999, V. L. Ginzburg published a seminal article on the challenges in physics and astrophysics, where he uniquely addressed the issue of totalitarianism in science. He specifically discussed Lysenkoism, which not only shaped a dominant scientific direction but also triggered strong opposition to emerging fields such as genetics and cybernetics, both of which were labeled "pseudosciences" and subjected to political persecution. A similar pattern has emerged in the case of cybernetics and biocybernetics. Today, the contributions of N. A. Bernstein, W. Weaver, and the Eskov-Zinchenko effect are often dismissed or ignored, despite their potential to advance biomedicine, psychology, and ecology beyond what genetics and cybernetics have achieved. What are these overlooked scientific facts? What new science did Weaver envision? And what is the key message of our study? We explore these questions and present our findings.

Keywords: stochastics, chaos, totalitarianism, the Eskov-Zinchenko effect.

Acknowledgements: this study is a part of the FNEF-2024-0001 government order contracted to the Scientific Research Institute for System Analysis of the National Research Centre “Kurchatov Institute”, project No. 1023032100070-3-1.2.1 Development and Implementation of Trusted Artificial Intelligence Systems Based on new Mathematical Methods and Algorithms, Fast Computing Models for Domestic Computing Systems.

Cite this article: Eskov V. M., Popov Yu. M., Vedyasova O. A., Korchina T. Ya. Modern Problems in Biocybernetics. *Russian Journal of Cybernetics*. 2025;6(1):108–115.

Original article submitted: 13.11.2024. *Revision submitted: 15.12.2024.*

Введение

На рубеже тысячелетий, в 1999 году, нобелевский лауреат В. Л. Гинзбург представил обзорную статью, которая обозначила основные проблемы физики и астрофизики [1]. В этой работе Гинзбург отдельно поднимает проблему тоталитаризма во всех науках. В первую очередь, это относится к наукам о жизни [1].

Действительно, развитие науки очень часто ограничивалось тоталитарными представителями. Это было со времен Коперника, Галилея и Бруно и многократно повторялось, в том числе и в 50-е годы в виде «лысенковщины» по отношению к генетике. При этом выделили и кибернетику, как реальную (буржуазную) лженауку. Начались тогда политические гонения в адрес генетиков и кибернетиков.

Гонения в адрес генетики и кибернетики были недолгими, но весьма «эффективными» для тех, кто их организовал. За ними стояла политическая система, что в современной России совершенно невозможно. Однако у «лысенковщины» есть вторая сторона: замалчивание научных фактов. Это тоже является тормозом для всей науки и активной формой догматизма в науке. Это вторая догма тоталитаризма.

Напомним, что в 1947 году Н. А. Бернштейн выдвинул гипотезу о «повторении без повторений» в биомеханике. Через год W. Weaver [2] предложил общую классификацию всех систем природы и выделил биосистемы в особый, третий, класс (тип) науки [2]. Доказательства правоты этих ученых мы смогли представить только через 50 лет, и это был эффект Еськова-Зинченко (ЭЕЗ), который сейчас игнорируется [3–11]. По нашим оценкам замалчивание — это тоже тоталитаризм.

Представления Бернштейна, Weaver и Гинзбурга — это революция в биомеханике и кибернетике.

В фундаментальной работе [1] нобелевский лауреат поднимал целый ряд проблем всей [1] науки. Это касается не только физики и астрофизики [1], но именно всей науки в целом. Он говорит о развитии науки, о проблемах догматизма во всех науках, а не только в физике. Это принципиальная статья для всей науки [3–9].

Действительно, догматизм (тоталитаризм) в науке погубил многие передовые идеи. В ряде случаев были уничтожены (унижены) и сами ученые — носители этих передовых идей. Уже в 20-м веке мы имеем тоталитаризм в науке, который затормозил науку в СССР, но не в мире. Это торможение активно поддерживалось политической системой и это было ошибкой (государство не должно вмешиваться в развитие науки).

Речь идет о гонении на генетику и кибернетику. Гинзбург говорит: «Однако, лишь победа тоталитаризма… может радикально помешать прогрессу в науке в результате возникновения явлений типа лысенковщины» [1]. О чем идет речь?

Такое «явление» имеет два аспекта. Во-первых, оно устанавливает доминирующее представление в науке, и это представление является «лженаукой». Во-вторых, лысенковщина активно борется с новыми направлениями в науке. Эта борьба может быть активной, в виде репрессий по отношению к генетикам и кибернетикам, а может быть пассивной.

Именно со второй формой мы столкнулись в середине 20-го века. Выдающийся биомеханик Н. А. Бернштейн в 1947 году в книге «О построении движений» выдвинул гипотезу о «повторении без повторений». Он утверждал, что любое движение неповторимо. В каком смысле (точно неповторимо или в рамках стохастики) эта неповторимость проявляется, он не сказал.

Напомним, что существует два типа «повторений»: в виде точного повторения любой фазовой траектории в m-мерном фазовом пространстве состояний, это мы имеем в теории динамических систем (ТДС), либо повторения в стохастике. В ТДС имеем совпадение по точкам, в стохастике говорим о распределениях [3–11], где точки могут не совпадать (как в ТДС).

В последнем случае мы имеем дело с выборками и функциями распределения. Существуют критерии приблизительного сравнения этих статистических функций распределения $F(x)$. Подчеркнем, в стохастике нет точных (точечных) совпадений, сравниваются выборки. Обычно при этом опыты многократно повторяют (и получают выборки), но точки (в выборках при сравнении) не совпадают.

Очень странно, но за последние 70 лет никто не пытался проверить эту гипотезу Бернштейна о «повторении без повторений». Подчеркнем, что он дал качественное доказательство для нее. Он говорил, что в организации «повторения» движений участвует не менее 5-ти разных систем (это системы A, B, C, D, E). Все они включаются, выключаются в процесс движений хаотически. Сила и длительность влияния этих систем на организацию движений тоже хаотическая (не регулируется сознанием, мозгом человека).

Ровно через год W. Weaver предложил общую классификацию систем [2]. В ней он выделял простые (детерминистские) системы — Simplicity (или системы 1-го типа — СПТ), системы 2-го типа — Disorganized Complexity (СВТ), т. е. стохастические системы. Отдельно он выделил биосистемы или системы 3-го типа — СТТ. Для СТТ Weaver предлагал создать новую (третью) науку.

Для СПТ он считал, организована (создана) теория динамических систем — ТДС. Для СВТ создана стохастика (теория вероятности, статистика и т.д.). Для СТТ нужна новая особая третья наука. Какими особыми свойствами обладают биосистемы (СТТ), Weaver тогда не указал. Это была только гипотеза, которая основана на самоорганизации СТТ (это идея Weaver) [2].

В итоге, за последние 70 лет мы наблюдаем классический тоталитаризм. Гипотезы Бернштейна и Weaver игнорируются (замалчиваются). Никто даже не пытается проверить отсутствие «повторений» в биомеханике или выявить особые свойства СТТ в науках о живых системах. Что это такое? Проявление «тоталитаризма» на мировом уровне?

Нам всем (нашей научной школе, а это более 200 кандидатов и докторов наук в Сургуте, Самаре, Туле, Москве и т.д.) непонятно, почему никто в медицине, биологии, психологии, экологии и т. д. не проверил устойчивость выборок параметров биосистем. Все были уверены: выборка, полученная на интервале времени Δt_{ij} легко (статистически) будет повторена и на интервале (следующим за Δt_1) Δt_2. Это было догмой (без доказательств) всех наук о живых системах последние 150 лет [11–18].

Почему никто за 150-200 лет даже не пытался проверить статистическую устойчивость выборок любых параметров в биомедицине? Это загадка для нашей научной школы. Поразительно, но это так очевидно: проверить совпадение двух соседних выборок кардиоинтервалов (КИ), электромиограмм (ЭМГ), теппинграмм (ТПГ), электроэнцефалограмм (ЭЭГ), треморограмм (ТМГ) для одного и того же человека [12–18].

Совпадают ли выборки на соседних интервалах времени Δt_1 и Δt_2 для одного и того же человека, находящегося в одном и том же психическом, физическом, физиологическом состоянии? Это тривиальный вопрос, который мог бы сразу доказать гипотезы Бернштейна и Weaver, но никто даже не пытался сделать это. Все были в плену догмы — тоталитаризма. Эта догма была более 150 лет: соседние выборки функций организма человека совпадают. Никто в этом не сомневался в биологии, биокибернетике, медицине и т. д.

Вторая глобальная догма: любая группа одинаковых по возрасту, полу, здоровых (или с одним заболеванием) людей всегда будет однородной. Это догма всей современной биологии, медицины, экологии, психологии и т. д. Миллионы биологов, врачей, психологов, экологов работали в рамках этой фундаментальной догмы. Никто за 150 лет не пытался это проверить [11–18].

Существенно, что нобелевский лауреат В. Л. Гинзбург в этом сомневался [1]. Он прямо говорит, что: «...гипотеза о редукции — возможности все живое объяснить на основе физики, уже известной физики... можно ли считать, что возможность редукции биологии и современной физики несомненна? ... дать положительный ответ было бы, как мне кажется, неправильным» [1, с. 436–437]. Именно такая «неправильность» сейчас господствует в физике, биофизике, биокибернетике, математике, во всех науках о живых системах. Это следствие из ЭЕЗ: если сам человек не может совпадать по выборкам, то тем более даже 2 человека не совпадут статистически по выборкам.

Может ли физика описывать биосистемы?

В своей фундаментальной статье В. Л. Гинзбург поднимает «три «великие» проблемы» физики [1]. Первая проблема «...о возрастании энтропии, необратимости и «стреле времени» [1, с. 435]. Сразу отметим, что эта проблема для биосистем (СТТ) имеет совершенно иной смысл. СТТ не подчиняются законам и моделям ТДС и всей стохастики. Энтропия для биосистем (локально) падает или не

изменяется [3, 4], но не по законам физики. Это отдельная проблема всей современной науки.

Вторая проблема — «это проблема интерпретации и понимание квантовой механики» [1]. Как показал М. Б. Менский, она связана с наблюдателем и квантовой теорией сознания (КТС). Сама КТС продолжает парадоксы H. Evertt и проблемы неопределенности в квантовой механике [12–15].

Наконец, третья «великая» проблема физики напрямую связанна с СТТ [1]. Она касается темы: «. . . это вопрос о связи физики с биологией и, конкретно, проблема редукционизма» [1]. Как мы отметили цитатой Гинзбурга выше, нобелевский лауреат не верит в редукционизм. По его мнению, надеяться на редукционизм «неправильно» [1].

Более того, Гинзбург подчеркивает, что «. . . вопрос о редукционизме — это одновременно великая физическая и биологическая проблема, и она, как я убежден, будет одной из центральных в науке XXI века» [1]. К большому нашему сожалению, выдающийся ученый ошибся. Заканчивается первая четверть 21-го века, но в биомедицине ничего не изменилось. Все догмы старые, никто даже не пытается проверить гипотезы Бернштейна и Weaver о биосистемах. Это реальный тоталитаризм.

ТДС и стохастика, как и прежде, описывают все биосистемы. Люди (ученые) получают нобелевские премии за модели в рамках ТДС и стохастики. При этом никто не осознает, что любая статистическая модель не имеет практического смысла. СТТ — уникальные системы, их нельзя описывать в рамках ТДС и стохастики. Такие модели имеют исторический характер, нет прогноза будущего [12–18].

В. Л. Гинзбург это чувствовал и понимал. Он говорил: «Пока дело не сделано, нельзя исключать возможность того, что мы даже на фундаментальном уровне еще не знаем чего-то необходимого для редукции» [1, с. 437]. Таким фундаментальным уровнем является эффект Еськова-Зинченко (ЭЕЗ). В нем доказана уникальность любой выборки любого параметра СТТ [12–19].

Это означает, что нет прогноза будущего для любой биосистемы (СТТ), любая модель в ТДС или в стохастике описывает прошлое СТТ на любом интервале времени Δt_1. Сказать что-либо определенное о состоянии СТТ на следующем Δt_2 мы не можем. Прогноз имеет 0,05-0,3 вероятности. Именно это доказано в ЭЕЗ, и это сейчас пренебрегается в физике, математике и во всех науках о жизни [12–19].

Из-за ЭЕЗ во всех науках возникли проблемы с описанием СТТ. Физические модели, столь общепринятые в биокибернетике и биофизике, имеют только качественный характер. Об этом еще говорил нобелевский лауреат I. R. Prigogine в своей последней монографии «The End of Certainty. . . » [20]. Нет в природе биосистем, которые описываются дифференциальными (и другими) уравнениями (имеется в виду их точное описание).

Более того, любая биосистема не может повторить фазовую траекторию модели в ТДС, которая якобы описывает эту биосистему. Именно об этом говорил Prigogine [20] и третий нобелевский лауреат R. Penrose [21]. Об этом прямо говорил и M. Gell-Mann [22].

Все эти четверо нобелевских лауреатов говорили о проблемах неопределенности для всех Complexity, но никто из них не отрицал возможности стохастики при описании всех биосистем. Prigogine верил в стохастику и теорию динамического хаоса (ТДХ). Однако мы доказали [11–19], что ТДХ не имеет никакого отношения к биосистемам (СТТ) в плане прогноза будущего.

Реальная Complexity для СТТ

На ТДХ мы остановимся подробнее в других наших сообщениях, а сейчас продемонстрируем несколько примеров из ЭЕЗ, которые доказывают статистическую неустойчивость выборок и потерю однородности любой группы испытуемых. Фактически речь идет о реальной Complexity, про которую пытались сказать Бернштейн, Weaver и Гинзбург.

Действительно, если мы 15 раз зарегистрируем по 5 минут у одного и того же человека выборки КИ, то можно построить матрицу парных сравнений этих 15-ти выборок КИ. В такие матрицы мы вносим критерий Вилкоксона p_{ij} для сравниваемой *i*-й и *j*-й выборок КИ (для данного испытуемого в покое).

В итоге мы построили несколько тысяч таких матриц парных сравнений выборок и нашли, что в них число пар k (с $p_{ij} \geq 0{,}05$) очень мало. Обычно $k \leq 20$ (из всех 105-ти пар сравнения). Для примера мы представляем такую матрицу в таблицу 1 для КИ. Очевидно, что при $p_{ij} < 0{,}05$ такая пара не имеет общую генеральную совокупность. Выборки статистически разны $k < 20$.

Подчеркнем, что в кардиологии, физиологии это стандарт: КИ нужно регистрировать не менее 5 минут. Считается, что такая выборка будет репрезентативна. Однако таблица 1 и тысячи ей подобных

Таблица 1

Уровни значимости (P) для попарных сравнений 15-ти выборок параметров кардиоинтервалов (КИ) испытуемой (без нагрузки, число повторов n=15), использовался критерий Вилкоксона (значимость p<0,05, число совпадений $k^1_{i,j} = 12$)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1			0,07	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00		0,00	0,00	0,00
2	0,00		0,00	0,00	0,94	0,00	0,00	0,00	0,00	0,00		0,00	0,00	0,00	0,00
3	**0,07**	0,00		0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,07	0,00	0,00	0,00
4	0,00	0,00	0,00		0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
5	0,00	**0,94**	0,00	0,00		0,00	0,00	0,00	0,00	0,00	0,94	0,00	0,00	0,00	0,00
6	0,00	0,00	0,00	0,00	0,00		0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
7	0,00	0,00	0,00	0,00	0,00	0,00		0,01	0,60	0,250	0,00	0,00	0,06	0,00	0,00
8	0,00	0,00	0,00	0,00	0,00	0,00	0,01		0,00	0,03	0,00	0,00	0,17	0,02	0,06
9	0,00	0,00	0,00	0,00	0,00	0,00	**0,60**	0,00		0,40	0,00	0,00	0,01	0,00	0,00
10	0,00	0,00	0,00	0,00	0,00	0,00	**0,250**	0,03	**0,40**		0,00	0,00	0,25	0,00	0,00
11	0,00		0,00	0,00	**0,94**	0,00	0,00	0,00	0,00	0,00		0,00	0,00	0,00	0,00
12		0,00	**0,07**	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00		0,00	0,00	0,00
13	0,00	0,00	0,00	0,00	0,00	0,00	**0,06**	**0,17**	0,01	**0,25**	0,00	0,00		0,00	0,04
14	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,02	0,00	0,00	0,00	0,00	0,00		0,83
15	0,00	0,00	0,00	0,00	0,00	0,00	0,00	**0,06**	0,00	0,00	0,00	0,00	0,04	**0,83**	

показали, что прогноз работы сердца будет правильным с вероятностью менее $\beta \leq 20\%$. В статистике это уже не прогноз (обычно требуют $\beta \geq 0{,}95$ и выше). Такие выборки описывают прошлое СТТ, но никак не будущее [12–19].

В целом, из-за ЭЕЗ получается, что любая выборка уникальна (статистически не повторима), и поэтому статистика не может дать прогноз будущего для любой биосистемы. Это крушение первой догмы современной науки (в стохастике в частности) в отношении СТТ. Но мы доказали разрушение второй догмы биомедицины. Подчеркнем, обе эти догмы сейчас господствуют во всех науках о живых системах [3–19].

Оказалось, что любая группа испытуемых не может быть однородной. Если взять 15 якобы одинаковых физиологически, психически, физически людей и у них зарегистрировать любой параметр организма, то их объединение (общая выборка) не будет однородной. Приведем один пример из тысяч других (подобных) примеров [12–19].

Если в группе из 15-ти одинаковых по возрасту, полу и т. д. испытуемых у каждого из них зарегистрировать по одной выборке КИ, а затем сравнить их между собой, то опять получим ЭЕЗ. Это означает, что все 15 испытуемых не могут иметь общую генеральную совокупность. Иными словами, по критерию Ньюмана-Кейлса (и др.) две любые выборки статистически не совпадают.

Если все 15 человек имеют свои особые генеральные совокупности, то их невозможно объединять в одну (якобы однородную) группу. Для примера представим табл. 2 в виде матрицы парных сравнений 15-ти выборок (разных людей) КИ по критерию Ньюмана-Кейлса.

Оказалось, что в таких матрицах число k_i пар выборок КИ, для которых этот критерий $p_{ij} \geq 0{,}05$, будет очень мало. Обычно $k_i < 20$. Это означает, что выборки все разные, они не имеют общую генеральную совокупность. Группа не однородна, а с такими группами невозможно работать, статистика это запрещает (неизвестно из-за чего они разнятся).

В итоге, мы имеем не только уникальные выборки для одного испытуемого, но и потерю однородности любых групп. Это касается не только КИ, но и ТМГ, ЭМГ, электроэнцефалограмм (ЭЭГ), электронейрограмм, теппинграмм (ТПГ) и т. д. Любой параметр функций организма человека статистически неустойчив. Любая группа людей не однородна!

Эти две фундаментальные догмы разрушились после доказательства ЭЕЗ. Это требует создания новой науки (как говорил Weaver [2]) для описания биосистем. Мы сейчас создаем такую науку. Это новая теория хаоса-самоорганизации — ТХС [12–19]. В ТХС мы создали новые понятия (псевдоаттрактор, неопределенность 1-го и 2-го типов и т.д.) и новые модели. В ТХС появились и новые инварианты для СТТ (несовпадения по точкам или по функциям f(x)).

Таблица 2

Уровни значимости (P) для попарных сравнений 15-ти выборок параметров КИ группы девушек без нагрузки с помощью непараметрического критерия Ньюмана-Кейлса, число совпадений k =20

	1 R:14 72,3	2 R:26 25,2	3 R:23 58,4	4 R:29 06,9	5 R:17 41,4	6 R:15 91,8	7 R:33 56,5	8 R:12 69,1	9 R:25 39,3	10 R:28 87,7	11 R:12 41,9	12 R:56 4,67	13 R:46 1,79	14 R:34 9,71	15 R:27 65,9
1		0,00	0,00	0,00	**0,57**	**1,00**	0,00	**1,00**	0,00	0,00	**1,00**	0,00	0,00	0,00	0,00
2	0,00		**0,62**	**0,38**	0,00	0,00	0,00	0,00	**1,00**	**0,70**	0,00	0,00	0,00	0,00	**1,00**
3	0,00	**0,62**		0,00	0,00	0,00	0,00	0,00	**1,00**	0,00	0,00	0,00	0,00	0,00	0,00
4	0,00	**0,38**	0,00		0,00	0,00	0,00	0,00	**0,02**	**1,00**	0,00	0,00	0,00	0,00	**1,00**
5	**0,57**	0,00	0,00	0,00		**1,00**	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
6	**1,00**	0,00	0,00	0,00	**1,00**		0,00	**0,09**	0,00	0,00	**0,03**	0,00	0,00	0,00	0,00
7	0,00	0,00	0,00	0,00	0,00	0,00		0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
8	**1,00**	0,00	0,00	0,00	0,00	**0,09**	0,00		0,00	0,00	**1,00**	0,00	0,00	0,00	0,00
9	0,00	**1,00**	**1,00**	0,02	0,00	0,00	0,00	0,00		**0,03**	0,00	0,00	0,00	0,00	**1,00**
10	0,00	**0,70**	0,00	**1,00**	0,00	0,00	0,00	0,00	0,03		0,00	0,00	0,00	0,00	**1,00**
11	**1,00**	0,00	0,00	0,00	0,00	0,03	0,00	**1,00**	0,00	0,00		0,00	0,00	0,00	0,00
12	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00		**1,00**	**1,00**	0,00
13	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	**1,00**		**1,00**	0,00
14	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	**1,00**	**1,00**		0,00
15	0,00	**1,00**	0,00	**1,00**	0,00	0,00	0,00	0,00	**1,00**	**1,00**	0,00	0,00	0,00	0,00	

Обсуждение

Последние 150-200 лет во всех науках о жизни господствуют две фундаментальные догмы. Эти догмы никто даже не пытался проверить. Никто не пытался проверить, совпадают ли статистически две соседние выборки любых параметров функций организма человека на интервалах времени Δt_1 и Δt_2. Оказалось, что частота таких совпадений от 0,05 до 0,3, что намного меньше $\beta \geq 0{,}95$.

Тем более никто не проверял статистическую устойчивость выборок одного испытуемого на многих интервалах $\Delta t_1, \Delta t_2, \ldots, \Delta t_n$. Все ученые мира более 150 лет считали, что такие выборки должны совпадать. Это было огромной ошибкой всей биомедицины, а также физики, биофизики, биокибернетики и математики. В целом, это полностью отрицает редукцию, о которой говорил Гинзбург [1] в трех «великих» проблемах. Это решает сразу все эти 3 великие проблемы для СТТ.

Все эти десятилетия наука была в иллюзии статистической устойчивости выборок. Более того, была и вторая догма статистики: любая группа (якобы физиологически одинаковых) испытуемых будет однородной. Это тоже огромная ошибка современной науки. Любая группа испытуемых не имеет общей генеральной совокупности. Они всегда статистически неоднородны, с такой группой статистика не работает.

В целом, ЭЕЗ разрушил две фундаментальные догмы биомедицины, психологии, экологии и других наук о живых системах. Но лысенковщина, о которой говорил Гинзбург [1], осталась. Особых гонений на ЭЕЗ и ТХС вроде и нет, но упорное замалчивание очевидно. Идет игнорирование ЭЕЗ и ТХС, как это было с гипотезами Бернштейна и Weaver. Игнорируют и высказывание Гинзбурга о невозможности редукции биосистем. Фактически это второй тип тоталитаризма (господство одного мнения).

Крайне необходимы существенные перемены в науке в аспекте признания конца догм о статистической устойчивости выборок параметров функций организма и потере однородностей любых групп. Все эти данные дают толчок для точного описания будущего для СТТ. Мы доказали реальность новых инвариант для СТТ. Показали, что у биосистем нет динамического хаоса Лоренца, т. к. показатели Ляпунова хаотически изменяются, а инвариантность мер (в псевдоаттракторах Еськова) отсутствует [12–19].

Более того, расчет энтропии Шеннона для СТТ ничего не дает. Обычно энтропия Шеннона не изменяется при изменении физического состояния биосистемы. Это специфика СТТ. Все законы физики к ним не могут быть применимы из-за ЭЕЗ, потери однородности групп и отсутствия инвариант (с позиции ТДС и статистики). Нужна новая наука для СТТ, об этом говорил W. Weaver еще в 1948 году. Однако идеи Бернштейна, Weaver и Гинзбурга просто замалчиваются (нет дискуссий в науке), а это уже признак догматизма.

Выводы

В 1999 году В. Л. Гинзбург предупреждал ученых об опасности догматизма и тоталитаризма в науке. Он говорил о лысенковщине, которая проявляется не только в гонении ученых, но и в элементарных замалчиваниях уже полученных результатов или гипотез (это было с генетикой и кибернетикой в СССР).

Так было и с гипотезами Н. А. Бернштейна (о повторении без повторений) и W. Weaver (биосистемы не объект современной науки). Эти гипотезы игнорируют и сейчас. Более того, нет реакции и на статью Гинзбурга о лысенковщине и трех «великих» проблемах физики. За 25 лет никто в науке не пытался обсудить эти проблемы (кроме нашей научной школы).

После открытия ЭЕЗ обрушились две фундаментальные догмы биологии, медицины, психологии, экологии и т. д. Любая выборка параметра функций организма человека уникальна (ее нельзя использовать для прогноза), и любая группа не однородна. Гинзбург говорил о конце редукционизма, но сейчас его тоже игнорируют.

Биосистемы не могут далее описываться в рамках современной ТДС и стохастики. Они требуют создания новой науки. Мы ее называем ТХС. В ней имеются новые понятия, новые модели и новые законы. То, что в современной науке называется покоем, в ТХС это движение. Наоборот, в ТХС изменения, а в современной науке это неизменность и т. д. Нужны новые термины, новые понятия и новые модели для описания и изучения биосистем.

ЛИТЕРАТУРА

1. Гинзбург В. Л. Какие проблемы физики и астрофизики представляются сейчас особенно важными и интересными (тридцать лет спустя, причем уже на пороге XXI века)? *Успехи физических наук*. 1999;169:419–441. DOI: 10.25367/UFNr.0169.199904d.0419.
2. Weaver W. Science and Complexity. *American Scientist*. 1948;36:536–544.
3. Воронюк Т. В., Музиева М. И., Гриценко И. А., Галимзянова А. Д. Стохастический анализ параметров кардиоинтервалов женщин, проживающих в разных климатических условиях. *Сложность. Разум. Постнеклассика*. 2023;2:12–27. DOI: 10.12737/2306-174X-2023-2-16-24.
4. Коннов П. Е., Мельникова Е. Г., Кухарева А. Три парадигмы естествознания. *Сложность. Разум. Постнеклассика*. 2023;2:28–37. DOI: 10.12737/2306-174X-2023-2-25-29.
5. Филатов М. А., Розенберг Г. С., Акопов Г. В. Типы регуляции (управления) в системах природы и общества. *Сложность. Разум. Постнеклассика*. 2023;2:47–55. DOI: 10.12737/2306-174X-2023-2-38-44.
6. Галкин В. А., Еськов В. М., Еськов В. В., Шамов К. А. Потеря эргодичности — фундаментальная математическая проблема всех наук о живых системах. *Сложность. Разум. Постнеклассика*. 2023;2:56–67. DOI: 10.12737/2306-174X-2023-2-45-53.
7. Коннов П. Е., Топазова О. В., Трофимов В. Н., Еськов В. В., Самойленко И. С. Нейросети в идентификации главных клинических признаков при актиническом дерматите. *Вестник новых медицинских технологий*. 2023;2:115–118. DOI: 10.24412/1609-2163-2023-2-115-118.
8. Шакирова Л. С., Кухарева А. Ю., Еськов В. М. Неопределенность первого типа параметров сердечно-сосудистой системы девочек Югры. *Вестник новых медицинских технологий.* 2023;2:111–114. DOI: 10.24412/1609-2163-2023-2-111-114.
9. Еськов В. В., Газя Г. В., Коннов П. Е. Фундаментальные проблемы биокибернетики из-за неустойчивости выборок биосистем. *Успехи кибернетики*. 2022;3:110–122. DOI: 10.51790/2712-9942-2022-3-4-13.
10. Кухарева А. Ю., Еськов В. В., Газя Н. Ф. Гипотеза Эверетта и квантовая теория сознания. *Успехи кибернетики*. 2023;4:65–71. DOI: 10.51790/2712-9942-2023-4-1-09.
11. Филатова О. Е., Филатов М. А., Воронюк Т. В., Музиева М. И. Квантовомеханический подход в электрофизиологии. *Успехи кибернетики*. 2023;4:68–77. DOI: 10.51790/2712-9942-2023-4-2-10.
12. Еськов В. В., Газя Г. В., Кухарева А. Ю. Потеря однородности группы — вторая «великая» проблема биомедицины. *Успехи кибернетики*. 2023;4:78–84. DOI: 10.51790/2712-9942-2023-4-2-11.
13. Газя Г. В., Кухарева А. Ю., Мельникова Е. Г., Газя Н. Ф. Проблема эргодичности — фундаментальная проблема всех наук о живых системах. *Успехи кибернетики*. 2023;4:55–64. DOI: 10.51790/2712-9942-2023-4-3-06.

14. Галкин В. А., Кухарева А., Мельникова Е. Г., Филатов М. А. Реально «великие» проблемы физики живых систем. *Сложность. Разум. Постнеклассика*. 2023;2:80–88. DOI: 10.12737/2306-174X-2023-2-68-77.
15. Кухарева А. Ю., Еськов В. В., Еськов В. М., Воронюк Т. В., Самойленко И. С. Энтропийный подход в биомеханике. *Вестник новых медицинских технологий*. 2023;4:122–126. DOI: 10.24412/1609-2163-2023-4-122-126.
16. Gazya G. V., Eskov V. V., Bashkatova Yu. V., Stratan N. F. Research of the Industrial Electromagnetic Field Influence on Heart State in Oil and Gas Workers of the Russian Federation. *Ecology and Industry of Russia*. 2022;26:55–59.
17. Gazya G. V., Eskov V. V., Filatov M. A. The State of the Cardiovascular System under the Action of Industrial Electromagnetic Fields. *International Journal of Biology and Biomedical Engineering*. 2021;15:249–253. DOI: 10.46300/91011.2021.15.250.
18. Filatova O. E., Maistrenko E. V., Boltaev A. V., Gazya G. V. The Influence of Industrial Electromagnetic Fields on Cardio-Respiratory Systems Dynamics of Oil-Gas Industry Complex Female Workers. *Ecology and Industry of Russia*. 2017;21:46–51.
19. Khadartsev A. A., Eskov V. V., Pyatin V. F., Filatov M. A. The Use of Tremorography for the Assessment of Motor Functions. *Biomedical Engineering*. 2021;54:388-392. DOI: 10.1007/s10527-021-10046-6.
20. Prigogine I. R. The End of Certainty: Time, Chaos, and the New Laws of Nature. 1996.
21. Penrose R. Newton, Quantum Theory and Reality. In: Hawking S. W., Israel W. *300 Years of Gravity*. Cambridge: Cambridge University Press; 1987.
22. Gell-Mann M. Fundamental Sources of Unpredictability. *Complexity*. 1997;3:13–19.

АНАЛИЗ ПРОБЛЕМЫ ЭФФЕКТИВНОСТИ КОНТРОЛЯ ТЕХНИЧЕСКОГО СОСТОЯНИЯ АВТОТРАНСПОРТА НА ПРЕДПРИЯТИЯХ В УСЛОВИЯХ ОГРАНИЧЕННОГО ВРЕМЕНИ

Д. В. Федотов[1,а], **Т. В. Гавриленко**[2,1,б]

[1] *Сургутский государственный университет, г. Сургут, Российская Федерация*

[2] *Сургутский филиал федерального государственного автономного учреждения «Федеральный научный центр Научно-исследовательский институт системных исследований Национального исследовательского центра «Курчатовский институт», г. Сургут, Российская Федерация*

[а] *ORCID: https://orcid.org/0009-0006-0447-5187, ✍ fedotov_dv@edu.surgu.ru*

[б] *ORCID: https://orcid.org/0000-0002-3243-2751, taras.gavrilenko@gmail.com*

Аннотация: в статье рассматриваются современные подходы к контролю технического состояния автотранспортных средств с акцентом на внедрение риск-ориентированных методов. Представлены основные проблемы, связанные с ограничением времени на осмотр, нехваткой данных о предыдущих проверках и координации действий между сотрудниками. Предложены методы оптимизации контроля, включающие использование алгоритмов прогнозирования отказов и автоматизированных систем диагностики. Проведен сравнительный анализ временных затрат и эффективности предложенных решений. Представленные выводы позволяют предприятиям минимизировать затраты на диагностику, повысить безопасность эксплуатации автотранспортных средств и улучшить производственные показатели.

Ключевые слова: риск-ориентированный контроль, автоматизированная диагностика, прогнозирование отказов, машинное обучение, оптимизация диагностики, цифровые технологии в автотранспорте, система технического мониторинга.

Благодарности: работа выполнена в рамках государственного задания НИЦ «Курчатовский институт» — НИИСИ по теме № FNEF-2024-0001 «Создание и реализация доверенных систем искусственного интеллекта, основанных на новых математических и алгоритмических методах, моделях быстрых вычислений, реализуемых на отечественных вычислительных системах» (1023032100070-3-1.2.1).

Для цитирования: Федотов Д. В., Гавриленко Т. В. Анализ проблемы эффективности контроля технического состояния автотранспорта на предприятиях в условиях ограниченного времени. *Успехи кибернетики*. 2025;6(1):116–127.

Поступила в редакцию: 12.01.2025. *В окончательном варианте: 09.03.2025.*

EFFICIENCY OF CONDUCTING TECHNICAL INSPECTIONS ON CORPORATE VEHICLE FLEETS WITHIN LIMITED TIMEFRAMES

D. V. Fedotov[1,a], **T. V. Gavrilenko**[2,1,b]

[1] *Surgut State University, Surgut, Russian Federation*

[2] *Surgut Branch of Scientific Research Institute for System Analysis of the National Research Centre "Kurchatov Institute", Surgut, Russian Federation*

[a] *ORCID: https://orcid.org/0009-0006-0447-5187, ✍ fedotov_dv@edu.surgu.ru*

[b] *ORCID: https://orcid.org/0000-0002-3243-2751, taras.gavrilenko@gmail.com*

Abstract: we reviewed modern approaches to monitoring the technical condition of motor vehicles (MV), focusing on risk-based methods. We identified the main problems related to time constraints for inspections, lack of data on previous inspections, and coordination issues between employees. We proposed methods for optimizing inspections, including the use of failure prediction algorithms and automated diagnostic systems. We conducted a comparative analysis of time costs and the efficiency of the proposed solutions. Our conclusions allow enterprises to minimize diagnostic costs, increase the safety of MV operations, and improve production performance.

Keywords: risk-oriented inspection, automated diagnostics, failure prediction, machine learning, diagnostics optimization, digital technologies in motor transport, condition monitoring system.

Acknowledgements: this study is a part of the FNEF-2024-0001 government order contracted to the Scientific Research Institute for System Analysis of the National Research Centre "Kurchatov Institute", project No. 1023032100070-3-1.2.1 Development and Implementation of Trusted Artificial Intelligence

Systems Based on new Mathematical Methods and Algorithms, Fast Computing Models for Domestic Computing Systems.

Cite this article: Fedotov D. V., Gavrilenko T. V. Efficiency of Conducting Technical Inspections on Corporate Vehicle Fleets within Limited Timeframes. *Russian Journal of Cybernetics*. 2025;6(1):116–127.

Original article submitted: 12.01.2025. *Revision submitted: 09.03.2025.*

Введение

Контроль технического состояния автотранспортных средств (АТС) является важным компонентом обеспечения безопасности дорожного движения и оптимизации производственных процессов транспортных предприятий. Особенно это актуально для компаний, занимающихся интенсивной эксплуатацией автопарков в условиях высоких нормативных требований. Введение современных риск-ориентированных методов контроля и автоматизированных систем позволяет минимизировать риски, связанные с неисправностями транспорта, и улучшить качество диагностики, одновременно снижая временные и трудовые затраты.

Несмотря на развитие технологий, многие предприятия сталкиваются с ограничениями, включая нехватку времени для полноценной проверки, отсутствие централизованной базы данных о проведенных осмотрах, а также низкую координацию между подразделениями. Это приводит к снижению эффективности контроля и повышает вероятность эксплуатационных рисков. Настоящее исследование направлено на выявление ключевых проблем контроля технического состояния АТС и поиск решений на основе риск-ориентированных подходов.

Современные транспортные предприятия, особенно те, которые занимаются грузовыми и пассажирскими перевозками, вынуждены выпускать на линию большое количество автомобилей ежедневно. В таких условиях на проверку одного транспортного средства зачастую выделяется меньше времени, чем необходимо для полноценной диагностики всех его узлов и агрегатов. Это приводит к тому, что проверка осуществляется выборочно или формально, в результате чего мелкие неисправности остаются незамеченными. Такие неисправности могут накапливаться, переходя в разряд серьезных дефектов, способных повлечь за собой поломку на линии или дорожно-транспортное происшествие [5, 9].

Особое значение проблема контроля технического состояния приобретает в условиях эксплуатации автопарков, включающих транспортные средства различных марок, годов выпуска и назначения. Разнообразие условий эксплуатации, от городских дорог с качественным покрытием до сельских трасс с отсутствием асфальта, создает дополнительные сложности в оценке состояния автомобилей. Транспортные средства, работающие в сложных условиях, подвергаются ускоренному износу, что требует более тщательной диагностики, но при этом сокращает доступное время для ее проведения.

Одной из главных проблем остается отсутствие единой системы учета данных о проверках. На многих предприятиях данные хранятся в бумажных журналах или в разрозненных электронных файлах, что затрудняет оперативный доступ к ним. Это приводит к пропуску важных этапов диагностики и увеличивает вероятность выпуска неисправных автомобилей на линию.

Сложившаяся ситуация требует детального анализа и поиска новых решений, способных повысить эффективность контроля технического состояния автотранспорта. Важной задачей является разработка риск-ориентированных методов, которые позволят сосредоточить внимание на наиболее критичных узлах и агрегатах, минимизировав при этом временные затраты на диагностику. Эти методы могут быть дополнены автоматизированными системами, которые помогут оптимизировать процесс проверки, обеспечив доступ к актуальной информации и улучшив координацию между различными участниками процесса.

На протяжении последнего десятилетия проблема обеспечения контроля технического состояния автотранспорта стала одной из ключевых в транспортной отрасли [1, 7]. Проведенные исследования показывают, что в большинстве случаев процесс осмотра ограничивается органолептическими методами, основанными на визуальной оценке и субъективных суждениях проверяющих. Это обусловлено дефицитом времени, человеческих ресурсов и недостатком технической оснащенности. Однако такой подход существенно снижает качество контроля и создает дополнительные риски.

Одной из наиболее острых проблем является несоответствие между объемом работ, необходимым для полной проверки автомобиля, и временем, которое имеется у проверяющего [2, 8]. Например, ежедневная проверка транспортного средства включает более десятка пунктов, начиная от оценки состояния шин и заканчивая проверкой тормозной системы [3, 4]. При этом некоторые операции требуют

значительных временных затрат. Так, проверка тормозной системы может занять до 10–15 минут, тогда как общий лимит времени на проверку одного автомобиля часто не превышает 20 минут. В результате проверяющие вынуждены концентрироваться только на наиболее очевидных и критических элементах, что неизбежно приводит к упущению менее заметных, но важных дефектов.

Кроме того, анализ работы контрольно-технических пунктов (КТП) выявил, что значительная часть времени уходит на поиск и анализ документации. Например, проверка срока годности аптечки, огнетушителя или состояния автошин требует сопоставления данных из разных источников. Отсутствие централизованной базы данных и автоматизированных инструментов приводит к тому, что такие проверки занимают неоправданно много времени и часто проводятся с ошибками. Это особенно актуально для крупных предприятий, где автопарк может насчитывать сотни автомобилей, а данные обновляются ежедневно.

Еще одной важной проблемой является низкая степень координации между сотрудниками, отвечающими за проверку. При сменной работе механиков и контролеров технического состояния информация о выявленных неисправностях передается неполно или теряется. Это приводит к дублированию проверок одних и тех же пунктов и пропуску других, что снижает общую эффективность процесса. В результате некоторые узлы автомобиля могут оставаться без внимания на протяжении недель или даже месяцев, что создает серьезные эксплуатационные риски.

Особое внимание следует уделить разнообразию эксплуатационных условий и их влиянию на состояние автомобилей. Например, автомобили, работающие в тяжелых условиях (поездки по бездорожью, перевозка тяжелых грузов), изнашиваются быстрее, чем машины, работающие в городских условиях. Однако текущая система проверки не учитывает этих различий, предлагая единый подход для всех транспортных средств. Это создает неравномерную нагрузку на проверяющих и приводит к неэффективному использованию ресурсов.

Исследования также показывают, что отсутствие систематического учета истории проверок и степени износа узлов приводит к принятию ошибочных решений. Например, неисправности, выявленные на одном этапе, могут не фиксироваться должным образом, что затрудняет их устранение в будущем. Это приводит к тому, что автомобили продолжают эксплуатироваться с накопившимися дефектами, что увеличивает вероятность выхода из строя.

Основной вклад данного исследования заключается в разработке риск-ориентированного подхода к диагностике технического состояния автотранспорта, адаптированного к ограниченным временным и ресурсным условиям эксплуатации. Предложены алгоритмы оптимизации проверки, которые учитывают индивидуальные особенности эксплуатации транспортных средств, такие как износ узлов и компоненты риска. Новизна работы заключается также в интеграции методов автоматизированного анализа данных с использованием предиктивной аналитики для снижения затрат на диагностику и минимизации эксплуатационных рисков. Впервые проведен системный анализ временных ограничений и их влияния на организацию контроля автопарков в реальных условиях российских предприятий.

Риск-ориентированный подход, используемый в смежных областях (например, в авиации и промышленной диагностике), позволяет сосредоточить усилия на наиболее критичных компонентах [7, 8]. Например, алгоритмы машинного обучения могут использоваться для прогнозирования отказов на основе данных о предыдущих неисправностях, что снижает трудозатраты и увеличивает точность диагностики. Однако их внедрение в автотранспортных системах пока ограничено.

Для повышения точности диагностики рекомендуется использовать предиктивную аналитику, основанную на методах машинного обучения, что позволяет прогнозировать износ узлов и сокращать эксплуатационные риски [11, 12].

Исследование временных затрат на диагностику автопарка из 100 автомобилей показало, что для проведения полной проверки всех узлов одного автомобиля требуется в среднем 75 минут. Это включает осмотр тормозной системы (15 минут), рулевого управления (10 минут), шин (10 минут) и других компонентов. Однако предприятия часто ограничивают время диагностики 20-45 минутами, что приводит к формальному подходу, при котором проверяются только наиболее критичные элементы.

Для количественной оценки был построен график накопительного времени проверки, отражающий зависимость между количеством автомобилей и временем их диагностики. График показал, что при текущих временных ограничениях проверяющий способен провести качественный осмотр только 8-10 автомобилей за смену. Это подчеркивает необходимость внедрения автоматизированных систем,

которые позволят сократить время анализа за счет автоматизации обработки данных.

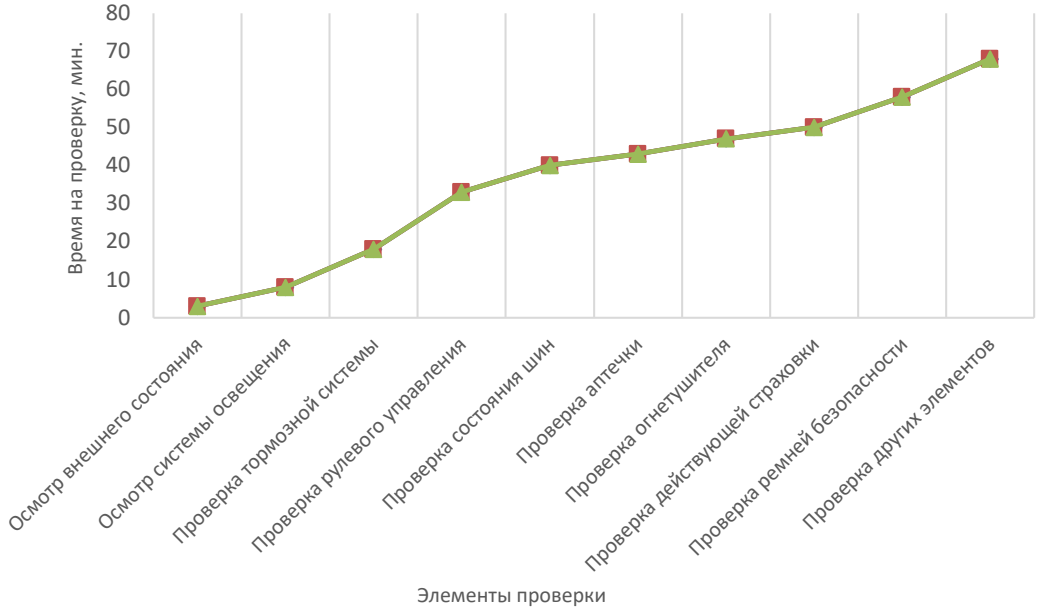


Рис. 1. *График накопительного времени на проверку 1 автомобиля*

График на рисунке 1 демонстрирует накопительное время, необходимое для выполнения последовательной проверки каждого элемента автомобиля. Логика графика такова: при проверке первого элемента затрачивается определенное время, и каждый последующий пункт добавляет к общему времени свои минуты. Таким образом, общий суммарный показатель времени увеличивается, но не отражает, что каждый пункт сам по себе требует больше времени.

Такой подход наглядно показывает, что качественная проверка всех элементов занимает значительное время, которое превышает нормативное окно для проведения технического осмотра на предприятии. При нормированной длительности рабочего дня проверяющие не имеют возможности уделять должное внимание всем узлам, что приводит к формальному подходу к проверке.

Формула для расчета времени проверки:

общее время T_{total} проверки определяется как:

$$T_{total} = \int_{i=1}^{n} T_i,$$

где T_i — время, затраченное на проверку i-го пункта, n — общее количество пунктов проверки.

Допустим, для качественной проверки всех 10 пунктов необходимо следующее время (в минутах): $T_{total} = 5 + 5 + 15 + 20 + 10 + 12 + 8 + 7 + 10 + 10 = 102$ минуты.

То есть полная проверка автомобиля занимает 102 минуты, что более чем в 2 раза превышает типичное нормативное время осмотра (например, 45 минут). Это вынуждает проверяющих сокращать проверку до минимальных пунктов, чаще всего сосредотачиваясь на наиболее критичных элементах, таких как тормозная и рулевая системы.

Пропуск проверок на ранних стадиях или поверхностный подход приводит к тому, что мелкие неисправности остаются незамеченными. Со временем они трансформируются в серьезные поломки, ремонт которых требует значительных затрат времени и средств.

Таким образом, суммарное время проверки существенно превышает возможности, доступные в рамках нормативного рабочего времени, что делает качественный осмотр практически невозможным. Решение этой проблемы заключается в оптимизации процессов проверки, внедрении автоматизированных систем и риск-ориентированных методов, позволяющих перераспределить ресурсы на критичные узлы и уменьшить временные затраты.

Ограниченное время, выделяемое на проверку технического состояния транспортных средств, вынуждает персонал на предприятиях принимать компромиссные решения. В условиях, когда на осмотр одного автомобиля отводится около 20 минут, проверяющие вынуждены концентрироваться

только на наиболее критичных элементах, напрямую влияющих на безопасность эксплуатации. В результате приоритет отдается узлам и агрегатам, которые можно быстро и визуально оценить, таким как тормозная система, рулевое управление и осветительные приборы. Эти элементы проверяются как визуально, так и минимальными инструментальными средствами, что позволяет оперативно вынести решение о возможности выпуска автомобиля на линию. Однако такой подход исключает из поля зрения множество менее очевидных, но также важных компонентов, чья неисправность может проявиться только со временем.

Практика показывает, что в подобных условиях диагностика носит, скорее, формальный характер. Проверка сводится к поверхностному осмотру, когда, например, состояние шин оценивается визуально, без замера глубины протектора, или работа двигателя слушается без нагрузки. Это экономит время, но снижает точность диагностики. Узлы, которые требуют длительного доступа или применения специализированных инструментов, часто остаются без должного внимания. Такие системы, как подвеска или детали подачи топлива, проверяются реже, а иногда и вовсе не осматриваются в рамках ограниченного времени. Это приводит к накоплению мелких неисправностей, которые на начальном этапе могли быть устранены в течение нескольких часов, но в дальнейшем перерастают в серьезные поломки, требующие сложного ремонта.

Пропуск регулярных проверок отдельных узлов порождает еще одну серьезную проблему – неравномерность контроля. Пункты, которые считаются менее важными, такие как аптечки или огнетушители, часто остаются без внимания на протяжении длительного времени. Их проверка, как правило, осуществляется только в случае инспекционных проверок или в критических ситуациях. Такой подход создает дисбаланс в системе контроля, что может привести к неожиданным последствиям на этапе эксплуатации.

Для сокращения времени проверки часто используется визуальный и быстрый подход, основанный на органолептических методах. Осмотр внешнего состояния автомобиля, проверка работоспособности освещения и визуальная оценка состояния шин занимают минимальное время. При этом минимальная инструментальная диагностика, такая как измерение давления в шинах или проверка уровня жидкости, ограничивается только базовыми операциями. Проверка документации также упрощается до сверки сроков годности аптечек и огнетушителей, а также соответствия путевого листа текущему автомобилю. Даже при таком упрощенном подходе, где осмотр внешнего состояния занимает 2 минуты, проверка освещения – 3 минуты, а быстрая сверка документов – 5 минут, общее время проверки достигает 20 минут. Однако этот подход оставляет вне поля зрения множество потенциальных дефектов, которые могут нести существенные риски.

Проблема ускоренного подхода к диагностике заключается в том, что он не позволяет выявить мелкие неисправности, такие как утечка масла или снижение давления в системе охлаждения. Эти дефекты накапливаются, приводя к сложным и дорогостоящим ремонтам, требующим длительного простоя транспортного средства. Кроме того, отсутствие систематизации данных о проверках усугубляет проблему. Если результаты осмотра не фиксируются и не передаются в общий регистр, информация о степени износа или выявленных ранее неисправностях может быть утеряна. Это создает ситуацию, в которой проверка проводится фактически «с нуля», а накопленный износ остается незамеченным.

Еще одной важной проблемой ускоренного осмотра является его зависимость от человеческого фактора. Когда проверяющий ориентируется только на собственный опыт и интуицию, вероятность ошибок и пропуска важных дефектов существенно возрастает. В условиях, где отсутствует автоматизированная система поддержки, эффективность проверки напрямую зависит от компетенций и внимательности конкретного сотрудника, что увеличивает риск критических упущений.

Таким образом, ограниченное время, выделяемое на осмотр, приводит к упрощению диагностики и пропуску потенциально серьезных неисправностей. Для минимизации этих рисков требуется внедрение системного подхода, включающего автоматизацию процессов, использование риск-ориентированных методов и систематизацию данных о состоянии транспортных средств. Только так можно гарантировать как высокую оперативность проверки, так и ее полноту, обеспечивая безопасность и минимизируя затраты на эксплуатацию автопарка.

В современных условиях эксплуатации автопарков одной из важнейших задач является обеспечение регулярной проверки транспортных средств, что напрямую влияет на их безопасность и эксплуатационные характеристики. В данной статье рассматривается процесс оценки времени, необходимого

для проверки автомобилей, а также оптимизация численности контролеров для эффективного обслуживания автопарка.

Процесс проверки одного транспортного средства включает в себя целый ряд процедур, включая осмотр технического состояния, проверку документации и других эксплуатационных характеристик. В среднем, на проверку одного автомобиля требуется 45 минут, что составляет 0,75 часа. Это значение может варьироваться в зависимости от типа транспорта, его состояния и наличия мелких неисправностей, однако для общей оценки мы принимаем указанное время как стандартное.

Для оценки времени, необходимого для проверки всего автопарка, предположим, что автопарк состоит из 100 единиц техники. Тогда общее время, необходимое для проведения проверок, составит:

$$0.75 \times 100 = 75 \text{ часов.}$$

Эти 75 часов могут быть распределены среди нескольких контролеров, что позволяет снизить время, необходимое для выполнения всей работы.

Для расчета числа автомобилей, которые может проверить один контролер за смену, необходимо учитывать продолжительность его рабочего дня. Обычная смена контролера составляет от 8 до 12 часов, что также влияет на общую эффективность работы.

Для смены длительностью 8 часов:

$$\frac{75}{8} = 9.4 \text{ автомобиля.}$$

Для смены длительностью 12 часов:

$$\frac{75}{12} = 6.3 \text{ автомобиля.}$$

Таким образом, за одну смену контролер может проверить от 6 до 10 автомобилей, в зависимости от длительности смены и других факторов, влияющих на процесс проверки.

Важно отметить, что общий процесс выпуска автомобилей из автопарка, включая как технические проверки, так и устранение мелких неисправностей, обычно не превышает 2-3 часов. В этом контексте ключевым является обеспечение оперативности выполнения проверок, что требует оптимизации численности контролеров и времени, затрачиваемого на каждый осмотр.

Для того чтобы эффективно покрыть весь автопарк за один рабочий день, необходимо обеспечить достаточное количество контролеров. В случае автопарка из 100 автомобилей потребуется:

$$\frac{100}{8} = 13 \text{ контролеров.}$$

Однако в реальных условиях даже при наличии необходимого числа персонала могут возникнуть задержки, связанные с необходимостью поиска документации, устранения неисправностей или выполнения других сопутствующих операций.

В условиях ограниченного рабочего времени важно также учитывать, сколько автомобилей контролер способен проверить за сутки. Например, при рабочем времени в 24 часа и стандартном времени проверки 0,75 часа на автомобиль, один контролер сможет проверить:

$$\mathrm{N} = \frac{\mathrm{T}_{total}}{\mathrm{T}_{single}} = \frac{24}{0.75} = 32,$$

где:

T_{total} — общее рабочее время;

T_{single} — время на проверку одного автомобиля.

Если время на проверку сокращается до 20 минут на автомобиль, то количество проверенных автомобилей за сутки увеличивается:

$$\mathrm{N} = \frac{\mathrm{T}_{total}}{\mathrm{T}_{single}} = \frac{24}{0.333} = 72.$$

Однако стоит отметить, что в большинстве случаев время на выпуск ограничено несколькими часами, что также накладывает ограничения на количество автомобилей, которые могут быть выведены на линию в течение дня. Например, за 3 часа контролер может выпустить на линию только 9 автомобилей:

$$N=\frac{T_{total}}{T_{single}} = \frac{3}{0.333} = 9.$$

Одной из ключевых проблем в процессе проверки автопарков является отсутствие необходимой информации о предыдущих проверках. В случае отсутствия данных о том, какие автомобили уже были проверены, контролер сталкивается с трудностью в принятии решения о приоритетности проверки. Документация, как правило, хранится в удаленных местах, что затрудняет доступ к ней в процессе проведения осмотра.

В реальных условиях работы даже при достаточном числе контролеров возможны задержки, вызванные необходимостью поиска документации, устранения мелких неисправностей и выполнения других сопутствующих операций. Оптимизация рабочего процесса требует учета всех этих факторов, а также внедрения эффективных систем управления и учета, которые позволят ускорить процесс проверки и выпуска автомобилей из автопарка.

Для покрытия автопарка за один рабочий день потребуется достаточное количество квалифицированных диагностов, а также четкая организация рабочего процесса, направленная на минимизацию задержек и ускорение процедуры проверки и выпуска транспортных средств.

В системах, где контроль множества объектов осуществляется с использованием чек-листов, часто возникает конфликт между временными ограничениями и необходимостью соблюдения регламента. Автотранспортные системы лишь один из примеров подобных процессов. Временные рамки, установленные для проведения проверки, редко позволяют проводить тщательный осмотр всех элементов. Это порождает дилемму: либо проверять только критически важные аспекты, жертвуя качеством диагностики, либо увеличивать затраты на персонал и время.

Проблема временных ограничений не уникальна для автотранспорта. В других отраслях также приходится сталкиваться с проверкой множества компонентов с набором характеристик, которые требуют регулярного контроля. Например:

- в авиационной индустрии регламент проверки самолетов часто расходится с доступным рабочим временем технического персонала;
- в медицинской практике тщательная проверка оборудования в условиях высокой загруженности ведет к упрощению протоколов диагностики;
- в промышленности контроль состояния оборудования по чек-листам часто затрудняется из-за необходимости обеспечивать бесперебойность производственных процессов.

Во всех этих случаях приходится искать компромисс между тщательностью и объемом выполняемых работ.

Для визуализации проблемы можно построить график, где ось времени отображает процесс проверки объектов, а «красная линия» обозначает регламентное время, доступное для выполнения задачи. Пересечение этой линии с кумулятивным временем проверки показывает моменты, когда качество проверки начинает страдать из-за нехватки ресурсов.

Создадим иллюстрацию для примера с автотранспортной системой:

- ось X: количество проверяемых автомобилей;
- ось Y: время на проверку;
- линия дозволенного: фиксированное время, выделяемое на проверку всех автомобилей.

Формализация проблемы

Пусть:

- n — количество проверяемых объектов;
- T_i — время на проверку i-го объекта;
- T_{total} — общее регламентное время, доступное для проверки.

Общее время проверки:

$$T_{total}=\int_{i=1}^{n} T_i.$$

Если $T_{total} > T$, то качество проверки автоматически снижается, так как часть объектов будет осмотрена поверхностно.

Оптимизация трудозатрат

Для обеспечения контроля всех объектов в рамках временных ограничений потребовалось бы увеличить численность проверяющего персонала до значения:

$$N = \frac{\sum_{i=1}^{n} t_i}{T_{shift}},$$

где T_{shift} – продолжительность смены одного проверяющего. Это приводит к парадоксальной ситуации: количество проверяющих стремится к числу объектов.

В реальных условиях, чтобы уложиться в регламентное время, проверяющие вынуждены выбирать только наиболее критичные пункты чек-листа. Однако без систематизации данных о предыдущих проверках и автоматизированных подсказок это приводит к потере информации о менее приоритетных, но потенциально важных аспектах.

Таким образом, систематизация данных, автоматизация процессов и внедрение риск-ориентированного подхода могут значительно повысить эффективность контроля в условиях ограниченного времени.

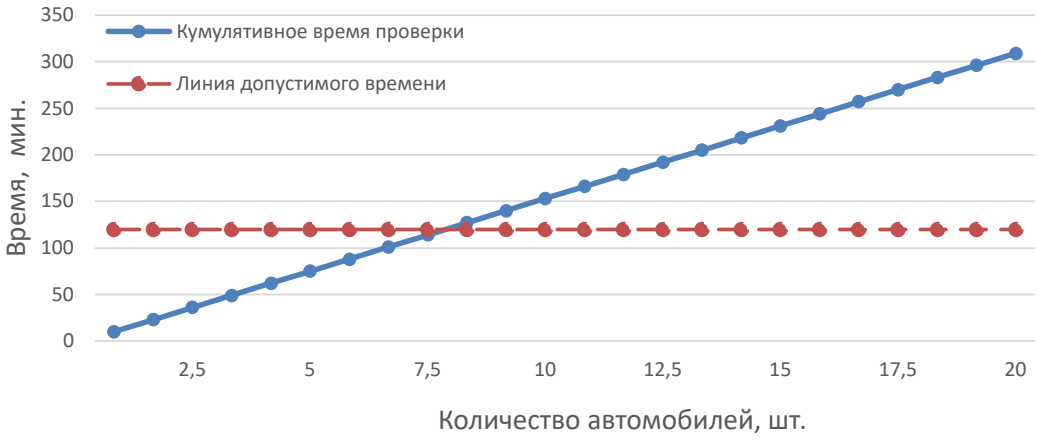


Рис. 2. *Графическое представление конфликта между временем на проверку и регламентными ограничениями*

На рисунке 2 мы видим график — визуализацию проблемы, возникающей в процессе проверки технического состояния множества автомобилей при ограниченном времени. График отражает соотношение между числом проверяемых объектов (автомобилей) и суммарным временем, необходимым для их осмотра, в сравнении с установленным временным регламентом.

На графике показано, как увеличивается общее время проверки автомобилей по мере роста их количества. Линия имеет линейный характер, что объясняется постоянным временем, затрачиваемым на проверку одного автомобиля. В нашем примере для осмотра каждого автомобиля требуется 15 минут. Таким образом, проверка 10 автомобилей занимает 150 минут, а 20 автомобилей – уже 300 минут. Линейная зависимость кумулятивного времени демонстрирует, что суммарные затраты времени напрямую зависят от числа объектов контроля.

Красная пунктирная линия — «Линия допустимого времени» представляет собой установленный регламент времени, отведенного на проверку всего автопарка. В данном случае он равен 120 минутам (например, 2 часа рабочего времени). Линия остается неизменной, так как выделенное время фиксировано и не зависит от числа проверяемых объектов.

Ключевой элемент графика — это точка пересечения синей и красной линий. Она обозначает критический момент, начиная с которого кумулятивное время проверки превышает установленный временной лимит. В данном случае это происходит после проверки 8 автомобилей. После этой точки становится очевидным, что при текущем регламенте времени невозможно провести осмотр всех элементов автомобилей без снижения качества проверки.

Область выше красной линии — это «зона недопустимости», где общее время проверки превышает регламентное. Автомобили, попадающие в эту зону, фактически не могут быть проверены в рамках установленного времени, что создает риск их выхода на линию без полного осмотра или с проверкой только наиболее критичных элементов.

Маркеры на синей линии выделяют точки, соответствующие числу автомобилей, которые не могут быть проверены в регламентное время. Это автомобили, проверка которых занимает больше времени, чем доступно по установленному лимиту. Начиная с 9-го автомобиля, все проверки будут выходить за пределы допустимого времени.

Анализ и интерпретация

График четко демонстрирует проблему несоответствия между временем, необходимым для проверки автопарка, и доступным рабочим временем. При текущем регламенте (2 часа) максимально возможно качественно проверить только 8 автомобилей. При этом любой дополнительный автомобиль, осмотр которого выходит за пределы регламентного времени, либо не будет проверен, либо будет осмотрен поверхностно, что снижает качество контроля.

Аналогичные проблемы могут возникать в других системах, где есть множество проверяемых объектов и строгие временные рамки. Для их решения требуется оптимизация процесса проверки, например, через автоматизацию или перераспределение приоритетов.

Данный график иллюстрирует необходимость пересмотра временного регламента или применения риск-ориентированного подхода, при котором внимание уделяется наиболее критичным элементам. Он также подчеркивает важность внедрения автоматизированных систем учета и планирования для повышения эффективности контроля в условиях дефицита времени.

Проведенное исследование демонстрирует ключевые проблемы, связанные с контролем технического состояния автотранспортных средств в условиях ограниченного времени и ресурсов. Анализ показал, что для качественной проверки всех элементов одного автомобиля требуется значительно больше времени, чем предусматривают существующие нормативы. Это приводит к необходимости выбора: либо ограничиваться поверхностной проверкой наиболее критичных узлов, либо увеличивать трудозатраты и персонал, что не всегда возможно.

Проведенный анализ и предложенные подходы демонстрируют, что переход к автоматизированным и риск-ориентированным методам контроля не только снизит временные и трудовые затраты, но и повысит общую безопасность эксплуатации транспортных средств.

Выводы

Можно провести экспертный анализ, в ходе которого выделить наиболее критичные компоненты, влияющие на безопасность и исправность автомобиля в целом.

На примере одного автотранспортного предприятия был проведен экспертный анализ проверяемого перед выпуском на линию перечня неисправностей, с которыми запрещена эксплуатация, взятого из протокола в соответствии с ГОСТ. Исследуемые (проверяемые) диагностические показатели были ранжированы по классу опасности, в том числе определено максимальное и минимальное время на проверку.

Проведенное исследование позволило выявить наиболее приоритетные и менее важные пункты. Исследование позволяет перераспределить пункты для проверки в соответствии с приоритетом.

В соответствии с существующими нормативными документами проверка автошин осуществляется инструментально при каждом выезде на линию, но периодичность исследования данного параметра можно изменить в зависимости от результатов предыдущих измерений. Если остаточный размер протектора существенно превосходит нормативный, то следующее измерение может быть отложено на более долгий срок. Когда остаток протектора приближается к минимальному значению, мы можем задавать периодичность проверки чаще. Таким образом будет освобождаться время для проверки других пунктов.

Есть параметры, у которых при известных сроках проверки, эксплуатации можно изменить график проверки (огнетушитель, аптечка).

В результате может быть получена методика изменения периодичности тех или иных параметров для сокращения рисков при соблюдении необходимых нормативных актов.

Таблица

Приоритетные компоненты и уровни вероятностей отказа

№ п/п	Категории поверок	Приоритет	Класс опасности	Опасность, %	Время MIN	Время MAX	Время среднее	Уровень опасности
1	Тормозная система	1	6	6,59	2	3	2,5	1
2	Рулевое управление	1	6	6,59	2	3	2,5	1
3	Звуковой сигнал	3	4	4,40	0,1	0,2	0,15	1
4	Внешние световые приборы и световозвращатели	3	4	4,40	0,1	0,5	0,3	1
5	Стеклоомыватели	3	4	4,40	0,2	0,5	0,35	1
6	Стеклоочистители	3	4	4,40	0,3	0,5	0,4	1
7	Тахограф	3	4	4,40	0,3	0,7	0,5	1
8	Аппаратура спутниковой навигации	3	4	4,40	0,3	0,7	0,5	1
9	Устройства (системы) вызова экстренных оперативных служб	3	4	4,40	0,3	0,7	0,5	1
10	Приказы, распоряжения	2	5	5,49	0,1	10	5,05	2
11	Герметичность систем, узлов и агрегатов транспортного средства, в том числе системы выпуска отработавших газов, а также дополнительно устанавливаемых на транспортное средство гидравлических устройств	4	3	3,30	0,5	3	1,75	2
12	Устройства фиксации подушки и спинки водительского сиденья	5	2	2,20	1	2	1,5	2
13	Исправность замков дверей кузова или кабины, запоров бортов грузовой платформы, запоров горловин цистерн и пробок топливных баков (кроме транспортных средств городского наземного электрического транспорта)	5	2	2,20	1	3	2	2
14	Тягово-сцепное устройство, а также страховочные тросы (цепи)	5	2	2,20	0,5	1	0,75	2
15	Держатель запасного колеса	5	2	2,20	0,1	0,3	0,2	2
16	Крепление колес, состояние шин	5	2	2,20	1	1,3	1,15	2
17	Наружные зеркала и государственные регистрационные знаки	5	2	2,20	1	2	1,5	2
18	Устройства обогрева и обдува	6	1	1,10	0,5	2	1,25	2
19	Грузовая платформа, навесное оборудование	6	1	1,10	0,5	2	1,25	2
20	Отсутствие внесенных в конструкцию транспортного средства изменений в нарушение установленного порядка, предусмотренного разделом 4 главы V Технического регламента, в части газобаллонного оборудования, кузовов транспортных средств, бортов грузовой платформы (для грузовых транспортных средств), дополнительных сидений, фар ближнего света, противотуманных фар (при наличии) и дневных ходовых огней (при наличии)	6	1	1,10	0,5	3	1,75	2
21	Отсутствие установленных на передней части транспортного средства световых приборов с огнями красного цвета или световозвращающих приспособлений красного цвета, на транспортном средстве устройств для подачи специальных световых или звуковых сигналов (за исключением охранной сигнализации) без соответствующего разрешения, а также незаконно нанесенных на наружные поверхности транспортного средства специальных цветографических схем автомобилей оперативных	6	1	1,10	0,5	3	1,75	2
22	При наличии конструктивных особенностей специальных и специализированных транспортных средств контролер должен проверить состояние и исправность других систем, механизмов, агрегатов, узлов и деталей, характерных для данных типов транспортных средств, влияющих на безопасность движения транспортных средств, на соответствие требованиям приложений 6 и 8 к Техническому регламенту	6	1	1,10	0,5	3	1,75	2
23	Внешний вид	6	1	1,10	1	3	2	2
24	Водительское удостоверение	2	5	5,49	0,5	2	1,25	3
25	Страховка	2	5	5,49	1	2	1,5	3
26	Колеса	2	5	5,49	2	3	2,5	3
27	Шины	2	5	5,49	2	3	2,5	3
28	Медицинская аптечка	5	2	2,20	2	4	3	3
29	Огнетушитель	5	2	2,20	2	4	3	3
30	Противооткатные упоры (для грузовых транспортных средств и автобусов)	6	1	1,10	2	4	3	3
Итого				**100,00**	**25,80**	**70,40**	**48,10**	

В таблице представлена подробная классификация неисправностей автотранспортных средств, которая включает данные о приоритетности, классе опасности, а также временных характеристиках устранения выявленных дефектов. Эти данные дают возможность оценить как критичность каждой неисправности, так и временные, и ресурсные затраты, необходимые для их устранения. Такой подход позволяет предприятиям анализировать текущую эффективность контроля технического состояния автотранспорта и выявлять узкие места в процессах диагностики и ремонта.

В таблице выделено несколько категорий поверок, включая поверки тормозной системы, рулевого управления, шин, световых приборов, аптечки и прочих элементов, имеющих ключевое значение для безопасной и стабильной эксплуатации транспортных средств. Например, тормозная система и рулевое управление занимают особое место, поскольку их неисправности имеют не только высокий приоритет (1), но и максимальный класс опасности (6), что указывает на их критическую важность для предотвращения аварий и обеспечения безопасности на дорогах. Класс опасности в процентах подтверждает это: тормозная система составляет 6,59% от общего уровня опасности, что делает ее главной целью профилактических осмотров и ремонтов.

В случае выявления неисправности в процессе проверки, по нормативам может быть определено время, необходимое для устранения неисправностей. Показатели нормативов времени, требуемого для устранения неисправностей, дополнительно подчеркивают сложность организации процесса контроля. Например, минимальное время устранения неисправностей тормозной системы составляет 2 часа, а максимальное – 3 часа, что отражает значительную временную нагрузку даже на базовые процедуры ремонта. В то же время для менее критичных элементов, таких как аптечки, временные затраты значительно меньше (от 0,3 до 0,7 часов).

Суммарное время на устранение всех неисправностей, представленное в итоговой строке таблицы, составляет 70,4 часа, что существенно превышает нормативные ограничения рабочего времени для одной смены.

Эти данные демонстрируют важность риск-ориентированного подхода, который позволяет перераспределить усилия на диагностику наиболее критичных узлов и агрегатов, минимизируя временные затраты на менее важные элементы. Например, в условиях ограниченного времени проверки фокусировка на элементах с наибольшим приоритетом, таких как тормоза и рулевое управление, позволит снизить риск отказов транспортных средств, однако при этом менее значительные неисправности могут остаться без внимания, что создает долгосрочные риски.

Таблица также подчеркивает проблему координации и систематизации данных. В условиях, когда информация о проверках и ремонтах фиксируется вручную или в разрозненных документах, анализ подобных данных становится сложным и ресурсоемким. В статье делается акцент на необходимости внедрения автоматизированных систем, которые могут оперативно обрабатывать такие таблицы, идентифицировать наиболее уязвимые элементы и формировать оптимальные планы ремонта. Особое значение приобретают автоматизированные системы, которые позволяют оперативно обрабатывать данные о состоянии автомобилей, минимизируя влияние человеческого фактора [6, 10]. Это особенно важно, поскольку отсутствие единого подхода к обработке данных о неисправностях увеличивает вероятность человеческих ошибок и снижает общую эффективность процессов контроля.

Таким образом, таблица не только иллюстрирует текущие проблемы организации контроля технического состояния автотранспорта, но и акцентирует внимание на ключевых направлениях их решения. Риск-ориентированный подход, предложенный в статье, может быть эффективным инструментом для интерпретации и использования подобных данных. Это позволит сосредоточиться на наиболее критичных неисправностях, сократить временные затраты на диагностику и повысить безопасность эксплуатации транспортных средств, что является основной целью данного исследования.

ЛИТЕРАТУРА

1. Aven T. On Some Foundational Issues Concerning the Relationship between Risk and Resilience. *Risk Analysis*. 2022;42:2062–2074. DOI: 10.1111/risa.13848.
2. Hore U. W., Wakde D. G. Intelligent Predictive Maintenance for Industrial Internet of Things (IIoT) Using Machine Learning Approach. *Intelligent Cyber Physical Systems and Internet of Things*. Editors: Hemanth J., Pelusi D., Chen J. Cham: Springer; 2023. DOI: 10.1007/978-3-031-18497-0_65.

3. Лушников А. А. Управление ремонтами в транспортных подразделениях ОАО «Сургутнефтегаз». *Вестник СурГУ*. 2014;2:106-111. Режим доступа: https://cyberleninka.ru/article/n/upravlenie-remontami-v-transportnyh-podrazdeleniyah-oao-surgutneftegaz.
4. Комова Ю., Хожай В., Соколова Т. Риск-ориентированный подход в управлении качеством. *Сборник научных трудов Уральского федерального университета*. 2021;31:202–210.
5. Шилкина А. Т., Варакина О. Е. Тенденции развития риск-ориентированного подхода в контексте индустрии 4.0. *Научно-технические ведомости СПбГПУ. Экономические науки*. 2019;12(1):9–20.
6. Макаренко Е. Н., Булгаков С. А. Исследование риск-ориентированного подхода и возможностей его применения во внутреннем контроле. *Учет и статистика*. 2021;2:19–28.
7. Anderson J., Lewis S. Decision Support Systems for Technical Diagnostics in Aerospace Engineering. *Aerospace Science and Technology*. 2019;92:292–303. DOI: 10.1016/j.ast.2019.04.010.
8. Campbell N., Gray J. Evaluation of Technical Conditions Using Advanced Diagnostic Tools. *Engineering Analysis with Boundary Elements*. 2019;109:98–109. DOI: 10.1016/j.enganabound.2019.05.006.
9. Fernandez R., Lopez M. Risk-Based Optimization of Technical Systems for Improved Reliability. *Reliability Engineering and System Safety*. 2019;124:256–270. DOI: 10.1016/j.ress.2019.01.014.
10. Taylor M. J., Wilson L. Application of Machine Learning in Risk-Based Inspection Systems. *Reliability Engineering and System Safety*. 2018;105:467–480. DOI: 10.1016/j.ress.2018.02.011.
11. Thomas L., Richardson S. Machine Learning Techniques for Predictive Maintenance. *Procedia Computer Science*. 2019;151:709–717. DOI: 10.1016/j.procs.2019.04.095.
12. Turner D., Hall M. Use of Artificial Intelligence in Predictive Maintenance Systems. *Applied Soft Computing*. 2019;85:105790. DOI: 10.1016/j.asoc.2019.105790.

МОДЕЛЬ ВЗАИМОДЕЙСТВИЯ И КОНКУРЕНЦИИ МЕЖДУ ДВУМЯ ВИДАМИ АВТОНОМНЫХ АГЕНТОВ

В. Г. Редько
Федеральное государственное автономное учреждение «Федеральный научный центр Научно-исследовательский институт системных исследований Национального исследовательского центра «Курчатовский институт», г. Москва, Российская Федерация
ORCID: http://orcid.org/0000-0002-3748-2379, ✍ vgredko@gmail.com

Аннотация: в этой работе строится и исследуется модель взаимодействия между двумя видами автономных агентов. В модели предполагается, что в клеточном мире «живут» два вида агентов: сильные агенты-хищники и относительно слабые мирные агенты. Мирные агенты в обычной жизни не нападают на других агентов, но они способны коллективно противодействовать отдельным хищникам. Мирные агенты занимают определенный участок освоенной ими территории, территории своего коллектива. Каждый агент имеет свой ресурс. Имеются порции пищи мирных агентов. Съедая порцию пищи, мирный агент пополняет свой ресурс. Агенты-хищники могут нападать на мирных агентов, убивать и съедать их. Если агент-хищник убивает и съедает мирного агента, то ресурс хищника значительно увеличивается. Мирные агенты могут большим коллективом одновременно угрожать агенту-хищнику, который пришел на участок коллектива мирных агентов. Для согласованности действий мирных агентов эти агенты выбирают в своем коллективе агента-лидера, который подает команды всем мирным агентам собраться в определенном месте, а затем коллективно угрожать агенту-хищнику, вытесняя его с участка мирных агентов. В настоящей работе продемонстрирована способность большого коллектива относительно слабых мирных агентов противостоять сильным агентам-хищникам.

Ключевые слова: взаимодействия между агентами, мирные агенты, агенты-хищники, угрожать агентам, уходить от угрожающих агентов.

Благодарности: работа выполнена в рамках государственного задания НИЦ «Курчатовский институт» — НИИСИ по теме № FNEF-2024-0001 «Создание и реализация доверенных систем искусственного интеллекта, основанных на новых математических и алгоритмических методах, моделях быстрых вычислений, реализуемых на отечественных вычислительных системах» (1023032100070-3-1.2.1). Автор благодарен Б. В. Крыжановскому за плодотворные дискуссии.

Для цитирования: Редько В. Г. Модель взаимодействия и конкуренции между двумя видами автономных агентов. *Успехи кибернетики*. 2025;6(1):128–136.

Поступила в редакцию: 08.11.2024. *В окончательном варианте: 11.12.2024.*

A MODEL OF INTERACTION AND COMPETITION BETWEEN TWO TYPES OF AUTONOMOUS AGENTS

V. G. Redko
Scientific Research Institute for System Analysis of the National Research Centre "Kurchatov Institute", Moscow, Russian Federation
ORCID: http://orcid.org/0000-0002-3748-2379, ✍ vgredko@gmail.com

Abstract: this work constructs and studies a model of interaction between two types of autonomous agents. The model assumes that strong, predatory agents and relatively weak defensive agents coexist in a cellular world. Defensive agents typically do not attack others but can collectively resist a predator. They occupy a specific area of the world known as the "defensive agent domain". Each agent has its resources, and defensive agents can replenish their resources by eating portions of food. Predator agents can attack, kill, and eat defensive agents, significantly increasing their resources when they succeed. A large group of defensive agents can pose a threat to a predator that enters their area. To coordinate their actions, defensive agents elect a leader. Upon encountering a predator, the leader commands the defensive agents to gather in a specific location and then collectively threaten the predator, driving it out of their area. This work demonstrates that a large group of relatively weak defensive agents can successfully resist strong predator agents.

Keywords: interactions between agents, defensive agents, predatory agents, threatening agents, escaping from threatening agents.

Acknowledgements: this study is a part of the government contract No. FNEF-2024-0001 Development and Implementation of Trusted AI Systems Using New Mathematical and Algorithmic Methods; Fast Computing Models on Domestic Hardware with the Kurchatov Institute (1023032100070-3-1.2.1). The author thanks B. V. Kryzhanovsky for helpful discussions.

Cite this article: Redko V. G. A Model of Interaction and Competition Between Two Types of Autonomous Agents. *Russian Journal of Cybernetics*. 2025;6(1):128–136.

Original article submitted: 08.11.2024. *Revision submitted: 11.12.2024.*

Введение

Модели взаимодействия и конкуренции между автономными агентами начали активно изучаться в начале 1990-х годов, когда возникло направление исследований «Искусственная жизнь» (английское название Artificial Life или ALife) [1, 2]. В этом направлении интенсивно развивались компьютерные модели автономных агентов. Охарактеризуем некоторые модели, в которых рассматривалось конкурентное взаимодействие между агентами.

В работе [3] была исследована многоагентная компьютерная модель PolyWorld, в которой рассматривалась эволюция популяции достаточно сложных агентов, имеющих нейронные сети в качестве систем внутреннего управления поведением агентов. Агенты жили в двумерной клеточной среде, в части клеток была пища агентов. Агенты выполняли следующие действия: питаться, двигаться, бороться, скрещиваться. В этой модели рассматривались агенты-хищники, которые могли нападать на других агентов, отнимать у них энергетический ресурс и убивать агентов, причем каркас умерших агентов превращался в пищу для других агентов. В процессе эволюции в модели PolyWorld формировались определенные стратегии поведения. Одну из стратегий можно условно назвать «ленивый каннибал»: агент крутится на месте, скрещиваясь или вступая в борьбу с каждым, кто приблизится (поедая каркас соперника в случае победы или погибая в случае поражения). Модель PolyWorld изучалась в ряде дальнейших работ, в которых детально анализировались структуры нейронных сетей рассматриваемых эволюционирующих агентов (см., например, [4]).

В работе [5] изучалась эволюция популяции агентов в двумерной клеточной среде, в которой были агенты-хищники с фиксированными свойствами, которые могли нападать на основных агентов, менявшихся в процессе обучения и эволюции. Основное внимание в [5] уделялось анализу взаимоотношения между обучением и эволюцией.

Модель Echo [6, 7] описывает эволюцию простых агентов, которые взаимодействуют между собой путем скрещивания, борьбы и торговли. Взаимодействия между агентами в этой модели приводило к формированию различных экологических систем: «войны миров», симбиозов и т.п.

В работе [8] изучалось поведение агентов с несколькими потребностями, в том числе с потребностью защиты от хищников.

В настоящей работе строится и исследуется модель конкурентного взаимодействия между двумя видами автономных агентов. Рассматриваются два вида агентов: сильные агенты-хищники и относительно слабые мирные агенты. Мирные агенты в обычной жизни не нападают на других агентов, но они способны коллективно противодействовать отдельным хищникам. Предшественником настоящей модели является работа [9], в которой исследовалась довольно сложная модель эволюционной самоорганизации и процессов формирования видов в популяции агентов. В частности, в некоторых случаях наблюдалась борьба коллектива сравнительно слабых агентов с более сильным агентом. В настоящей работе свойства такой борьбы анализируются в более четком и явном виде. По сравнению с охарактеризованными выше работами мы сосредоточимся на свойствах борьбы мирных агентов с сильными агентами-хищниками.

Описание модели

Рассматриваем одномерный клеточный мир, в котором находятся агенты. Число клеток в мире равно N. Имеется два вида агентов: 1) мирные агенты и 2) агенты-хищники. В каждой клетке мира могут находиться несколько агентов. Время t дискретно: $t = 1,2,\ldots$ Каждый агент имеет свой ресурс, ресурсы мирных агентов и агентов-хищников обозначаем $r(t)$ и $R(t)$, соответственно.

В клетках мира имеются порции пищи мирных агентов. При этом в части мира плотность пищи выше, эта часть мира предпочтительна для мирных агентов. Считаем, что предпочтительная для мирных агентов часть мира занимает половину клеток мира, это клетки с номерами $1,2,\ldots,N/2$. Эту

предпочтительную для мирных агентов часть мира будем называть участок мирных агентов. Число клеток с порцией пищи в предпочтительной части мира равно M_1, в остальной части мира число клеток с порцией пищи равно $M_2 (M_2 < M_1 < N/2)$. В начальный момент времени ($t = 1$) порции пищи распределены по клеткам случайно. При $t = 1$ все мирные агенты находятся в случайных клетках своего участка мира.

Опишем действия агентов.

Мирные агенты могут выполнять следующие действия: 1) питаться, 2) перемещаться по своему участку мира, 3) угрожать агенту другого вида, если этот чужой агент пришел на участок мирных агентов, 4) нападать на агента другого вида. При питании агента ресурс этого агента пополняется. При выполнении других действий ресурс агента понижается.

Действие питание состоит в следующем. Если в той клетке, в которой находится мирный агент, имеется порция пищи, то агент может съесть эту порцию. При этом он съедает всю порцию пищи сразу. При съедании агентом пищи в одной клетке, новая порция пищи появляется в случайной свободной от еды клетке в той части мира, в которой находится агент. То есть, число порций пищи в каждой части мира постоянно.

Действие перемещение состоит в передвижении на соседнюю клетку в случайном направлении. Если мирный агент видит агента-хищника, то он может либо убегать от хищника (т.е. перемещаться в сторону, противоположную направлению на хищника), либо угрожать агенту-хищнику или даже нападать на хищника. Понятно, что нападение на хищника имеет смысл, когда мирные слабые агенты нападают на агента-хищника одновременно достаточно большим коллективом.

Агент-хищник может выполнять следующие действия: 1) перемещаться на соседние клетки, 2) нападать на мирного агента, 3) убивать и съедать мирного агента, 4) уходить от коллектива нападающих мирных агентов.

Прирост ресурса мирного агента при съедании одной порции пищи равен Δr_1. Расходы ресурса мирного агента на перемещение на одну клетку, на угрозу агенту-хищнику, на удар по агенту-хищнику равны Δr_2, Δr_3, Δr_4, соответственно. Считаем, что $\Delta r_2 < \Delta r_3 < \Delta r_4$.

Прирост ресурса агента-хищника при съедании убитого мирного агента равен ΔR_1. Расходы ресурса агента-хищника на перемещение на одну клетку, на удар по мирному агенту, на убийство мирного агента равны ΔR_2, ΔR_3, ΔR_4, соответственно. Считаем, что $\Delta R_2 < \Delta R_3 < \Delta R_4$.

Считаем, что каждый агент видит других агентов во всех клетках мира. Также считаем, что скорость перемещения агентов-хищников существенно выше скорости перемещения мирных агентов.

Охарактеризуем детали взаимодействий между агентами. Рассматриваем два режима взаимодействия 1) обычный режим, когда мирные агенты не угрожают хищникам и не нападают на хищников, 2) режим с угрозами, когда мирные агенты угрожают одному или нескольким агентам-хищникам, а могут и нападать на хищников.

Сначала рассмотрим обычный режим. Когда агент-хищник видит мирного агента в соседней клетке, то хищник заходит в эту клетку и убивает его (за счет разницы скоростей перемещения мирный агент не успевает убежать от хищника). Если хищник видит нескольких мирных агентов в соседних клетках, то он выбирает в качестве жертвы одного агента случайным образом. Если мирный агент видит хищников (на расстоянии больше одной клетки), то он перемещается в сторону, противоположную направлению на ближайшего хищника.

Теперь рассмотрим режим с угрозами. Когда агент-хищник (или несколько агентов-хищников) приходит на участок мирных агентов (или находится на этом участке), то мирные агенты коллективно угрожают хищнику и стремятся вытеснить хищника со своего участка. То есть, каждый из мирных агентов приближается к хищнику и выполняет действие «угрожать». Если число угрожающих велико, то обычно хищник постепенно перемещается в сторону, противоположную угрожающим, и, в конце концов, уходит с участка мирных агентов.

Возможно, что хищник сочтет, что его силы достаточно велики и остановится в клетке на участке мирных агентов (например, это может случиться, если ресурс хищника достаточно велик). Тогда мирные агенты приблизятся к хищнику и из соседней с ним клетки будут наносить удары по агенту-хищнику. Но агент-хищник может в ответ наносить удары по мирным агентам. Считаем, что сила удара пропорциональна ресурсу ударяющего агента (для мирного агента введем небольшое уточнение: если ресурс мирного агента большой, то его сила удара ограничена, подробней см. ниже). А

уменьшение ресурса ударяемого агента пропорционально силе удара. Если ресурс какого-либо агента истощился (стал меньше нуля), то такой агент погибает.

Борьбе коллектива мирных агентов с хищниками может способствовать наличие в коллективе агента-лидера, который дает команду мирным агентам на коллективную угрозу агенту-хищнику. А если агент-хищник не уходит с участка мирных агентов при наличии коллективной угрозы, то агент-лидер дает команду на одновременное нападение на хищника, и все мирные агенты одновременно наносят удары по агенту-хищнику. Агент-лидер может выбираться коллективом мирных агентов в соответствии с максимальным ресурсом среди этих агентов. Если агентов с максимальным ресурсом оказалось несколько, то лидер выбирается из этих агентов случайным образом. Наличие агента-лидера в коллективе мирных агентов рассматривается в отдельном варианте модели.

Результаты анализа и моделирования

Опорные параметры моделирования

Предполагаются следующие опорные параметры моделирования. Рассматриваем сначала эти опорные параметры, которые качественно характеризуют свойства агентов, при моделировании эти параметры могут варьироваться. Размер всего мира равен $N = 40$ клеток. Размер участка мирных агентов равен $N_1 = 20$ клеток. Число клеток с порцией пищи на участке мирных агентов равно $M_1 = 8$, в остальной части мира число клеток с порцией пищи $M_2 = 4$. Начальный ресурс каждого мирного агента $r_0 = 1$. Начальный ресурс каждого хищника $R_0 = 10$. Текущее число мирных агентов и агентов-хищников обозначаем $n_1(t)$ и $n_2(t)$. Начальное число мирных агентов равно 10: $n_1(1) = 10$.

Прирост ресурса мирного агента при съедании одной порции пищи $\Delta r_1 = 0.1$. Расход ресурса мирного агента на перемещение на одну клетку $\Delta r_2 = 0.01$. Расход ресурса мирного агента на угрозу агенту-хищнику $\Delta r_3 = 0.02$. Расход ресурса мирного агента на удар по агенту-хищнику $\Delta r_4 = 0.05$.

Прирост ресурса агента-хищника при съедании убитого мирного агента $\Delta R_1 = 1.0$. Расход ресурса агента-хищника на перемещение на одну клетку $\Delta R_2 = 0.02$. Расход ресурса агента-хищника на удар по мирному агенту $\Delta R_3 = 0.1$. Расход ресурса агента-хищника на убийство мирного агента $\Delta R_4 = 0.5$.

При борьбе между агентами двух видов сила удара зависит от ресурса ударяющего агента. Полагаем, что при ударе агента-хищника эта сила равна $F = k_1 R$. А так как ресурс мирного агента может сильно возрастать (см. рис. 1), то считаем при ударе мирного агента сила удара равна $f = k_1 r$ при малом ресурсе агента ($r < 3$) и достигает постоянной величины $f = 3k_1$, при больших r ($r > 3$), здесь k_1 — коэффициент пропорциональности, R и r — ресурсы хищника и мирного агента, соответственно. Если хищника ударяют сразу несколько мирных агентов, то силы ударяющих агентов суммируются. Так если ресурсы всех ударяющих мирных агентов больше 3, то их общая сила удара равна $f = 3k_1 n_f$, n_f — число ударяющих агентов.

Результаты анализа и компьютерного моделирования

Рассмотрим результаты анализа и компьютерного моделирования для нескольких характерных случаев.

Случай без хищников

Сначала рассмотрим случай, когда в мире нет агентов-хищников. В этом случае мирные агенты случайно блуждают по своему участку, питаются, когда встречают клетки с порциями пищи, и постепенно накапливают ресурс. Прирост ресурса при съедании одной порции пищи мирным агентом равен $\Delta r_1 = 0.1$. Расход ресурса на случайное перемещение на одну клетку по своему участку Δr_2 варьировался. Основной расчет проводился для указанных выше параметров N_1 и M_1: $N_1 = 20$ и $M_1 = 8$. Усредненная зависимость ресурса мирных агентов от времени (полученная в результате компьютерного моделирования) для $\Delta r_2 = 0.01$ и $\Delta r_2 = 0.02$ представлена на рис. 1. Усредненная зависимость числа мирных агентов от времени для тех же параметров представлена на рис. 2. Видно, что с ростом расхода ресурса на перемещение скорость роста ресурса мирного агента уменьшается, а также при этом происходит некоторое вымирание числа агентов. Если еще увеличивается расход ресурса на перемещение, то агенты вымирают. Например, если расход ресурса на случайное перемещение на одну клетку Δr_2 равен 0.03, то мирным агентам не хватает ресурса на перемещение, и все они достаточно быстро (примерно за 1200 тактов времени) вымирают.

Отметим, что можно было бы рассмотреть размножение мирных агентов (механизм такого размножения мог бы быть таким: когда ресурс мирного агента превышает определенный порог, то этот агент рождает агента потомка и передает ему часть своего ресурса), но поскольку основная цель настоящей работы — анализ взаимодействия между разными видами агентов, то размножение агентов в этой работе не рассматривалось.

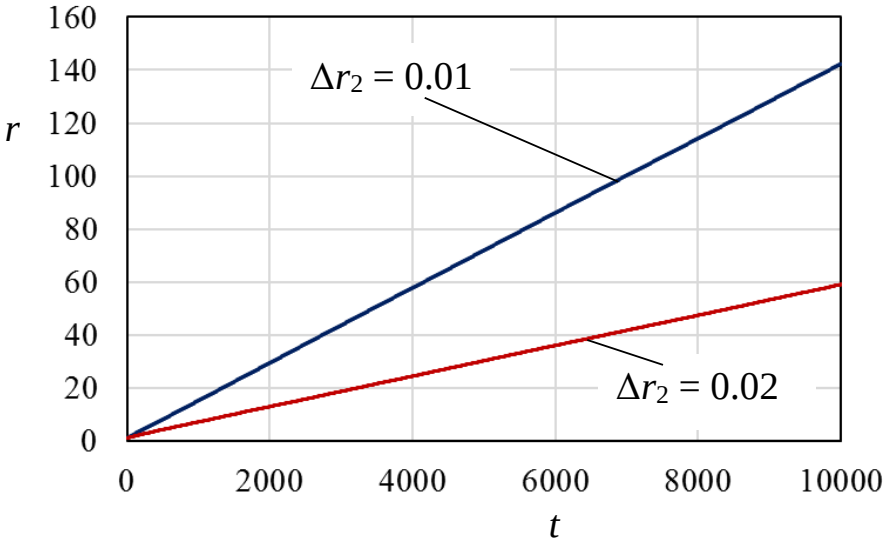


Рис. 1. *Усредненная зависимость ресурса мирных агентов r от времени t. Размер участка мирных агентов равен N_1 = 20 клеток. Результаты усреднены по 300 различным расчетам*

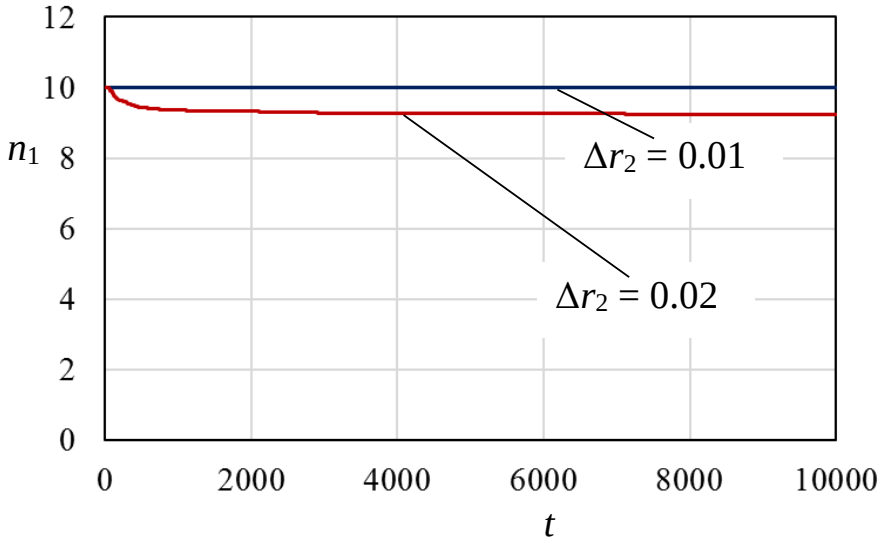


Рис. 2. *Усредненная зависимость числа мирных агентов n_1 от времени t. Размер участка мирных агентов равен N_1 = 20 клеток. Результаты усреднены по 300 различным расчетам*

При моделировании в этом случае была обнаружена интересная особенность поведения мирных агентов. Если участок мирных агентов был увеличен в 5 раз, и число порций пищи на этом участке было также увеличено в 5 раз (т.е. было N_1 = 100 и M_1 = 40), так что плотность порций пищи оставалась такой же, как и для рассмотренного выше расчета при N_1 = 20, то ресурс агентов стал неожиданно уменьшаться. Остальные параметры расчетов были такими же, как и для N_1 = 20 и M_1 = 8. Усредненные зависимости ресурса и числа агентов от времени для N_1 = 100 представлены на рис. 3 и рис. 4 соответственно.

Таким образом, при увеличении участка мирных агентов (при N_1 = 100 и M_1 = 40) поведение агентов резко отличается от их поведения на меньшем участке (при N_1 = 20 и M_1 = 8). Анализ резуль-

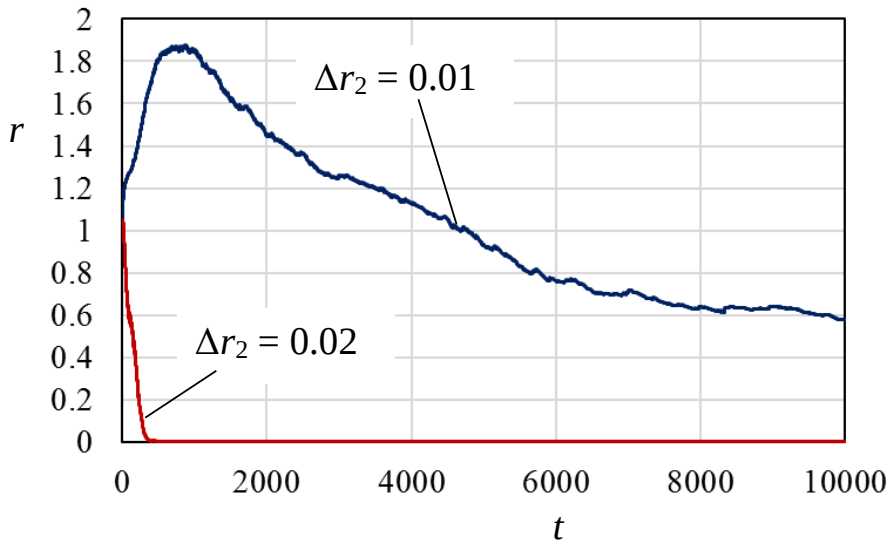


Рис. 3. *Усредненная зависимость ресурса мирных агентов r от времени t. Размер участка мирных агентов равен N_1 = 100 клеток. Результаты усреднены по 300 различным расчетам*

татов компьютерных расчетов при N_1 = 100 показал, что при достаточно большом времени на участке мирных агентов порции пищи оказались явно неравномерно распределены: 1) были фрагменты цепочек клеток (длиной примерно 10 клеток), на которых порций пищи не было совсем на ряде смежных клеток фрагмента, 2) были фрагменты цепочек клеток, на которых порции пищи находились в каждой из смежных клеток фрагмента. Для анализа возникновения такой неравномерности расположения порций пищи была сделана следующая оценка. Для каждой клетки участка мирных агентов подсчитывалось число порций пищи на следующих 10 клетках $k(i)$, i — номер рассматриваемой клетки. То есть величины $k(i)$ характеризовали локальную плотность порций пищи в проведенных компьютерных расчетах.

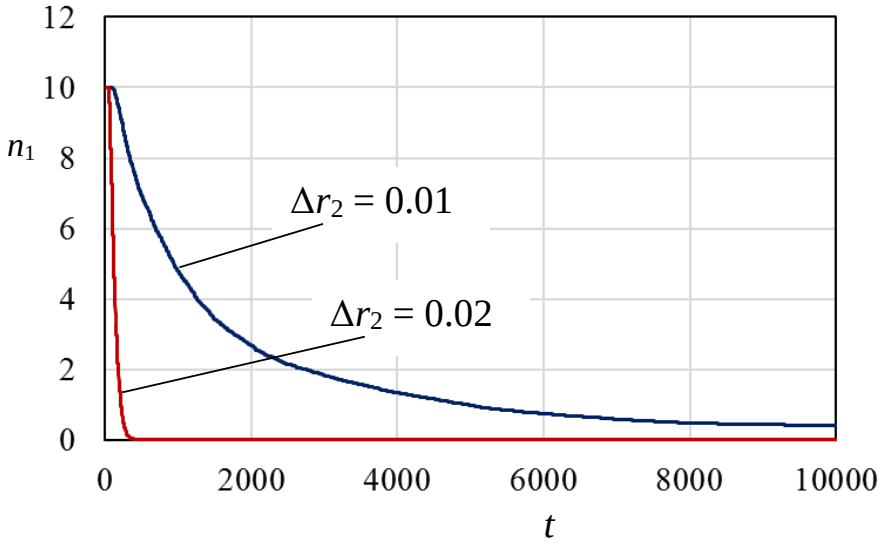


Рис. 4. *Усредненная зависимость числа мирных агентов n_1 от времени t. Размер участка мирных агентов равен N_1 = 100 клеток. Результаты усреднены по 300 различным расчетам*

В среднем величина $k(i)$ для рассматриваемых расчетов k_m была равна 4 (8 порций пищи в 20 клетках). И подсчитывалось среднеквадратическое отклонение $\sigma(t)$ величины $k(i)$ от k_m для рассмотренных расчетов (для основного расчета при N_1 = 20 и M_1 = 8 и для расчета с неожиданным результатом при N_1 = 100 и M_1 = 40) для рассматриваемых моментов времени t. Если рассматриваемая клетка была на расстоянии, меньшем 10 от конца участка мирных агентов, то расчет величины $k(i)$ и

среднего значения этой величины k_m велся по меньшему числу клеток. Зависимости $\sigma(t)$ от времени для этих расчетов приведены на рис. 5.

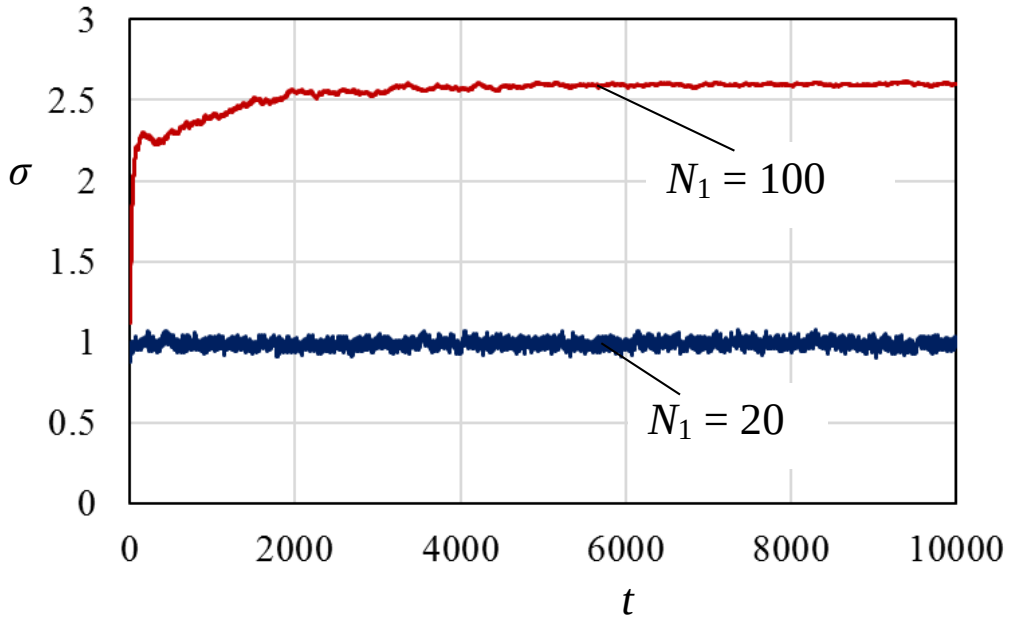


Рис. 5. *Усредненная зависимость среднеквадратического отклонения σ величины $k(i)$ от k_m от времени t для основного расчета при N_1 = 20 и для расчета с неожиданным результатом при N_1 = 100. Для этих расчетов считалось, что величина расхода ресурса при перемещении мирного агента на одну клетку равна Δr_2= 0.01. Результаты усреднены по 300 различным расчетам*

Величина $\sigma(t)$ характеризует неравномерность распределения порций пищи на участке мирных агентов. Видно, что для основного расчета (N_1 = 20) неравномерность есть, но эта неравномерность невелика и практически не меняется со временем. А для расчета при большом числе клеток на участке мирных агентов (N_1 = 100) эта неравномерность возрастает со временем и становится значительно больше, чем для основного расчета. Эта большая неравномерность прямо подтверждается отмеченным выше непосредственным наблюдением за числом порций пищи для разных фрагментов цепочек клеток при N_1 = 100. Интуитивно понятно, что такая неравномерность может быть связана со спецификой локального перемещения агентов (перемещение состоит в движении на одну из соседних клеток) и агенты способны формировать области с малой плотностью порций пищи и довольно долго оставаться на этих областях, в то же время как на других областях плотность порций пищи может быть достаточно большой. Для дополнительной проверки выбора основного расчета были проведены вспомогательные расчеты при нескольких выборах параметров. Вспомогательные расчеты подтвердили целесообразность выбора основного расчета с параметрами $N_1 = 20$, $M_1 = 8$.

Случай с хищником без агента-лидера

Теперь рассмотрим случай одного хищника без наличия агента-лидера мирных агентов. Считаем, что мирные агенты достаточно долго жили на своем участке, так что их ресурс довольно большой. Ресурс хищников тоже велик. Поэтому малое уменьшение ресурса агентов при их движении не существенно. В рассматриваемом случае хищник часто нападает на мирных агентов, убивает и поедает их. Хотя мирные агенты видят приближение хищника, но поскольку скорость перемещения агента-хищника выше скорости мирных агентов, то мирные агенты не успевают уйти от хищника, подвергаются нападению и погибают. Ресурс агента-хищника при этом возрастает. В этом случае агент-хищник может полностью съесть мирных агентов. Понятно, что при большей численности агентов-хищников, вымирание мирных агентов будет усиливаться. В принципе, можно рассмотреть барьеры на границах мира, например, ввести барьер в клетке номер 1, не позволяющий мирным агентам выходить из своего участка. Тогда возможна такая ситуация, что все мирные агенты, убегая от хищника, соберутся в клетке номер 1. Эти агенты могут коллективно угрожать хищнику и даже наносить удары по нему. Хотя без агента-лидера их действия вряд ли будут одновременными, согласованными между собой. Кроме того,

введение барьера и сбор «загнанных» хищником мирных агентов на границе мира у барьера выглядит несколько искусственным. Поэтому целесообразно рассмотрение более естественного случая мира без барьеров при наличии агента-лидера.

Случай с хищником и с агентом-лидером

Далее рассмотрим случай одного агента-хищника при наличии агента-лидера мирных агентов. Считаем, что хищник приходит в мир через границу в клетке номер N, видит мирных агентов и приближается к ним. В этом случае при появлении агента-хищника в мире агент-лидер дает команду всем мирным агентам собраться в одном месте (в середине участка мирных агентов), а при переходе хищника на участок мирных агентов дает команду: всем мирным агентам коллективно угрожать агенту-хищнику. При этом обычно агент-хищник уходит с участка мирных агентов. Возможно, что агент-хищник может переоценить свои силы и при коллективной угрозе может остановиться, тогда все мирные агенты одновременно нанесут удары по агенту-хищнику и ресурс его значительно уменьшится.

Рассмотрим детали поведения агентов в данном случае. Считаем, что параметры, характеризующие ситуацию, таковы. Число клеток участка мирных агентов и число порций пищи на этом участке равны $N_1 = 20$ и $M_1 = 8$, соответственно. Расход ресурса мирного агента на перемещение на одну клетку равен $\Delta r_2 = 0.01$. Мирные агенты уже провели достаточное большое время на своем участке, питаясь и перемещаясь на соседние клетки, так что ресурс одного агента большой: $r > 10$ (см. рис. 1), число этих агентов равно $n_1 = 10$ (см. верхнюю линию на рис. 2). Агент-хищник в определенный момент времени t_P появляется в N-й клетке (наиболее удаленной клетке от участка мирных агентов), видит мирных агентов и начинает двигаться к ним. Скорость этого движения выше, чем скорость мирных агентов, считаем, что, двигаясь к мирным агентам, хищник проходит две клетки за один такт времени. Сразу же после появления агента-хищника на краю мира (при $t = t_P$) агент-лидер мирных агентов подает команду своим мирным агентам всем собраться в центре участка (в 10-й клетке). На этот сбор потребуется 10 тактов времени, за это же время агент-хищник подойдет к участку мирных агентов. После сбора всех мирных агентов, агент-лидер подает команду двигаться в сторону агента-хищника, угрожая ему. Теперь агент-хищник начинает отступать от коллектива мирных агентов, двигаясь уже медленней, со скоростью одна клетка за такт времени. И через 20 тактов времени агент-хищник покинет рассматриваемый мир. Весь процесс вытеснения агента-хищника займет всего 30 тактов времени. Расход ресурса мирного агента при этом составит не более, чем $10\Delta r_2 + 30\Delta r_3 = 0.7$, что значительно меньше накопленного им до этого ресурса. Расход ресурса агента-хищника также небольшой, примерно 0.8. То есть расходы обоих видов агентов значительно меньше, чем имеющиеся у них ресурсы.

Если агент-хищник вдруг остановится и решит бороться с коллективом мирных агентов, то агент-лидер мирных агентов подает команду всем мирным агентам нанести одновременно удары по хищнику. При этом общая сила ударов по хищнику составит $f = 30k_1$, а возможная сила удара хищника равна приближенно (в пренебрежении малыми расходами его ресурса на движение) $10k_1$. То есть коллективные удары мирных агентов значительно сильнее отдельных ударов агента-хищника. Эти коллективные удары значительно уменьшат ресурс хищника и он, естественно, предпочтет уйти от коллектива мирных агентов и покинуть рассматриваемый мир.

Итак, относительно слабые мирные агенты в этом случае коллективно вытесняют сильного агента-хищника со своего участка.

Заключение

Таким образом, в настоящей работе построена и исследована модель взаимодействия между агентами двух видов: 1) мирными относительно слабыми агентами и 2) хищными более сильными агентами. Показано, что мирные агенты могут коллективно вытеснять сильных агентов-хищников со своей хорошо освоенной территории. Необходимо отметить, что для такого вытеснения необходима четкая согласованность большого коллектива агентов. Такую согласованность может обеспечить выбранный мирными агентами агент-лидер, подающий команды на организацию коллективной угрозы агенту-хищнику. Отметим, что подобное коллективное вытеснение хищника происходило, когда многочисленная трубящая стая гусей атаковала лису, вытесняя ее со своей территории [10]. Аналогичная защита своей стаи наблюдалась, когда огромная стая скворцов очень плотно окружала ястреба-перепелятника, лишая его возможности летать [10]. После многократного применения согласованного

коллективного вытеснения сильных агентов со своей территории у мирных агентов может формироваться стереотип атакующего поведения. По-видимому, такой стереотип формировался у скворцов, атакующих хищников.

Итак, основной вывод настоящей работы таков: коллектив относительно слабых агентов способен защищаться от сильных врагов и вытеснять хищников со своей территории.

ЛИТЕРАТУРА

1. *Artificial Life:* The Proceedings of an Interdisciplinary Workshop on the Synthesis and Simulation of Living Systems. Redwood City CA: Addison-Wesley; 1989. 655 p.
2. *Artificial Life II:* Proceedings of the Second Artificial Life Workshop. Redwood City CA: Addison-Wesley; 1992. 854 p.
3. Yaeger L. S. Computational Genetics, Physiology, Metabolism, Neural Systems, Learning, Vision, and Behavior or PolyWorld: Life in a New Context. *Proceedings of the Artificial Life III Conference*. Addison-Wesley; 1994;17:263–298.
4. Yaeger L., Sporns O., Williams S., Shuai X., Dougherty S. Evolutionary Selection of Network Structure and Function. *Artificial Life XII:* Proceedings of the Twelfth International Conference on the Simulation and Synthesis of Living Systems. Cambridge, Massachusetts, London, England: The MIT Press; 2010:313–320. Режим доступа: https://dn790008.ca.archive.org/0/items/ArtificialLifeXII/Artificial%20Life%20XII.pdf.
5. Ackley D., Littman M. Interactions between Learning and Evolution. *Artificial Life II:* Proceedings of the Second Artificial Life Workshop. Redwood City CA: Addison-Wesley; 1992:487–509.
6. Holland J. H. Echoing Emergence: Objectives, Rough Definitions, and Speculations for Echo-class Models. *Complexity: Metaphors, Models and Reality*. Addison-Wesley; 1994:309–342.
7. Hraber P. T., Jones T., Forrest S. The Ecology of Echo. *Artificial Life*. 1997;3(3):165–190. DOI: 10.1162/artl.1997.3.3.165.
8. Коваль А. Г., Редько В. Г. Поведение модельных организмов, обладающих естественными потребностями и мотивациями. *Математическая биология и биоинформатика*. 2012;7(1):266–273. DOI: 10.17537/2012.7.266.
9. Burtsev M., Turchin P. Evolution of Cooperative Strategies from First Principles. *Nature*. 2006;440(7087):1041–1044. DOI: 10.1038/nature04470.
10. Lorenz K. *On Aggression*. London, New York: Routledge Classics; 2002. 306 p.

О ВОССТАНОВЛЕНИИ КУЛОНОВСКОГО ПОЛЯ ПОТЕНЦИАЛЬНЫХ ТЕЧЕНИЙ НА ОСНОВЕ ИСКУССТВЕННЫХ НЕЙРОННЫХ СЕТЕЙ

А. Д. Смородинов[1,2]

[1] *Сургутский филиал федерального государственного автономного учреждения «Федеральный научный центр Научно-исследовательский институт системных исследований Национального исследовательского центра «Курчатовский институт», г. Сургут, Российская Федерация*

[2] *Сургутский государственный университет, г. Сургут, Российская Федерация*

ORCID: https://orcid.org/0000-0002-9324-1844, ✍ smorodinov.alexsandr@yandex.ru

Аннотация: в статье рассматривается возможность применения искусственной нейронной сети для решения задачи Коши на примере моделирования течения тяжелой примеси в потенциальных течениях несжимаемой жидкости, порожденных гармоническим потенциалом. Конструируется искусственная нейронная сеть, не требующая обучения, для представления градиента кулоновского потенциала, показывающая высокую точность приближения. Показывается связь итерационных методов решения задачи Коши с ИНС, конструируется ИНС, реализующая метод Рунге–Кутты, в частности, 4-го порядка точности, не требующая обучения, в которой весовые коэффициенты установлены на этапе конструирования. Проводится тестовый расчет, в котором сравниваются результаты моделирования динамики примеси с использованием разработанной ИНС и программного комплекса для решения задачи Коши с различными конфигурациями полей начальных данных и визуализации результатов математического моделирования течения тяжелой примеси. Кроме того, показывается возможность практического применения ИНС для поиска местоположений источника кулоновского потенциала.

Ключевые слова: искусственные нейронные сети, теорема Колмогорова-Арнольда, задача Коши, метод Рунге–Кутты, динамика примеси.

Благодарности: работа выполнена в рамках государственного задания НИЦ «Курчатовский институт» — НИИСИ по теме № FNEF-2024-0001 «Создание и реализация доверенных систем искусственного интеллекта, основанных на новых математических и алгоритмических методах, моделях быстрых вычислений, реализуемых на отечественных вычислительных системах» (1023032100070-3-1.2.1).

Для цитирования: Смородинов А. Д. О восстановлении кулоновского поля потенциальных течений на основе искусственных нейронных сетей. *Успехи кибернетики*. 2025;6(1):137–149.

Поступила в редакцию: 27.02.2025. *В окончательном варианте: 09.03.2025.*

NEURAL NETWORK APPLICATIONS TO THE RECONSTRUCTION OF COULOMB FIELD IN POTENTIAL FLOWS

A. D. Smorodinov[1,2]

[1] *Surgut Branch of Scientific Research Institute for System Analysis of the National Research Centre "Kurchatov Institute", Surgut, Russian Federation*

[2] *Surgut State University, Surgut, Russian Federation*

ORCID: https://orcid.org/0000-0002-9324-1844, ✍ smorodinov.alexsandr@yandex.ru

Abstract: we apply an artificial neural network (ANN) to solve the Cauchy problem by modeling the flow of heavy impurities in potential flows of an incompressible fluid generated by a harmonic potential. We construct an ANN that does not require training to represent the gradient of the Coulomb potential, demonstrating high approximation accuracy. We establish the connection between iterative methods for solving the Cauchy problem and ANNs. Specifically, we design an ANN that implements the Runge-Kutta method, including a fourth-order accurate version, where weight coefficients are predefined at the design stage, eliminating the need for training. We perform a test calculation to compare impurity dynamics simulated using the developed ANN with results from a software package for solving the Cauchy problem. We analyze different initial data configurations and visualize the simulated flow of heavy impurities. Additionally, we demonstrate the practical application of the ANN for locating sources of the Coulomb potential.

Keywords: artificial neural networks, Kolmogorov-Arnold theorem, Cauchy problem, Runge-Kutta method, impurity dynamics.

Acknowledgements: this is a part of the FNEF-2024-0001 Development and Deployment of Trusted AI Systems based on New Mathematical and Algorithmic Approaches and Fast Computing Models Compatible with Domestic Computer Hardware government contract (1023032100070-3-1.2.1) granted to the NRC "Kurchatov Institute" – SRISA.

Cite this article: Smorodinov A. D. Neural Network Applications to the Reconstruction of Coulomb Field in Potential Flows. *Russian Journal of Cybernetics*. 2025;6(1):137–149.

Original article submitted: 27.02.2025. *Revision submitted: 09.03.2025.*

Введение

Искусственные нейронные сети (ИНС) в настоящее время показывают хорошие результаты при решении некоторых задач. Это обусловлено тем, что часто для разработки систем искусственного интеллекта используют предобученные ИНС, в которых уже были установлены функции активации, размерность ИНС и определены основные значения весовых коэффициентов. В таких случаях исследователю остается только дообучить ИНС на своем наборе данных для решения более узкоспециализированной задачи.

В данной работе представляется предварительно сконструированная ИНС, весовые коэффициенты которой установлены исходя из аналитических соображений, основанных на итерационном методе решения задачи Коши. Такой подход может быть использован для решения задачи, которая возникает у экологической полиции: предположим, что есть информация о загрязнении некоторого водоема вследствие сброса загрязняющих веществ в воду. Необходимо, имея информацию о движении загрязнений (примеси) в некоторой части водоема, найти местоположение и силу сброса. Покажем создание описанной ИНС, а также применение ИНС для решения задачи. По сути, будет сконструирована предобученная ИНС, которую необходимо будет дообучить для определения местоположения источников на конкретной задаче.

Постановка задачи математического моделирования

Рассмотрим возможность применения ИНС к задаче гидродинамики на примере задачи моделирования движения безынерционных недиффундирующих примесей. Для этого рассмотрим следующее поле скоростей (1), которое является точным решением системы уравнений Навье–Стокса:

$$\vec{V} = \nabla\varphi\left(\vec{x},t\right), \tag{1}$$

где $\varphi\left(\vec{x},t\right)$ — скалярное поле, которое удовлетворяет уравнению Лапласа:

$$\Delta_x\varphi(\vec{x},t) \equiv 0. \tag{2}$$

Выберем в качестве скалярного поля поле следующего вида:

$$\varphi\left(\vec{x},t\right) = \frac{q_i}{\|\vec{x} - M_1\|}, \tag{3}$$

$$\vec{V}\left(\vec{x}\right) = \nabla\left(\sum_{i=1}^{K}\frac{q_i}{\|\vec{x} - M_i\|}\right), \tag{4}$$

где знаменатель есть евклидова норма:

$$\|\vec{x} - M_1\| = \sqrt{\left(x_1 - M_1^1\right)^2 + \left(x_2 - M_1^2\right)^2 + \left(x_3 - M_1^3\right)^2}, \tag{5}$$

где K — количество потенциальных полей, q_i — мощность i-го источника, M_i — местоположение i-го источника.

Тогда моделирование динамики примеси, движущейся в потенциальных течениях, задаваемых кулоновскими потенциалами, описывается уравнением:

$$\frac{\partial N}{\partial t} + \left(\vec{V}\cdot\nabla\right)N = \sigma^2\cdot\Delta N, \tag{6}$$

где N – концентрация примеси, т.к. рассматривается движение безынерционных недиффундирующих примесей, то коэффициент диффузии $\sigma = 0$, а вектор скорости движения примеси совпадает с вектором скорости движения жидкости $\vec{V}(\vec{x})$, и тогда уравнение (6) примет следующий вид:

$$\frac{\partial N}{\partial t} = -\left(\vec{V} \cdot \nabla\right) N. \tag{7}$$

Следовательно, для получения траекторий движения частиц примеси необходимо решить задачу Коши в трехмерном пространстве.

В статье [1] приведена полная постановка задачи, а также исследована структура течений и представлена серия вычислительных экспериментов по моделированию динамики примеси во встречных потенциальных потоках одинаковой мощности, а также сравнение с натурным экспериментом столкновения двух встречных струй с использованием программного обеспечения [17], описанного в [2]. Покажем, что данная задача может быть решена на основе ИНС.

Конструирование ИНС для решения задачи Коши

Обратимся к теореме Колмогорова–Арнольда о суперпозиции функций многих переменных функцией одной переменной [3].

Теорема Колмогорова–Арнольда: при любом целом $n \geq 2$ существуют такие определенные на единичном отрезке $E^1 = [0;1]$ непрерывные действительные функции $\psi^{pq}(x)$, что каждая определенная на n-мерном единичном кубе E^n непрерывная действительная функция $f(x_1, \ldots, x_n)$ представима в виде:

$$f(x_1, \ldots, x_n) = \sum_{q=1}^{q=2n+1} \chi_q \left[\sum_{p=1}^{n} \psi^{pq}(x_p) \right], \tag{8}$$

где функции $\chi_q(y)$ действительны и непрерывны.

Фактически в данной теореме показано, что любую функцию многих переменных можно представить в виде суперпозиции функции одной переменной и сложения. В более поздних работах показано, что внутренние функции ψ можно заменить одной [5]. Применительно к ИНС похожая модификация теоремы предложена в работе [6]. Строго говоря, теоремы, представленные в работах [5, 6], не повторяют теорему Колмогорова-Арнольда в точности в отношении суперпозиции, а говорят, что такие функции можно подобрать в зависимости от наперед заданной точности приближения.

На данную теорему можно смотреть с точки зрения расширения теоремы Вейерштрасса [4], в которой показано, что любая непрерывная функция на некотором замкнутом интервале действительных чисел может быть представлена абсолютно и равномерно сходящимся рядом полиномов.

Данная теорема является теоремой существования, из которой, строго говоря, не следует, что данная конфигурация является оптимальной с точки зрения времени обучения или что используемые функции являются легко реализуемыми, поэтому чаще всего для решения подобных задач применяются многослойные перцептроны с большим количеством нейронов в каждом слое. В связи с этим при конструировании ИНС будем опираться на данную теорему, но не будем требовать полного совпадения структуры полученной ИНС с теоремой.

Покажем, что поле скоростей (4) представимо в виде суперпозиции функций одной переменной и может быть реализовано как ИНС. Для этого распишем оператор набла в (4) и получим поле $\vec{V}(\vec{x})$ в покомпонентной записи:

$$V_i(x_1, x_2, x_3) = -\sum_{k=1}^{K} \left[q_k \left\{ \left(x_1 - M_k^1\right)^2 + \left(x_2 - M_k^2\right)^2 + \left(x_3 - M_k^3\right)^2 \right\}^{-\frac{1}{3}} (x_i - M_k^i) \right], \tag{9}$$

где $i = 1, \ldots, 3$ — количество компонент, M_k^i — i-ая координата k-го источника.

Покажем, что (9) представимо в виде классической многослойной ИНС, особенностью которой будет отсутствие необходимости проводить поиск весовых коэффициентов сети. Фактически в данной задаче искусственная нейронная сеть будет представлять собой вычислительный ориентированный ациклический граф.

Для этого введем следующие обозначения для элементарных функций:

$$G^2(x) = x^2, \tag{10}$$

$$G^{-\frac{3}{2}}(x) = x^{-\frac{3}{2}}, \tag{11}$$

$$G(x) = x. \tag{12}$$

Формулы (10)–(12) будут являться функциями активации, а константы в формуле (8) будут весовыми коэффициентами. Тогда формулы (9) преобразуются в следующий вид:

$$V_i(x_1,x_2,x_3) = -\sum_{k=1}^{K}\left[q_k G^{-\frac{3}{2}}\left\{G^2\left(x_1 - M_k^1\right) + G^2\left(x_2 - M_k^2\right) + G^2\left(x_3 - M_k^3\right)\right\} G\left(x_i - M_k^i\right)\right]. \tag{13}$$

Упростим запись формул (13) и аргумент функции $\mathrm{G}^{-\frac{3}{2}}$ запишем под знаком суммы, тогда V_i будет иметь следующую запись:

$$V_i(x_1,x_2,x_3) = -\sum_{k=1}^{K}\left[q_k G^{-\frac{3}{2}}\left\{\sum_{j=1}^{3} G^2(x_j - M_k^j)\right\} G(x_i - M_k^i)\right]. \tag{14}$$

В классических ИНС нет умножения функций, а только сложение, потому распишем умножение функции $G^{-\frac{3}{2}}$ на G, используя следующую формулу, основанную на формуле сокращенного умножения:

$$x_1 x_2 = \frac{(x_1 + x_2)^2 - x_1^2 - x_2^2}{2}. \tag{15}$$

Заметим, что данная формула легко представима с помощью функций (10) и (12):

$$x_1 \cdot x_2 = \frac{G^2\left(\sum\limits_{m=1}^{2} G(x_m)\right) - \sum\limits_{m=1}^{2} G^2(x_m)}{2}. \tag{16}$$

Тогда, подставив вместо x_1 функцию $G^{-\frac{3}{2}}$, вместо x_2 функцию G и результат в (9), вместо умножения соответствующих формул получаем:

$$V_i(x_1,x_2,x_3) = \\ -\sum_{k=1}^{K}\left[q_k \frac{G^2\left(G^{-\frac{3}{2}}\left\{\sum\limits_{j=1}^{3} G^2\left(x_j - M_k^j\right)\right\} + G\left(x_i - M_k^i\right)\right) - G^2\left(G^{-\frac{3}{2}}\left\{\sum\limits_{j=1}^{3} G^2\left(x_j - M_k^j\right)\right\}\right) - G^2\left(G\left(x_i - M_k^i\right)\right)}{2}\right]. \tag{17}$$

Данная формула показывает, что (9) представима в виде ИНС.

Примем гипотезу из некоторой априорной информации о кулоновском виде потенциальных течений в виде суперпозиции конечного числа точечных источников и стоков.

В несколько этапов спроектируем данную ИНС. Для начала сконструируем ИНС для умножения двух чисел. Особенностью данной ИНС является ее сложная, составная структура, которую можно представить в виде вычислительного ациклического графа посредством библиотеки для реализации ИНС Keras для языка программирования Python.

Модель искусственной нейронной сети для умножения двух чисел

Покажем реализацию ИНС для сети (17) с реализацией части, отвечающей за умножение двух чисел. Данная модель ИНС описывается равенством (15), приведем схему ИНС, реализующей данную модель (рисунок 1).

На рисунке 1 линии (ребра) означают наличие связи, т.е. то, что соответствующий весовой коэффициент отличен от нуля, если ребро не подписано, то весовой коэффициент равен 1, иначе равняется значению подписи. В нейронах указаны используемые функции активации, x_1, x_2 — входные данные модели.

Формирование модели происходит в 3 этапа:

1 этап. Формирование слоев.

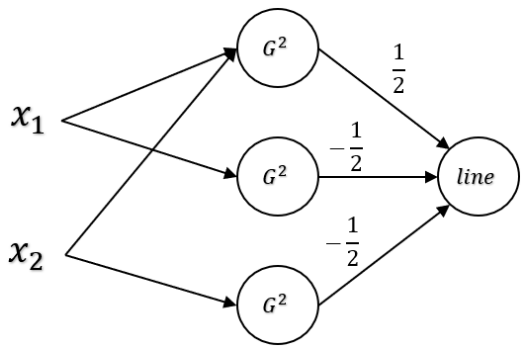


Рис. 1. *Схема ИНС, реализующей умножение двух чисел*

2 этап. Создание модели, в том числе ее компиляция.

3 этап. Установка нужных весовых коэффициентов ИНС.

Для тестирования ИНС проведем сравнение точности умножения классическим методом и методом с использованием вышеописанной ИНС. Для этого будем генерировать пары чисел и сравнивать по модулю разность умножения двумя способами. Средняя ошибка на 10000 пар для целых чисел составила 0, для вещественных чисел с 4 знаком после запятой средняя ошибка на 10000 пар составила 0.0317059. При детальном исследовании проблем точности оказывается, что при слишком больших или слишком маленьких результатах вычислений появляются проблемы округления. Т.е. если получаемый результат больше, чем 10^7 (меньше 10^{-7}), то ИНС начинает выдавать погрешность вычисления.

Реализация модели ИНС для градиента потенциального поля

Покажем теперь непосредственно реализацию модели ИНС для (17), схема ИНС, реализующей градиент потенциального поля для одной координаты, представлена на рисунке 2.

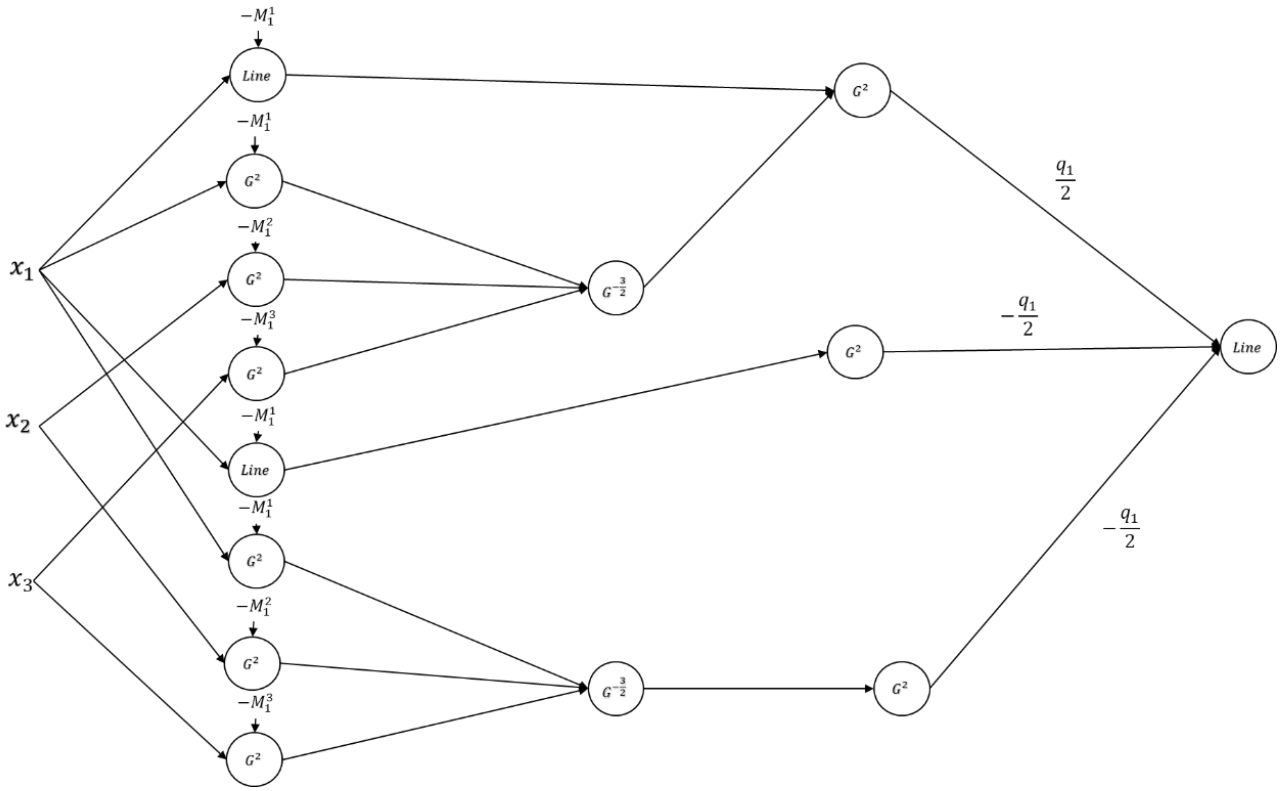


Рис. 2. *Схема ИНС, реализующей градиент потенциального поля для одного источника, для* x_1 *координаты*

На представленной схеме ИНС обозначения аналогичны рисунку 1. Обозначение в виде стрелки сверху с подписью $-M_1^i$ означает смещение (биас) нейрона, который равняется местоположению

источника. Отсутствие стрелки сверху означает, что смещение данного нейрона равняется 0. Весовые коэффициенты перед выходным слоем, равные $|\frac{q_1}{2}|$, отвечают за мощность источника. На вход ИНС подаются координаты примеси, на выходе ожидается V_1. Для получения V_2, V_3 необходимо незначительно изменить весовые коэффициенты первого слоя, а именно для первого и пятого нейрона первого слоя обнулить веса, отвечающие за первый вход, и установить веса, равные 1, для второго входа для реализации V_2. Для V_3 необходимо для тех же нейронов обнулить веса, отвечающие за первый вход, и установить веса, равные 1, для третьего входа. Ввиду минимальных отличий в схеме соответствующие схемы для V_2, V_3 не приводятся.

Приведем результаты тестирования расчета стандартным методом и методом с использованием ИНС, максимальное отклонение расчетов составляет 2.4798^{-07}, тестирование проводилось для источника мощностью $q = 2$, местоположение источника $M = (0{,}0, -2)$, точки, для которых проводился расчет, были сформированы в виде сетки с шагом 0.1, от $(-1, -1, -1)$ до $(1{,}1{,}1)$.

Данная схема производит расчет только для одного источника, для расчета V_i с K источниками необходимо объединить K моделей, описанных на рисунке 2, изменив соответствующим образом весовые коэффициенты, отвечающие за местоположение и мощность источника. После добавить еще один слой, состоящий из одного нейрона, в котором происходит суммирование выходов из каждой модели. Схематично обозначение данной модели представлено на рисунке 3.

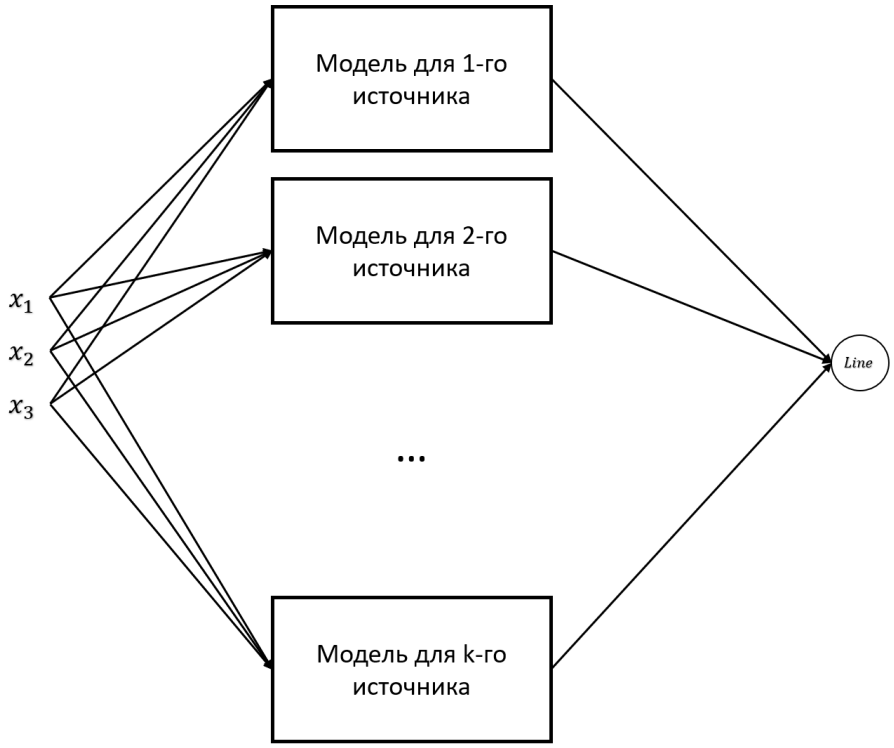


Рис. 3. *Схема ИНС для одного V_i с K источниками*

Под каждым квадратом на рисунке 3 подразумевается модель, представленная на рисунке 2. На выходе ожидается значение для определенной координаты V_i.

Для реализации схемы ИНС, реализующей все V_i , для каждого i необходимо объединить ИНС, представленные на рисунке 3, в ансамбль моделей и изменить весовые коэффициенты на описанные выше, тогда каждый выход ИНС будет отвечать за свой V_i.

Максимальное уклонение расчета градиента потенциального поля, полученного с использованием модели ИНС и без использования ИНС, не превышало 10^{-7}. Расчетные точки формировались в виде сетки с шагом 0.1, от $(-1, -1, -1)$ до $(1{,}1{,}1)$, количество источников варьировалось от 1 до 5, мощность изменялась от 0.01 до 5. Общее количество проведенных тестов с различными конфигурациями местоположений и мощности источников составило 20.

Продемонстрировав возможности реализации (17) в виде модели ИНС, покажем, что данный подход можно использовать для решения задачи Коши.

Реализация метода Рунге–Кутты в виде модели ИНС

Покажем, что с использованием ИНС можно реализовать итерационный численный метод решения задачи Коши на примере реализации метода Рунге–Кутты 4-го порядка точности. Строго говоря, итерационный метод лучшим образом представим в виде рекуррентной ИНС, например, в виде ИНС Джордана, описанной в работах [7–9]. Схема ИНС Джордана представлена на рисунке 4 [10].

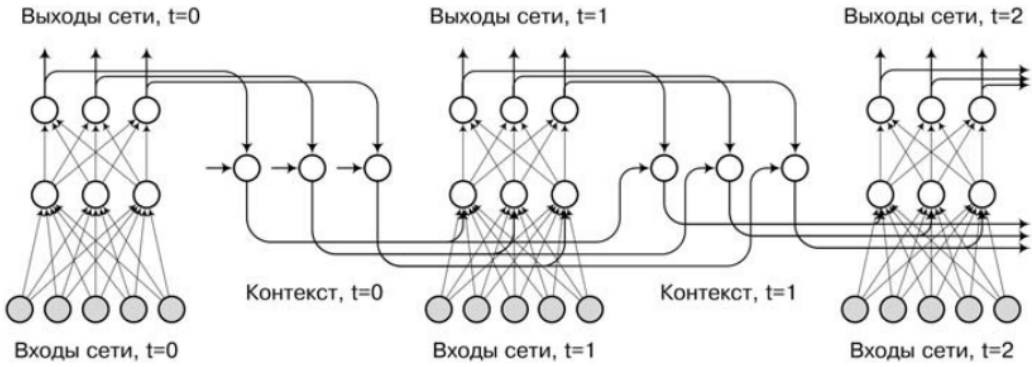


Рис. 4. *Рекуррентная нейронная сеть Джордана [10]*

Для применения рекуррентной нейронной сети Джордана для задачи Коши необходимо на вход подавать Н.У. в момент времени $t_0 = 0$, а также шаг по времени h. На выходе сети, подписанном на схеме как «выход сети, $t = 0$», получаем значение примеси в момент времени $t_1 = t_0 + h$. Контекстом в данном случае называется положение примеси в момент времени t_1, на вход в данном случае необходимо подавать текущее время и шаг по времени. На выходе будет получено местоположение примеси в момент времени $t_2 = t_1 + h$, после чего процедура повторяется требуемое количество итераций.

Стоит заметить, что современные рекуррентные ИНС с долговременной памятью очень похожи на k-шаговые методы Адамса для численного решения задачи Коши, который часто используют при расчете траекторий движения небесных тел.

Суть данных методов заключается в том, что они используют информацию из k-предыдущих решений. Т.е. для вычисления шага $k + 1$ необходимо знать решения в предыдущих k-шагах, в то время как одношаговые методы требуют информации только о предыдущем местоположении точки. Если же говорить в терминах ИНС, то сеть (решатель) должна «помнить» (учитывать) предыдущие k «предсказаний». В ИНС с долговременной памятью используются «вентиль забывания», целью которого является определение, насколько значим шаг по времени, причем для каждого обрабатываемого шага по времени устанавливается свой коэффициент в ходе обучения сети. Данный алгоритм напоминает явный метод Адамса–Башфорта или неявный Адамса–Моултона, которые были представлены в работах [14–16].

Рассмотрим метод Рунге–Кутты 4-го порядка точности и реализуем ИНС, которая проводит вычисления по данному методу. Для этого приведем формулы метода Рунге–Кутты 4-го порядка точности [11]:

$$\vec{y}_{n+1} = \vec{y}_n + \frac{1}{6}h\left(\vec{k}_1^n + 2\vec{k}_2^n + 2\vec{k}_3^n + \vec{k}_4^n\right), \tag{18}$$

$$\vec{k}_1^n = \vec{f}\left(t^n, x_1^n, x_2^n, x_3^n\right), \tag{19}$$

$$\vec{k}_2^n = \vec{f}\left(t^n + \frac{h}{2}, x_1^n + \frac{h}{2}k_1^n, x_2^n + \frac{h}{2}k_1^n, x_3^n + \frac{h}{2}k_1^n\right), \tag{20}$$

$$\vec{k}_3^n = \vec{f}\left(t^n + \frac{h}{2}, x_1^n + \frac{h}{2}k_2^n, x_2^n + \frac{h}{2}k_2^n, x_3^n + \frac{h}{2}k_2^n\right), \tag{21}$$

$$\vec{k}_4^n = \vec{f}\left(t^n + h, x_1^n + hk_3^n, x_2^n + hk_3^n, x_3^n + hk_3^n\right). \tag{22}$$

Если посмотреть на формулу 18, то видно, что там стоит сумма функции правой части, рассчитанная с разными аргументами, т.е. фактически для решения задачи Коши с использованием ИНС необходимо использовать каскад ИНС. Т.е. фактически решение трехмерной задачи Коши представимо в виде суперпозиции функции одной переменной и сложения при условии, что правая часть Д.У. представима в виде суперпозиции функций одной переменной.

Разработанная ИНС «предсказывает» положение примеси на основе полносвязной ИНС. Рекуррентная связь в ИНС реализована путем направления выхода ИНС на вход через цикл, такой подход выбран в целях упрощения реализации ИНС.

Проведем вычислительный эксперимент по моделированию динамики примеси в сталкивающихся потоках. Для этого зададим два источника мощностью $q_1 = q_2 = 1$ и местоположением $M_1(1{,}0{,}0)$ и $M_2(-1{,}0{,}0)$, в некоторой области около каждого источника будем формировать начальное положение примеси.

Примесью заполнялась цилиндрическая область со следующими параметрами:

1) основание цилиндра по проекции на yOz при $x = 0.45$ и $x = 0.95$ для первого источника и проекция на yOz при $x = -0.45$ и $x = -0.95$ для второго источника;

2) радиус цилиндра $r = 0.125$, ось цилиндра совпадала с осью абсцисс;

3) примесь заполняла сечение цилиндра с шагом 0.01 по оси цилиндра, полученное сечение заполнялось начальными точками по следующим формулам:

$$R(i) = 0.1 + 0.1 \cdot i, \tag{23}$$

$$DF(i) = 8 + 2 \cdot i, \tag{24}$$

$$Y = R(i) \cdot \cos((DF(i) \cdot j + (i \cdot j + 1))\pi)r, \tag{25}$$

$$Z = R(i) \cdot \sin((DF(i) \cdot j + (i \cdot j + 1))\pi)r, \tag{26}$$

$$j = 0, \ldots, 9, \qquad i = 0, \ldots, 9. \tag{27}$$

Данные формулы использованы для того, чтобы с удалением формирования начальных условий от оси цилиндра увеличивалось количество частиц примеси. Использование генератора случайных чисел для заполнения случайным образом примесью указанных цилиндрических областей вызывает трудности ввиду того, что для моделирования динамики примеси используется различное программное обеспечение, разработанное на разных языках программирования.

Всего начальных условий 22 500 — по 11 250 для каждой из областей, начальное распределение частиц примеси представлено на рисунке 5. Шаг по времени составлял $dt = \frac{1}{2000}$, количество шагов по времени 6000. Моделирование проводилось на вычислительной машине со следующими характеристиками: процессор AMD Ryzen Threadripper 2990WX 32-Core-Processor, 32 ГБ оперативной памяти, с частотой 2 400 МГц и графическим процессором NVIDIA Titan V с 5 120 вычислительными Cuda-ядрами и 12 ГБ видеопамяти.

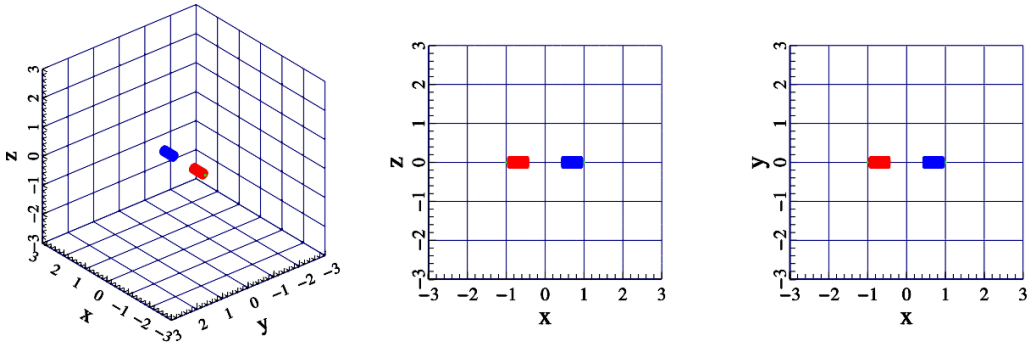


Рис. 5. *Начальная конфигурация частиц примеси*

Для моделирования примеси методом Рунге–Кутты 4-го порядка точности применялся программный комплекс, описанный в [2], а также зарегистрированный в ФИПС [17], который проводит параллельные вычисления с использованием технологии OpenMP, а для визуализации — библиотеку mathGL. Кроме того, данный программный комплекс использовался авторами в [1, 12, 13].

Моделирование примеси с использованием описанной выше ИНС происходило в два этапа: на 1 этапе производился расчет динамики примеси с использованием искусственной полносвязной нейронной сети, весовые коэффициенты которой были рассчитаны и установлены на этапе проектирования сети, ИНС реализована на языке программирования Python с использованием библиотеки Keras,

с автоматическим распараллеливанием на видеокарту. После чего полученные данные были записаны в файл и визуализированы с использованием программного комплекса, описанного в [2]. Использование указанного программного комплекса обусловлено тем, что проводилось не только количественное сравнение, но и качественное, а при использовании другой системы визуализации сравнение было бы не таким наглядным. Кроме того, в программном комплексе [2] реализовано распараллеливание визуализации по ядрам процессора, что значительно ускорило процесс визуализации решения.

На рисунках 6–8 представлены изображения для качественного сравнения результатов. Если же говорить о количественном сравнении результатов вычисления, то уклонение метода Рунге–Кутты 4-го порядка точности от метода Рунге–Кутты 4-го порядка точности, реализованного с использованием ИНС, составляет не более 10^{-6}.

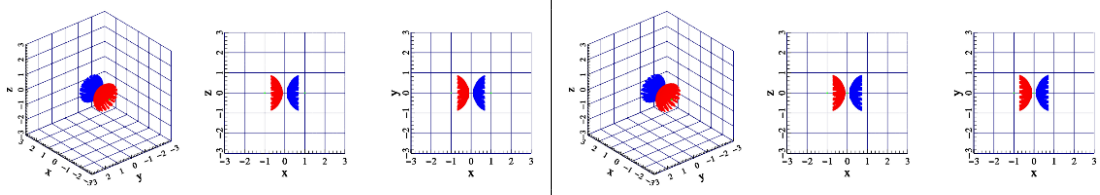
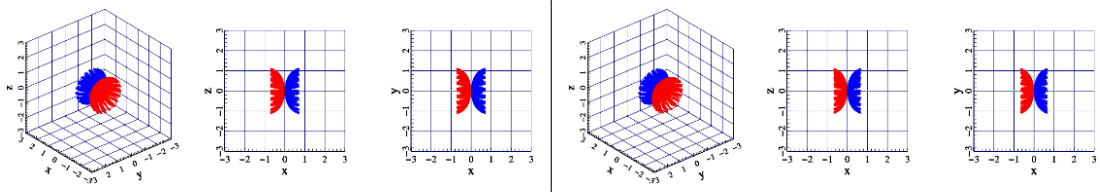
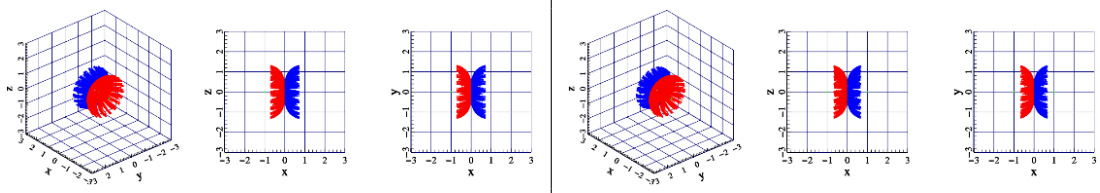
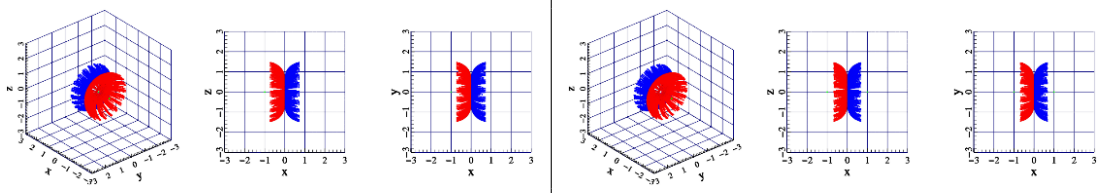
Время	Расчет динамики примеси с использованием ИНС	Расчет динамики примеси без использования ИНС
$t = 0.25$		
$t = 0.5$		
$t = 0.75$		
$t = 1.0$		

Рис. 6. *Сравнение моделирования динамики примеси с использованием ИНС и без,* $0.25 \leqslant t \leqslant 1.0$

Обсуждение результатов

Несмотря на то, что скорость моделирования динамики примеси с использованием ИНС ниже по сравнению с программным комплексом [2], возможность применения ИНС для решения задачи Коши имеет большой практический потенциал. Рассмотрим обратную задачу. Такая задача, как было сказано выше, может возникать у экологической полиции для поиска места сброса загрязнений. Пусть есть решения задачи Коши в виде некоторого набора траекторий динамики примеси либо части траекторий в некоторые моменты времени от t_1 до некоторого t_2, но неизвестны точные характеристики течения, такие как местоположение источника и его мощность. Тогда применение классических алгоритмов обучения ИНС позволит решить данную задачу в ходе обучения, когда на вход ИНС будут поданы местоположения примеси в моменты времени t_i, а на выходе из ИНС будут ожидаться местоположения примеси в моменты времени t_{i+1}. В ходе обучения необходимо будет подобрать весовые коэффициенты только для 4K-параметров, где K — количество источников, информация о которых будет выбираться исходя из некоторой априорной информации для конкретной задачи.

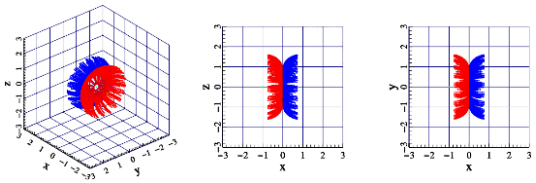
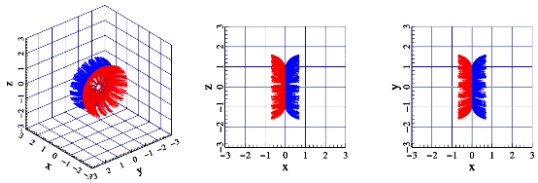
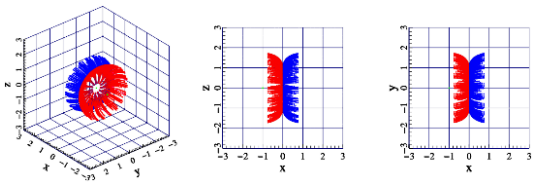
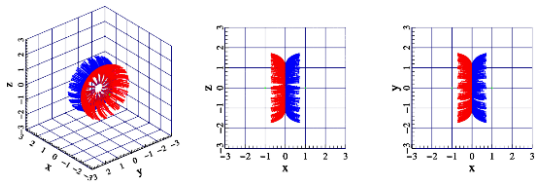
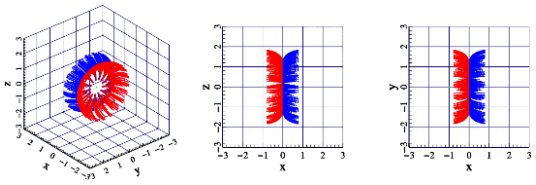
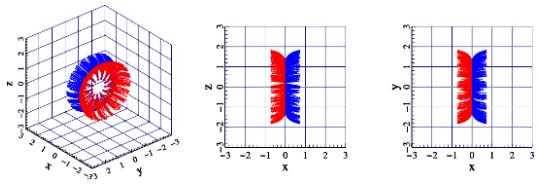
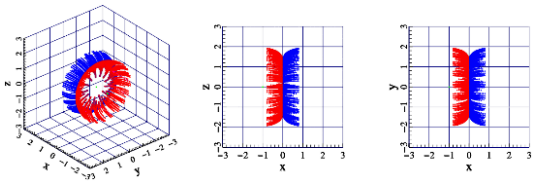
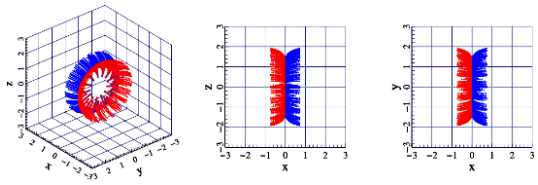
Время	Расчет динамики примеси с использованием ИНС	Расчет динамики примеси без использования ИНС
$t = 1.25$		
$t = 1.5$		
$t = 1.75$		
$t = 2.0$		

Рис. 7. *Сравнение моделирования динамики примеси с использованием ИНС и без,* $1.25 \leqslant t \leqslant 2.0$

Кроме того, использование именно ИНС для решения поставленной задачи позволит решить сразу несколько проблем: избежать необходимости реализации методов градиентного спуска, получить возможность использовать наиболее подходящие модификации данных методов. При использовании библиотек для ИНС с реализованными вычислениями на видеокарте потенциально должно сократиться время на решение обратной задачи поиска конфигурации течения. Кроме того, в дальнейшем можно использовать данный метод для решения сходных задач по поиску конфигураций области течения, например в цилиндрических областях, описанных в работах [12, 13].

На данный момент первые эксперименты показывают возможность применения описанного выше подхода для поиска местоположения источника. Так, были проведены следующие эксперименты по поиску местоположения двух источников с координатами:

$$M'_1\,[0{,}0,-1] \text{ и } M'_2\,[0{,}0{,}1]\,, \qquad q_1 = q_2 = 1.$$

В качестве обучающих данных использовалась информация о динамике 300 из 22 500 частиц примеси с 1 000 отсчетов по времени, итого обучающих данных было 300 000. В качестве метода обучения использовался стандартный стохастический градиентный спуск без модификаций. Функция потерь (минимизируемая функция) — средняя абсолютная процентная ошибка. Для обучения считалось, что есть априорная информация о том, что сброс примеси происходит с двух источников.

Обучение ИНС № 1

Весовые коэффициенты, отвечающие за местоположение источника и его мощность, инициализированные случайными значениями, равны:

$$M_0\colon\ [\,0.04\ \ 0.025\ \ -0.044],\qquad q = -0.01,$$
$$M_1\colon\ [-0.059\ \ -0.0135\ \ 0.034],\qquad q = -0.11.$$

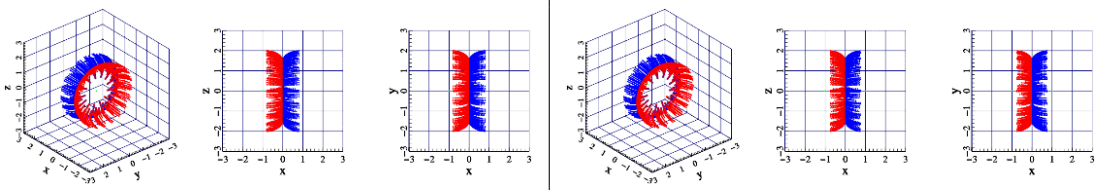
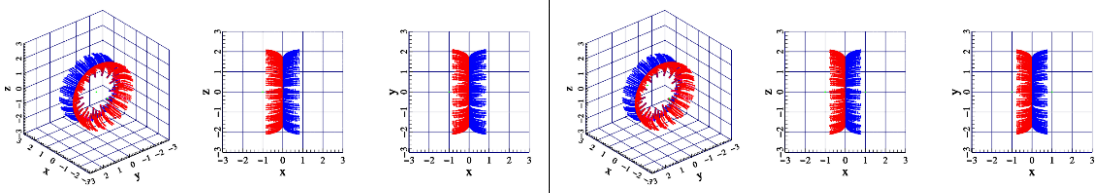
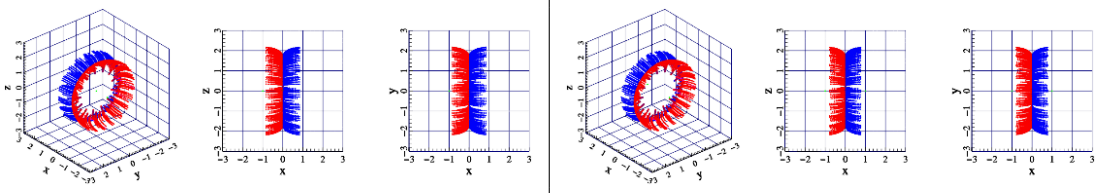
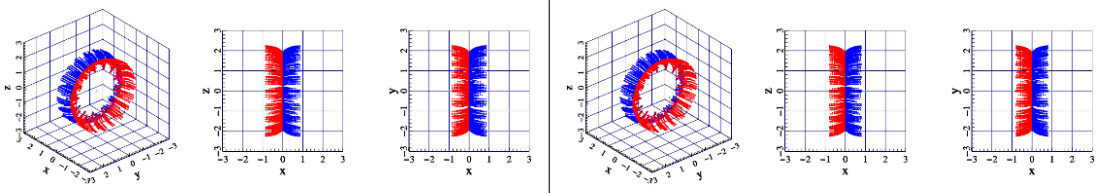
Время	Расчет динамики примеси с использованием ИНС	Расчет динамики примеси без использования ИНС
$t = 2.25$		
$t = 2.5$		
$t = 2.75$		
$t = 3.0$		

Рис. 8. *Сравнение моделирования динамики примеси с использованием ИНС и без,* $2.25 \leqslant t \leqslant 3.0$

Результат поиска, который длился 5 эпох по 3.5 сек каждая.

$$M_0\colon\ [\,0.021\quad -0.0014\quad -1.0008],\qquad q = 1.009,$$
$$M_1\colon\ [0.0088\quad -0.00284\quad 1.004],\qquad q = 1.0057.$$

В ходе первого обучения получены достаточно хорошие результаты по поиску местоположения. Проведем еще одно обучение, но теперь укажем 1000 эпох с возможностью автоматической остановки обучения при минимальных изменениях потерь (если изменение значения функции потерь будет меньше 10^{-11}).

Обучение ИНС № 2

Весовые коэффициенты, отвечающие за местоположения источников и их мощности, инициализированные случайными значениями, равны:

$$M_0\colon\ [\,0.0158\quad 0.0856\quad -0.0456],\qquad q = -0.006,$$
$$M_1\colon\ [0.0034\ 0.004\quad 0.041],\qquad q = -0.025.$$

Результат поиска, который длился 75 эпох по 3.5 секунды каждая. Обучение автоматически остановилось ввиду минимальных изменений весов (местоположений и мощностей источника).

$$M_0\colon\ [\,-2.68^{-09}\quad 3.11^{-09}\quad -1.00],\qquad q = 1.0,$$
$$M_1\colon\ [-7.66^{-09}\ 2.09^{-08}\quad 9.99^{-01}],\qquad q = 0.999.$$

Результаты показывают, что в ходе обучения ИНС точно показывает местоположения источников, поэтому проведем еще одно обучение ИНС, в ходе которого у нас будет «ошибочная» информация о количестве источников, а именно будем считать, что идет поиск местоположения 5 источников.

Обучение ИНС № 3

$$
\begin{aligned}
&M_0\colon\ [0.0492\ \ 0.045\ \ 0.04], && q = 0.001,\\
&M_1\colon\ [0.14\ \ \ 0.022\ \ -0.01], && q = -0.058,\\
&M_2\colon\ [0.105\ \ -0.003\ \ -0.0298], && q = 0.003,\\
&M_3\colon\ [-0.019\ \ -0.063\ \ -0.049], && q = 0.007,\\
&M_4\colon\ [0.017\ \ 0.027\ \ 0.119], && q = -0.0047.
\end{aligned}
$$

Результат поиска:

$$
\begin{aligned}
&M_0\colon\ [-0.023\ \ 0.006\ \ 0.165], && q = 1.15^{-06},\\
&M_1\colon\ [-0.005\ \ \ 0.012\ \ -0.213], && q = -3.49^{-06},\\
&M_2\colon\ [5.8^{-05}\ \ -5.51^{-05}\ \ -9.91^{-01}], && q = 0.471,\\
&M_3\colon\ [-5.54^{-05}\ \ 5.31^{-05}\ \ -1.01], && q = 0.529,\\
&M_4\colon\ [3.35^{-07}\ \ -9.36^{-08}\ \ 1.00], && q = 1.00.
\end{aligned}
$$

Обучение длилось на протяжении 90 эпох по 4.5 секунды и автоматически остановилось ввиду минимальных изменений весов (местоположений и мощностей источников).

По результатам обучения ИНС показала корректные местоположения источников, т.к. источники M_0 и M_1 имеют мощность, близкую к 0, а следовательно, можно ими пренебречь. Источники M_2 и M_3 имеют местоположения, близкие к искомому местоположению M_1', и в сумме их мощность равняется 1. А источник M_4 с достаточно хорошей точностью близок к источнику M_2' как по местоположению, так и по мощности.

Видно, что использование метода обучения ИНС для поиска конфигурации течения имеет смысл, но необходимо дальнейшее исследование для поиска условий применения, степени значимости априорной информации о течении, а также поиска необходимого количества начальных данных, достаточных для решения задачи.

ЛИТЕРАТУРА

1. Галкин В. А., Смородинов А. Д., Моргун Д. А. Решение уравнения Навье–Стокса для сталкивающихся потоков. *Успехи кибернетики.* 2023;4(2):8–15. DOI: 10.51790/2712-9942-2023-4-2-01.
2. Смородинов А. Д., Гавриленко Т. В., Дубовик А. О., Моргун Д. А. Реализация программного обеспечения для решения задачи Коши с различными конфигурациями полей начальных данных и визуализации результатов математического моделирования течения тяжелой примеси. *Успехи кибернетики.* 2024;5(2):35–45. DOI: 10.51790/2712-9942-2024-5-2-04.
3. Колмогоров А. Н. О представлении непрерывных функций нескольких переменных в виде суперпозиций непрерывных функций одного переменного и сложения. *Докл. АН СССР.* 1957;114(5):953–956.
4. Weierstrass K. Uber die analystische Darstellbarkeit sogenannter willkűrlicher Funktionen reeller Argumente. *Sitzungsbericht der Kőnig. Academie der Wissenschaften*. 1885.
5. Sprecher D. On the Structure of Continuous Functions of Several Variables. *Transactions of the American Mathematical Society*. 1965;115(3):340–355. DOI: 10.2307/1994273.
6. Cybenko G. Approximation by Superpositions of a Sigmoidal Function. *Mathematics of Control, Signals, and Systems.* 1989;2(4):303–314. CiteSeerX: 10.1.1.441.7873. DOI: 10.1007/BF02551274.
7. Jordan M. I. *Serial Order: A Parallel Distributed Processing Approach*. ICS Report 8604.Institute for Cognitive Science, University of California, San Diego, 1986.
8. Jordan M. I. Attractor Dynamics and Parallelism in a Connectionist Sequential Machine. *Proceedings of the Annual Meeting of the Cognitive Science Society*. 1986:8
9. Jordan M. I. Serial Order: A Parallel Distributed Processing Approach. *Advances in Psychology*. 1997;121:471–495.
10. Николенко С., Кадурин А., Архангельская Е. *Глубокое обучение*. СПб.: Питер, 2018. 480 с. ISBN: 978-5-496-02536-2.

11. Калиткин Н. Н. *Численные методы.* СПб.: БХВ-Петербург; 2011. 592 с.
12. Галкин В. А., Дубовик А. О., Смородинов А. Д. Моделирование и визуализация седиментации в течении вязкой несжимаемой жидкости. *Научная визуализация*. 2024;16(3):71–78. DOI: 10.26583/sv.16.3.07. EDN: KNUTBP.
13. Галкин В. А., Дубовик А. О., Смородинов А. Д. Моделирование течения вязкой несжимаемой жидкости с учетом наличия примеси в поле течения. *Информатика, управляющие системы, математическое и компьютерное моделирование (ИУСМКМ-2024).* XV Международная научно-техническая конференция в рамках X Международного научного форума Донецкой Народной Республики. Донецк: Донецкий национальный технический университет; 2024:132–139. EDN: WNJHRZ.
14. Butcher J. C. *Numerical Methods for Ordinary Differential Equations*. 2 edition. New Zealand: Wiley, 2003. ISBN: 0-471-96758-0.
15. Hairer E., Nørsett S. P., Wanner G. *Solving Ordinary Differential Equations I: Nonstiff Problems.* 2 edition. Berlin: Springer-Verlag; 2008. ISBN: 978-3-540-56670-0.
16. Cash J. R., Karp Alan H. A Variable Order Runge–Kutta Method for Initial Value Problems with Rapidly Varying Right-Hand Sides. *ACM Trans. Math. Softw*. 1990;16(3):201–222.
17. *Программный комплекс для решения задачи Коши с различными конфигурациями полей начальных данных и визуализации результатов математического моделирования течения тяжелой примеси*: свидетельство о государственной регистрации программы для ЭВМ № 2024689709, Российская Федерация / Смородинов А. Д., Гавриленко Т. В., Дубовик А. О., Моргун Д. А. EDN: XBPJPY.

ПРОБЛЕМА ОТСУТСТВИЯ ПСИХОЭМОЦИОНАЛЬНОЙ СОСТАВЛЯЮЩЕЙ В РАЗРАБОТКЕ УЧЕБНЫХ ТЕКСТОВ

А. Р. Гончаров[1,а], **С. А. Лысенкова**[1,б]
[1] *Сургутский государственный университет, Сургут, Российская Федерация*
[а] *ORCID: https://orcid.org/0009-0004-5338-3110, ✍ mr.tesanta@ya.ru*
[б] *ORCID: https://orcid.org/0000-0003-1007-7610, lsa1108@mail.ru*

Аннотация: в статье сформулирована проблема при разработке учебных материалов: не уделяется должное внимание психоэмоциональной составляющей как характеристике, влияющей на восприятие учебного текста. Значимость данной проблемы подчеркивается на этапе распространения личностно ориентированного и самостоятельного подходов к образованию с применением генеративных языковых моделей в формировании образовательного контента. Произведен анализ традиционной и инновационной технологии разработки учебных текстов, где выявлена необходимость в методах тонального анализа для формирования определенного окраса учебных текстов. Применение системы анализа эмоциональных окрасов учебных текстов позволяет адаптировать учебные материалы под требуемые условия и повысить их качество восприятия учащимися, устранив психоэмоциональные барьеры в образовательной деятельности.

Ключевые слова: учебный текст, технология разработки текстов, психоэмоциональная составляющая, эмоциогенность текста, тональный анализ, генеративный искусственный интеллект.

Для цитирования: Гончаров А. Р., Лысенкова С. А. Проблема отсутствия психоэмоциональной составляющей в разработке учебных текстов. *Успехи кибернетики*. 2025;6(1):150–157.

Поступила в редакцию: 01.12.2024. *В окончательном варианте: 21.12.2024.*

LACK OF PSYCHOEMOTIONAL CONTENT IN TRAINING TEXTS

A. R. Goncharov[1,a], **S. A. Lysenkova**[1,b]
[1] *Surgut State University, Surgut, Russian Federation*
[a] *ORCID: https://orcid.org/0009-0004-5338-3110, ✍ mr.tesanta@ya.ru*
[b] *ORCID: https://orcid.org/0000-0003-1007-7610, lsa1108@mail.ru*

Abstract: this article addresses a key issue in the development of educational materials: the lack of attention to the psychoemotional component, which affects how students perceive educational texts. This issue becomes particularly significant with the rise of personalized and independent learning approaches that use generative language models to create educational content. We analyze both traditional and innovative methods for developing educational texts and highlight the need for tone analysis techniques to shape their emotional impact. Implementing a system for analyzing the emotional tone of educational texts enables the adaptation of materials to specific conditions, improving their readability and helping students overcome psychoemotional barriers in learning.

Keywords: educational text, text development technology, psychoemotional component, emotionality of text, tonal analysis, generative artificial intelligence.

Cite this article: Goncharov A. R., Lysenkova S. A. Lack of Psychoemotional Content in Training Texts. *Russian Journal of Cybernetics*. 2025;6(1):150–157.

Original article submitted: 01.12.2024. *Revision submitted: 21.12.2024.*

Введение

Поиск путей и методов улучшения образовательной системы остается и будет оставаться актуальной исследовательской задачей, так как эффективное и качественное обучение подрастающего поколения играет ключевую роль в стратегическом развитии любого независимого государства. В настоящее время существует достаточно работ [1–4], посвященных описанию текущей ситуации в образовательной деятельности с выявлением проблем и предложений по ее улучшению, среди которых выделяют необходимость в модернизации образования под новые реалии с акцентом на личностно ориентированный подход [5, 6] где в том числе формируется задача о необходимости формирования новых учебных материалов.

Сформированные в советское время учебные пособия до сих пор обсуждаются, и делаются выводы об их содержании и эффективности [7], что не всегда можно сказать о современных образовательных материалах. В недолгом постсоветском периоде разработка учебных текстов происходила традиционно со строгой верификацией содержания и цели трансляции необходимых моделей поведения, норм, ценностей, идеалов подрастающему поколению. Распространение информационных технологий и дальнейшее развитие компьютерных сетей изменили ситуацию: огромный, растущий массив информации, возможность создания открытых платформ и доступность каждому транслировать в массы знания, мнения и суждения усложнили ситуацию верификации данных. Появление государственных образовательных платформ и стандартов призваны сохранить качество, и актуальность познавательной информации. Тем не менее, дальнейшее развитие методов и средств педагогики и внедрение инновационных технологий, в том числе сегодняшнее внедрение систем искусственного интеллекта с генеративными языковыми моделями [8] значительно усложнили ситуацию разработки учебных текстов, так как формирование таких материалов опирается на тот самый открытый массив данных с не всегда гарантированно достоверными сведениями. В интенсивном темпе развития и внедрения таких технологий выявляется тенденция, что в скором будущем мы «утонем» в этих самых сгенерированных текстах, которые будут претендовать на образовательный контент. Это создает угрозу достоверности восприятия информации. Поэтому, при внедрении инновационных технологий в образование необходимо задуматься о дополнительной проверке качества материалов, в том числе и оценивать их валидность — степень пригодности теста для его использования в заданном контексте [9].

Разработка и оценка учебных материалов происходят через призму объективности, где валидность текстов проверяется с учетом измерения успеваемости учащихся и их усвоения знаний непосредственно через контрольные тесты или другие оценочные работы в зависимости от множества субъективных факторов [10]. С распространением тенденции дистанционного образования в качестве нового фактора, влияющего на образовательную деятельность, выявлена психоэмоциональная составляющая. Так, показатели самочувствия, активности и настроения учащихся влияют на результаты тестирования, возникновение переутомления и тревожности, формирование самооценки и заинтересованности в обучении [11, 12]. Кроме того, эмпатийные способности педагогов [13] и эмоциогенность текста [14] (эмоциональное влияние на адресата) входят в комплексную «психоэмоциональную составляющую» и вместе оказывают влияние на успеваемость и результативность образовательной деятельности.

Если принять данность, что прочитанный текст вызывает определенные эмоции, тогда учебный текст в условиях ориентированного подхода и самостоятельного образования становится одним из факторов формирования психоэмоционального состояния у учащихся. Оно в последующем может образовать диссонанс восприятия материала, делая обучение менее эффективным. В психологии эмоций под словом «диссонанс» подразумевают такое эмоционально-отрицательное состояние человека, которое он стремится преодолеть путем пересмотра ожиданий или приобретения новой информации, преодолевая тем самым внутренние дискомфорт и противоречия [15]. Поэтому важно, чтобы разработчики учебных материалов учитывали эти психоэмоциональные аспекты.

Психоэмоциональной составляющей, как характеристике влияния на восприятие учебных текстов, не уделяется должное внимание в процессе разработки материалов, так как сложно измерить и оценить эмоциональные аспекты человека с научной точки зрения. Они включают субъективные переживания, требующие специализированных методологий и инструментов измерения, таких как психометрические тесты, а их составление и применение может быть трудоемким процессом. В качестве альтернативного подхода к предварительной оценке психоэмоциональной составляющей учебного текста в контексте его восприятия могут выступать методы тонального анализа [16, 17]. Определяя тональность учебного текста (эмоционально классифицируя его), можно в дальнейшем вычислить его эмоциогенность и при помощи языковых моделей [18] изменить эмоциональный окрас текста для повышения восприимчивости.

Таким образом, тональный анализ позволяет выявить, являются ли тексты позитивными, негативными или нейтральными, и насколько сильно выражены те или иные эмоции, что полезно для выявления потенциальных проблемных участков текста, которые могут вызвать стресс или снизить мотивацию у учащихся.

Материалы и методы

Рассмотрим существующие технологии разработки учебного текста, среди которых выделяют традиционный и инновационный подходы.

Текст, призванный называться «учебным», представляет из себя не просто подготовленный материал для осуществления образовательного процесса, но также является «культурным» продуктом, раскрывающим необходимые значения и смыслы, заложенные создателем [19]. То есть, текст представляет из себя набор фактов, сведений, практик, обладающий строгой формализацией и несущий тот самый компонент необходимых знаний. Его необходимо предварительно адаптировать или преобразовать, снабдив его авторским видением, опытом и определенной культурой. Так создаются сегодняшние образовательные материалы в дидактических целях, формируя предпосылки для представления поэтапного процесса создания учебных текстов, в основу которого закладывается преобразование первичного (исходного) текста во вторичный (учебный).

Несмотря на то, что создание учебного текста в основном является творческой деятельностью, существуют ряд моделей и технологий педагогического проектирования [20], среди которых распространена ADDIE модель разработки образовательных программ и соответствующих материалов. Так, создание учебных текстов из заранее недетерминированного процесса сводится к некоторой заданной установке (с нестрогой привязанностью), позволяющей моделировать процесс разработки и анализировать связанные подпроцессы.

Большинство исследовательских работ, посвященных разработке учебных текстов [21–24], отражают личный опыт в проектировании материалов различных образовательных направлений (особенного гуманитарных) и предлагают теоретические аспекты по их улучшению в новых образовательных условиях с применением цифровых платформ. Резюмируя существующий опыт и рекомендации педагогического проектирования, можно представить общую схему традиционной технологии разработки учебных текстов, основываясь на существующих первичных текстах (рис. 1). Под традиционной разработкой учебных текстов подразумевается линейная организация содержания с обязательным соответствием федеральным государственным образовательным стандартам (ФГОС), одним из требований которых является обеспечение контроля и оценки освоения материала учащимися [25].

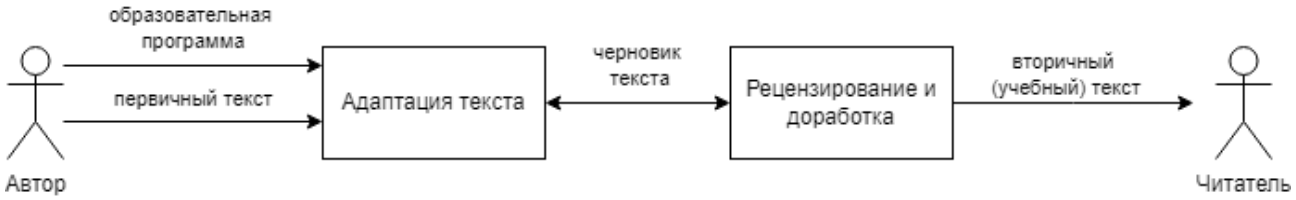


Рис. 1. *Традиционная технология разработки учебного текста*

Традиционная разработка учебного текста силами автора (или авторов) подразумевает адаптацию первичного текста с последующим его рецензированием. Первичный текст — основной источник информации, используемый как базис в разработке учебного материала. Им может выступать научная статья, учебник, доклад и любой другой фактологический документ, содержащий сведения по необходимой теме. Из данного источника извлекаются необходимые сведения, синтезируются с последующей адаптацией для создания нового учебного материала. Процесс адаптации подразумевает переработку содержания с учетом образовательных целей, концепций и требований. Работа автора может основываться как на существующей образовательной программе, так и на новых концепциях под требуемый уровень образования (школьное, среднее специальное, высшее) и специализации (гуманитарные науки, естественные науки и т.п.) при наличии компетенций, с последующей ее аккредитацией в соответствующих государственных органах. Немаловажным этапом в традиционной разработке любых учебных материалов является проведение рецензирования с получением обратной связи от экспертов, в качестве которых могут выступать как другие авторы, так и заранее созданная редакционная коллегия. Рецензирование проводится с целью не только выявления языковых ошибок, но и проверки фактической точности соответствия материала заданным стандартам с дальнейшей рекомендацией по улучшению стиля изложения и структуры текста. Так, традиционная технология разработки учебных материалов требует ответственности и значительного внимания к деталям с глубоким пониманием педагогических подходов для обеспечения стандартизированного и достоверного учебного текста.

В настоящее время, когда окружающий нас мир обретает качества нестабильности и противоречивости, становясь все более изменчивым, образовательная деятельность вынуждена подстраиваться под новые условия для адекватного постижения человеком реальности. Поэтому, появление новых пе-

дагогических подходов улучшения педагогической деятельности [26], а также модернизация учебных материалов являются необходимыми мерами для осуществления образования в заданных условиях. Под влиянием развития информационных технологий, внедряются инновационные средства и методы, например, генеративные языковые модели, получившие широкое внимание и обсуждение применения в сфере образования [27, 28]. Такие модели способны генерировать текст, улучшая и адаптируя его под конкретные требования. Следовательно, данная инновационная технология способна автоматизировать процесс разработки учебного текста (рис. 2). Важно отметить, что под понятием инновационная технология понимается новый или улучшенный подход, способный обеспечивать более эффективный результат, чем его предыдущая версия. В случае образования, инновационная технология должна также акцентировать внимание на развитии механизмов интеллектуальной рефлексии и коммуникационных способностей [29] желательно всех заинтересованных сторон в образовательном процессе.

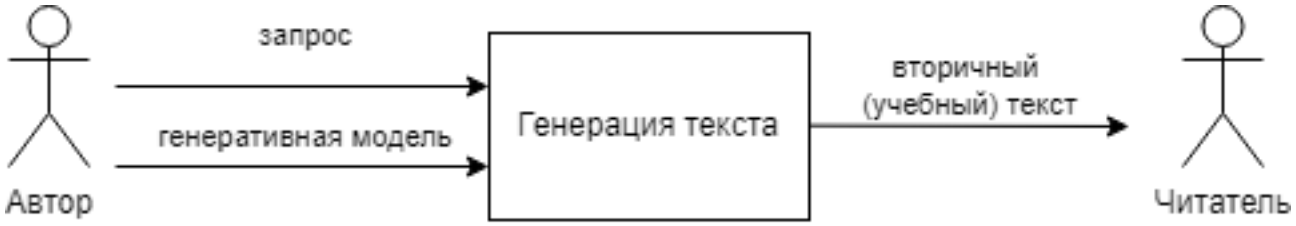


Рис. 2. *Инновационная технология разработки учебного текста*

Инновационная разработка учебного текста подразумевает автоматическое создание и адаптацию учебного материала, соответствующего образовательным стандартам и индивидуальным потребностям автора или конечного потребителя, на основе заданной образовательной программы, которая представляется в виде запросов для модели с описанием целей, задач и концепций требуемого итогового текста. Генерация текста происходит при помощи генеративной языковой модели, заранее обученной на доступных базах данных и ресурсах (научные статьи, учебники и т.п.). Автор задает запрос, устанавливая предварительно тему, содержание, уровень сложности текста, указывая включение конкретного материала и другие аспекты, которые должны быть раскрыты в материале. Основную работу по созданию текстов выполняет генеративная модель. Автор должен проявить творческий подход к формулировке запроса, чтобы итоговый результат полностью раскрывал его первоначальные цели в создании учебного материала.

Любой запрос Q к генеративной языковой модели можно представить в виде совокупности параметров (настроек) $T = (t_1, t_2, \ldots t_n)$, которые модель должна воспринимать и учитывать при генерации текста. Например, при использовании правила «T-A-G» (Задача-Действие-Цель) запрос Q можно представить в виде $Q = (t_1, t_2, t_3)$, где t_1 — описание задачи для решения модели, t_2 — действия, которые модель должна совершить над данными, t_3 — описание данных, которые необходимо получить.

Таким образом, генеративная модель должна учитывать специфические требования автора и применять заданные параметры T в задаче генерации текста (структурированность, стиль изложения, глубина информации и т.п.). Важно отметить, что в концепции инновационной разработки учебных текстов на основе генеративных языковых моделей создание и отслеживание промежуточных результатов не предусмотрено. Так, процесс разработки текстов представляется «черным ящиком» для авторов и тем более экспертов, в связи с этим процесс рецензирования осуществим лишь только на финальной стадии разработки на готовом сформированном материале. Несмотря на представленное ограничение, данный подход обеспечивает высокую скорость разработки и адаптации учебных материалов под определенные требования, что делает его перспективным.

Результаты и обсуждение

При создании учебных материалов авторы используют как традиционные, так и инновационные подходы, включая использование генеративного искусственного интеллекта. Оба подхода разработки учебных текстов имеют свои преимущества и недостатки (таблица 1), однако в них отсутствует компонент анализа психоэмоциональной составляющей.

В традиционных подходах к созданию учебных материалов основное внимание уделяется структуре и содержанию текста, соблюдению дидактических принципов и соответствию образователь-

ным стандартам. Однако такие материалы могут иногда быть «сухими» или сложными, что снижает уровень вовлеченности учащихся и может вызвать у них чувство скуки. Инновационный подход на основе генеративного искусственного интеллекта позволяет создавать адаптивные и интерактивные учебные материалы, которые проще персонализировать под потребности конкретного учащегося. С помощью искусственного интеллекта можно быстро генерировать различные вариации текстов, учитывающие уровень подготовки и интересы учеников. Тем не менее, без анализа эмоциональной составляющей такие материалы могут оставаться недостаточно эффективными.

Таблица 1

Анализ технологий в разработке учебных текстов

	Традиционная технология	**Инновационная технология**
Преимущества	• Разработка в соответствии со стандартами • Гарантированная достоверность информации • Валидность материалов	• Высокая адаптивность и индивидуализация материалов • Уменьшение времени на разработку • Поддержка различных форматов материалов
Недостатки	• Больше времени на разработку • Ограниченность в актуализации материала • Отсутствие компонента индивидуализации	• Отсутствие стандартизации • Проблемы с достоверностью информации • Сложность реализации разработки

Для учета эмоциональной составляющей, например, в инновационной технологии разработки учебных текстов, было бы целесообразно интегрировать системы анализа эмоционального окраса и стиля предложений или отдельных слов на этапах разработки учебных материалов (рис. 3). Такие системы должны оценивать текст или его определенные блоки по заданным параметрам эмоциональной тональности и давать рекомендации по улучшению их эмоциональной «привлекательности». Такие аспекты, как выбор менее агрессивных формулировок, добавление ярких примеров, избегание чрезмерно сложной лексики, способной вызвать недоумение и чувство подавленности у учащихся в сформированных учебных текстах.

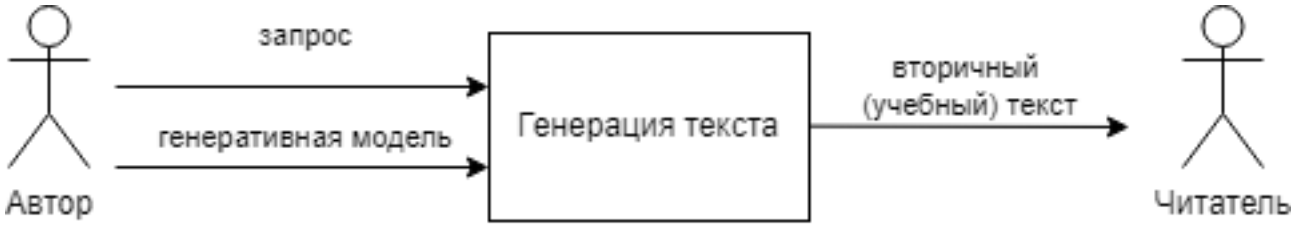


Рис. 3. *Анализ эмоционального окраса текста в инновационной технологии разработки*

Задачу анализа эмоционального окраса текста можно представить следующим образом. Пусть $W = \{w_1, w_2, \dots w_n\}$ — множество текстовых блоков, где каждый w_i представляет собой отдельный фрагмент общего анализируемого текста, $E = \{e_1, e_2, \dots\ e_m\}$ — множество возможных эмоциональных окрасов (например, позитивный, нейтральный, негативный и т.д.). Установим отображение $F : W \to E$, соотнеся каждый текстовый блок с его тональностью из множества E. Так, модель для анализа текста может быть выражена в виде определения пар элементов:

$$(w_i, F(w_i)) \to \left(w_i, e_j\right), \quad \forall\, w_i \in W,$$

где:

w_i — анализируемый блок текста,

e_j — тональность данного блока текста.

Эта модель предполагает, что имеется алгоритм или модуль для автоматического определения тональности текста любого текстового блока. Реализация такой модели может быть основана на методах обработки естественного языка (NLP) и машинного обучения, таких как классификация текста с использованием предварительно обученных языковых моделей.

Пусть, для текстового блока w заданы несколько вариаций содержания с разной тональностью, тогда формируются пары возможных вариантов $\{(w,e_1),(w,e_2)\ldots(w,e_n)\}$ написания данного блока под разный эмоциональный окрас: $V = \{v_1,v_2,\ldots v_n\}$. Система может оперировать данными парами вариантов написания и выбирать (заменять) содержимое текста под заданную целевую установку автора.

Рассматривая учебное пособие по информационным технологиям в цифровой экономике [30], в содержании текста можно выделить абзац (текстовый блок), содержащий определение понятия «информация». Если переформулировать текст под определенный окрас при помощи генеративного искусственного интеллекта, то можно получить совершенно разное восприятие данного определения (таблица 2).

Таблица 2

Примеры изменения текста под разный эмоциональный окрас

	Текст
Исходный текст	Информация — это совокупность знаний о предметах, фактах, явлениях и процессах, представленных в определенной форме, которые позволяют людям расширять и развивать свои знания, повышать уровень образования и принимать обоснованные решения в различных областях деятельности
Позитивный окрас	Информация — это богатый источник знаний о предметах, фактах, явлениях и процессах, оформленный так, чтобы вдохновлять людей на расширение их кругозора и развитие. Она способствует повышению уровня образования и помогает принимать обоснованные решения в разнообразных сферах деятельности, приводя к успешным и плодотворным результатам
Негативный окрас	Информация — это сложная совокупность данных о предметах, фактах, явлениях и процессах, представленная в форме, которая может перегружать и сбивать с толку. Вместо того чтобы определенно способствовать развитию знаний и образовательного уровня, она может создавать сложности в принятии обоснованных решений, усложняя понимание в различных областях деятельности

Важно отметить, что в работах тонального анализа различных текстов чаще всего идет процесс классификации тональности с точки зрения автора текста (например, задача анализа комментария или отзыва о товаре, услуге). Однако, в контексте анализа учебных текстов необходимо идентифицировать предполагаемую тональность с точки зрения читателя (учащихся), так как в образовании авторы стараются соблюдать «эмоциональный нейтралитет» — учебные материалы должны быть тщательным образом адаптированы, чтобы исключить личные предубеждения и обеспечить объективную подачу информации. Такой нейтралитет ведет к возникновению «мертвых текстов» — отсутствию стимуляции интереса и эмоционального вовлечения. Тем не менее это не дает установку на то, чтобы наделять тексты как можно более позитивной тональностью и устранять негативные. Необходимо соблюдать баланс и варьировать стиль и тон подачи информации в зависимости от контекста и целевой аудитории, конечно, учитывая особенности восприятия и психологические аспекты учащихся. Например, описывая исторические войны в текстах можно вызывать у учащихся эмоциональное отторжение определенных факторов, чтобы закрепить значимость данных событий. Такой подход начинает включать концепции инклюзивного языка и задает практическую ценность и актуальность текста, а это способствует лучшему усвоению материала.

Заключение

Инновационные технологии в разработке учебных текстов имеют значительный потенциал для модернизации образовательного процесса под новые реалии, делая его более адаптивным и эффективным. Интеграция психологических знаний и практик в процесс разработки и оценки учебных материалов является важным шагом на пути к созданию более эффективной образовательной системы. В ней успехи учащихся должны зависеть от их знаний и умений, а не от различных психоэмоциональных барьеров: от недостаточной мотивации и неверных педагогических стратегий до когнитивных и эмоциональных трудностей. Включив психологическую составляющую в процесс разработки учебных материалов можно существенно улучшить их доступность и эффективность.

Современные технологии на основе генеративных моделей позволяют реализовать системы анализа эмоционального окраса учебных текстов при создании нового контента. Они могут быть задействованы в анализе уже существующих материалов, что может значительно обогатить образовательный процесс и повысить качество восприятия информации учащимися.

Важно отметить, что помимо субъективных эмоций еще одним важным аспектом является культурный контекст. Эффективные учебные материалы должны учитывать культурные различия, которые могут влиять на восприятие и интерпретацию текста. А также вопрос о достоверности генерируемых текстов, где модели обучаются на данных из различных источников.

Если предположить, что вопрос о достоверности данных решен, то интеграция психоэмоционального анализа в процесс создания учебных материалов может стать стандартной практикой, значительно повысив эффективность обучения. Это поможет быстро создавать материалы, которые не только информативны, но и вызывают определенный эмоциональный отклик у учащихся, способствующий нелинейному обучению и личностному развитию.

ЛИТЕРАТУРА

1. Новожилов А. М. Проблема образования и квалификации как философская проблема информационного общества. *Позиция. Философские проблемы науки и техники*. 2024;21:132–140. EDN: OBOOHI.
2. Швецова Н. А., Хорошева Е. И. Проблема ценностей и ценностного отношения к образованию в условиях информатизации общества. *Евразийский гуманитарный журнал*. 2021;1:101–111. EDN: TJADND.
3. Хивренко А. В., Соляников В. С. Дистанционное образование как форма современного образования. *Инновационные научные исследования*. 2021;11–1(13):118–126. DOI: 10.5281/zenodo.5732706.
4. Старикова А. В. Познавательные барьеры в обучении студентов. *Вестник Московского университета. Серия 20: Педагогическое образование*. 2014;1:106–114. DOI: 10.51314/2073–2635-2014-1-106–114.
5. Учурова С. А. Современный педагог: новый мир – новая педагогика – новые роли. *Вопросы педагогики*. 2022;3–2:242–246. EDN: CPTTIB.
6. Аванесов В. Применение педагогических измерений и новых образовательных технологий для модернизации образования. *Педагогические измерения*. 2015;1:3–28. EDN: TUFWCB.
7. Дорогавцева И. С. Советский учебник как объект ностальгии в современной российской культуре. *Ученые записки Забайкальского государственного университета*. 2024;1(19):80–90. DOI: 10.21209/2658–7114-2024-19-1-80–90.
8. Вишнеков А. В., Ерохина Е. А., Иванова Е. М., Трубочкина Н. К. Особенности учебного процесса подготовки it-специалистов в условиях возрастания возможностей генеративного искусственного интеллекта. *Инженерное образование*. 2023;34:123–135. DOI: 10.54835/18102883_2023_34_11.
9. Мороз Л. С. Методы определения надежности и валидности тестов для контроля знаний. *Труды БГТУ. Физико-математические науки и информатика*. 2010;6:176–179. EDN: TGUDXD.
10. Пастухова Е. В., Романова Ю. С., Афанасьева О. В. Анализ эффективности социально-образовательных учреждений: методы и модели мониторинга успешности учащихся. *Научно-методический электронный журнал «Концепт»*. 2023;12:104–119. DOI: 10.24412/2304–120X-2023-11129.
11. Найденова Е. В., Морозова А. С. Взаимосвязь между адаптированностью и психическим состоянием в учебной деятельности у студентов 1 курса. *Инновационная наука: психология, педагогика, дефектология*. 2023;4(6):99–108. DOI: 10.23947/2658–7165-2023-6-4-99–108.

12. Усынина К. М., Михайлова К. В. Особенности психоэмоционального состояния обучающегося в условиях дистанционного обучения. *Инновационная наука*. 2021;7:141–144. EDN: ZQGYHO.
13. Сизова В. В., Крижовецкая О. М. Педагогическая эмпатия как фактор успеваемости учащихся. *Вестник Тверского государственного технического университета. Серия: Науки об обществе и гуманитарные науки*. 2021;2(25):48–53. DOI: 10.46573/2409–1391-2021-2-48–53.
14. Пиотровская Л. А. Эмоциогенность текста: психологический и лингвистический аспекты. *Вестник Волгоградского государственного университета. Серия 2: Языкознание*. 2023;1(22):28–38. DOI: 10.15688/jvolsu2.2023.1.3.
15. Митина Г. В. *Психология эмоций и мотивации*. Уфа: БГПУ; 2020. 110 с.
16. Лукашевич Н. В. Автоматический анализ тональности текстов: проблемы и методы. *Интеллектуальные системы. Теория и приложения*. 2022;1(26):50–61. EDN: IIBYWS.
17. Майорова Е. В. О сентимент-анализе и перспективах его применения. *Социальные и гуманитарные науки. Отечественная и зарубежная литература. Серия 6: Языкознание*. 2020;4:78–86. EDN: TAGOBD.
18. Мусаев Р. Р., Гериханов З. А. Языковые модели в образовании. *Современные тенденции развития информационных технологий в научных исследованиях и прикладных областях:* сборник докладов IV Международной научно-практической конференции. Владикавказ. 2023:90–94. EDN: BDTIKT.
19. Сенько Ю. В., Мясникова О. В. Создание учебных текстов в процессе обучения. *Мир науки, культуры, образования*. 2016;4(59):71–73. EDN: WWHNVJ.
20. Карамзина А. Г., Сильнова С. В. Системный анализ и моделирование процесса разработки программ дополнительного профессионального образования. *Вестник Воронежского государственного университета. Серия: Системный анализ и информационные технологии*. 2021;2:94–108. DOI: 10.17308/sait.2021.2/3507.
21. Сысоев П. В. Этапы разработки учебных материалов для предметно-языкового интегрированного обучения. *Вестник Московского университета. Серия 19. Лингвистика и межкультурная коммуникация*. 2020;3:169–178. EDN: KVBRCL.
22. Матвеева Е. И. Этапы создания программ и учебников по литературному чтению (из опыта работы 2000-х годов). *Проблемы современного образования*. 2014;6:201–218. EDN: TFQCZH.
23. Пластинина Н. А., Григорьева Е. С. Создание базового образовательного контента для дистанционного обучения. *Вестник Нижневартовского государственного университета*. 2021;1(53):48–55. DOI: 10.36906/2311–4444/20-4/07.
24. Одинокая М. А., Баринова Д. О., Андреева А. А. Принципы разработки учебных дидактических материалов нового типа при иноязычной подготовке магистрантов инженерного профиля. *Азимут научных исследований: педагогика и психология*. 2020;3(9):197–200. DOI: 10.26140/anip-2020-0903-0044.
25. Ефимов В. Н., Ефимова Н. Ю., Лазурцева Е. А. Теоретические основания проектирования образовательного процесса, ориентированного на ФГОС, ГИА и ЕГЭ. *Муниципальное образование: инновации и эксперимент*. 2012;4:8–14. EDN: PBTILF.
26. Раковская О. Л. Пути улучшения качества образовательного процесса. *Alma Mater (Вестник высшей школы)*. 2014;4:113–117. EDN: SARGML.
27. Паскова А. А. Возможности интеграции технологий генеративного искусственного интеллекта в процессы формирующего оценивания в высшем образовании. *Вестник Майкопского государственного технологического университета*. 2024;2(16):98–109. DOI: 10.47370/2078–1024-2024-16-2-98–109.
28. Паскова А. А. Практические аспекты применения ChatGPT в высшем образовании. *Вестник Майкопского государственного технологического университета*. 2023;3(15):67–74. DOI: 10.47370/2078–1024-2023-15-3-67–74.
29. Апаева С. Х., Каирбекова Б. Д., Цой В. И., Мельничук Ю. С. Инновационные инструменты как средства педагогического образования и деятельности. *Вестник Инновационного Евразийского университета*. 2024;1(93):24–30. DOI: 10.37788/2024-1/24–30.
30. Тихонов Д. В. *Информационные технологии в цифровой экономике*. М.: Финансовый университет; 2024. 144 с. ISBN: 978-5-6048031-7-2.

БИБЛИОТЕКА PYPHARM КАК ИНСТРУМЕНТ ДЛЯ РЕШЕНИЯ ЗАДАЧ ФАРМАКОКИНЕТИКИ

Р. С. Крашенинников[а], И. И. Митричев[б], М. А. Усова[в]
Федеральное государственное бюджетное образовательное учреждение высшего образования «Российский химико-технологический университет имени Д. И. Менделеева», г. Москва, Российская Федерация
[а] ORCID: https://orcid.org/0009-0007-3434-031X, ✍ krasheninnikov.r.s@muctr.ru
[б] ORCID: https://orcid.org/0000-0001-5733-3226, mitrichev.i.i@muctr.ru
[в] ORCID: https://orcid.org/0009-0002-9034-7165, 204137@muctr.ru

Аннотация: в последние годы заметно возросла популярность моделирования при создании медицинской продукции. Особое значение приобретает изучение фармакокинетики разрабатываемых лекарственных средств. Определение параметров таких моделей представляет собой задачу многомерной оптимизации, которую успешно решают эволюционные алгоритмы. В данном исследовании представлена новая открытая библиотека на языке Python, предназначенная для фармакокинетического моделирования и оптимизации параметров фармакокинетических моделей с использованием эволюционных алгоритмов.

Ключевые слова: оптимизация, эволюционный алгоритм, генетический алгоритм, вдохновленный природой алгоритм, фармакокинетика, камерная модель, разработка лекарственной продукции.

Для цитирования: Крашенинников Р. С., Митричев И. И., Усова М. А. Библиотека PyPharm как инструмент для решения задач фармакокинетики. *Успехи кибернетики*. 2025;6(1):158–169.

Поступила в редакцию: 25.11.2024. *В окончательном варианте: 16.12.2024.*

THE PYPHARM LIBRARY FOR SOLVING PHARMACOKINETIC PROBLEMS

R. S. Krasheninnikov[a], I. I. Mitrichev[b], M. A. Usova[c]
Mendeleev Russian University of Chemical Technology, Moscow, Russian Federation
[a] ORCID: https://orcid.org/0009-0007-3434-031X, ✍ krasheninnikov.r.s@muctr.ru
[b] ORCID: https://orcid.org/0000-0001-5733-3226, mitrichev.i.i@muctr.ru
[c] ORCID: https://orcid.org/0009-0002-9034-7165, 204137@muctr.ru

Abstract: recently, simulation has gained significant popularity in the creation of medical products, particularly in the study of pharmacokinetics for developed drugs. Determining the parameters of these models involves a multidimensional optimization problem, which evolutionary algorithms can effectively solve. This study introduces a new open-source Python library designed for pharmacokinetic modeling and optimizing the parameters of pharmacokinetic models using evolutionary algorithms.

Keywords: optimization, evolutionary algorithm, genetic algorithm, nature-inspired algorithm, pharmacokinetics, compartment model, drug design.

Cite this article: Krasheninnikov R. S., Mitrichev I. I., Usova M. A. The PyPharm Library for Solving Pharmacokinetic Problems. *Russian Journal of Cybernetics*. 2025;6(1):158–169.

Original article submitted: 25.11.2024. *Revision submitted: 16.12.2024.*

С каждым годом моделирование все чаще применяется в разработке лекарственных препаратов, что подтверждается ростом числа упоминаний о нем в различных источниках [1–2]. Эффективность такого подхода, в частности, подтверждается экономической выгодой [3–5].

Весь процесс разработки лекарственного препарата можно разделить на три этапа: стадия исследования (дизайн молекулы), стадия тестирования (проверка оптимальной дозы и безопасности лекарства) и стадия рынка (коммерциализация лекарства) [6]. Зачастую именно на разработку тратится большая часть бюджета и времени до выхода лекарства на рынок (срок может доходить до 10 лет) [6, 7]. Чтобы ускорить данный процесс применяются методы математического моделирования.

При тестировании лекарственного средства математическое моделирование применяется для двух основных целей: исследования фармакокинетики (ФК) и фармакодинамики (ФД). Фармакокинетика занимается изучением скорости распространения лекарственного средства и/или его метаболитов

в организме и силы, действующей на этот процесс [8]. В свою очередь, фармакодинамика изучает положительное или негативное влияние лекарства на организм [7]. Другими словами, ФК можно описать как «то, что организм делает с лекарством», а ФД — как «то, что лекарство делает с организмом».

В данной работе мы сообщаем о разработке нового программного обеспечения для решения задач фармакокинетики: библиотеки PyPharm с открытым исходным кодом на языке Python. Хотя совместное рассмотрение ФК/ФД дает большее количество информации о применимости и качестве лекарственного средства, на данный момент библиотека ограничена моделированием ФК. ФК модели позволяют определить неизвестные параметры и проводить интерполяцию или экстраполяцию результатов, что, в свою очередь, способствует выбору оптимального режима дозирования препарата [9, 10].

Структура статьи следующая: сначала мы рассмотрим, какие модели вообще используются в ФК-исследований, затем приведем обзор программ-аналогов. Далее следует объяснение основного функционала библиотеки PyPharm для программиста (API) и пример ее использования. Заметим, что для использования библиотеки требуется лишь базовое понимание языка Python, который достаточно прост и сейчас широко используется в научных исследованиях.

Типы моделей ФК

В литературе выделяют два основных типа моделей фармакокинетики, а именно камерная (компартментная) модель и физиологически обоснованная фармакокинетическая модель (PBPK) [11].

Камерная модель описывает распространение лекарства через различные камеры. В каждой камере поддерживается состояние кинетической однородности, обеспечивающее равномерное и с постоянной скоростью распределение лекарственного препарата [11].

PBPK модель, в отличие от камерной модели, связывает ткани и органы с кровотоком, основываясь на физиологических и анатомических особенностях организма. Она также подробно описывает структурные элементы, используя такие параметры, как объем ткани, кровоток через ткань, коэффициент распределения между тканью и плазмой и другие [11].

Хотя модель PBPK более точно описывает фармакокинетику лекарственного средства в организме, эта модель требует большого количества входных параметров, собранных эмпирически. Также размер модели увеличивает сложность вычислений. В свою очередь, камерная модель создана с большими допущениями и, соответственно, не может идеально моделировать фармакокинетический процесс, но сбор данных, построение и решение модели являются более доступными задачами. Стоит также отметить, что для многих лекарственных препаратов моделирование с помощью камерной модели дает достаточно точные решения [12].

Виды камерных моделей

Простейший вид камерной модели представляет собой однокамерную модель, где весь организм представлен в виде гомогенной камеры, которая обменивается лекарством со внешней средой. Данная модель возможна при больших допущениях и приводит к простому математическому описанию [13].

Более сложные модели строятся из нескольких камер и называются многокамерными моделями, одна из них представлена на рис. 1. Данные камеры в большинстве случаев представляют собой имитацию какого-либо органа или ткани в организме. Однако иногда они выступают в роли абстрактных структур, не имеющих физиологического смысла.

Принцип камерных моделей может быть продемонстрирован на наиболее распространенном их виде — двухкамерной модели (рис. 1). Она может быть описана математически в виде системы дифференциальных уравнений (1)–(2) [14]

$$V_1 \frac{dc_1}{dt} = -k_{12} V_1 c_1 + k_{21} V_2 c_2 - k_{10} V_1 c_1, \tag{1}$$

$$V_2 \frac{dc_2}{dt} = k_{12} V_1 c_1 - \mathrm{k}_{21} V_2 c_2, \tag{2}$$

где c_1 — концентрация вещества в центральной камере (нг/мл), V_1 — объем центральной камеры (мл), c_2 — концентрация в периферической камере (нг/мл), V_2 — объем периферической камеры (мл), k_{12} — константа скорости переноса фармакологического вещества из центральной камеры в периферическую

(мин$^{-1}$), k_{21} — константа скорости переноса фармакологического вещества из периферической камеры в центральную (мин$^{-1}$), k_{10} — константа скорости элиминации фармакологического вещества из центральной камеры (мин$^{-1}$), все константы скорости имеют первый порядок.

Стоит отметить, что основным параметром математического моделирования фармакокинетики является концентрация. Концентрацией называют отношение количества лекарства (измеряется в мг или в молях) к объему камеры. Именно концентрациями оперируют в фармакокинетике из-за легкости их измерения в крови.

Для описания фармакокинетики препарата организм представляют как набор отдельных камер, содержащих часть введенной дозы препарата за исключением элиминированной (выведенной дозы), а скорость перемещения препарата из одной камеры в другую и выведения вещества из организма определяется константами скорости [15].

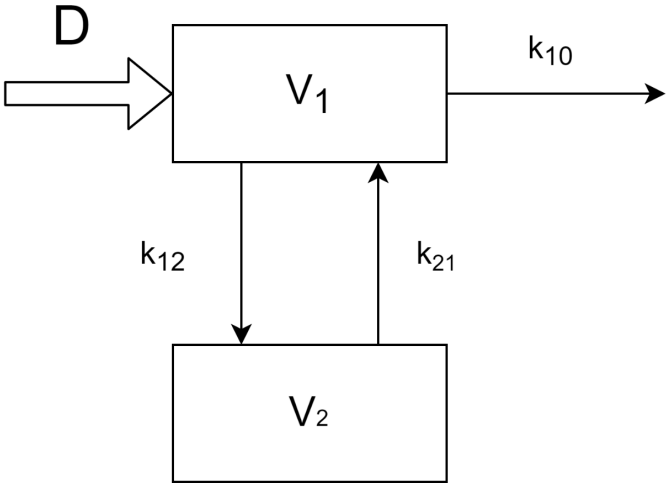


Рис. 1. *Двухкамерная модель [14]*

Математические выражения, описывающие камерную модель, получаются путем записи соответствующих балансов масс через концентрации и решения результирующей системы дифференциальных уравнений [13]. Решая данную систему, можно получить зависимости концентрации лекарственного средства в камере в зависимости от времени (t) или концентрацию лекарственного препарата в камере в определенный момент t. В камерных моделях рассматривается одно уравнение баланса массы для каждой камеры, охватываемой моделью. Таким образом, в многокамерной модели изначально требуется столько уравнений баланса масс, сколько используется камер.

Программное обеспечение

Камерные модели отличаются простотой, поэтому исследователи часто создают свое внутреннее программное обеспечение для моделирования ФК. Однако существуют и общедоступные коммерческие программы для построения фармакокинетических моделей с высокой точностью, ниже перечислены наиболее известные из них:

Simcyp® (www.simcyp.com) — это коммерческая программа для моделирования процессов ADME (всасывания, распределения, метаболизма и выведения лекарств). Simcyp® объединяет экспериментальные данные с демографической, физиологической и генетической информацией для прогнозирования взаимодействия лекарств и вариабельности в популяциях. Программа умеет моделировать ситуации перорального приема лекарств с учетом таких факторов, как прием пищи и анатомические изменения. Так же Simcyp® для моделирования распределения лекарства использует различные модели для учета различных физиологических аспектов. Цель данной программы состоит в том, чтобы поддержать разработку лекарств, предоставляя информацию о том, как лекарства действуют на разных людей, и прогнозируя эффективность и безопасность взаимодействия лекарств. Из-за своего функционала Simcyp® является самым используемым программным обеспечением в области фармакокинетики [17, 18].

GastroPlus^{TM} (www.simulations-plus.com) является коммерческим программным обеспечением, которое помогает моделировать, как лекарственные препараты всасываются, распределяются, метабо-

лизируются и выводятся из организма. Данное программное обеспечение известно своим простым в использовании интерфейсом и высококачественной графикой. Функционал программы заключается в способности использовать настраиваемые модели, в том числе модели для детей и различных животных (крысы, собаки, мыши, обезьяны). GastroPlus^{TM} является одним из самых востребованных программных обеспечений в области разработок лекарств [17, 18].

PK-Sim® (www.open-systems-pharmacology.org) — это программное обеспечение с открытым исходным кодом, используемое для комплексного моделирования, как лекарственные препараты перемещаются и взаимодействуют в организме. Оно объединяет анатомические и физиологические параметры человека и животных для создания надежных моделей. PK-Sim® работает совместно с другим программным обеспечением под названием MoBi® для расширенного физиологического моделирования, которое помогает определить воздействие лекарств на организм. Это позволяет исследователям прогнозировать, как лекарства поведут себя в различных сценариях, в том числе у разных видов животных и при различных условиях (например, голодание или кормление). PK-Sim® известен своей способностью изучать поведение лекарств у пациентов с особыми состояниями, такими как цирроз печени [17, 18].

Intellipharm® PK (www.intellipharm.com) является коммерческим программным обеспечением, разработанным для моделирования концентрации лекарственных средств в плазме крови при пероральном приеме препарата. В Intellipharm® PK используется модель, в которой желудок и кишечник рассматриваются как один большой резервуар. Программное обеспечение связывает эту модель всасывания с более простыми фармакокинетическими моделями (одно- или двухкамерными), что помогает симулировать, как лекарства растворяются и перемещаются с течением времени. Но данная модель не полностью учитывает сложную природу ЖКТ и не может учитывать различия между отдельными людьми в всасывании, распределении, метаболизме и выведении лекарственных средств без клинических данных [17].

Cloe® PK (www.cyprotex.com) — это коммерческий инструмент моделирования, способный прогнозировать всасывание при приеме внутрь, а также распределение и выведение из организма. Эта модель представляет просвет желудочно-кишечного тракта в виде ряда из пяти отделов. Cloe® PK объединяет физико-химические данные и данные ADME для прогнозирования продолжительности действия лекарств в крови, тканях и органах [17].

PKQuest (www.pkquest.com) является бесплатным программным обеспечением, используется в образовательных целях и для исследований в области PK. PKQuest разработан на языке программирования Java, и включает в себя возможности различных методов моделирования на основе клинических данных ФК человека или человека. Основное преимущество программы — это минимальное количество вводимых пользователем параметров, по средствам классификации лекарственных препаратов и предварительно запрограммированные оптимизированные данные ФК человека или крысы [18].

MaxSim2 — программа так же разработана в образовательных целях для обучения ФК и ФД, а также для доклинического моделирования. MaxSim2 предоставляет распространенные модели ФК и ФД и позволяет изменять параметры моделей в режиме реального времени, чтобы визуализировать графики концентрации и времени отклика [18].

Большинство из представленных программ являются дорогостоящими и зарубежными, что ограничивает возможности исследователей на территории России. Единственным отечественным продуктом в этой области является программное обеспечение «Симург» (Simurg) от компании «Симург-Фарм» (simurgpharm.ru). Оно предназначено для математического моделирования доклинических и клинических исследований лекарственных средств.

Кроме того, некоторые лекарственные средства вводятся с использованием носителя, и вещество постепенно высвобождается из него уже в организме. Описание ФК при внутривенном введении наночастиц с лекарством представляет собой значительный интерес, а большая часть программных продуктов не имеет возможностей моделирования ФК в таком случае.

В связи с этим было принято решение создать продукт с открытым исходным кодом, который позволит создавать и оптимизировать параметры фармакокинетических моделей. В основу ПО была заложена созданная для этой цели общедоступную библиотека на языке Python. А поиск параметров ФК моделей был произведен при помощи вдохновленных природой эволюционных алгоритмов оптимизации.

Библиотека PyPharm

Созданная в ходе работы библиотека *PyPharm* [19] способна производить моделирование ФК с использованием камерных моделей, а также производить поиск параметров модели по экспериментальным данным.

Основным модулем библиотеки является модуль *models*, который на текущий момент имеет в себе три типа моделей, отраженных в соответствующих классах:

- *BaseCompartmentModel* — классическая камерная модель;
- *MagicCompartmentModel* — классическая камерная модель с дополнительно введенным параметром *magic_coefficient*, отвечающим за преобразование единиц измерения входной концентрации в единицы измерения выходной концентрации. Этот параметр необходим, в тех случаях, когда доза препарата задана в одних единицах измерения, а концентрация в органах в других, и нет явного способа преобразования;
- *ReleaseCompartmentModel* — камерная модель, учитывающая высвобождение лекарственного вещества из носителя по заданному закону.

Библиотека построена на принципах объектно-ориентированного программирования (ООП) и подразумевает наследование моделей друг от друга для избежания повторного написания кода при разработке. Так два последних класса наследуются от *BaseCompartmentModel*, поэтому основные методы классов подобны и отличаются лишь набором входных параметров. Далее будет описана суть их работы.

Метод __init__

Данный метод отвечает за инициализацию параметров модели, и в базовом варианте принимает следующие аргументы:

- configuration_matrix — настроечная матрица модели, отображающая константы перехода между камерами. Значение a_{ij} отражает константу перехода вещества их i-й камеры в j-ю;
- outputs — вектор констант перехода во вне камер. В данном случае a_i отражает константу элиминации из i-й камеры;
- volumes — вектор объемов камер;
- numba_option — флаг необходимости использования библиотеки *numba* при оптимизации параметров;
- use_shared_memory — флаг необходимости использования общей памяти при оптимизации параметров.

Все неизвестные параметры модели в настроечных матрицах и векторах, пользователь может задать значением None, тогда их поиск будет производиться при оптимизации модели.

Для класса *MagicCompartmentModel* добавляется дополнительные параметры:

- magic_coefficient — параметр, отвечающий за преобразование данных. Простое число, на которое необходимо умножить входные данные, чтобы преобразовать их в выходные. Так же может быть задан None, если неизвестен;
- exclude_compartments — список номеров камер, которые нужно исключить из процесса преобразования.

Дополнительные параметры имеются и для класса *ReleaseCompartmentModel:*

- *v_release* — объем гипотетической камеры, из которой происходит высвобождение. Может быть задан None, если неизвестен;
- *release_parameters* — список параметров функции высвобождения. Данные параметры передаются в функцию высвобождения и могут быть заданы *None*, если неизвестны;
- *release_compartment* — номер камеры, в которую происходит высвобождение;
- *release_function* — функция высвобождения;
- *with_accumulate* — флаг необходимости учесть аккумуляцию вещества;
- *accumulation_parameters* — список параметров функции высвобождения.;
- Данные параметры передаются в функцию высвобождения и могут быть заданы *None*, если неизвестны;
- *accumulation_function* — функция аккумуляции.

Метод __call__

Метод *__call__* позволяет вызывать объект, классов моделей, как функцию и на выходе дает численное решение системы дифференциальных уравнений, описывающей камерную модель. В базовом варианте функция принимает следующие аргументы:

- t_max — предельное время расчета;
- *c0* — вектор начальных концентраций (может не передаваться, если передано d и $compartment\_number$);
- d — вводимая доза (может не передаваться, если передано $c0$);
- $compartment\_number$ — номер камеры, в которую вводится доза (может не передаваться, если передано *c0*);
- max_step — максимальный шаг при решении СДУ;
- t_eval — вектор временных точек, в которых необходимо получить решение.

На выходе метод возвращает результат работы функции *solve_ivp* [20] библиотеки numpy.

Для класса *ReleaseCompartmentModel* смысл параметров немного видоизменяется, и $compartment\_number$ больше не требуется, а $c0$ обозначает начальную концентрацию в гипотетической камере.

Методы load_optimization_data и optimize

Данные методы используются для поиска параметров камерной модели и используются в связке.

Метод *load_optimiztion_data* позволяет загрузить в модель экспериментальные данные, для ее дальнейшей оптимизации. Он принимает следующие параметры:

- $teoretic\_x$ — вектор значений времени замера экспериментальных данных;
- $teoretic\_y$ — матрица с экспериментальными значениями по известным камерам;
- $know\_compartments$ — вектор с номерами камер, по которым есть данные;
- $c0$ — вектор начальных концентраций (может не передаваться, если передано d и $compartment\_number$);
- d — вводимая доза (может не передаваться, если передано *c0*);
- $compartment\_number$ — номер камеры, в которую вводится доза (может не передаваться, если передано $c0$);
- w — матрица весов, отражающая достоверность данных, по которым производится оптимизация.

Для класса *ReleaseCompartmentModel*, передаваемые параметры меняются по аналогии с методом *__call__*.

Метод optimize производит непосредственный поиск неизвестных параметров с использованием нескольких алгоритмов. Для всех классов моделей он принимает одни и те же параметры:

- $method$ — метод, который будет использован для оптимизации. Может принимать строковые значения: country_optimization (использование метода взаимодействующих стран [10, 15, 16]), GA (использование генетического алгоритма) или иные, тогда оптимизация производится функцией minimize [20] библиотеки scipy;
- $user\_method$ — принимает объект функции или класс, метод, который пользователь хочет использовать для оптимизации;
- $method\_is\_func$ — флаг является ли пользовательский метод функцией (по умолчанию истина);
- $optimization\_func\_name$ — метод, который нужно вызвать у объекта пользовательского класса (по умолчанию *__call__*);
- max_step — максимальный шаг при численном решении системы дифферинциальных уравнений;
- $metric$ — метрика, определяющая целевую функцию. Может принимать строковые значения: *R2* (используется показатель R^2), norm (используется норма матрицы отклонений);
- $**kwargs$ — именованные аргументы, которые необходимо передать для настройки оптимизационного алгоритма.

Выбранный в методе *optimize* подход позволяет оптимизировать данные не только способами, встроенным библиотеку, но и выбранными пользователем, что значительно повышает исследовательский потенциал.

Методы plot_model и get_kinetic_params

Два данных метода позволяют вывести информацию о модели.

Метод *plot_model* позволяет вывести графики распределения вещества в камерах и в базовом варианте принимает аргументы:

- $compartment\_numbers$ — список камер, в которых нужно построить графики распределения;
- $compartment\_names$ — словарь имен камер, где ключ — это номер камеры, а значение — отображаемое имя;
- $left$ — левая граница по времени на графиках;
- $right$ — правая граница по времени на графиках;
- y_lims — словарь ограничений по оси y для камер. Ключом в данном случае является номер камеры, а значением кортеж (y_{min}, y_{max});
- $**kwargs$ — именованные аргументы, передаваемые в метод *__call__*.

Для класса *ReleaseCompartmentModel* добавляется флаг $plot\_accumulation$ отражающий необходимость построения данных в теоретической камере.

Дополнительные особенности библиотеки

Как упоминалось ранее, библиотека PyPharm опционально способна использовать библиотеку numba для своих вычислений. Numba [21] — это JIT-компилятор для CPython 2, самой популярной реализации языка Python, который, как следует из названия, написан на C. Numba реализован в виде библиотеки, которую можно загрузить в программы, запущенные в интерпретаторе CPython, без необходимости обновления или замены самого интерпретатора. Изначально он был создан для языка Python, активно использующего массивы и числовые скаляры в циклах, чтобы пользователям не приходилось переписывать свой код на низкоуровневых языках для повышения производительности. В рамках библиотеки PyPharm numba позволяет ускорить работу кода за счет компиляции такого функционала, как вычисление целевой функции.

Необходимо также упомянуть и о целевой функции, которая лежит в основе оптимизации фармакокинетических моделей в рамках библиотеки. Для количественной оценки точности модели, а также для подбора параметров используется функция (3):

$$F = \sum_{i=1}^{k} \sum_{j=1}^{n} \frac{\left(C_{ij}^{calc} - C_{ij}^{ex}\right)^2}{\left(\overline{C_k^{ex}} - C_{ij}^{ex}\right)^2} \tag{3}$$

где k — количество известных наборов данных, n — количество данных в наборе, C_{ij}^{calc} — вычисленное значение для i-го набора в j-й точке, C_{ij}^{ex} — экспериментальное значение для i-го набора в j-й точке, $\overline{C_k^{ex}}$ — среднее значение по набору экспериментальных данных.

Стоит упомянуть о возможности работы библиотеки с общей памятью (shared memory) [22]. Данная технология доступна в библиотеке начиная с версии 1.3.0, она удобна для получения информации о ходе процесса оптимизации извне запущенного кода. Данное действие возможно при помощи обращения к памяти по имени, которое хранится в поле *memory_name* объекта модели. Обращение может быть осуществлено при помощи стандартного класса *ShareableList* библиотеки *multiprocessing*.

Пример: поиск параметров модели, описывающей ФК $[^{25}Mg^{2+}]$ порфирин-фуллереновых наночастиц в организме крысы при однократном, внутривенном введении

Далее будет рассмотрен пример использования библиотеки для поиска параметров модели, описывающей ФК $[^{25}Mg^{2+}]$ порфирин-фуллереновых наночастиц в организме крысы при однократном, внутривенном введении. Данный лекарственный препарат представляет интерес в роли нейропротектора, способного помочь при ишемическом инсульте.

В статье [23], авторами предлагается пятикамерная (рис. 2) модель для описания распространения данного лекарственного вещества в организме человека.

В оригинальной статье не учитывался объем отдельных камер, однако для демонстрации воз-

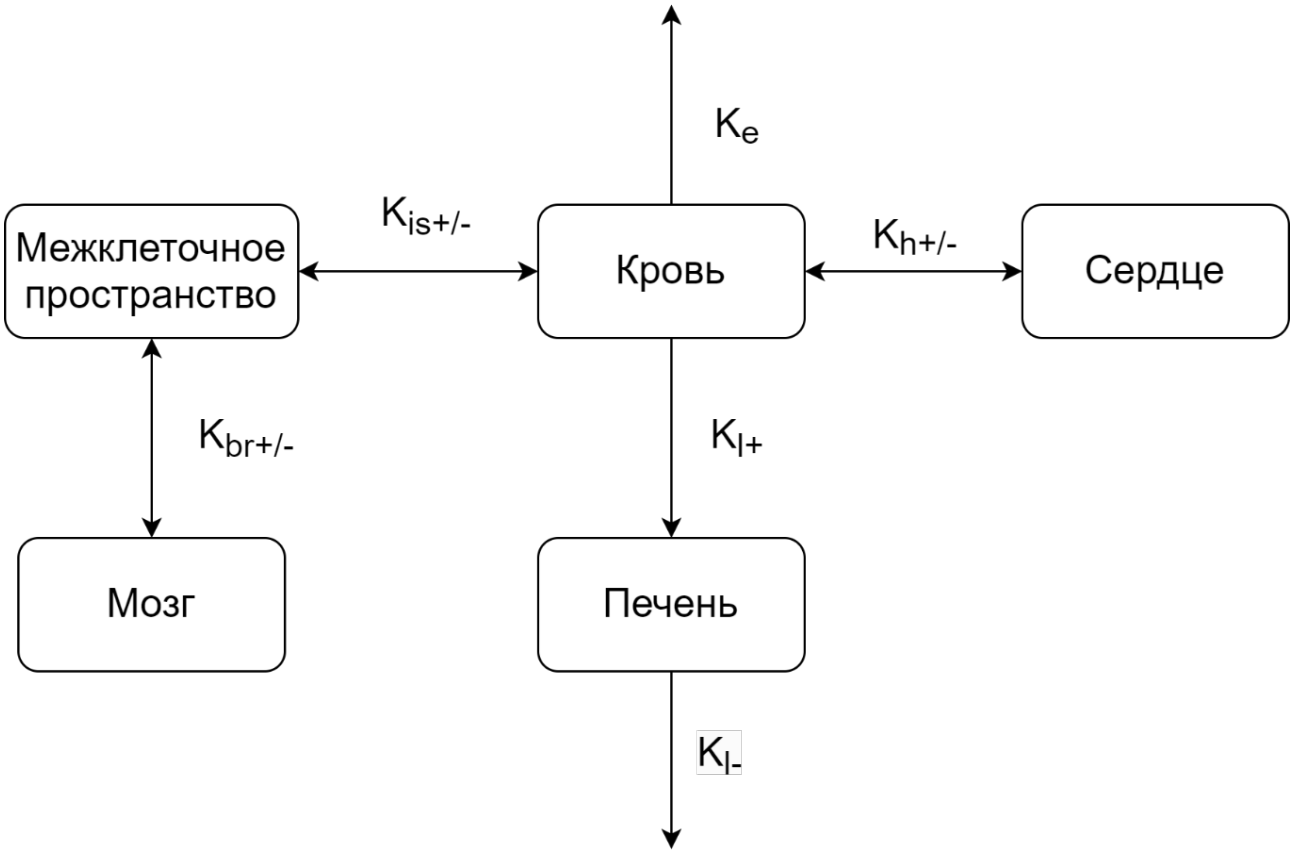


Рис. 2. *Схематичное отображение пятикамерной модели*

можностей библиотеки, можно учесть его, тогда модель может быть записана в виде системы ОДУ (4):

$$
\begin{aligned}
V_b\frac{dC_b}{dt} &= -k_eC_bV_b - k_{l+}C_bV_b - k_{h+}C_bV_b - k_{is+}C_bV_b + k_{h-}C_hV_h + k_{is-}C_{is}V_{is}\\
V_l\frac{dC_l}{dt} &= k_{l+}C_bV_b - k_{l-}C_lV_l\\
V_h\frac{dC_h}{dt} &= k_{h+}C_bV_b - k_{h-}C_hV_h\\
V_{is}\frac{dC_{is}}{dt} &= k_{is+}C_bV_b - k_{br+}C_{is}V_{is} - k_{is-}C_{is}V_{is} + k_{br-}C_{is}V_{is}\\
V_{br}\frac{dC_{br}}{dt} &= k_{br+}C_{is}V_{is} - k_{br-}C_{br}V_{br}
\end{aligned}
\tag{4}
$$

где C_b, C_l, C_h, C_{is}, C_{br} отражают концентрации лекарственного вещества в крови, печени, сердце, межклеточном пространстве мозга и клетках мозга соответственно.

В целом, механизм использования библиотеки PyPharm может быть описан в виде схемы (3):

Работа начинается с получения экспериментальных данных по ФК вещества в организме человека. Далее выбирается камерная модель, которая по мнению исследователя наилучшим образом будет описывать экспериментальные данные. Затем неизвестные параметры модели подбираются исходя из минимизации отклонения рассчитанных на модели данных от экспериментальных значений. Результатом работы библиотеки является массив данных, описывающий изменение концентрации лекарственного вещества в той или иной камере во времени. Данные для удобства могут быть представлены в виде графиков.

Если поэтапно рассматривать описанный выше процесс для модели $[^{25}Mg^{2+}]$ порфирин-фуллереновых наночастиц, работа с библиотекой PyPharm будет выглядеть следующим образом (рис. 4–5):

1) так как входные и выходные концентрации в источнике данных выражены в разных единицах

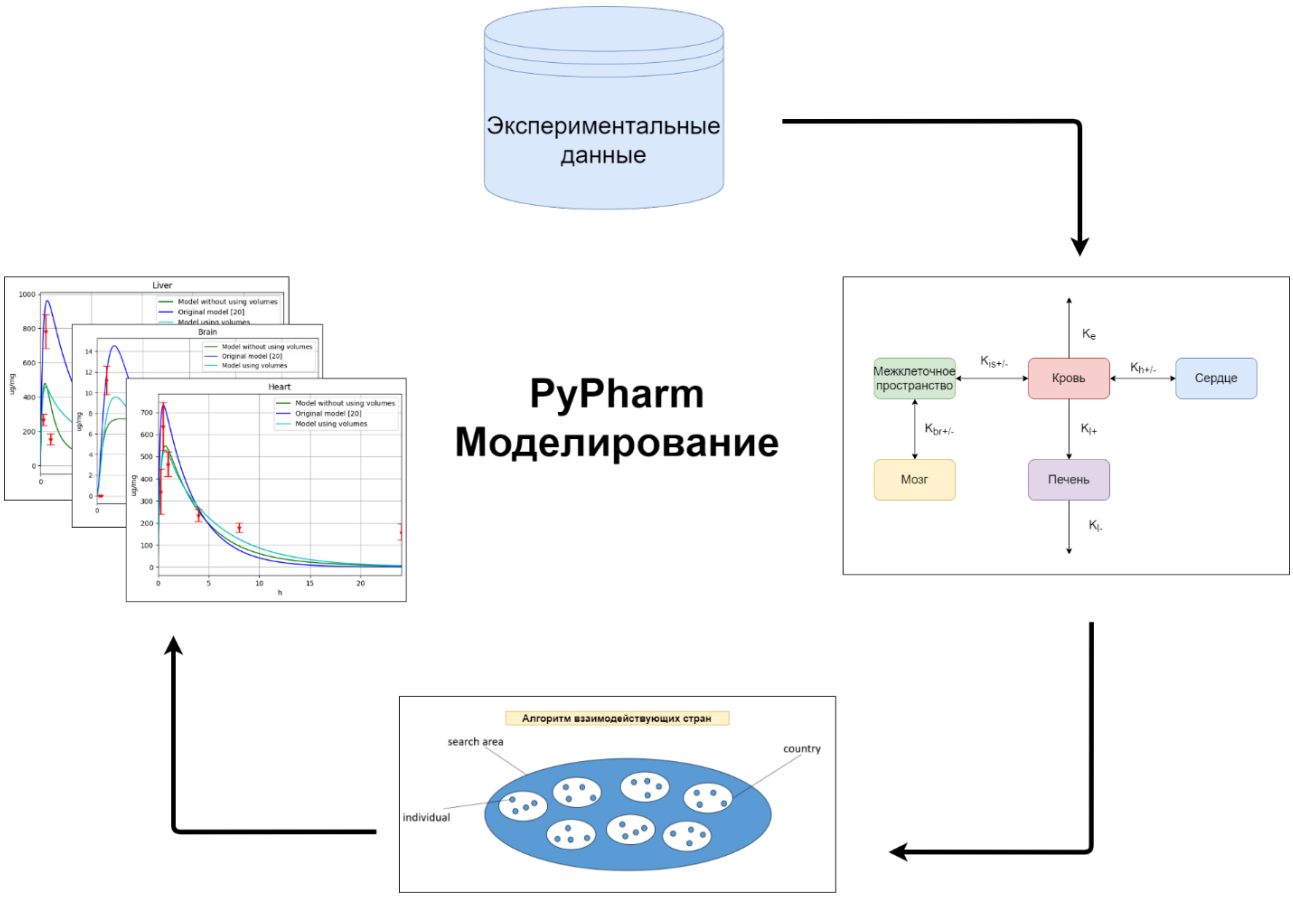


Рис. 3. *Схематичное описание работы библиотеки PyPharm*

измерения (таких, что нельзя явно преобразовать одни в другие), необходимо создать объект класс *MagicCompartmentModel*;

2) на этапе инициализации объекта необходимо задать неизвестные объемы и коэффициенты как None для их дальнейшего поиска. Кроме того, нужно задать параметр *exclude_compartments*, исключив в данном случае камеру, отвечающую за кровь, так как концентрация лекарственного вещества в ней выражена в других единицах измерения. При этом переменная *matrix* отвечает за конфигурацию модели, а *outputs* за элиминацию, в данном случае она будет происходить только через печень;

3) далее необходимо произвести загрузку экспериментальных данных, используя *load_optimization_data* и запустить оптимизацию функцией *optimize*. Параметры M, N, n, p, m, k, l и ep являются настроечными параметрами для алгоритма взаимодействующих стран [21];

4) в завершении можно использовать *plot_model* для построения графиков высвобождения ЛВ в отдельных камерах, либо же использовать полученные данные для дальнейшего форматирования и вывода.

Пример результатов, полученных для данной модели с использованием библиотеки PyPharm приведен на рис. 6. Стоит заметить, что по результатам моделирования можно заключить, что модели, настроенные при помощи библиотеки PyPharm существенно превосходят авторскую [23] модель (значение целевой функции уменьшилось примерно на 67%).

Таким образом, мы сообщаем о разработке библиотеки *PyPharm*, способной решать задачи построения фармакокинетических моделей. На данный момент библиотека имеет версию 1.3.6 и доступна в репозитории PyPi https://pypi.org/project/pypharm/. Там же имеются примеры использования библиотеки и описание ее возможностей.

Библиотека способна строить три вида камерных моделей и уже была опробована и хорошо показала себя в различных исследованиях [14, 19]. В дальнейшем планируется расширение PyPharm, с добавлением в нее методов машинного обучения и расширением списка доступных моделей, например, PBPK-моделей.

```
from PyPharm import MagicCompartmentModel


# Задание матрицы, определяющей конфигурацию модели
matrix = [[0, None, 0, 0, 0],
          [None, 0, None, 0, 0],
          [0, None, 0, None, None],
          [0, 0, 0, 0, 0],
          [0, 0, None, 0, 0]]
# Задание вектора, определяющего конфигурацию выводов
outputs = [0, 0, 0.077, None, 0]

model = MagicCompartmentModel( #Создание объекта модели
    matrix, outputs,
    volumes=[None, None, None, None, None], #Неизвестные параметры, везде
    magic_coefficient=None,                 #задаются None
    exclude_compartments=[2],
    numba_option=True
)
model.load_optimization_data( #загружка данных для оптимизации
    teoretic_x=[0.25, 0.5, 1, 4, 8, 24],
    teoretic_y=[[0, 0, 11.2, 5.3, 5.42, 3.2],
                [268.5, 783.3, 154.6, 224.2, 92.6, 0],
                [342, 637, 466, 235, 179, 158]],
    know_compartments=[0, 3, 4],
    d=20 * 0.2 * 1000,
    compartment_number=2
)
```

Рис. 4. *Использование библиотеки PyPharm для построения конфигурации модели и загрузки экспериментальных данных*

```
model.optimize( #Оптимизация параметров методом взаимодействующих стран
    method='country_optimization',
    Xmin=[0.5, 0.001, 0.001, 0.00001, 0.01, 0.01, 1, 20, 20, 20, 20, 20, 1],
    Xmax=[5, 2, 2.5, 0.3, 1, 1, 10, 100, 100, 100, 100, 100,10],
    M=10,
    N=25,
    n=[1, 10],
    p=[0.00001, 2],
    m=[1, 8],
    k=8,
    l=3,
    ep=[0.2, 0.4],
    tmax=300,
    max_step=0.1,
    printing=True,
)

model.plot_model( #Построение кривых распределения
    d=20 * 0.2 * 1000,
    compartment_number=2,
    t_max=24,
    compartment_numbers=[3, 4, 0],
    compartment_names={3: "Печень", 4: "Сердце", 0: "Мозг"}
)
```

Рис. 5. *Использование библиотеки PyPharm для оптимизации параметров модели и построения кривых распределения*

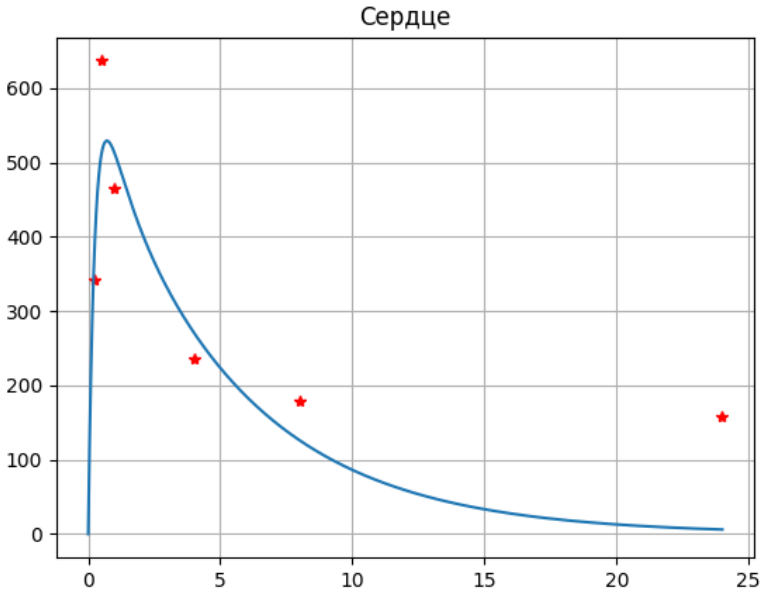


Рис. 6. *Пример результатов работы*

ЛИТЕРАТУРА

1. *Guideline on the Qualification and Reporting of Physiologically Based Pharmacokinetic (PBPK) Modelling and Simulation*. Режим доступа: https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-reporting-physiologically-based-pharmacokinetic-pbpk-modelling-and-simulation_en.pdf.
2. *Physiologically Based Pharmacokinetic Analyses – Format and Content Guidance for Industry*. Режим доступа: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/physiologically-based-pharmacokinetic-analyses-format-and-content-guidance-industry.
3. Ivy S. P., Siu L. L., Garrett-Mayer E., Rubinstein L. Approaches to Phase 1 Clinical Trial Design Focused on Safety, Efficiency, and Selected Patient Populations: a Report from the Clinical Trial Design Task Force of the National Cancer Institute Investigational Drug Steering Committee. *Clin Cancer Res.* 2010;16(6):1726–1736. DOI: 10.1158/1078-0432.CCR-09-1961.
4. Hill-McManus D., Marshall S., Soto E., Hughes D. A. Integration of Pharmacometrics and Pharmacoeconomics to Quantify the Value of Improved Forgiveness to Nonadherence: A Case Study of Novel Xanthine Oxidase Inhibitors for Gout. *Clin Pharmacol Ther.* 2019;106(3):652-660. DOI: 10.1002/cpt.1454.
5. Hill-McManus D., Marshall S., Soto E., Lane S., Hughes D. Impact of Non-Adherence and Flare Resolution on the Cost-Effectiveness of Treatments for Gout: Application of a Linked Pharmacometric/Pharmacoeconomic Model. *Value Health.* 2018;21(12):1373–1381. DOI: 10.1016/j.jval.2018.06.002.
6. Singh N., Vayer P., Tanwar S., Poyet J.-L., Tsaioun K. and Villoutreix B. O. Drug Discovery and Development: Introduction to the General Public and Patient Groups. *Frontiers in Drug Discovery.* 2023; 3:1201419. DOI: 10.3389/fddsv.2023.1201419.
7. Mortensen S. B., Jónsdóttir A. H., Klim S. and Madsen H. *Introduction to PK/PD Modelling - with Focus on PK and Stochastic Differential Equations*. Lyngby: Technical University of Denmark, DTU Informatics; Building 321. Technical Report, No. 2008-16. Режим доступа: https://accio.github.io/AMIDD/assets/2019/08/2008-Mortensen-IntroductionToPKPDModelling.pdf.
8. Сергиенко В. И., Джеллифф Р., Бондарева И. Б., *Прикладная фармакокинетика: основные положения и клиническое применение*. М.: Изд-во РАМН; 2003. 208 с.
9. Bhavatharini P., Deepalakshmi M., Arun K. Pharmacometrics: The Science Applied from Bench to Bedside. *J Appl Pharm Sci*; 2022;12(1):55–64. DOI: 10.7324/JAPS.2021.120104.
10. Yoon E., Babar A., Choudhary M., Kutner M., Pyrsopoulos N. Acetaminophen-Induced Hepatotoxicity:

a Comprehensive Update. *J Clin Transl Hepatol*. 2016;4(2):131-142. DOI: 10.14218/JCTH.2015.00052.

11. Ryu H.-J., Kang W.-H., Kim T., Kim J. K., Shin K.-H., Chae J.-W. and Yun H.-Y. A Compatibility Evaluation between the Physiologically Based Pharmacokinetic (PBPK) Model and the Compartmental PK Model Using the Lumping Method with Real Cases. *Front. Pharmacol*. 2022;13:1–12. DOI: 10.3389/fphar.2022.964049.
12. Talevi A., Bellera C. L. One-Compartment Pharmacokinetic Model. *The ADME Encyclopedia.* Springer, Cham. 2021:1–8. DOI: 10.1007/978-3-030-51519-5_58-1.
13. Talevi A., Bellera C. L. Compartmental Pharmacokinetic Models. *ADME Processes in Pharmaceutical Sciences.* Springer, Cham. 2024:173–192. DOI: https://doi.org/10.1007/978-3-031-50419-8_8.
14. Крашенинников Р. С., Митричев И. И. Применение алгоритма взаимодействующих стран для поиска констант фармакокинетической модели. *Математические методы в технологиях и технике.* 2023;4:12–15. DOI: 10.52348/2712-8873_MMTT_2023_2_12.
15. Самородов П. О. Особенности построения математической модели фармакокинетики при многократном введении препарата. *Теоретико-методологические аспекты преподавания математики в современных условиях:* материалы III Международной научно-практической конференции (1–7 июня 2020 г., г. Луганск). Луганск: Книта; 2020:78–83.
16. Jamei M., Dickinson G. L., Rostami-Hodjegan A. A Framework for Assessing Inter-Individual Variability in Pharmacokinetics Using Virtual Human Populations and Integrating General Knowledge of Physical Chemistry, Biology, Anatomy, Physiology and Genetics: a Tale of ‘bottom-up’ vs ‘top-down’ Recognition of Covariates. *Drug Metabolism and Pharmacokinetics*. 2009;24(1):53–75. DOI: 10.2133/dmpk.24.53.
17. Rowland M. Physiologically-Based Pharmacokinetic (PBPK) Modeling and Simulations Principles, Methods, and Applications in the Pharmaceutical Industry. *CPT Pharmacometrics Syst Pharmacol*. 2013;2(7):e55. DOI: 10.1038/psp.2013.29.
18. Rostami-Hodjegan A., Bois F. Y. Opening a Debate on Open-Source Modeling Tools: Pouring Fuel on Fire Versus Extinguishing the Flare of a Healthy Debate. *CPT Pharmacometrics Syst Pharmacol*. 2021;10(5):420–427. DOI: 10.1002/psp4.12615.
19. Крашенинников Р. С., Митричев И. И. Разработка библиотеки Python для решения задач фармакокинетики. *Свободное программное обеспечение в высшей школе:* материалы девятнадцатой конференции, Переславль-Залесский, 28–30 июня 2024 года. Москва: ООО «МАКС Пресс»; 2024:166–170.
20. *Документация метода solve_ivp*. Режим доступа: https://docs.scipy.org/doc/scipy/reference/generated/scipy.integrate.solve_ivp.html.
21. Крашенинников Р. С., Митричев И. И. Оптимизационный алгоритм взаимодействующих стран. *Информационные технологии в моделировании и управлении: подходы, методы, решения:* материалы VI Всероссийской научной конференции с международным участием, Тольятти, 18–20 апреля 2023 года. Тольятти: Тольяттинский государственный университет; 2023:184–191.
22. Крашенинников Р. С., Митричев И. И. Оптимизационный алгоритм взаимодействующих стран и его применение для решения задач фармакокинетики. *Прикладная математика и вопросы управления*. 2023;4:26–135. DOI: 10.15593/2499-9873/2023.4.08.
23. Фурсов В. В., Зинченко Д. И., Наместникова Д. Д., Кузнецов Д. А. In silico-моделирование в оптимизации алгоритмов фармакокинетических исследований $[^{25}Mg^{2+}]$ порфирин-фуллереновых наночастиц. *Вестник Российского государственного медицинского университета*. 2022;4:65–71. DOI: 10.24075/vrgmu.2022.037.

СИСТЕМНЫЙ ПОДХОД К РЕШЕНИЮ ЗАДАЧИ ПОДДЕРЖКИ ПРИНЯТИЯ ВРАЧЕБНЫХ РЕШЕНИЙ

А. И. Пахтусов
Сургутский государственный университет, г. Сургут, Российская Федерация
ORCID: https://orcid.org/0009-0008-8276-2901, ✍ pahtusov@bk.ru

*Аннотация:*в статье рассматриваются возможности применения в медицинских учреждениях программного решения для поддержки принятия врачебных решений и оценки рисков развития заболеваний, которое охватывает различные наборы алгоритмов диагностики заболеваний и их интеграцию с информационными системами медицинских учреждений.

Описанное программное решение предполагает автоматическое информирование врачей о возможном наличии рисков развития заболеваний у обследуемых пациентов на основании имеющихся медико-социальных данных с применением наборов алгоритмов диагностики заболеваний даже в тех случаях, когда пациент пришел на прием по другим нозологиям.

Ключевые слова: алгоритмы дифференциальной диагностики заболеваний, экспертные системы, системы поддержки принятия врачебных решений, интеллектуальные системы, системы на основе знаний.

Для цитирования: Пахтусов А. И. Системный подход к решению задачи поддержки принятия врачебных решений. *Успехи кибернетики*. 2025;6(1):170–174.

Поступила в редакцию: 13.01.2025. *В окончательном варианте: 02.02.2025.*

SYSTEM APPROACH TO DECISION SUPPORT IN MEDICINE

A. I. Pakhtusov
Surgut State University, Surgut, Russian Federation
ORCID: https://orcid.org/0009-0008-8276-2901, ✍ pahtusov@bk.ru

*Abstract:*we examine the application of a software solution in medical institutions to support medical decision-making and assess disease risks. This solution integrates various diagnostic algorithms with medical information systems.

The proposed software automatically informs doctors about potential disease risks in examined patients based on available medical and social data. It applies diagnostic algorithms even when patients seek medical attention for unrelated conditions.

Keywords: differential diagnosis algorithms, expert systems, medical decision support systems, intelligent systems, knowledge-based systems.

Cite this article: Pakhtusov A. I. System Approach to Decision Support in Medicine. *Russian Journal of Cybernetics*. 2025;6(1):170–174.

Original article submitted: 13.01.2025. *Revision submitted: 02.02.2025.*

Введение

Сейчас действует государственная программа Российской Федерации «Развитие здравоохранения», утвержденная постановлением Правительства Российской Федерации от 26 декабря 2017 г. № 1640 «Об утверждении государственной программы Российской Федерации «Развитие здравоохранения» (далее соответственно — Госпрограмма, Постановление) [1].

Госпрограмма в соответствии с Постановлением с 2022 года реализуется в т.ч. по направлениям:

1. Совершенствование оказания медицинской помощи, включая профилактику заболеваний и формирование здорового образа жизни.
2. Развитие и внедрение инновационных методов диагностики, профилактики и лечения, а также основ персонализированной медицины.
3. Информационные технологии и управление развитием отрасли.

Несмотря на действие указанной Госпрограммы, одной из основных проблем развития тяжелых заболеваний остается поздняя диагностика. К примеру, эта причина является основной для повышенной смертности от рака в России [2]. На примере онкологического диагноза можно выделить следующие причины поздней диагностики [2]:

1. Со стороны врача — отсутствие онкологической настороженности. В связи с этим не проводится достаточно полного обследования пациентов, методы специальной диагностики применяются несвоевременно.

2. Со стороны пациента — невнимательное отношение к своему здоровью, недооценка симптомов заболевания, страх перед возможным диагнозом, нехватка времени на посещение врачей, попытки самостоятельного лечения. Статистика показывает, что в России около 10% больных раком отказываются от лечения по причине отсутствия средств. Еще 20% пациентов запаздывают с визитом к врачу, так как просто боятся обращаться в медицинские центры. В результате у них диагностируется рак на поздних стадиях, что затрудняет адекватное лечение.

3. Скрытое течение онкологического заболевания на ранних стадиях, объективные трудности диагностики, сочетание нескольких заболеваний, запутанность симптомов заболевания и так далее.

Существует большой спектр биомаркеров, которые направлены на помощь врачу в ранней диагностике многих заболеваний [3].

Объем медицинских данных постоянно растет. Также растет число разнообразных медицинских информационных систем для работы с такими данными. К примеру, системы поддержки принятия врачебных решений (СППВР). С одной стороны, СППВР может помогать врачам, а с другой стороны, — СППВР, как правило, узкоспециализированы. Поэтому возможности их применения сильно ограничены небольшим перечнем задач, в т.ч. по диагностике заболеваний. Дополнительно постоянно растут объем требований к деятельности врачей и их нагрузка. На основании изложенного можно выделить следующие проблемы:

1. Высокая нагрузка на медицинских работников и исполнение приказов Министерства здравоохранения Российской Федерации в области применения стандартов специализированной медицинской помощи [4] уменьшает возможность индивидуального подхода к обследованию, диагностике и лечению заболеваний у пациентов.

2. В мире постоянно появляется множество медицинских информационных систем [5], которые могут быть не связаны между собой. Это приводит к сложности массового внедрения в медицинских учреждениях, которое ограничивается, как правило, единичными примерами. А впоследствии, без активного их развития, может вообще перестать применяться.

3. Врачи могут применять системы поддержки принятия врачебных решений, как правило, точечно, так как они в большинстве своем предназначены для решения узконаправленных задач в медицинской сфере.

4. Требуется комплексное организационное и программное решение, позволяющее решить большинство перечисленных проблем.

Использование новых методов и подходов позволяет существенно упростить раннюю диагностику заболеваний.

Следует отметить, что на Госуслугах в личных кабинетах пользователей с 1 сентября 2022 года идет накопление медицинских документов, поступающих из различных медицинских учреждений: протоколы исследования, выписки, справки и т.п. [6] Если по согласованию с пациентами медицинским учреждениям предоставить доступ к их медицинским документам на Госуслугах, то при помощи AI-инструментов [7] возможно будет их распознать, получить данные из них и загрузить в медицинскую информационную систему медицинского учреждения. Это позволит увеличить охват диагностических показателей по пациентам, которые были получены в различных медицинских организациях, и может способствовать более качественному ведению пациентов.

Описанные выше проблемы и подходы к их решению могут быть выполнены на основе современной микросервисной архитектуры. В частности, решение должно включать в себя следующий минимально необходимый набор микросервисов (рис. 1):

1. Микросервис «Взаимодействие с Госуслугами». Предназначен для выгрузки по пациентам с Госуслуг социально-медицинских данных, полученных в различных медицинских организациях.

2. Микросервис «Анализ данных». Предназначен для анализа медицинских данных на наличие рисков развития (выявления) заболеваний на основе имеющихся алгоритмов. В качестве примера алгоритма диагностики заболеваний выступает алгоритм обследования и дифференциальной диагностики образований яичников, описанный в статье «Системный анализ процесса диагностики опухолей яичников» [8].

3. Микросервис «Поддержка принятия врачебных решений». Предназначен для поддержки принятия решений врачом посредством представления результатов анализа медицинских данных на наличие рисков развития (выявления) заболеваний.

4. Микросервис «Оповещение врачей». Предназначен для автоматического оповещения в фоновом режиме врачей по результатам анализа данных по пациентам, независимо от направления профиля деятельности врача и причины посещения пациентом врача. Предложения по дополнительным направлениям обследования.

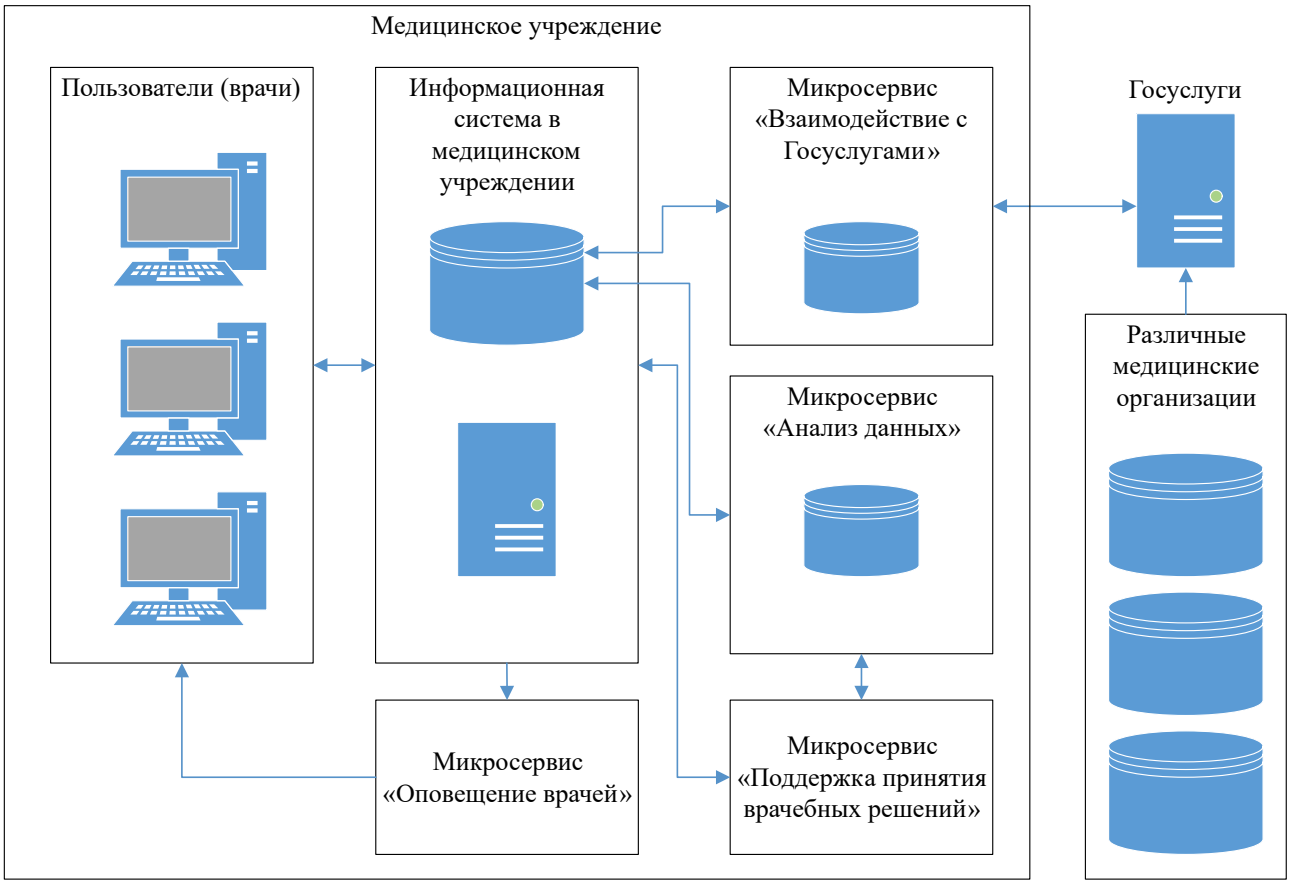


Рис. 1. *Архитектура решения*

Разрабатываемая архитектура позволяет:

1. Агрегировать медицинские данные из различных медицинских организаций.
2. Проводить анализ данных в автоматическом режиме.
3. Оповещать в фоновом режиме врачей о наличии рисков развития (выявления) заболеваний на основе имеющихся алгоритмов.

Микросервисы работают:

1. Периодически.
2. По факту поступления данных.
3. По факту изменения/дополнения методики (алгоритмов) диагностики заболеваний.

На рисунке 2 представлен упрощенный бизнес-процесс обследования, диагностики и лечения заболеваний, отражающий текущую ситуацию при симптоматическом лечении, т.е. это комплекс лечебных мероприятий, направленных на устранение отдельных симптомов заболевания (например, применение анальгетиков при боли, жаропонижающих препаратов при высокой температуре и т.п.).

По результатам работы указанного решения врач будет оповещен о наличии рисков либо маркеров, связанных с различными заболеваниями. Это может существенно увеличить число диагностируемых заболеваний на ранних стадиях.

На рисунке 3 представлен бизнес-процесс обследования, диагностики и лечения заболеваний с учетом предлагаемого решения, который предполагает, помимо симптоматического лечения, еще применение программного решения для поддержки принятия врачебных решений и оценки рисков разви-

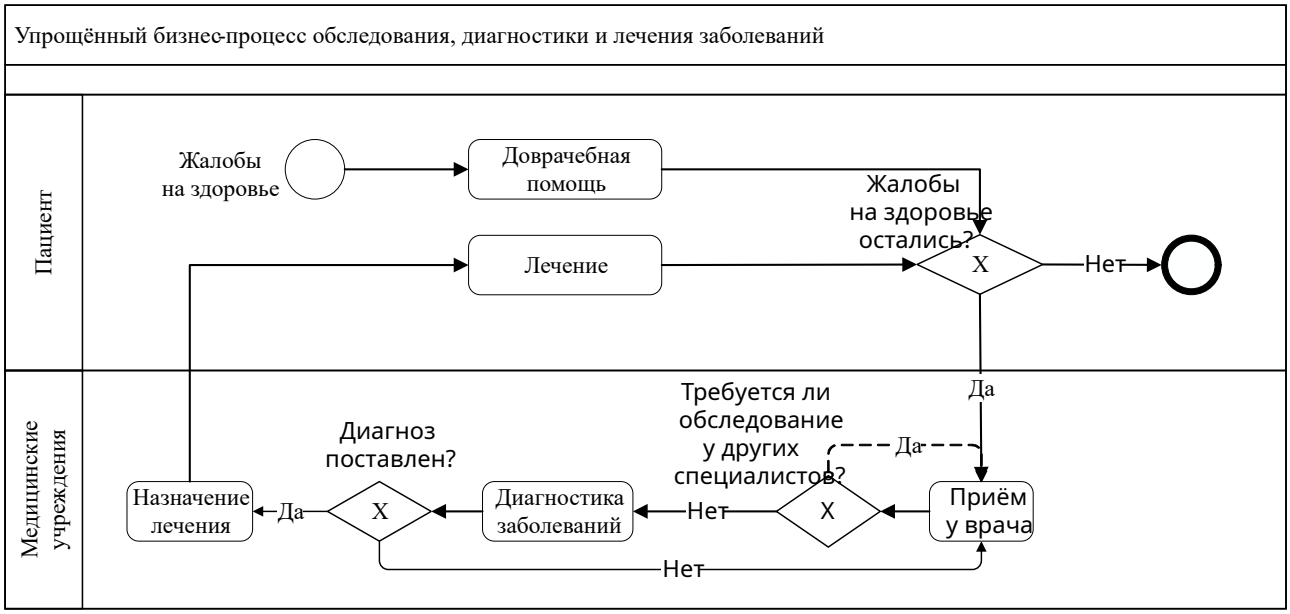


Рис. 2. *Упрощенный бизнес-процесс обследования, диагностики и лечения заболеваний*

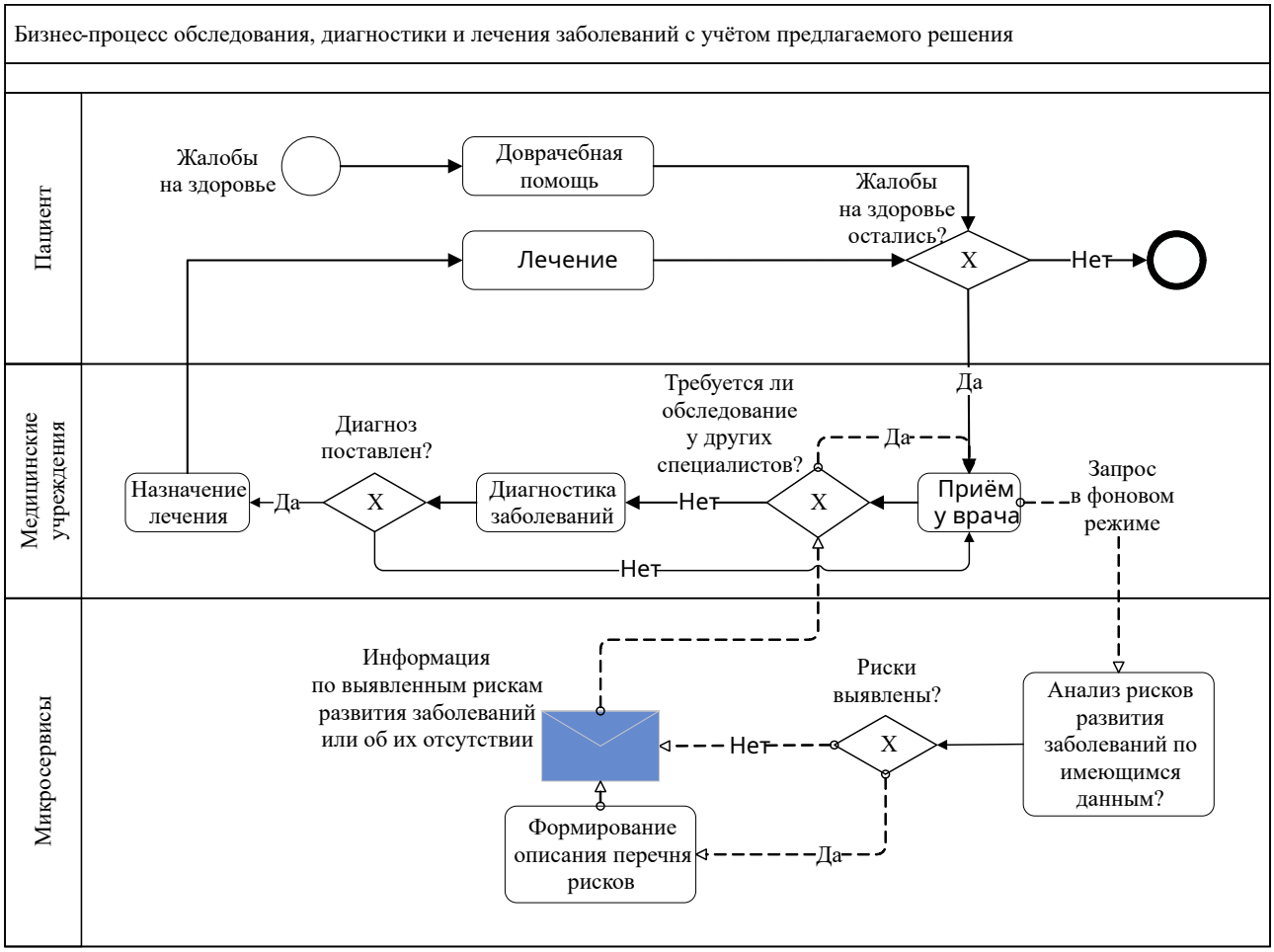


Рис. 3. *Бизнес-процесс обследования, диагностики и лечения заболеваний с учетом предлагаемого решения*

тия заболеваний. Данный подход может улучшить раннюю диагностику многих заболеваний на основе имеющихся медико-социальных данных с применением наборов алгоритмов диагностики заболеваний даже в тех случаях, когда пациент пришел на прием по другим нозологиям.

С учетом большой загруженности врачей (большое количество приемов и их короткая длительность), врачам сложно в полной мере обследовать пациента. Поэтому подсказки по выявленным рискам, которые будут направляться в фоновом режиме микросервисами, могут позволить обратить внимание на важные особенности анамнеза пациента. Это повысит качество обследования пациентов и/или позволит направить их дополнительно на обследования к другим специалистам.

Заключение

1. Проанализирована общая ситуация по профилактике заболеваний со стороны государства и медицинских организаций. Выявлены проблемы, которые, несмотря на описанные меры, продолжают оказывать негативное влияние на раннюю диагностику тяжелых заболеваний, что приводит к дорогостоящему лечению, высокой смертности и, соответственно, экономическим потерям со стороны государства.

2. Предложен вариант пополнения медицинских информационных систем данными из медицинских документов пациентов на Госуслугах, что может помочь увеличить объем данных для более качественного ведения пациентов.

3. Изучен общий подход к созданию информационных систем поддержки врачебных решений (далее – ИСПВР). Как правило, ИСПВР направлены на узкоспециализированные задачи, отсутствует интеграция с медицинскими информационными системами, что затрудняет их массовое практическое применение.

4. Предложен подход к решению задачи поддержки принятия врачебных решений, а именно предложено программное решение, которое будет интегрировано с медицинскими информационными системами, включать в себя набор микросервисов, обеспечивающих загрузку и анализ медицинских данных, автоматическое оповещение в фоновом режиме врачей по выявленным рискам развития заболеваний у пациентов, даже если пациент пришел к непрофильному врачу по другим нозологиям. Формирование дорожной карты обследования пациентов.

5. В целом, предложенное решение направлено на точечное решение проблем ранней диагностики заболеваний в соответствии с алгоритмами и помощь врачам, что может снизить тяжесть заболеваний, сохранить здоровье пациентам и, как следствие, избежать государству дополнительных трат на лечение.

ЛИТЕРАТУРА

1. Развитие здравоохранения. *Государственная программа Российской Федерации.* Режим доступа: https://minzdrav.gov.ru/ministry/programms/health/info.
2. Ройтберг Г. Е., Попова И. Е., Дорош Ж. В., Шархун О. О., Сластникова И. Д. *Онконастороженность в работе врача первичного звена: Профилактика. Скрининг. Ранняя диагностика*: *Учебное пособие.* М.: ФГАОУ ВО РНИМУ имени Н.И. Пирогова Минздрава России; 2020. Режим доступа: https://inuclear.ru/upload/iblock/186/31bb1b25ttiiwzsit2czivai6odpis08.pdf.
3. Дон Е. С., Тарасов А. В., Эпштейн О. И., Тарасов С. А. Биомаркеры в медицине: поиск, выбор, изучение и валидация. *Клиническая лабораторная диагностика*. 2017;62(1):52–59. Режим доступа: https://cyberleninka.ru/article/n/biomarkery-v-meditsine-poisk-vybor-izuchenie-i-validatsiya.
4. О стандартах медицинской помощи. *Официальный сайт Портала «Государственные услуги»*. Режим доступа: https://med-p-kgbuz-kgp14-r04.gosweb.gosuslugi.ru/informatsiya-dlya-spetsialistov/o-standartah-meditsinskoy-pomoschi/.
5. Монаков Д. М., Алтунин Д. В. Медицинские информационные системы: современные реалии и перспективы. *Российский журнал телемедицины и электронного здравоохранения*. 2022;8(4):46–53. DOI: 10.29188/2712-9217-2022-8-4-46-53.
6. Какие медицинские документы доступны на Госуслугах. *Официальный сайт Портала «Государственные услуги»*. Режим доступа: https://www.gosuslugi.ru/help/faq/medical_docs/102280.
7. Киселев И. Н. О применении искусственного интеллекта в распознавании текстов. *Вестник ВНИИДАД*. 2024;1:84–95. Режим доступа: https://journals.rcsi.science/2619-1601/article/view/267010.
8. Пахтусов А. И., Логарева Е. В., Парсаданян А. М., Каспарова А. Э., Федоров Д. А. Системный анализ процесса диагностики опухолей яичников. *Вестник кибернетики*. 2023;4:33–41. DOI: 10.35266/1999-7604-2023-4-5.

СРАВНЕНИЕ СОВРЕМЕННЫХ ТЕХНОЛОГИЙ РАСПОЗНАВАНИЯ ОБРАЗОВ

Т. С. Акобян
Сургутский государственный университет, г. Сургут, Российская Федерация
ORCID: https://orcid.org/0009-0000-6207-2374, ✍ ackobyan.tigran2018@yandex.ru

Аннотация: в представленной работе выполнен обзор методов и технологий, применяемых при построении искусственных нейронных сетей. Биометрические технологии, основанные на уникальных физических и поведенческих характеристиках человека, становятся ключевым инструментом для идентификации личности. Настоящее исследование посвящено сравнительному анализу различных моделей распознавания лиц, включая алгоритмы каскадов Хаара, dlib, MTCNN и FaceNet. Проведен анализ по девяти критериям, включая точность, полноту, время обработки и потребление вычислительных ресурсов. Для оценки эффективности алгоритмов использовался набор из 250 изображений, классифицированных по различным условиям. Полученные результаты демонстрируют, что каждый алгоритм обладает своими преимуществами и недостатками в зависимости от конкретных задач и условий эксплуатации.

Ключевые слова: нейронные сети, распознавание образов, сравнение технологий, искусственные нейронные сети, каскады Хаара.

Для цитирования: Акобян Т. С. Сравнение современных технологий распознавания образов. *Успехи кибернетики*. 2025;6(1):175–179.

Поступила в редакцию: 12.02.2025. *В окончательном варианте: 18.03.2025.*

MODERN IMAGE RECOGNITION TECHNOLOGIES: A REVIEW

T. S. Akobyan
Surgut State University, Surgut, Russian Federation
ORCID: https://orcid.org/0009-0000-6207-2374, ✍ ackobyan.tigran2018@yandex.ru

Abstract: in this paper, we provided an overview of the methods and technologies used in the construction of artificial neural networks (ANN). Biometric technologies based on a person's unique physical and behavioral characteristics have become a key tool for personal identification. We conducted a comparative analysis of various face recognition models, including Haar cascades, dlib, MTCNN, and FaceNet algorithms. We based the analysis on nine criteria, including accuracy, completeness, processing time, and computing resource consumption. To evaluate the effectiveness of the algorithms, we used a set of 250 images classified according to various conditions. The results demonstrated that each algorithm has its own advantages and disadvantages, depending on the specific tasks and operating conditions.

Keywords: neural networks, pattern recognition, technology comparison, artificial neural networks, Haar cascades.

Cite this article: Akobyan T. S. Modern Image Recognition Technologies: A Review. *Russian Journal of Cybernetics*. 2025;6(1):175–179.

Original article submitted: 12.02.2025. *Revision submitted: 18.03.2025.*

Биометрия базируется на свойствах человеческого организма, которые не подвержены изменениям (в отличие от физических паролей, пин-кодов, карт). Биометрические системы используют данные о последовательном поведении человека (речь, походка), его физических параметрах (палец, глаз, лицо), биометрических особенностях (отпечаток пальца, подпись) для идентификации и проверки личности [1].

Распознавание лица — это один из наиболее прогрессивных способов авторизации личности на сегодняшний день. Он позволяет быстро и точно идентифицировать человека, необходимого для выполнения определенных функций, например, распознавание и доступ к контролируемым зонам. Эти технологии стали неотъемлемой частью безопасности в ряде отраслей, включая государственную и корпоративную безопасность, медицинскую и банковскую сферы, аэропорты и другие крупные объекты, где безопасность является первоочередной задачей [2].

Цель исследования состоит в том, чтобы провести сравнительный анализ различных моделей распознавания лиц с целью выявления их преимуществ и недостатков в различных условиях и задачах, а также реализовать новый метод на базе комбинирования лучших технологий. Это поможет как исследователям, так и практикам в выборе наиболее подходящей модели для конкретных приложений.

Для достижения этой цели необходимо:

1. Найти, собрать и проанализировать данные для стартового набора.
2. Выбрать модели и методы.
3. Провести оценки производительности.
4. Проанализировать результат эксперимента.
5. Реализовать на базе лучших технологий комбинированный метод распознавания.

Исследование имеет практическую значимость для разработчиков систем безопасности, развития биометрических технологий, а также для обеспечения приватности и конфиденциальности личных данных.

Одним из основных методов распознавания лиц является биометрическое сопоставление, которое основано на уникальных характеристиках лица человека, таких как расстояние между глазами, форма и размер носа и т. д. Другие методы включают в себя использование глубокого обучения и нейронных сетей для анализа и сопоставления изображений лиц [3].

Для выбора оптимального алгоритма были выполнены тесты с набором данных.

Сам набор представляет из себя 250 изображений, разделенных на 5 групп:

1. Манекены и рисунки из мультфильмов с изображением лиц.
2. Лица людей в масках.
3. Лица людей с разных ракурсов.
4. Белокожие люди.
5. Темнокожие люди.

В каждой группе по 50 изображений и на каждой картинке есть как минимум одно лицо. Все изображения были подготовлены и размечены, то есть лица на изображениях были отмечены человеком — это было сделано для сравнения результата конкретного метода с результатом, который сделал человек.

Алгоритмы будут сравниваться по девяти следующим критериям:

1. Количество найденных лиц (может быть больше, чем лиц на изображении).
2. Количество ненайденных лиц.
3. Количество ложных срабатываний.
4. Точность (Precision) — процент найденных и ненайденных лиц, т.е. полнота работы системы без ложных срабатываний.
5. Полнота (Recall) — процент найденных лиц из общего количества лиц в системе.
6. Пересечение над объединением (IOU) — критерий, предполагающий сравнение области, найденной методом, с тем, что отметил человек на этапе формирования набора данных.
7. Время работы алгоритма — это требуемое время для распознавания лица в изображении.
8. CPU — процент производительности, который требуется от центрального процессора для выполнения задачи.
9. GPU — процент производительности, который требуется от графического процессора для выполнения задачи.

В ходе проведения тестов были получены следующие результаты (см. рис. 1): каскады Хаара распознали лишь треть результатов, возможно, это связано с тем, что находить людей в масках и с разным профилем является тяжелой задачей для данного алгоритма, но нужно учесть, что он является лучшим в производительности алгоритмом. DBFace же, наоборот, оказался лучшим по распознаванию, но худшим в производительности. dlib показал неплохой результат в распознавании, а также хороший результат в производительности. FaceNet показал результат, аналогичный DBFace, т.е. хорошее распознавание, но для данных действий требуется много ресурсов.

Выбрать, какой алгоритм является лучшим, достаточно тяжело, ведь критериев много, и каждый алгоритм показывает как хорошие, так и плохие результаты. В таком случае стоит воспользоваться комбинированным методом принятия решений.

Определим веса критериев для данного метода, для этого нужны оценки экспертов по ранее

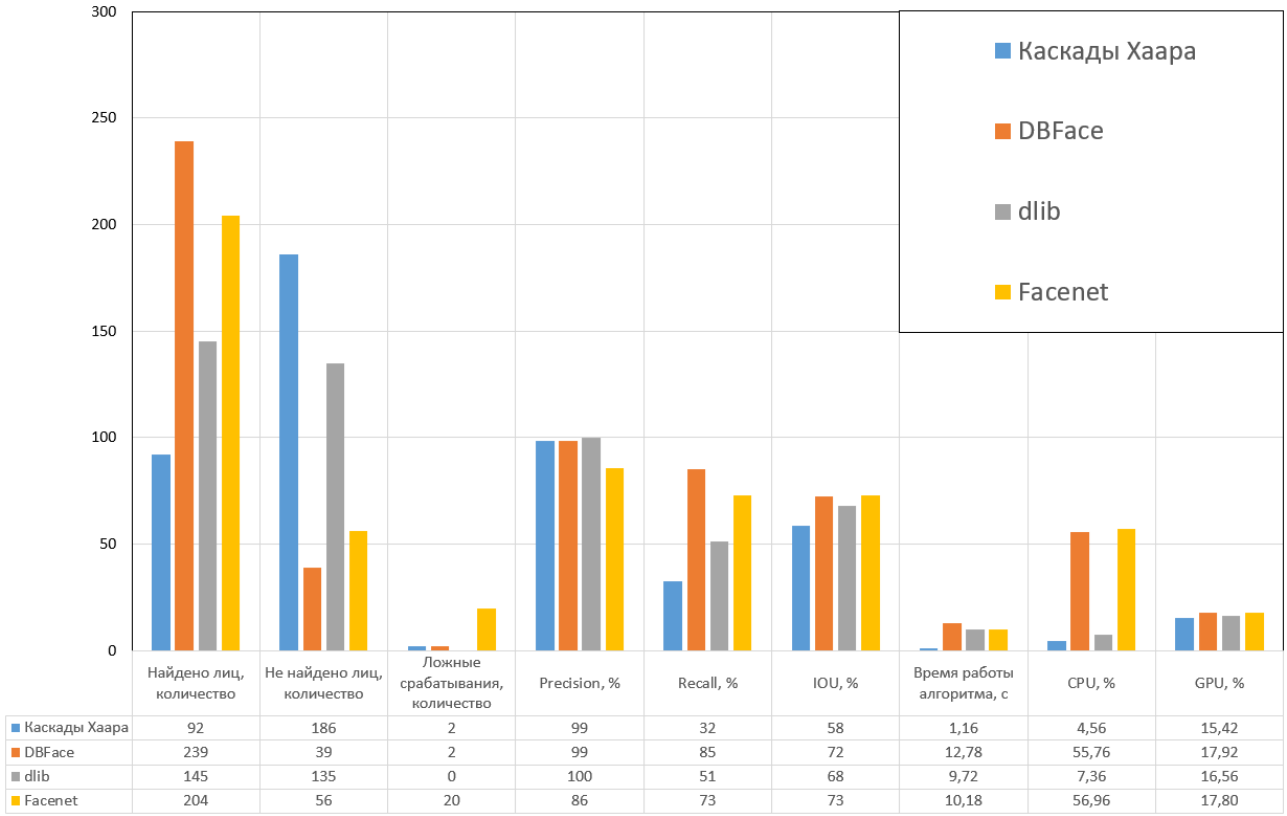


Рис. 1. *Результат применения методов после проведения тестов*

озвученным 9 критериям, по этому поводу был проведен опрос шести независимых экспертов (см. таблицу 1). Эти оценки были нормализованы, а потом вычислены веса критериев. Далее были определены суммы рангов у всех экспертов, они были равными, что говорило о согласованности между экспертами.

Таблица 1

Оценка независимых экспертов

Эксперт	Критерии									Сумма рангов
	Y1	Y2	Y3	Y4	Y5	Y6	Y7	Y8	Y9	
Э1	6	6	4	4	5	5	2	2	2	36
Э2	5	5	5	3	5	6	2	4	1	36
Э3	7	6	6	1	6	5	3	1	1	36
Э4	5	5	6	5	6	5	1	2	1	36
Э5	4	5	5	6	5	6	2	2	1	36
Э6	7	5	4	5	5	4	4	1	1	36

В комбинированном методе использовалось ограничение по максимину (для критериев: ложные срабатывания, точность, полнота и время работы алгоритма), а для выбора оптимального алгоритма критерий относительной уступки (для критериев: найдено лиц, не найдено лиц, пересечение над объединением, CPU, GPU). Нормализуем данные с метрик, использовав метод нормализации критерием Сэвиджа для минимизации максимально возможных потерь, а потом относительную нормализацию для более точного сравнения и оценки альтернатив. Далее использовали ограничение максимина для исключения решения, которое не защищает от потенциальных неблагоприятных ситуаций. В ходе расчета данным ограничением были исключены алгоритмы DBFace и FaceNet. В конечном итоге, использовав критерий относительной уступки, мы определили, что лучшей альтернативой будет являться алгоритм dlib. Хотелось скомбинировать возможности двух лучших методов — dlib и каскады Хаара — для реализации нового гибридного метода распознавания образов.

Обоснование выбора между последовательным и параллельным методами программирования

для распознавания лиц важно по нескольким причинам, поскольку каждый из этих методов имеет свои преимущества и недостатки. Чтобы понять, какой подход будет более эффективным, определим их основные характеристики и сравним.

Последовательный метод:

Преимущества:

1. Последовательная проверка результатов с использованием dlib повышает точность. Если каскад Хаара ошибается, dlib может помочь откорректировать его ошибки.
2. Меньше параллельных вычислений, что снижает вероятность ошибок из-за несовпадений в результатах.

Недостатки: потери по времени из-за двух этапов обработки: сначала каскады Хаара, затем dlib.

Параллельный метод:

Преимущества:

1. Может быть более быстрым, так как оба метода выполняются одновременно, что снижает общее время работы программы.
2. Потенциально более эффективен для обработки нескольких изображений.

Недостатки:

1. Большие требования к ресурсам (память и процессор).
2. Высокая вероятность ошибок при некорректной синхронизации или пересечении результатов двух методов.

Если говорить о математическом обосновании, то можно рассматривать задачи с точки зрения скорости (время обработки) и точности (вероятность ошибки).

Время обработки:

Параллельный метод будет эффективнее с точки зрения времени, если $T_{\text{послед}} > T_{\text{параллель}}$, где:

$$T_{\text{послед}} = T_{\text{Хаара}} + T_{\text{dlib}}$$

$$T_{\text{параллель}} = \max(T_{\text{Хаара}},\ T_{\text{dlib}}).$$

Так как оба метода работают одновременно, то время обработки ограничено временем самого долгого метода, а не суммой времени обоих.

Точность:

1. В случае с параллельным методом вероятность ошибки (например, ложные положительные срабатывания) может быть выше, так как два метода могут не всегда точно совпадать в результатах.
2. В случае последовательного метода dlib будет служить фильтром для корректировки результатов, что снижает количество ложных срабатываний [4].

Вывод: так как приоритетом является точность, а не скорость, то будет реализован последовательный метод.

Для распознавания лиц с помощью комбинированного метода каскадов Хаара и библиотеки dlib необходимо использовать два этапа. На первом этапе, с помощью каскадов Хаара, будет происходить быстрое и предварительное распознавание лиц в изображении, а на втором этапе метод из библиотеки dlib уточнит результаты, проверив найденные области.

Шаги решения:

1. Применение каскадов Хаара: каскады Хаара являются быстрым методом для обнаружения объектов на изображении. Это основано на обучении классификаторов, которые с помощью простых признаков (например, разница в яркости пикселей в разных областях изображения) определяют, является ли область изображения лицом. Каскады Хаара используют набор слабых классификаторов, которые комбинируются в более мощный классификатор [5].
2. Использование библиотеки dlib: после того как каскады Хаара распознают лица, мы применяем dlib для уточнения положения лиц. Это происходит путем применения метода поиска лица с помощью модели HOG (Histogram of Oriented Gradients), который дает более точные результаты в случае, когда каскады Хаара могут ошибиться.

Алгоритм:

1. Каскады Хаара: алгоритм Хаара использует признаки, которые вычисляются через суммы значений пикселей в прямоугольных областях. Признаки выглядят как простые «окна» для каждого объекта, которые обучаются на наборе лиц и фона. С помощью классификаторов, обученных на этих признаках, каскад Хаара может быстро искать объекты на изображении.

Если рассматривать это как математическое преобразование, то задача заключается в том, чтобы для каждого окна вычислить сумму пикселей внутри и вокруг него, а затем принять решение, что это лицо (или нет), используя классификатор.

2. Обнаружение лиц с помощью dlib: метод из библиотеки dlib основан на использовании градиентов изображений, которые вычисляются как:

$$G_x = \frac{\partial I}{\partial x}, G_y = \frac{\partial I}{\partial y},$$

где I — это изображение, а G_x и G_y — градиенты изображения по осям x и y соответственно. Эти градиенты характеризуют изменения яркости в разных направлениях и помогают выделить контуры, которые могут быть признаками лиц.

3. Комбинированный метод: идея комбинированного подхода заключается в том, что каскады Хаара дают быстрый, но не всегда точный результат, а метод dlib помогает уточнить положение лиц, уменьшая количество ложных срабатываний.

Заключение

В ходе исследования был проведен сравнительный анализ алгоритмов распознавания лиц: каскадов Хаара, dlib, MTCNN и FaceNet. Эксперименты показали, что точность и полнота выше у нейронных сетей (MTCNN, FaceNet), однако они требуют значительных вычислительных ресурсов. В то же время каскады Хаара демонстрируют высокую скорость работы, но уступают в качестве распознавания в сложных условиях. Алгоритм dlib показал сбалансированные результаты по точности и производительности, что делает его универсальным решением.

Полученные результаты подтверждают, что выбор алгоритма зависит от конкретных задач и условий эксплуатации, а также требований к скорости обработки и точности. Настоящее исследование может быть полезно разработчикам систем безопасности и биометрических технологий при выборе оптимального решения.

ЛИТЕРАТУРА

1. Частикова В. А., Титова А. А., Войлова Д. О. Аналитический обзор методов идентификации личности на основе биометрических характеристик. *Вестник Адыгейского государственного университета. Серия: Естественно-математические и технические науки.* 2022;1(296):92–112. DOI: 10.53598/2410-3225-2022-1-296-92-112.
2. Ковшов Е. Е., Завистовская Т. А. Разработка автоматизированной системы контроля и управления доступом на основе анализа динамики носогубных мышц лица человека. *Современные проблемы науки и образования.* 2013;2:185.
3. Иванова Е. В., Струева А. Ю. Система учета посещаемости студентов на основе распознавания лиц. *Вестник Южно-Уральского государственного университета. Серия: Вычислительная математика и информатика.* 2021;4:60–73. DOI: 10.14529/cmse210404.
4. Калинин Н. Н., Скурыдин Ю. Г., Скурыдина Е. М. Оценка производительности моделей машинного обучения, основанных на библиотеках tensorflow и dlib при определении состояния сонливости на Raspberry Pi. *Высокопроизводительные вычислительные системы и технологии.* 2023;1:106–114.
5. Хорошева А. А., Федько М. Д., Юсупов А. Р. Применение каскада Хаара для разработки системы распознавания птиц. *Молодой исследователь Дона.* 2020;2(23):166–170.

Сетевое издание. Полные тексты статей размещаются на официальном сайте издания jcyb.ru.

Адрес учредителя и издателя:
Федеральное государственное автономное учреждение «Федеральный научный центр
Научно-исследовательский институт системных исследований Национального исследовательского
центра «Курчатовский институт»,
117218, г. Москва, Нахимовский просп., дом 36, корпус 1. Телефон: +7 (495) 718-21-10.

Дата выхода в свет 31.03.2025.
Формат 60 × 84/8.
Усл. печ. л. 12,3. Уч.-изд. л. 18,2.
Цена свободная.