

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

ВЕСТНИК ТОМСКОГО ГОСУДАРСТВЕННОГО УНИВЕРСИТЕТА

УПРАВЛЕНИЕ,
ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА
И ИНФОРМАТИКА

TOMSK STATE UNIVERSITY
JOURNAL OF CONTROL AND COMPUTER SCIENCE

Научный журнал

2025

№ 73

Зарегистрирован в Федеральной службе по надзору в сфере
массовых коммуникаций, связи и охраны культурного наследия
(свидетельство о регистрации ПИ № ФС 77-29497 от 27 сентября 2007 г.)

Подписной индекс в объединённом каталоге «Пресса России» 44031

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»,
Высшей аттестационной комиссии

16+

Учредитель – Национальный исследовательский Томский государственный университет (НИ ТГУ)

РЕДАКЦИОННАЯ КОЛЛЕГИЯ

Замятин Александр Владимирович – гл. редактор, проф., д-р техн. наук, НИ ТГУ. Тел. +73822529496
Смагин Валерий Иванович – зам. гл. редактора, проф., д-р техн. наук, НИ ТГУ. Тел. +73822529485
Андреева Валентина Валерьевна – ответственный секретарь, доц., канд. техн. наук, НИ ТГУ. Тел. +73822529496
E-mail: vestnik_uvti@mail.tsu.ru

Воробейчиков Сергей Эрикович – д-р физ.-мат. наук, НИ ТГУ
Вишневский Владимир Миронович – проф., д-р техн. наук, Институт проблем управления РАН (г. Москва)
Гордов Евгений Петрович – проф., д-р физ. мат. наук, Институт мониторинга климатических и экологических систем СО РАН (г. Томск)
Кошкин Геннадий Михайлович – проф., д-р физ.-мат. наук, НИ ТГУ
Костюк Юрий Леонидович – проф., д-р техн. наук, НИ ТГУ
Матросова Анжела Юрьевна – проф., д-р техн. наук, НИ ТГУ
Назаров Анатолий Андреевич – проф., д-р техн. наук, НИ ТГУ
Нежелская Людмила Алексеевна – проф., д-р физ.-мат. наук, НИ ТГУ
Пашинская Татьяна Юрьевна – д-р физ.-мат. наук, НИ ТГУ
Перепелкин Евгений Александрович – проф., д-р техн. наук, Санкт-Петербургский государственный университет аэрокосмического приборостроения (г. Санкт-Петербург)
Пчелинцев Евгений Анатольевич – д-р физ.-мат. наук, НИ ТГУ
Самуйлов Константин Евгеньевич – проф., д-р техн. наук, РУДН (г. Москва)
Семенкин Евгений Станиславович – проф., д-р техн. наук, СибГУ им. акад. М.Ф. Решетнева (г. Красноярск)
Сушенко Сергей Петрович – проф., д-р техн. наук, НИ ТГУ
Фархадов Маис Паша оглы – д-р техн. наук, Институт проблем управления РАН (г. Москва)
Харин Юрий Семенович – академик НАНБ, проф., д-р физ.-мат. наук, БГУ, Республика Беларусь (г. Минск)
Цициашвили Гурами Шалвович – проф., д-р физ.-мат. наук, ИПМ ДВО РАН (г. Владивосток)

РЕДАКЦИОННЫЙ СОВЕТ

Александр Николаевич Дудин – проф., д-р физ.-мат. наук, БГУ, Минск, Республика Беларусь
Ана Роза Ковали (Ana Rosa Cavalli) – проф., д-р. философии (PhD), University VII Paris, Франция
Енцо Орсингер (Enco Orsinger) – проф., д-р. философии (PhD), University of Rome, Италия
Паоло Принетто (Paolo Prinetto) – проф. Politecnico Institute, Torino, Италия
Нина Владимировна Евтушенко – проф., д-р. техн. наук, ИСП РАН им. Иваницова В.П., Москва, Россия
Ервант Зориан (Yervant Zorian) – д-р. философии (PhD), Synopsys, Mountain View, США

JOURNAL INFO

Tomsk State University Journal of Control and Computer Science is an independent peer-reviewed research journal that welcomes submissions from across the world.

Tomsk State University Journal of Control and Computer Science is issued four times per year, and can be subscribed to in the Russian Press Joint Catalogue (Subscription Index 44031)

The publication in the journal is free of charge and may be in Russian or in English.

The topics of the journal are the following:

- *control of dynamical systems,*
- *mathematical modeling,*
- *data processing,*
- *informatics and programming,*
- *discrete function and automation,*
- *designing and diagnostics of computer systems.*

Rules of registration articles are given in a site:

<http://journals.tsu.ru/informatics/>

ISSN 2311-2085 (Online), ISSN 1998-8605 (Print).

О ЖУРНАЛЕ

Журнал «Вестник Томского государственного университета.

Управление, вычислительная техника и информатика» выходит ежеквартально и распространяется по подписке

Статьи публикуются на русском и английском языках.

Тематика публикаций журнала:

- *управление динамическими системами,*
- *математическое моделирование,*
- *обработка информации,*
- *информатика и программирование,*
- *дискретные функции и автоматы,*
- *проектирование и диагностика вычислительных систем.*

Журнал входит в систему Российского Индекса Научного Цитирования (РИНЦ).

Правила оформления статей приведены на сайте:

<http://journals.tsu.ru/informatics/>

ISSN 2311-2085 (Online), ISSN 1998-8605 (Print).

СОДЕРЖАНИЕ

УПРАВЛЕНИЕ ДИНАМИЧЕСКИМИ СИСТЕМАМИ

Алиева С.Т. Необходимые условия оптимальности для разностных уравнений дробного порядка с запаздыванием	4
Зубов Н.Е., Рябченко В.Н. О редукции модальных тестов на управляемость и наблюдаемость МИМО-системы	15
Перепелкин Е.А. О задаче управления качеством переходных процессов по выходу в матричной системе второго порядка с обратной связью по состоянию	23

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ

Малкин А.Ю., Марков Н.Г. Исследование устойчивости моделей полносверточных нейронных сетей с робастными функциями потерь к импульсным помехам на изображениях	30
Цициашвили Г.Ш. Нестационарные потоки в СеМО без очереди и с детерминированным временем обслуживания	41

ОБРАБОТКА ИНФОРМАЦИИ

Базилевский М.П. О связи между многослойной модульной регрессионной моделью и регрессией в виде производственной функции Леонтьева	46
Гервас Н.В., Зеркаль С.М. Вычислительная кинематическая диагностика сред с возможностью линеаризации скорости зондирующего сигнала в условиях томографического эксперимента	58
Dmitriev Yu.G., Koshkin G.M. Nonparametric evaluation of continuous r -year deferred m -year term life annuity using information on probabilistic characteristics of lifetime	65
Добронет Б.С., Попова О.А. Численная вероятностная арифметика для обобщенных кусочно-полиномиальных функций	75
Жиравок А.Н., Зуев А.В., Протченко А.А., Горностаев И.В., Сердюкова А.Ю. Метод идентификации дефектов в датчиках с использованием скользящих наблюдателей высокого порядка	81
Игнатьев Н.А. Регуляризаторы по наборам обобщенных оценок	90
Исаева О.С., Исаев С.В., Кулясов Н.В. Подход к выделению значимых признаков сетевой активности устройств Интернета вещей	100
Нежельская Л.А., Пономаренко В.Д. Рекуррентный обобщенный асинхронный поток событий с продлевающимся мертвым временем в особом случае	110

ИНФОРМАТИКА И ПРОГРАММИРОВАНИЕ

Обходский А.В., Кульбакин Д.Е., Обходская Е.В., Лаконкин В.С., Родионов Е.О., Сачков В.И., Чернов В.И., Чойнзонов Е.Л. Нейронные сети в диагностике злокачественных новообразований по выдыхаемому воздуху	122
Прокудина Ю.А., Моисеев А.Н. Концепция архитектуры приложения для построения аналитической аппроксимации распределения вероятностей числа заявок в СМО на основе имитационного моделирования	134

CONTENTS

CONTROL OF DYNAMICAL SYSTEMS

Aliyeva S.T. Necessary optimality conditions for fractional-order difference equations with delay	4
Zubov N.E., Ryabchenko V.N. On reduction of modal tests on controllability and observability of MIMO system	15
Perepelkin E.A. On the problem of transient process quality control of output in a matrix second-order system with state feedback	23

MATHEMATICAL MODELING

Malkin A.Yu., Markov N.G. Research of resistance of fully convolutional neural network models with robust loss functions to impulse noise in images	30
Tsitsiashvili G.Sh. Non-stationary flows in queuing networks without queue and with deterministic service time	41

DATA PROCESSING

Bazilevskiy M.P. On the relationship between the multilayer modular regression model and regression in the form of Leontief's production function	46
Gervas N.V., Zerkal S.M. Computational kinematic diagnostics of media with the possibility of linearizing the probing signal velocity in the context of a tomographic experiment	58
Dmitriev Yu.G., Koshkin G.M. Nonparametric evaluation of continuous r -year deferred m -year term life annuity using information on probabilistic characteristics of lifetime	65
Dobronets B.S., Popova O.A. Numerical probabilistic arithmetic for generalized piecewise polynomial functions	75
Zhirabok A.N., Zuev A.V., Protsenko A.A., Gornostaev I.V., Serduykova A.Yu. Method of sensor fault identification based on higher-order sliding mode observers	81
Ignatev N.A. Regularizers on sets of generalized estimates	90
Isaeva O.S., Isaev S.V., Kulyasov N.V. Approach to identifying significant features of IoT device network activity	100
Nezhel'skaya L.A., Ponomarenko V.D. A recurrent generalized asynchronous event flow with extended dead time in a special case	110

INFORMATICS AND PROGRAMMING

Obkhodskiy A.V., Kulbakin D.E., Obkhodskaya E.V., Lakonkin V.S., Rodionov E.O., Sachkov V.I., Chernov V.I., Choyznzonov E.L. Neural networks in the diagnosis of malignant neoplasms by exhaled air	122
Prokudina Y.A., Moiseev A.N. Concept of application architecture for constructing analytical approximation of the probability distribution of the number of customers in queuing systems based on simulation modeling	134

УПРАВЛЕНИЕ ДИНАМИЧЕСКИМИ СИСТЕМАМИ

CONTROL OF DYNAMICAL SYSTEMS

Научная статья

УДК 517.977; 62-50

doi: 10.17223/19988605/73/1

Необходимые условия оптимальности для разностных уравнений
дробного порядка с запаздыванием

Саадат Тофик кызы Алиева

*Бакинский государственный университет, Баку, Азербайджан;
Институт систем управления науки и образования Азербайджана, Баку, Азербайджан;
Азербайджанский государственный экономический университет, Баку, Азербайджан;
Университет Азербайджан, Баку, Азербайджан, saadat.t.aliyeva@au.edu.az*

Аннотация. Исследуется терминальная дискретная задача оптимального управления с запаздыванием дробного порядка. С применением одного из вариантов метода приращений доказаны необходимые условия оптимальности первого порядка в форме аналога дискретного принципа максимума, линеаризованного условия максимума и уравнения Эйлера.

Ключевые слова: разностное уравнение дробного порядка с запаздыванием; формула приращения; дискретный принцип максимума; оптимальное управление; необходимые условия оптимальности; аналог линеаризованного условия максимума; аналог уравнения Эйлера.

Для цитирования: Алиева С.Т. Необходимые условия оптимальности для разностных уравнений дробного порядка с запаздыванием // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2025. № 73. С. 4–14. doi: 10.17223/19988605/73/1

Original article

doi: 10.17223/19988605/73/1

Necessary optimality conditions for fractional-order difference equations with delay

Saadat T. Aliyeva

*Baku State University, Baku, Azerbaijan;
Institute of Control Systems, Science and Education of Azerbaijan, Baku, Azerbaijan;
Azerbaijan State University of Economics, Baku, Azerbaijan;
Azerbaijan University, Baku, Azerbaijan, saadat.t.aliyeva@au.edu.az*

Abstract. The paper studies a terminal discrete optimal control problem with fractional-order delay. Using one of the variants of the increment method, the necessary conditions for first-order optimality are proved in the form of an analogue of the discrete maximum principle, a linearized maximum condition, and the Euler equation.

Keywords: fractional-order difference equations with delay; increment formula; discrete maximum principle; optimal control; necessary conditions for optimality; an analogue of the linearized maximum condition; an analogue of the Euler equation.

For citation: Aliyeva, S.T. (2025) Necessary optimality conditions for fractional-order difference equations with delay. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 73. pp. 4–14. doi: 10.17223/19988605/73/1

Введение

В последние годы дробные задачи вариационного исчисления и оптимального управления привлекают большее внимание многих авторов (см., напр.: [1–3] и ссылки в них). Но имеющиеся результаты касаются только случая непрерывного времени. Дробные задачи оптимального управления, которые могут быть представлены обыкновенными дробными дифференциальными уравнениями, можно рассматривать как развитие или расширение обычных задач оптимального управления.

Принцип максимума Понтрягина является ключевым результатом в теории необходимых условий оптимальности первого порядка. Первоначально он был доказан для задач оптимального управления, включающих обыкновенные дифференциальные уравнения [4]. Впоследствии были выведены различные условия оптимальности для ряда систем, охватывающих условия как первого, так и более высокого порядка [5].

Необходимые условия оптимальности для дробных дифференциальных уравнений с запаздыванием включают комбинацию дробного дискретного исчисления, вариационного исчисления и адаптации принципов из теории оптимального управления. Полученные условия создают основу для разработки оптимальных стратегий управления в системах с эффектами памяти и временными задержками, которые часто встречаются в реальных приложениях.

Хорошо известно, что дискретные аналоги дифференциальных уравнений могут быть очень полезны в приложениях [6, 7] и что дробные дифференциальные уравнения Эйлера–Лагранжа чрезвычайно трудно решить, поскольку необходимо их дискретизировать [2, 8].

Дискретное исчисление всегда предпочтительнее, особенно когда компьютеры используются для изучения свойств определенных динамических задач управления. Широко исследованы свойства разных обыкновенных разностных уравнений. Однако теория дробно-разностных уравнений все еще разработана мало.

В работах [9, 10] рассмотрена одна задача оптимального управления, описываемая обыкновенным разностным уравнением дробного порядка и получен ряд необходимых условий оптимальности первого и второго порядка.

Многие сложные процессы описываются обыкновенными разностными уравнениями с запаздыванием дробного порядка [11–14]. Разностные уравнения дробного порядка с запаздыванием представляют собой значительную и быстро развивающуюся область исследований в более широкой области дробного исчисления. Традиционные разностные уравнения моделируют системы, где изменения происходят на дискретных временных шагах, в то время как дробное исчисление расширяет концепцию производных и интегралов до нецелых порядков, охватывая эффекты памяти и зависимость на больших расстояниях, часто наблюдаемые в явлениях реального мира. Объединяя эти две концепции, разностные уравнения дробного порядка с запаздыванием включают временные задержки в динамику системы, еще больше повышая их способность реалистично моделировать сложные системы.

В данной статье исследуется задача оптимального управления для модели, описываемой разностными уравнениями с запаздыванием дробного порядка.

Цель работы – при различных предположениях получить ряд необходимых условий оптимальности.

Установлены аналоги дискретного принципа максимума, линеаризованного принципа максимума, получен аналог классического уравнения Эйлера.

1. Основные понятия и постановка задачи

Рассмотрим некоторые необходимые определения и понятия.

Определение 1 [15–17]. Расширенный биномиальный коэффициент $\binom{a}{n}$ определяется следующим образом:

$$\binom{a}{n} = \begin{cases} \frac{\Gamma(a+1)}{\Gamma(a-n+1)\Gamma(n+1)}, & n > 0, \\ 1, & n = 0, \\ 0 & n < 0. \end{cases}$$

Определение 2 [18]. Для любого $x, y \in R$ $x^{(y)} = \frac{\Gamma(x+1)}{\Gamma(x+1-y)}$, где Γ – гамма функция, для которой выполняется $\Gamma(x+1) = x\Gamma(x)$.

Пусть a – произвольное действительное число, и $b = k + a$, здесь $k \in N, k \geq 2$; $T = \{a, a+1, \dots, b\}$, $T^k = \{a, a+1, \dots, b-1\}$, а \mathbb{T} – множество функций определенных на T .

Определение 3. Пусть $f \in \mathbb{T}$. Выражения

$${}_a\Delta_t^{-\alpha} f(t) = \frac{1}{\Gamma(\alpha)} \sum_{s=a}^t (t+\alpha-\sigma(s))^{(\alpha-1)} f(s), \quad {}_t\Delta_b^{-\alpha} f(t) = \frac{1}{\Gamma(\alpha)} \sum_{s=t+a}^b (s+\alpha-\sigma(t))^{(\alpha-1)} f(s)$$

называются соответственно левыми и правыми дробными суммами порядка $\alpha > 0$.

Определение 4. Пусть $0 < \alpha \leq 1$ и $\mu = 1 - \alpha$, тогда для функции $f \in \mathbb{T}$ левые и правые дробные суммы порядка α определяются в виде:

$$\begin{aligned} {}_a\Delta_t^{\alpha} f(t) &= \Delta({}_a\Delta_t^{-\mu} f(t)), \\ {}_t\Delta_b^{\alpha} f(t) &= -\Delta({}_t\Delta_b^{-\mu} f(t)). \end{aligned}$$

Приведем некоторые известные свойства дробной суммы и дробной разности из [15–17]:

1. $\Delta^{\alpha} \Delta^{\beta} f(t) = \Delta^{\alpha+\beta} f(t)$;
2. $\Delta^{-\alpha} \Delta^{\alpha} f(t) = f(t) - f(0)$;
3. $\Delta^{\alpha} \Delta^{-\alpha} f(t) = f(t)$;
4. $\Delta^{\alpha} f(0) = 0$ и $\Delta^{\alpha} f(1) = f(1) - f(0) = \Delta f(1)$.

Перейдем к формулировке постановки задачи. Пусть управляемый процесс описывается системой нелинейных разностных уравнений с запаздыванием дробного порядка α

$$\Delta^{\alpha} x(t+1) = f(t, x(t), x(t-h), u(t)), \quad t \in T = \{t_0, t_0+1, \dots, t_1-1\}, \quad (1)$$

с начальным условием

$$x(t) = \varphi(t), \quad t \in E_{t_0} = \{t_0-h, t_0-h+1, \dots, t_0-1\}, \quad x(t_0) = x_0. \quad (2)$$

Здесь $x(t)$ – n -мерный вектор фазовых переменных, t_0, t_1 – заданные числа, x_0 – заданный постоянный вектор, $f(t, x, y, u)$ – заданная n -мерная вектор-функция, дискретная по t и непрерывная по (x, y, u) вместе с $f_x(t, x, y, u)$, $f_y(t, x, y, u)$, h – заданное натуральное число (запаздывание), $x_0, \varphi(t), t \in E_{t_0}$ заданы, $\Delta^{\alpha} x(t)$ ($0 < \alpha < 1$) – дробный оператор порядка α , а $u(t)$ – r -мерный дискретный вектор управляющих воздействий со значениями из заданного непустого и ограниченного множества U , т.е.

$$u(t) \in U \subset R^r, \quad t \in T. \quad (3)$$

Такие управляющие функции называем допустимыми управлениями.

Цель состоит в том, чтобы минимизировать функционал

$$S(u) = \Phi(x(t_1)), \quad (4)$$

определенный на решениях системы (1), порожденных всевозможными допустимыми управлениями. Здесь $\Phi(x)$ – заданная непрерывно дифференцируемая скалярная функция.

Задача (1)–(4) называется задачей дискретного терминального управления с запаздыванием дробного порядка.

Допустимое управление $u(t)$, доставляющее минимум функционалу (4) при ограничениях (1)–(3), называется оптимальным управлением, а допустимый процесс $(u(t), x(t))$ – оптимальным процессом.

2. Формула приращения критерия качества и необходимое условие оптимальности

Пусть $(u(t), x(t))$ – фиксированный допустимый процесс, а $(\bar{u}(t) = u(t) + \Delta u(t), \bar{x}(t) = x(t) + \Delta x(t))$ – произвольный допустимый процесс. Тогда ясно, что $\Delta x(t)$ будет решением следующей системы:

$$\begin{aligned} \Delta^\alpha (\Delta x(t+1)) &= f(t, \bar{x}(t), \bar{x}(t-h), \bar{u}(t)) - f(t, x(t), x(t-h), u(t)), \quad t \in T, \\ \Delta x(t_0 - h) &= 0, \dots, \Delta x(t_0) = 0. \end{aligned} \quad (5)$$

Пусть $\psi(t)$ – пока неизвестный n -мерный дискретный вектор-столбец,

$$H(t, x, y, u, \psi) = \psi'(t) f(t, x, y, u)$$

дискретный аналог функции Гамильтона–Понтрягина, где $y(t) = x(t-h)$.

Умножая обе части тождества (5) слева скалярно на пока неизвестную вектор-функцию $\psi(t)$, затем суммируя обе части полученного тождества по t от t_0 до $t_1 - 1$ и при этом учитывая вид функции Гамильтона–Понтрягина, получим

$$\sum_{t=t_0}^{t_1-1} \psi'(t) \Delta^\alpha (\Delta x(t+1)) = \sum_{t=t_0}^{t_1-1} [H(t, \bar{x}(t), \bar{x}(t-h), \bar{u}(t), \psi(t)) - H(t, x(t), x(t-h), u(t), \psi(t))].$$

Поэтому приращение функционала $S(u)$ можно записать в виде:

$$\begin{aligned} \Delta S(u) &= S(\bar{u}) - S(u) = \phi(\bar{x}(t_1)) - \phi(x(t_1)) + \sum_{t=t_0}^{t_1-1} \psi'(t) \Delta^\alpha (\Delta x(t+1)) - \\ &- \sum_{t=t_0}^{t_1-1} [H(t, \bar{x}(t), \bar{x}(t-h), \bar{u}(t), \psi(t)) - H(t, x(t), x(t-h), u(t), \psi(t))]. \end{aligned} \quad (6)$$

Используя формулу Тейлора, приращение функционала (6) представим в виде:

$$\begin{aligned} \Delta S(u) &= S(\bar{u}) - S(u) = \phi_x(x(t_1)) \Delta x(t_1) + \sum_{t=t_0}^{t_1-1} \psi'(t) \Delta^\alpha (\Delta x(t+1)) - \\ &- \sum_{t=t_0}^{t_1-1} \frac{\partial H'(t, x(t), y(t), u(t), \psi(t))}{\partial x(t)} \Delta x(t) - \sum_{t=t_0}^{t_1-1} \frac{\partial H'(t, x(t), y(t), u(t), \psi(t))}{\partial y(t)} \Delta y(t) - \\ &- \sum_{t=t_0}^{t_1-1} \frac{\partial H_{\bar{u}}'(t, x(t), y(t), u(t), \psi(t))}{\partial x(t)} \Delta x(t) - \sum_{t=t_0}^{t_1-1} \frac{\partial H_{\bar{u}}'(t, x(t), y(t), u(t), \psi(t))}{\partial y(t)} \Delta y(t) - \\ &- \sum_{t=t_0}^{t_1-1} \Delta_{\bar{u}} H'(t, x(t), y(t), u(t), \psi(t)) - \sum_{t=t_0}^{t_1-1} o_2(\|\Delta z(t)\|) + o_1(\|\Delta x(t_1)\|), \end{aligned} \quad (7)$$

где по определению

$$\Delta_{\bar{u}} H[t] = H(t, x(t), y(t), \bar{u}(t), \psi(t)) - H(t, x(t), y(t), u(t), \psi(t)),$$

$$H_x[t] = H_x(t, x(t), y(t), u(t), \psi(t)),$$

$$H_y[t] = H_y(t, x(t), y(t), u(t), \psi(t)),$$

$$\Delta z = (\Delta x, \Delta y)',$$

$$\Delta_{\bar{u}} H_x[t] = H_x(t, x(t), y(t), \bar{u}(t), \psi(t)) - H_x(t, x(t), y(t), u(t), \psi(t)).$$

Выполним преобразование формулы приращения (7).

Сначала рассмотрим выражение

$$\sum_{t=t_0}^{t_1-1} \psi'(t) \Delta^\alpha (\Delta x(t+1)).$$

Сделав в нем замену переменных $(t+1) = s$ и учитывая начальное условие (6), получим

$$\begin{aligned} \sum_{t=t_0}^{t_1-1} \psi'(t) \Delta^\alpha (\Delta x(t+1)) &= \sum_{t=t_0+1}^{t_1} \psi'(t-1) \Delta^\alpha (\Delta x(t)) = \\ &= \psi'(t_1-1) \Delta^\alpha (\Delta x(t_1)) - \psi'(t_0-1) \Delta^\alpha (\Delta x(t_0)) + \sum_{t=t_0}^{t_1-1} \psi'(t-1) \Delta^\alpha (\Delta x(t)) = \\ &= \psi'(t_1-1) \Delta^\alpha (\Delta x(t_1)) + \sum_{t=t_0}^{t_1-1} \psi'(t-1) \Delta^\alpha (\Delta x(t)). \end{aligned} \quad (8)$$

Из свойства дробной суммы получаем, что

$$\Delta x(t) = \Delta^{-\alpha} \Delta^\alpha (\Delta x(t)) = \frac{1}{\Gamma(\alpha)} \sum_{s=t_0}^t (t + \alpha - \sigma(s))^{(\alpha-1)} \Delta^\alpha (\Delta x(s)), \quad (9)$$

$$\begin{aligned} \Delta x(t_1) &= \Delta^{-\alpha} \Delta^\alpha (\Delta x(t_1)) = \frac{1}{\Gamma(\alpha)} \sum_{s=t_0}^{t_1} (t_1 + \alpha - \sigma(s))^{(\alpha-1)} \Delta^\alpha (\Delta x(s)) = \\ &= \Delta^\alpha (\Delta x(t_1)) + \frac{1}{\Gamma(\alpha)} \sum_{s=t_0}^{t_1-1} (t_1 + \alpha - \sigma(s))^{(\alpha-1)} \Delta^\alpha (\Delta x(s)). \end{aligned} \quad (10)$$

Теперь, принимая во внимание тождества (8)–(10), из формулы приращения (7) получим

$$\begin{aligned} \Delta S(u) &= \phi_x'(x(t_1)) \left(\Delta^\alpha (\Delta x(t_1)) + \frac{1}{\Gamma(\alpha)} \sum_{s=t_0}^{t_1-1} (t_1 + \alpha - \sigma(s))^{(\alpha-1)} \Delta^\alpha (\Delta x(s)) \right) + \\ &+ \psi'(t_1-1) \Delta^\alpha (\Delta x(t_1)) + \sum_{t=t_0}^{t_1-h-1} \psi'(t-1) \Delta^\alpha (\Delta x(t)) + \sum_{t=t_1-h}^{t_1-1} \psi'(t-1) \Delta^\alpha (\Delta x(t)) - \\ &- \frac{1}{\Gamma(\alpha)} \sum_{t=t_0}^{t_1-h-1} \left[\sum_{s=t}^{t_1-1} (s + \alpha - \sigma(t))^{(\alpha-1)} \frac{\partial H'(s, x(s), y(s), u(s), \psi(s))}{\partial x(s)} \right] \Delta^\alpha (\Delta x(t)) - \\ &- \frac{1}{\Gamma(\alpha)} \sum_{t=t_1-h}^{t_1-1} \left[\sum_{s=t}^{t_1-1} (s + \alpha - \sigma(t))^{(\alpha-1)} \frac{\partial H'(s, x(s), y(s), u(s), \psi(s))}{\partial x(s)} \right] \Delta^\alpha (\Delta x(t)) - \\ &- \frac{1}{\Gamma(\alpha)} \sum_{t=t_0}^{t_1-h-1} \left[\sum_{s=t}^{t_1-1} (s + \alpha - \sigma(t))^{(\alpha-1)} \frac{\partial H'(s, x(s), y(s), u(s), \psi(s))}{\partial y(s)} \right] \Delta^\alpha (\Delta x(t)) - \\ &- \sum_{t=t_0}^{t_1-1} \frac{\partial H_{\bar{u}}'(t, x(t), y(t), u(t), \psi(t))}{\partial x(t)} \Delta x(t) - \sum_{t=t_0}^{t_1-1} \frac{\partial H_{\bar{u}}'(t, x(t), y(t), u(t), \psi(t))}{\partial y(t)} \Delta y(t) - \\ &- \sum_{t=t_0}^{t_1-1} \Delta_{\bar{u}} H(t, x(t), y(t), u(t), \psi(t)) - \sum_{t=t_0}^{t_1-1} o_2(\|\Delta z(t)\|) + o_1(\|\Delta x(t_1)\|). \end{aligned} \quad (11)$$

Предположим, что $\psi(t)$ является решением следующей системы:

$$\begin{aligned} \psi(t-1) &= - \sum_{s=t}^{t_1-1} (s + \alpha - \sigma(t))^{(\alpha-1)} \frac{\partial H'(s, x(s), y(s), u(s), \psi(s))}{\partial x(s)}, \quad t = t_1 - h, \dots, t_1 - 1, \\ \psi(t-1) &= -\phi_x'(x(t_1)) \frac{1}{\Gamma(\alpha)} \sum_{s=t_0}^{t_1-1} (t_1 + \alpha - \sigma(s))^{(\alpha-1)} + \\ &+ \sum_{s=t}^{t_1-1} (s + \alpha - \sigma(t))^{(\alpha-1)} H_x[s] + \sum_{s=t}^{t_1-1} (s + \alpha - h - \sigma(t))^{(\alpha-1)} H_y[s], \quad t = t_0, \dots, t_1 - h - 1, \end{aligned} \quad (12)$$

$$\psi(t_1-1) = -\phi_x'(x(t_1)). \quad (13)$$

Задачу (12)–(13) называем сопряженной системой для задачи управления (1)–(4).

Учитывая задачу (12)–(13), из формулы приращения (11) получим

$$\Delta S(u) = - \sum_{t=t_0}^{t_1-1} \Delta_{\bar{u}} H[t] + \eta(u, \Delta u), \quad (14)$$

где по определению

$$\eta(u, \Delta u) = o_1(\|\Delta x(t_1)\|) - \sum_{t=t_0}^{t_1-1} o_2(\|\Delta z(t)\|) - \sum_{t=t_0}^{t_1-1} \Delta_{\bar{u}} H_x[t] \Delta x(t) - \sum_{t=t_0}^{t_1-1} \Delta_{\bar{u}} H_y[t] \Delta y(t) - \sum_{t=t_0}^{t_1-1} \Delta_{\bar{u}} H_u[t] \Delta u(t). \quad (15)$$

Нам понадобится оценка для $\|\Delta x(t)\|$ $\left(\|\Delta x(t)\| = \sum_{i=1}^n |\Delta x_i(t)| \right)$.

Применяя $\Delta^{-\alpha}$ обеим сторонам уравнения (1), имеем

$$\Delta^{-\alpha} \Delta^{\alpha} x(t+1) = \Delta^{-\alpha} f(t, x(t), y(t), u(t)).$$

Займемся преобразованием этой формулы. С этой целью рассмотрим выражение

$$\Delta^{-\alpha} \Delta^{\alpha} x(t+1).$$

Учитывая свойства операторов дробной суммы и дробной разности, проведем следующие преобразования:

$$\begin{aligned} \Delta^{-\alpha} \Delta^{\alpha} x(t+1) &= \Delta^{-\alpha} (\Delta^{1-\mu} x(t+1)) = \Delta^{-\alpha} (\Delta^{-\mu} (\Delta x(t+1))) = \\ &= \Delta^{-1} (\Delta x(t+1)) = \sum_{j=t_0}^t (x(t+1) - x(j)) = x(t+1) - x(t_0). \end{aligned}$$

Правая сторона:

$$\begin{aligned} \Delta^{-\alpha} f(t, x(t), y(t), u(t)) &= \frac{1}{\Gamma(\alpha)} \sum_{j=t_0}^t (t - \sigma(j))^{(\alpha-1)} f(j, x(j), y(j), u(j)) = \\ &= \sum_{j=t_0}^t \binom{t-j+\alpha-1}{t-j} f(j, x(j), y(j), u(j)) = \sum_{j=t_0}^t A_{\alpha}(t, j) f(j, x(j), y(j), u(j)). \end{aligned}$$

Здесь

$$A_{\alpha}(t, j) = \binom{t-j+\alpha-1}{t-j}.$$

Таким образом, доказали, что

$$x(t+1) = x(t_0) + \sum_{j=t_0}^t A_{\alpha}(t, j) f(j, x(j), y(j), u(j)).$$

Отсюда, переходя к норме и используя условие Липшица, получаем, что

$$\begin{aligned} \|\Delta x(t)\| &= \sum_{j=t_0}^{t-1} A_{\alpha}(t-1, j) \|f(j, x(j) + \Delta x(j), y(j) + \Delta y(j), u(j) + \Delta u(j)) - \\ &- f(j, x(j), y(j), u(j))\| = \sum_{j=t_0}^{t-1} A_{\alpha}(t-1, j) \|f(j, x(j) + \Delta x(j), y(j) + \Delta y(j), u(j) + \Delta u(j)) - \\ &- f(j, x(j), y(j), u(j)) + \Delta_{\bar{u}} f(j, x(j), y(j), u(j))\| \leq \\ &\leq L_1 \sum_{j=t_0}^{t-1} A_{\alpha}(t-1, j) \left(\|\Delta x(j)\| + \|\Delta y(j)\| + \sum_{j=t_0}^t A_{\alpha}(t-1, j) \|\Delta_{\bar{u}} f[j]\| \right) \leq \\ &\leq L_1 + \sum_{j=t_0}^{t-1} A_{\alpha}(t-1, j) \|\Delta_{\bar{u}} f[j]\|, \quad t \in \{t_0, t_0+1, \dots, t_1-1\}. \end{aligned}$$

Применяя к последнему неравенству лемму Гронуолла–Беллмана дробного порядка (см., напр.: [15]), получим справедливость оценки

$$\|\Delta x(t)\| \leq L_2 \prod_{j=t_0}^{t-1} (1 + A_{\alpha}(t, j) \|\Delta_{\bar{u}} f[j]\|), \quad t = t_0, t_0+1, \dots, t_1. \quad (16)$$

3. Необходимые условия оптимальности

Теперь предположим, что вдоль допустимого процесса $(u(t), x(t))$ множество допустимых скоростей системы (1), т.е. множество

$$f(t, x(t), y(t), U) = \{\beta : \beta = f(t, x(t), y(t), v), v \in U\}, \quad (17)$$

выпукло. Тогда специальное приращение управления $u(t)$ можно определить по формуле

$$\Delta u_\varepsilon(t) = v(t, \varepsilon) - u(t), \quad t \in T, \quad (18)$$

здесь $\varepsilon \in [0, 1]$ – произвольное число, а $v(t, \varepsilon) \in U$ – такой вектор, что

$$f(t, x(t), y(t), v(t, \varepsilon)) = \varepsilon f(t, x(t), y(t), v(t)) + (1 - \varepsilon) f(t, x(t), y(t), u(t)). \quad (19)$$

Через $\Delta x_\varepsilon(t)$ обозначим специальное приращение траектории $x(t)$, отвечающее приращению $\Delta u_\varepsilon(t)$ управления $u(t)$.

С учетом оценки (16) получаем, что

$$\|\Delta x_\varepsilon(t)\| \leq L_3 \varepsilon, \quad t \in T \cup t_1, \quad L_3 = \text{const} > 0. \quad (20)$$

В силу этой оценки получим

$$\eta(u; \Delta u_\varepsilon(t)) = o(\varepsilon).$$

Тогда из формулы приращения (14) следует, что

$$\Delta S_\varepsilon(u) = S(u + \Delta u_\varepsilon) - S(u) = - \sum_{t=t_0}^{t_1-1} \Delta_{v(t, \varepsilon)} H[t] + o(\varepsilon) = -\varepsilon \sum_{t=t_0}^{t_1-1} \Delta_{v(t)} H[t] + o(\varepsilon). \quad (21)$$

Если предположить, что в задаче (1)–(4) управление $u(t)$ оптимальное, то из разложения (21) получим следующее неравенство:

$$-\varepsilon \sum_{t=t_0}^{t_1-1} \Delta_{v(t)} H[t] + o(\varepsilon) \geq 0. \quad (22)$$

Из этого неравенства в силу произвольности ε следует неравенство

$$\sum_{t=t_0}^{t_1-1} \Delta_{v(t)} H[t] \leq 0. \quad (23)$$

Таким образом, доказана следующая

Теорема 1 (дискретный принцип максимума для системы с запаздыванием). Если в задаче (1)–(2) вдоль допустимого процесса $(u(t), x(t))$ множество допустимых скоростей системы (19) выпукло, то для оптимальности управления $u(t)$ в рассматриваемой задаче необходимо, чтобы неравенство (23) выполнялось для любого $v(t) \in U$, $t \in T$.

Непосредственным следствием этого утверждения является

Теорема 2 (поточечный дискретный принцип максимума [19]). При выполнении условия теоремы 1 для оптимальности допустимого управления $u(t)$ в рассматриваемой задаче необходимо, чтобы условие

$$\max_{v \in U} H(\theta, x(\theta), y(\theta), v, \psi(\theta)) = H(\theta, x(\theta), y(\theta), u(\theta), \psi(\theta)) \quad (24)$$

выполнялось для всех $\theta \in T$.

Условие (24) является поточечным дискретным условием максимума для рассматриваемой задачи. Для доказательства условия (24) достаточно в неравенстве (23) $v(t)$ определить по формуле

$$v(t) = \begin{cases} v, & t = \theta \in T, v \in U, \\ u(t), & t \neq \theta \in T. \end{cases}$$

Также легко доказывается, что если вдоль процесса $(u(t), x(t))$ выполняется соотношение (24), то вдоль этого же процесса выполняется также неравенство (23). Все эти рассуждения показывают, что условия оптимальности (23) и (24) равносильны.

Теперь предположим, что вектор-функция $f(t, x, y, u)$ непрерывна по совокупности переменных вместе с частными производными по (x, u) , а множество U выпукло.

Далее, считая $(u(t), x(t))$ оптимальным процессом, специальное приращение оптимального управления определим следующим образом:

$$\Delta u(t; \varepsilon) = \varepsilon(v(t) - u(t)). \quad (25)$$

Здесь $\varepsilon \in [0, 1]$ – произвольное число, а $v(t) \in U, t \in T$ – произвольный вектор.

В силу выпуклости множества U специальное приращение $\Delta u(t; \varepsilon)$ оптимального управления $u(t)$, определяемое формулой (25), будет допустимым. В самом деле:

$$\bar{u}_\varepsilon(t) = u(t) + \Delta u(t; \varepsilon) = u(t) + \varepsilon[v(t) - u(t)] = \varepsilon v(t) + (1 - \varepsilon)u(t) \in U.$$

Через $\Delta x(t; \varepsilon)$ обозначим специальное приращение оптимальной траектории $x(t)$, отвечающей специальному приращению управления $u(t)$, определяемое формулой (25).

Используя работу [20], по аналогии с ней доказывается оценка

$$\|\Delta x(t+1)\| \leq L_1 \prod_{j=t_0}^{t-1} (1 + A_\alpha(t, j) \|\Delta u(j)\|),$$

здесь $L_1 = \text{const} > 0, t \in T \cup t_1$.

Из этого неравенства получаем, что

$$\|\Delta x(t; \varepsilon)\| \leq L_1 \prod_{j=t_0}^{t-1} (1 + A_\alpha(t, j) \|\Delta u(j; \varepsilon)\|) \leq \varepsilon L_2 (\|v(j) - u(j)\|),$$

отсюда следует

$$\|\Delta x(t; \varepsilon)\| \sim \varepsilon. \quad (26)$$

Учитывая формулу (25) и оценку (26), из формулы приращения (14) получим справедливость следующей теоремы.

Теорема 3. Пусть множество U выпукло, а $f(t, x, y, u)$ непрерывно по совокупности переменных вместе с частными производными по (x, u) . Тогда для оптимальности допустимого управления $u(t)$ необходимо, чтобы соотношение

$$\sum_{t=t_0}^{t_1-1} H_u'[t](u(t) - v(t)) \leq 0 \quad (27)$$

выполнялось для любого $u(t) \in U, t \in T$.

Соотношение (27) является аналогом линеаризованного принципа максимума.

Теперь предположим, что U – заданное непустое, ограниченное и открытое множество, а векторная функция $f(t, x, y, u)$ непрерывна относительно множества переменных вместе с частными производными по (x, u) .

Поскольку, по предположению, множество U открыто, особое приращение допустимого управления $u(t)$ можно определить по формуле

$$u_\varepsilon(t) = \varepsilon \delta u(t). \quad (28)$$

Здесь ε достаточно малое по абсолютной величине число, а $\delta u(t)$ – произвольная r -мерная векторная функция со значениями из R^r .

Через $\Delta x_\varepsilon(t)$ обозначим особое приращение допустимой траектории $x(t)$, соответствующее особому приращению управления $u(t)$, определяемому формулой (28).

Учитывая оценку (16), получаем, что

$$\|\Delta x_\varepsilon(t)\| \leq L_2 \varepsilon, t \in T \cup t_1, L_2 = \text{const} > 0. \quad (29)$$

Кроме того, для $\Delta x_\varepsilon(t)$ справедливо следующее разложение:

$$\Delta x_\varepsilon(t) = \varepsilon \delta x(t) + o(\varepsilon; t). \quad (30)$$

Отметим, что $\delta u(t)$ называется вариацией управления $\delta u(t)$, а $\delta x(t)$ называется вариацией траектории $x(t)$.

Используя формулы (28) и (30), а также оценку (29), из формулы приращения (25) получаем справедливость следующего утверждения.

Теорема 4. Пусть множество U открыто, а $f(t, x, u)$ непрерывна по множеству переменных вместе с частными производными по (x, u) . Тогда для оптимальности допустимого управления $u(t)$ необходимо, чтобы соотношение

$$H_u[t] = 0 \quad (31)$$

выполнялось для любого $t \in T$.

Соотношение (31) называется аналогом уравнения Эйлера для рассматриваемой задачи.

4. Пример проверки необходимого условия оптимальности

Пусть требуется минимизировать функционал

$$S(u) = -x(3) \rightarrow \min$$

при ограничениях

$$\Delta^\alpha x(t+1) = 3x(t) + x(t-1)u(t), \quad T = \{0, 1, 2\},$$

$$x(-1) = 0, \quad x(0) = 1, \quad |u(t)| \leq 2, \quad t \in T \setminus 2.$$

Используя определения 2, 3 можно доказать, что

$$S(u) = -(2 - \alpha - \frac{\alpha}{\Gamma(\alpha)}) - (\alpha + 1)u(1) - u(2).$$

Отсюда получаем, что $u^*(t) = (u^*(1), u^*(2)) = (2, 2)$ может быть оптимальным управлением.

Составим функцию Гамильтона–Понтрягина следующим образом:

$$H(t, x(t), x(t-1), u(t), \psi(t)) = \psi(t)(3x(t) + x(t-1)u(t)).$$

Сопряженная система имеет вид:

$$\psi(t-1) = -3 \left[\sum_{s=t}^{t_1-1} (s + \alpha - \sigma(t))^{(\alpha-1)} \psi(s) \right], \quad t = 2, 3,$$

$$\begin{aligned} \psi(t-1) = & -\varphi_x'(x(t_1)) \frac{1}{\Gamma(\alpha)} \sum_{s=t_0}^{t_1-1} (t_1 + \alpha - \sigma(s))^{(\alpha-1)} + \\ & + 3 \sum_{s=t}^{t_1-1} (s + \alpha - \sigma(t))^{(\alpha-1)} \psi(s) + \sum_{s=t}^{t_1-1} (s + \alpha - h - \sigma(t))^{(\alpha-1)} \psi(s)u(s), \quad t = t_0, \dots, t_1 - h - 1, \quad t = 0, 1. \end{aligned}$$

Отсюда получаем, что

$$\psi(2) = 1, \quad \psi(1) = 3\Gamma(\alpha).$$

По теореме 1 проверим, что управление $u^*(t)$ является оптимальным.

Получаем, что условие

$$H(t, x(t), x(t-1), v(t), \psi(t)) - H(t, x(t), x(t-1), u(t), \psi(t)) = \psi(t)x(t-1)(v(t) - 2) \leq 0$$

выполняется для любого $v(t) \in [-2, 2]$, $t = 0, 1, 2$.

Заключение

В работе ставится и исследуется дискретная терминальная задача оптимального управления, описываемая нелинейным разностным уравнением дробного порядка.

В рассмотрении введены функции типа Понтрягина, а также аналог сопряженной задачи, при сделанных предположениях построена общая формула приращения функционала качества.

В случае выпуклости множества допустимых скоростей рассматриваемой системы уравнений, используя построенную формулу приращения, доказан дискретный аналог принципа максимума Л.С. Понтрягина. В случае выпуклости области управления и непрерывной дифференцируемости правой части уравнения по управлению доказан аналог линеаризованного условия максимума, а в случае открытости области управления установлено необходимое условие оптимальности в форме аналога классического уравнения Эйлера.

Список источников

1. Agrawal O.P. Formulation of Euler-Lagrange equations for fractional variational problems // J. Math. Anal. Appl. 2002. V. 272 (1). P. 368–379.
2. Agrawal O.P. A general finite element formulation for fractional variational problems // J. Math. Anal. Appl. 2008. V. 337 (1). P. 1–12.
3. El-Nabulsi R.A., Torres D.F.M. Necessary optimality conditions for fractional action-like integrals of variational calculus with Riemann-Liouville derivatives of order (α, β) // Math. Methods Appl. Sci. 2007. V. 30 (15). P. 1931–1939. URL: <https://arxiv.org/abs/math-ph/0702099>
4. Понтрягин Л.С., Болтянский В.Г., Гамкрелидзе Р.В., Мищенко Е.М. Математическая теория оптимальных процессов. М. : Наука, 1984. 364 с.
5. Gabasov R., Kirillova F.M. High-order necessary conditions for optimality // SIAM J. Control. 1972. V. 10. P. 127–168.
6. Baleanu D., Jarad F. Difference discrete variational principles // Mathematical Analysis and Applications. Melville, NY : Amer. Inst. Phys., 2006. P. 20–29.
7. Jarad F., Baleanu D. Discrete variational principles for Lagrangians linear in velocities // Rep. Math. Phys. 2007. V. 59 (1). P. 33–43.
8. Baleanu D., Deftleri O., Agrawal O.P. A central difference numerical scheme for fractional optimal control problems // J. Vib. Control. 2009. V. 15 (4). P. 583–597.
9. Алиева С.Т. Принцип максимума Понтрягина для нелинейных разностных уравнений дробного порядка // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2021. № 54. С. 4–11. doi: 10.17223/19988605/54/1
10. Алиева С.Т. Необходимые условия оптимальности первого и второго порядков в одной задаче управления, описываемой нелинейными разностными уравнениями дробного порядка // Автоматика и телемеханика. 2023. № 2. С. 54–65. doi: 10.31857/S0005231023020034
11. Bhalekar S., Daftardar-Gejji V., Baleanu D., Magin R. Fractional Bloch equation with delay // Computers and Mathematics with Applications. 2011. V. 61 (5). P. 1355–1365.
12. Magin R.L. Fractional calculus models of complex dynamics in biological tissues // Computers and Mathematics with Applications. 2010. V. 59 (5). P. 1586–1593.
13. Bahaa G.M. Fractional optimal control problem for differential system with delay argument // Advanc. Differen. Equat. 2017. Art. 69. P. 1–19. doi: 10.1186/s13662-017-1121-6
14. Si-Ammour A., Djennoune S., Bettayeb M. A sliding mode control for linear fractional systems with input and state delays // Communications in Nonlinear Science and Numerical Simulation. 2009. V. 14 (5). P. 2310–2318.
15. Miller K., Ross B. An introduction to the fractional calculus and fractional differential equations. New York :Wiley, 1993. 366 p.
16. Samko S.G., Kilbas A.A., Marichev O.I. Fractional Integrals and Derivatives – Theory and Applications. Linghorne, PA : Gordon and Breach, 1993. 976 p.
17. Podlubny I. Fractional Differential Equations. San Diego, CA : Academic Press, 1999. 340 p.
18. Kilbas A., Srivastava M.H., Trujillo J.J. Theory and Application of Fractional Differential Equations. Elsevier Science, 2006. 541 p. (North Holland Mathematics Studies; v. 204).
19. Мансимов К.Б. Дискретные системы. Баку : Изд-во Бакинского гос. ун-та, 2013. 151 с.
20. Алиева С.Т., Мансимов К.Б. Аналог линеаризованного принципа максимума для задачи оптимального управления нелинейными разностными уравнениями дробного порядка // Вестник Пермского университета. Математика, механика и информатика. 2021. Вып 1 (52). С. 9–15. doi: 10.17072/1993-0550-2021-1-9-15

References

1. Agrawal, O.P. (2002) Formulation of Euler-Lagrange equations for fractional variational problems. *Journal of Mathematical Analysis and Applications*. 272(1). pp. 368–379.
2. Agrawal, O.P. (2008) A general finite element formulation for fractional variational problems. *Journal of Mathematical Analysis and Applications*. 337(1). pp. 1–12.
3. El-Nabulsi, R.A. & Torres, D.F.M. (2007) Necessary optimality conditions for fractional action-like integrals of variational calculus with Riemann-Liouville derivatives of order (α, β) . *Mathematical Methods in the Applied Sciences*. 30(15). pp. 1931–1939. [Online] Available from: <https://arxiv.org/abs/math-ph/0702099>
4. Pontryagin, L.S., Boltyanskiy, V.G., Gamkrelidze, R.V. & Mishchenko, E.M. (1984) *Matematicheskaya teoriya optimal'nykh protsessov* [Mathematical Theory of Optimal Processes]. Moscow: Nauka.

5. Gabasov, R. & Kirillova, F.M. (1972) High-order necessary conditions for optimality. *SIAM Journal on Control*. 10. pp. 127–168.
6. Baleanu, D. & Jarad, F. (2006) Difference discrete variational principles. In: *Mathematical Analysis and Applications*. Melville, NY: Amer. Inst. Phys. pp. 20–29.
7. Jarad, F. & Baleanu, D. (2007) Discrete variational principles for Lagrangians linear in velocities. *Reports on Mathematical Physics*. 59(1). pp. 33–43.
8. Baleanu, D., Deftterli, O. & Agrawal, O.P. (2009) A central difference numerical scheme for fractional optimal control problems. *Journal of Vibration and Control*. 15(4). pp. 583–597.
9. Alieva, S.T. (2021) Pontryagin's maximum principle for nonlinear difference equations of fractional order. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 54. pp. 4–11. doi: 10.17223/19988605/54/1
10. Alieva, S.T. (2023) Neobkhodimye usloviya optimal'nosti pervogo i vtorogo poryadkov v odnoy zadache upravleniya, opisyyvayemy nelineynymi raznostnymi uravneniyami drobnogo poryadka [Necessary conditions for first- and second-order optimality in one control problem described by nonlinear difference equations of fractional order]. *Avtomatika i telemekhanika*. 2. pp. 54–65. doi: 10.31857/S0005231023020034
11. Bhalekar, S., Daftardar-Gejji, V., Baleanu, D. & Magin, R. (2011) Fractional Bloch equation with delay. *Computers and Mathematics with Applications*. 61(5). pp. 1355–1365.
12. Magin, R.L. (2010) Fractional calculus models of complex dynamics in biological tissues. *Computers and Mathematics with Applications*. 59(5). pp. 1586–1593.
13. Bahaa, G.M. (2017) Fractional optimal control problem for differential system with delay argument. *Advances in Difference Equations*. Art. 69. pp. 1–19. doi: 10.1186/s13662-017-1121-6
14. Si-Ammour, A., Djennoune, S. & Bettayeb, M. (2009) A sliding mode control for linear fractional systems with input and state delays. *Communications in Nonlinear Science and Numerical Simulation*. 14(5). pp. 2310–2318.
15. Miller, K. & Ross, B. (1993) *An Introduction to the Fractional Calculus and Fractional Differential Equations*. New York: Wiley.
16. Samko, S.G., Kilbas, A.A. & Marichev, O.I. (1993) *Fractional Integrals and Derivatives – Theory and Applications*. Linghorne, PA: Gordon and Breach.
17. Podlubny, I. (1999) *Fractional Differential Equations*. San Diego, CA: Academic Press.
18. Kilbas, A., Srivastava, M.H. & Trujillo, J.J. (2006) *Theory and Application of Fractional Differential Equations*. (North Holland Mathematics Studies, vol. 204). Elsevier Science.
19. Mansimov, K.B. (2013) *Diskretnye sistemy* [Discrete Systems]. Baku: Baku University.
20. Alieva, S.T. & Mansimov, K.B. (2021) Analog linearizovannogo printsipa maksimuma dlya zadachi optimal'nogo upravleniya nelineynymi raznostnymi uravneniyami drobnogo poryadka [An Analog of the Linearized Maximum Principle for an Optimal Control Problem with Nonlinear Fractional Order Difference Equations]. *Vestnik Permskogo universiteta. Matematika, mekhanika i informatika*. 1(52). pp. 9–15. doi: 10.17072/1993-0550-2021-1-9-15

Информация об авторе:

Алиева Саадат Тофик кызы – доцент, кандидат физико-математических наук, Бакинский государственный университет; Институт систем управления Министерства науки и образования Азербайджана; Азербайджанский государственный экономический университет; Университет Азербайджан (Баку, Азербайджан). E-mail: saadat.t.aliyeva@au.edu.az

Автор заявляет об отсутствии конфликта интересов.

Information about the author:

Aliyeva Saadat T. (Associate Professor, Candidate of Physical and Mathematical Sciences, Baku State University; Institute of Control Systems of the Ministry of Science and Education of Azerbaijan; Azerbaijan State University of Economics; Azerbaijan University, Baku, Azerbaijan). E-mail: saadat.t.aliyeva@au.edu.az

The author declares no conflicts of interests.

Поступила в редакцию 23.04.2025; принята к публикации 02.12.2025

Received 23.04.2025; accepted for publication 02.12.2025

Научная статья
УДК 681.5:517
doi: 10.17223/19988605/73/2

О редукции модальных тестов на управляемость и наблюдаемость ММО-системы

Николай Евгеньевич Zubov¹, Владимир Николаевич Рябченко²

^{1, 2} Московский государственный технический университет им. Н.Э. Баумана, Москва, Россия

¹ nik.zubov@gmail.com

² ryabchenko.vn@mail.ru

Аннотация. Показано, что анализ управляемости и наблюдаемости линейной стационарной системы с многими входами и многими выходами путем редукции моделей системы сводится к анализу управляемости и наблюдаемости систем с существенно меньшей размерностью пространства состояний. Предельным случаем редукции являются скалярные системы.

Ключевые слова: линейные стационарные системы; управляемость; наблюдаемость; редукция систем.

Для цитирования: Zubov N.E., Ryabchenko V.N. О редукции модальных тестов на управляемость и наблюдаемость ММО-системы // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2025. № 73. С. 15–22. doi: 10.17223/19988605/73/2

Original article
doi: 10.17223/19988605/73/2

On reduction of modal tests on controllability and observability of MIMO system

Nikolay E. Zubov¹ Vladimir N. Ryabchenko²

^{1, 2} Bauman Moscow State Technical University, Moscow, Russian Federation

¹ nik.zubov@gmail.com

² ryabchenko.vn@mail.ru

Abstract. It is shown that the analysis of controllability and observability of a linear stationary system with many inputs and many outputs by reduction of the system models is reduced to the analysis of controllability and observability of systems with substantially lower dimensionality of the state space. The limiting case of reduction is scalar systems.

Keywords: linear stationary systems; controllability; observability; reduction of systems.

For citation: Zubov, N.E., Ryabchenko, V.N. (2025) On reduction of modal tests on controllability and observability of MIMO system. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naja tehnika i informatika – Tomsk State University Journal of Control and Computer Science*. 73. pp. 15–22. doi: 10.17223/19988605/73/2

Введение

Рассмотрим линейную стационарную систему с многими входами и многими выходами (ММО-систему)

$$\begin{aligned}\dot{x}(t) &= Ax(t) + Bu(t), \\ y(t) &= Cx(t),\end{aligned}\tag{1}$$

где $x(t) \in \mathbb{R}^n$ – вектор состояния; $u(t) \in \mathbb{R}^r$ – вектор управления; $y(t) \in \mathbb{R}^m$ – вектор выхода; $\text{rank } B = r$, $\text{rank } C = m$.

Известными (модальными) критериями управляемости и наблюдаемости (1) являются тесты Попова–Белевича–Хотиса (РВН-tests) [1. С. 265]. Согласно этим тестам для управляемости и наблюдаемости ММО-системы необходимо и достаточно, чтобы

$$\forall \lambda \in \mathbb{C}: \text{rank} \begin{bmatrix} A - \lambda I_n & B \end{bmatrix} = n, \quad (2)$$

$$\forall \lambda \in \mathbb{C}: \text{rank} \begin{bmatrix} A - \lambda I_n \\ C \end{bmatrix} = n. \quad (3)$$

Заметим, что условие $\forall \lambda \in \mathbb{C}$ может быть заменено на $\forall \lambda \in \text{eig}(A)$, где множество собственных значений матрицы A

$$\text{eig}(A) = \{ \lambda_i \mid \det(A - \lambda_i I_n) = 0 \}.$$

В практических задачах анализа управляемости и наблюдаемости больших линейных многомерных систем (Large Scale Systems), которые характерны для электроэнергетики [2–4], зачастую встречаются ситуации, когда размерность векторов управления и выхода сопоставима с размерностью пространства состояний, т.е. $r \approx n$, $m \approx n$. В этом случае стандартные тесты на управляемость системы (1) представляют собой высокоразмерные плохо обусловленные задачи.

Сходные трудности возникают в ММО-системах, где некоторые элементы представлены неопределенными параметрами (так называемые параметризованные системы).

Цель данной работы – распространение предложенного в работе [2] подхода для энергетических систем в общем случае на ММО-системы, который основан на построении цепочки преобразований (редукций), что позволяло бы выносить суждение об управляемости и наблюдаемости системы (1) на основе изучения управляемости и наблюдаемости систем существенно меньшей размерности состояний.

1. Редукция тестов при анализе управляемости

Известно, что любую числовую матрицу $B \in \mathbb{R}^{n \times r}$ ранга r можно привести к виду [5, 6]

$$\begin{bmatrix} I_r \\ \dots \\ 0_{(n-r) \times r} \end{bmatrix}$$

путем невырожденного преобразования T вида

$$T = \begin{bmatrix} B^+ \\ \dots \\ L_0 \end{bmatrix}, \quad (4)$$

где B^+ – псевдообратная по Муру–Пенроузу матрица [7], L_0 – максимальное решение однородного уравнения [8, 9]

$$L_0 B = 0_{(n-r) \times r}.$$

Используя (4), осуществим преобразование матрицы $\begin{bmatrix} A - \lambda I_n & B \end{bmatrix}$ по типу

$$T \begin{bmatrix} A - \lambda I_n & B \end{bmatrix} = \begin{bmatrix} B^+ \\ \dots \\ L_0 \end{bmatrix} \begin{bmatrix} A - \lambda I_n & B \end{bmatrix}. \quad (5)$$

Раскрывая правую часть (5), получим

$$\begin{bmatrix} B^+ \\ \dots \\ L_0 \end{bmatrix} \begin{bmatrix} A - \lambda I_n & B \end{bmatrix} = \begin{bmatrix} B^+ (A - \lambda I_n) & I_r \\ \dots & \dots \\ L_0 (A - \lambda I_n) & 0_{(n-r) \times r} \end{bmatrix},$$

при этом в силу невырожденности матрицы (4)

$$\text{rank} \begin{bmatrix} A - \lambda I_n & B \end{bmatrix} = \text{rank} \begin{bmatrix} B^+ (A - \lambda I_n) & I_r \\ \dots & \dots \\ L_0 (A - \lambda I_n) & 0_{(n-r) \times r} \end{bmatrix}. \quad (6)$$

Как следует из структуры (6), подматрица

$$\begin{bmatrix} B^+ (A - \lambda I_n) & I_r \end{bmatrix}$$

при любых λ имеет ранг r . Поэтому для выполнения условия (2) необходимо и достаточно, чтобы ранг подматрицы $L_0(A - \lambda I_n)$ удовлетворял требованию

$$\forall \lambda \in \mathbb{C}: \text{rank } L(A - \lambda I_n) = n - r.$$

Введем в рассмотрение невырожденную матрицу

$$T_1 = \begin{bmatrix} L^+ & B \end{bmatrix},$$

как видно, удовлетворяющую уравнению

$$L_0 T_1 = L \begin{bmatrix} L^+ & B \end{bmatrix} = \begin{bmatrix} I_{n-r} & 0_{(n-r) \times r} \end{bmatrix}.$$

Осуществим далее невырожденное преобразование подматрицы $L_0(A - \lambda I_n)$ по типу

$$L_0(A - \lambda I_n) T_1 = L_0(A - \lambda I_n) \begin{bmatrix} L^+ & B \end{bmatrix}. \quad (7)$$

Раскроем правую часть (7):

$$L(A - \lambda I_n) \begin{bmatrix} L^+ & B \end{bmatrix} = \begin{bmatrix} LAL^+ - \lambda I_{n-r} & LAB \end{bmatrix}. \quad (8)$$

При этом, как и в предыдущем случае (6),

$$\text{rank } L_0(A - \lambda I_n) = \text{rank} \begin{bmatrix} L_0 AL_0^+ - \lambda I_{n-r} & L_0 AB \end{bmatrix}. \quad (9)$$

Сравнивая правые части из (2) и (9), приходим к справедливости следующего утверждения.

Лемма 1. МИМО-система (1) управляема, если и только если выполняются эквивалентные условия:
A: управляема система

$$\dot{x}_1(t) = L_0 AL_0^+ x_1(t) + L_0 AB u_1(t), \quad (10)$$

где $x_1(t) \in \mathbb{R}^{n-r}$ – вектор состояния; $u_1(t) \in \mathbb{R}^r$ – вектор управления;

B:

$$\forall \lambda \in \text{eig}(LAL^+): \text{rank} \begin{bmatrix} L_0 AL_0^+ - \lambda I_{n-r} & L_0 AB \end{bmatrix} = n - r, \quad (11)$$

где $\text{eig}(L_0 AL_0^+)$ – множество собственных значений матрицы $L_0 AL_0^+$.

Как видно, в результате проведенных преобразований произошла редукция размерности пространства состояний с n до величины $n - r$. Более того, если

$$\text{rank } L_0 AB = n - r \leq r,$$

то (10) является управляемой независимо от вида и свойств матрицы LAL^+ .

Отметим также, что в общем случае очевидным является соотношение $\text{eig}(LAL^+) \not\subset \text{eig}(A)$.

Введем новые обозначения:

$$A_1 = L_0 AL_0^+, \quad B_1 = L_0 AB, \quad n - r = n_1, \quad \text{rank } L_0 AB = r_1. \quad (12)$$

С учетом (12) условие управляемости системы (11) примет вид:

$$\forall \lambda \in \Lambda(A_1): \text{rank} \begin{bmatrix} A_1 - \lambda I_{n_1} & B_1 \end{bmatrix} = n_1.$$

Далее нам понадобятся максимальные решения следующих матричных уравнений:

$$\begin{aligned} L_1 B_1 &= 0_{(n_1-r_1) \times r_1}, \\ B_1 R_1 &= 0_{n_1 \times (r-r_1)}, \\ J_1^L B_1 J_1^R &= I_{n_1}, \end{aligned} \quad (13)$$

относительно матриц L_1, J_1^L, R_1, J_1^R . Отметим, что уравнения (13) разрешимы для любой ненулевой матрицы над \mathbb{R} , при этом существуют (неединственные) невырожденные блочные матрицы [5]

$$T_1^L = \begin{bmatrix} J_1^L \\ L_1 \end{bmatrix}, \quad T_1^R = \begin{bmatrix} J_1^R & R_1 \end{bmatrix}. \quad (14)$$

Как следует из (13), преобразование вида

$$T_1^L B_1 T_1^R$$

с учетом (14) приводит к тождеству

$$T_1^L B_1 T_1^R = \begin{bmatrix} J_1^L \\ L_1 \end{bmatrix} B_1 \begin{bmatrix} J_1^R & R_1 \end{bmatrix} = \begin{bmatrix} I_{r_1} & 0_{r_1 \times (r-r_1)} \\ 0_{(n_1-r_1) \times r_1} & 0_{(n_1-r_1) \times (r-r_1)} \end{bmatrix}. \quad (15)$$

Применяя данное преобразование к матрице $\begin{bmatrix} A_1 - \lambda I_{n_1} & B_1 \end{bmatrix}$, с учетом (15) получим

$$\begin{bmatrix} J_1^L \\ L_1 \end{bmatrix} \begin{bmatrix} A_1 - \lambda I_{n_1} & B_1 \end{bmatrix} \begin{bmatrix} J_1^R & R_1 \end{bmatrix} = \begin{bmatrix} J_1^L (A_1 - \lambda I_{n_1}) J_1^R & J_1^L (A_1 - \lambda I_{n_1}) R_1 & I_{r_1} & 0_{r_1 \times (r-r_1)} \\ L_1 (A_1 - \lambda I_{n_1}) J_1^R & L_1 (A_1 - \lambda I_{n_1}) R_1 & 0_{(n_1-r_1) \times r_1} & 0_{(n_1-r_1) \times (r-r_1)} \end{bmatrix}. \quad (16)$$

Анализ (16) показывает, что в данном случае справедлива цепочка ранговых условий

$$\begin{aligned} \text{rank} \begin{bmatrix} J_1^L (A_1 - \lambda I_{n_1}) J_1^R & J_1^L (A_1 - \lambda I_{n_1}) R_1 & I_{r_1} & 0_{r_1 \times (r-r_1)} \\ L_1 (A_1 - \lambda I_{n_1}) J_1^R & L_1 (A_1 - \lambda I_{n_1}) R_1 & 0_{(n_1-r_1) \times r_1} & 0_{(n_1-r_1) \times (r-r_1)} \end{bmatrix} = \\ = \text{rank} \begin{bmatrix} J_1^L (A_1 - \lambda I_{n_1}) J_1^R & J_1^L (A_1 - \lambda I_{n_1}) R_1 & I_{r_1} & 0_{r_1 \times (r-r_1)} \end{bmatrix} + \\ + \text{rank} \begin{bmatrix} L_1 (A_1 - \lambda I_{n_1}) J_1^R & L_1 (A_1 - \lambda I_{n_1}) R_1 & 0_{(n_1-r_1) \times r_1} & 0_{(n_1-r_1) \times (r-r_1)} \end{bmatrix} = \\ = r_1 + \text{rank} \begin{bmatrix} L_1 (A_1 - \lambda I_{n_1}) J_1^R & L_1 (A_1 - \lambda I_{n_1}) R_1 \end{bmatrix} = \\ = r_1 + \text{rank} L_1 (A_1 - \lambda I_{n_1}) \begin{bmatrix} J_1^R & R_1 \end{bmatrix} = \\ = r_1 + \text{rank} L_1 (A_1 - \lambda I_{n_1}). \end{aligned} \quad (17)$$

Как видно, структура матрицы $L_1 (A_1 - \lambda I_{n_1})$ из (17) в точности соответствует структуре матрицы $L(A - \lambda I_n)$ из (8).

Это позволяет нам сформулировать еще одну лемму.

Лемма 2. МИМО-системы (1), (10) управляемы, если и только если выполняются эквивалентные условия:

A: управляема система

$$\dot{x}_2(t) = L_1 A_1 L_1^+ x_2(t) + L_1 A_1 R_{L_1} u_2(t),$$

где $x_2(t) \in \mathbb{R}^{n_1-r_1}$ – вектор состояния; $u_2(t) \in \mathbb{R}^{r_1}$ – вектор управления; R_{L_1} – максимальное решение уравнения

$$L_1 R_{L_1} = 0_{(n_1-r_1) \times r_1};$$

B:

$$\forall \lambda \in \text{eig}(L_1 A_1 L_1^+): \text{rank} \begin{bmatrix} L_1 A_1 L_1^+ - \lambda I_{n_1-r} & L_1 A_1 R_{L_1} \end{bmatrix} = n_1 - r_1,$$

где $\text{eig}(L_1 A_1 L_1^+)$ – множество собственных значений матрицы $L_1 A_1 L_1^+$.

Продолжая рассуждения по индукции, приходим к следующей теореме, справедливость которой нами фактически доказана.

Теорема 1. МИМО-система (1) управляема, если и только если управляемо множество МИМО-систем

$$\dot{x}_i(t) = L_{i-1} A_{i-1} L_{i-1}^+ x_i(t) + L_{i-1} A_{i-1} R_{L_{i-1}} u_i(t), \quad i = \overline{1, n-r},$$

где $x_i(t) \in \mathbb{R}^{n_i-r_i}$ – вектор состояния i -й системы; $u_i(t) \in \mathbb{R}^{r_i}$ – вектор управления i -й системы; $n_i = n_{i-1} - r_{i-1}$, $r_i = \text{rank} L_{i-1} A_{i-1} R_{L_{i-1}}$; L_{i-1} , $R_{L_{i-1}}$ – максимальные решения соответственно уравнений

$$L_{i-1} R_{L_{i-1}} = 0_{(n_{i-1}-r_{i-1}) \times r_{i-1}},$$

$$L_{i-1} R_{L_{i-1}} = 0_{(n_{i-1}-r_{i-1}) \times r_{i-1}},$$

$$A_0 = A, \quad R_0 = B, \quad n_0 = n, \quad r_0 = r.$$

Из теоремы вытекает легко доказываемое следствие.

Следствие 1. МИМО-система (1), где $n > 1$, $u(t) \in \mathbb{R}^1$ – скаляр, управляема, если и только если управляема скалярная система

$$\dot{x}_{n-1}(t) = L_{n-2}A_{n-2}L_{n-2}^+x_{n-1}(t) + L_{n-2}A_{n-2}R_{L_{n-2}}u_{n-1}(t), \quad (18)$$

т.е. в (18) скаляр $L_{n-1}A_{n-1}R_{L_{n-1}} \neq 0$.

В качестве методического числового примера рассмотрим две СИМО-системы, каждая из которых является предельным вариантом МИМО-системы (1) с матрицами

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 2 & 3 \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \quad (19)$$

$$A = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 2 & 3 \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}. \quad (20)$$

Первая из данных систем – управляема, а вторая – нет.

В результате преобразований, выполненных по теореме 1, приходим к двум скалярным системам

$$\dot{x}_2 = u_2(t), \quad (21)$$

$$\dot{x}_2 = x_2(t). \quad (22)$$

Система (21) соответствует редукции состояния (19), а система (22) – (20). Очевидно, что (21) – управляема, а (22) – неуправляема. Это и требовалось показать.

2. Редукция тестов при анализе наблюдаемости

Для решения задачи редукции размерности состояния при анализе наблюдаемости МИМО-системы (1) воспользуемся преобразованиями матрицы (3), дуализированными к выполненным в предыдущем разделе преобразованиям матрицы (2). В результате придем к теореме.

Теорема 2. МИМО-система (1) наблюдаема, если и только если наблюдаемо множество МИМО-систем

$$\dot{x}_i(t) = R_{i-1}^+A_{i-1}R_{i-1}x_i(t),$$

$$y_i(t) = L_{R_{i-1}}A_{i-1}R_{i-1}x_i(t), \quad i = \overline{1, n-m},$$

где $x_i(t) \in \mathbb{R}^{n_i-m_i}$ – вектор состояния i -й системы; $y_i(t) \in \mathbb{R}^{m_i}$ – вектор выхода i -й системы;

$n_i = n_{i-1} - m_{i-1}$, $m_i = \text{rank } L_{R_{i-1}}A_{i-1}R_{i-1}$; R_{i-1} , $L_{R_{i-1}}$ – максимальные решения соответственно уравнений

$$L_{i-1}R_{i-1} = 0_{m_{i-1} \times (n_{i-1}-m_{i-1})}, \quad (23)$$

$$L_{R_{i-1}}R_{i-1} = 0_{m_{i-1} \times (n_{i-1}-m_{i-1})}, \quad (24)$$

$$A_0 = A, \quad L_0 = C, \quad n_0 = n, \quad m_0 = m.$$

Следствие 2. МИМО-система (1), где $n > 1$, $y(t) \in \mathbb{R}^1$ – скаляр, наблюдаема, если и только если наблюдаема скалярная система

$$\dot{x}_{n-1}(t) = R_{n-2}^+A_{n-2}R_{n-2}x_{n-1}(t), \quad (25)$$

$$y_{n-1}(t) = L_{R_{n-2}}A_{n-2}R_{n-2}x_{n-1}(t),$$

т.е. в (25) скаляр $L_{R_{n-2}}A_{n-2}R_{n-2} \neq 0$.

На основе данной теоремы осуществим анализ наблюдаемости следующей модели, описывающей поведение электроэнергетической системы [10]:

$$A = \begin{bmatrix} -G_1 M_1^{-1} & 0 & -M_1^{-1} & 0 & 0 & M_1^{-1} & 0 & 0 \\ 0 & -G_2 M_2^{-1} & M_1^{-1} & 0 & 0 & 0 & 0 & M_2^{-1} \\ T_{12} & -T_{12} & 0 & 0 & 0 & 0 & 0 & 0 \\ -E_1 T_{g_1}^{-1} & 0 & 0 & -T_{g_2}^{-1} & 0 & 0 & 0 & 0 \\ 0 & -E_2 T_{g_2}^{-1} & 0 & 0 & -T_{g_2}^{-1} & 0 & 0 & 0 \\ 0 & 0 & 0 & T_{t_1}^{-1} & 0 & -T_{t_1}^{-1} & 0 & 0 \\ 0 & 0 & 0 & 0 & T_{t_2}^{-1} & 0 & -T_{t_2}^{-1} & 0 \\ 0 & 0 & 0 & 0 & -2D_2^{-1} & 0 & 2(D_2^{-1} + T_{t_2}^{-1}) & -2D_2^{-1} \end{bmatrix}, \quad (26)$$

$$C = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} = [I_3 \mid 0_{3 \times 5}]. \quad (27)$$

Определим решения однородных уравнений (23), (24) для $i = 1$. Сначала рассмотрим уравнение

$$CR_0 = 0_{3 \times 5}$$

с матрицей C (27). Получим

$$R_0 = \begin{bmatrix} 0_{3 \times 5} \\ I_5 \end{bmatrix}. \quad (28)$$

Подставим матрицу (28) в уравнение (24):

$$L_{R_0} R_0 = 0_{3 \times 5},$$

и решим его относительно матрицы L_{R_0} . Будем иметь

$$L_{R_0} = [I_3 \mid 0_{3 \times 5}] = C.$$

Из (28) также следует, что

$$R_0^+ = R_0^T = [0_{5 \times 3} \mid I_5].$$

Теперь можно свести анализ наблюдаемости (26), (27) к анализу наблюдаемости системы

$$\dot{x}_1(t) = R_0^+ A R_0 x_1(t),$$

$$y_1(t) = L_{R_0} A R_0 x_1(t),$$

где

$$R_0^+ A R_0 = A_1 = \begin{bmatrix} -T_{g_2}^{-1} & 0 & 0 & 0 & 0 \\ 0 & -T_{g_2}^{-1} & 0 & 0 & 0 \\ T_{t_1}^{-1} & 0 & -T_{t_1}^{-1} & 0 & 0 \\ 0 & T_{t_2}^{-1} & 0 & -T_{t_2}^{-1} & 0 \\ 0 & -2D_2^{-1} & 0 & 2(D_2^{-1} + T_{t_2}^{-1}) & -2D_2^{-1} \end{bmatrix},$$

$$L_{R_0} A R_0 = C_1 = \begin{bmatrix} 0 & 0 & M_1^{-1} & 0 & 0 \\ 0 & 0 & 0 & 0 & M_2^{-1} \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Осуществляя далее аналогичные преобразования для $i = 2$, придем к анализу наблюдаемости системы

$$\dot{x}_2(t) = R_1^+ A_1 R_1 x_2(t),$$

$$y_2(t) = L_{R_1} A_1 R_1 x_2(t),$$

где

$$R_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \quad R_1^+ = R_1^T, \quad L_{R_1} = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

и

$$R_1^+ A_1 R_1 = A_2 = \begin{bmatrix} -T_{g_2}^{-1} & 0 & 0 \\ 0 & -T_{g_2}^{-1} & 0 \\ 0 & T_{t_2}^{-1} & -T_{t_2}^{-1} \end{bmatrix}, \quad (29)$$

$$L_{R_1} A_1 R_1 = C_2 = \begin{bmatrix} T_{t_1}^{-1} & 0 & 0 \\ 0 & -2D_2^{-1} & 2(D_2^{-1} + T_{t_2}^{-1}) \end{bmatrix}. \quad (30)$$

Итак, анализ наблюдаемости МИМО-системы с матрицами (26), (27) и $x(t) \in \mathbb{R}^8$, $y(t) \in \mathbb{R}^3$ сведен к анализу наблюдаемости МИМО-системы с матрицами (29), (30) и $x_2(t) \in \mathbb{R}^3$, $y_2(t) \in \mathbb{R}^2$.

Если осуществить еще одну редукцию, то получим МИМО-систему

$$\begin{aligned} \dot{x}_3(t) &= R_2^+ A_2 R_2 x_3(t), \\ y_3(t) &= L_{R_2} A_2 R_2 x_3(t), \end{aligned} \quad (31)$$

где

$$R_2 = \begin{bmatrix} 0 \\ \frac{2D_2^{-1}}{2(D_2^{-1} + T_{t_2}^{-1})} \\ 1 \end{bmatrix}, \quad R_2^+ = \frac{1}{1 + \left(\frac{2D_2^{-1}}{2(D_2^{-1} + T_{t_2}^{-1})} \right)^2} \begin{bmatrix} 0 & \frac{2D_2^{-1}}{2(D_2^{-1} + T_{t_2}^{-1})} & 1 \end{bmatrix},$$

$$L_{R_2} = \begin{bmatrix} 0 & 1 & -\frac{2D_2^{-1}}{2(D_2^{-1} + T_{t_2}^{-1})} \\ 1 & 0 & 0 \end{bmatrix},$$

и

$$R_2^+ A_2 R_2 = A_3 \in \mathbb{R}^1 \neq 0, \quad L_{R_2} A_2 R_2 = C_3 \in \mathbb{R}^2 \neq 0. \quad (32)$$

Как видно, МИМО-система (31) с параметрами (32) всегда наблюдаема, поэтому в силу теоремы 2 МИМО-система (1) с матрицами (26), (27) также является наблюдаемой.

Заключение

В работе доказаны утверждения, согласно которым анализ управляемости и наблюдаемости исходной МИМО-системы путем редукции может быть сведен к анализу управляемости и наблюдаемости МИМО-систем с существенно меньшей размерностью пространства состояний. В предельном случае анализ исходной МИМО-системы сводится к анализу скалярных систем.

Список источников

1. Поляк Б.Т., Щербаков П.С. Робастная устойчивость и управление. М. : Наука, 2002.
2. Мисриханов М.Ш., Рябченко В.Н. Редукция размерности состояния при анализе управляемости и наблюдаемости линейных моделей энергосистем // Известия ТРГУ. 2005. Т. 55, № 11. С. 45–54.

3. Гуссейнов Ф.Г. Упрощение расчетных схем электрических систем. М. : Энергия. 1978.
4. Kundur P. Power system stability and control. McGraw-Hill, Inc., 1994.
5. Зубов Н.Е., Микрин Е.А., Рябченко В.Н. Матричные методы в теории и практике систем автоматического управления летательных аппаратов. М. : Изд-во МГТУ им. Н.Э. Баумана, 2016.
6. Мисриханов М.Ш. Ленточные критерии управляемости и наблюдаемости линейных динамических систем // Вестник ИГЭУ. 2002. Вып. 3. С. 61–69.
7. Гантмахер Ф.Р. Теория матриц. М. : Наука, 1987.
8. Уонем М. Линейные многомерные системы управления: геометрический подход. М. : Наука, 1980.
9. Тауфер И. Решение граничных задач для систем линейных дифференциальных уравнений. М. : Наука, 1981.
10. Christensen G.S., El-Hawary M.E., Soliman S.A. Optimal Control Applications Electric Power Systems. New York ; London : Plenum Press, 1987.

References

1. Polyak, B.T. & Shcherbakov, P.S. (2002) *Robustnaya ustoychivost' i upravlenie* [Robust Stability and Control]. Moscow: Nauka.
2. Misrikhanov, M.Sh. & Ryabchenko, V.N. (2005) *Reduktsiya razmernosti sostoyaniya pri analize upravlyaemosti i nablyudaemosti lineynykh modeley energosistem* [State Dimension Reduction in the Analysis of Controllability and Observability of Linear Power System Models]. *Izvestiya TRGU*. 55(11). pp. 45–54.
3. Guseynov, F.G. (1978) *Uproshchenie raschetnykh skhem elektricheskikh sistem* [Simplification of calculation schemes of electrical systems]. Moscow: Energiya.
4. Kundur, P. (1994) *Power System Stability and Control*. McGraw-Hill, Inc.
5. Zubov, N.E., Mikrin, E.A. & Ryabchenko, V.N. (2016) *Matrichnye metody v teorii i praktike sistem avtomaticheskogo upravleniya letatel'nykh apparatov* [Matrix methods in the theory and practice of automatic control systems of aircraft]. Moscow: Bauman Moscow State Technical University.
6. Misrikhanov, M.Sh. (2002) *Lentochmye kriterii upravlyayemosti i nablyudayemosti lineynykh dinamicheskikh sistem* [Tape criteria of controllability and observability of linear dynamic systems]. *Vestnik IGEU*. 3. pp. 61–69.
7. Gantmacher, F.R. (1987) *Teoriya matrits* [The Theory of Matrices]. Moscow: Nauka.
8. Wonham, W.M. (1980) *Lineynye mnogomernye sistemy upravleniya: geometricheskii podkhod* [Linear Multivariable Control: A Geometric Approach]. Translated from English. Moscow: Nauka.
9. Tauffer, I. (1981) *Reshenie granichnykh zadach dlya sistem lineynykh differentsial'nykh uravneniy* [Solution of Boundary Problems for Systems of Linear Differential Equations]. Moscow: Nauka.
10. Christensen, G.S., El-Hawary, M.E. & Soliman, S.A. (1987) *Optimal Control Applications Electric Power Systems*. New York & London: Plenum Press.

Информация об авторах:

Зубов Николай Евгеньевич – профессор, доктор технических наук, декан факультета «Ракетно-космическая техника», профессор кафедры «Системы автоматического управления» МГТУ им. Н.Э. Баумана (Москва, Россия). E-mail: nik.zubov@gmail.com

Рябченко Владимир Николаевич – доцент, доктор технических наук, профессор кафедры «Системы автоматического управления» МГТУ им. Н.Э. Баумана (Москва, Россия). E-mail: ryabchenko.vn@mail.ru

Вклад авторов: все авторы сделали эквивалентный вклад в подготовку публикации. Авторы заявляют об отсутствии конфликта интересов.

Information about the authors:

Zubov Nikolay E. (Professor, Doctor of Technical Sciences, Dean of Rocket and Space Techniques Faculty, Professor of Department of Automatic Control Systems at Bauman Moscow State Technical University, Moscow, Russian Federation). E-mail: nik.zubov@gmail.com

Ryabchenko Vladimir N. (Associate Professor, Doctor of Technical Sciences, Professor of Department of the Automatic Control Systems at Bauman Moscow State Technical University, Moscow, Russian Federation). E-mail: ryabchenko.vn@mail.ru

Contribution of the authors: the authors contributed equally to this article. The authors declare no conflicts of interests.

Поступила в редакцию 21.04.2025; принята к публикации 02.12.2025

Received 21.04.2025; accepted for publication 02.12.2025

Научная статья
УДК 681.51
doi: 10.17223/19988605/73/3

О задаче управления качеством переходных процессов по выходу в матричной системе второго порядка с обратной связью по состоянию

Евгений Александрович Перепелкин

*Санкт-Петербургский государственный университет аэрокосмического приборостроения,
Санкт-Петербург, Россия, perepelkin@guap.ru*

Аннотация. Решается задача управления качеством переходных процессов по выходу в матричной системе второго порядка с обратной связью по состоянию. Качество переходных процессов задается эталонной моделью выхода в виде системы дифференциальных уравнений второго порядка. Описывается алгоритм модального синтеза обратной связи, при которой выход системы точно соответствует динамике эталонной модели. При этом замкнутая обратная связь система является асимптотически устойчивой. Исследуются условия существования решения задачи. Указываются ограничения, накладываемые на размеры векторов входа и выхода системы и на свойства матриц системы. Приводятся численный пример и результаты моделирования.

Ключевые слова: матричная система второго порядка; качество переходных процессов; эталонная модель.

Для цитирования: Перепелкин Е.А. О задаче управления качеством переходных процессов по выходу в матричной системе второго порядка с обратной связью по состоянию // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2025. № 73. С. 23–29. doi: 10.17223/19988605/73/3

Original article
doi: 10.17223/19988605/73/3

On the problem of transient process quality control of output in a matrix second-order system with state feedback

Evgenii A. Perepelkin

*Saint-Petersburg State University of Aerospace Instrumentation,
Saint-Petersburg, Russian Federation, perepelkin@guap.ru*

Abstract. The article solves the problem of transient process quality control of output in a matrix second-order system with state feedback. The transient process quality is specified by a reference output model in the form of second-order differential equations system. The paper describes modal synthesis algorithm, in which the system output exactly matches to the dynamics of the reference model. At the same time, the closed loop system is asymptotically stable. The conditions for the existence of a solution to the problem are investigated. The restrictions imposed on the sizes of the system input and output vectors and on the properties of the system matrices are indicated. A numerical example and simulation results are given.

Keywords: matrix second-order system; transient process quality; reference model.

For citation: Perepelkin, E.A. (2025) On the problem of transient process quality control of output in a matrix second-order system with state feedback. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naja tehnika i informatika – Tomsk State University Journal of Control and Computer Science*. 73. pp. 23–29. doi: 10.17223/19988605/73/3

Введение

Проблема качества переходных процессов в системах автоматического управления является одной из основных в теории и практике управления сложными динамическими объектами. Известные методы синтеза регуляторов [1, 2] не всегда могут обеспечить желаемое качество переходных процессов. Модальные, линейно-квадратичные, оптимальные и робастные по критериям H_2 / H_∞ регуляторы частично решают эту проблему, поскольку сложно определить влияние полюсов замкнутой системы или весовых матриц в интегральных критериях качества на временные характеристики переходных процессов в многосвязных системах.

В настоящей работе решается задача управления качеством переходных процессов по выходу для матричной системы второго порядка. Матричные системы дифференциальных уравнений второго порядка применяются для описания объектов управления в механике, акустике, электротехнике, робототехнике. Для решения задач управления такого рода системами применяют классические методы теории автоматического управления, в том числе методы модального и оптимального управления [3, 4].

В данной работе описывается метод синтеза обратной связи по состоянию в матричной системе второго порядка, при которой переходный процесс по выходу точно соответствует переходному процессу эталонной модели. Такой подход позволяет решить задачу управления качеством переходных процессов по выходу при условии асимптотической устойчивости системы с обратной связью.

Синтез на основе эталонной модели является классическим методом решения задач управления динамическими объектами. Данный метод применяется в теории автоматического управления для решения широкого круга задач, включая задачи адаптивного и робастного управления [5–10].

Работа содержит постановку задачи, предварительные сведения о решении задачи модального управления матричной системой второго порядка, алгоритм синтеза обратной связи с применением эталонной модели выхода, численный пример и результаты моделирования.

1. Постановка задачи

Рассмотрим линейную матричную систему второго порядка, поведение которой описывается уравнениями

$$\ddot{\mathbf{x}} + \mathbf{A}_1 \dot{\mathbf{x}} + \mathbf{A}_2 \mathbf{x} = \mathbf{B} \mathbf{u}, \quad \mathbf{y} = \mathbf{C} \mathbf{x}, \quad (1)$$

где $\mathbf{x} \in \mathbb{R}^n$ – вектор состояния, $\mathbf{u} \in \mathbb{R}^m$ – вектор управления, $\mathbf{y} \in \mathbb{R}^l$ – вектор выхода, матрицы системы $\mathbf{A}_1, \mathbf{A}_2 \in \mathbb{R}^{n \times n}$, $\mathbf{B} \in \mathbb{R}^{n \times m}$, $\mathbf{C} \in \mathbb{R}^{l \times n}$, $l < m < n$. Будем считать, что матрицы \mathbf{B} и \mathbf{C} полного ранга, $\text{rank } \mathbf{B} = m$, $\text{rank } \mathbf{C} = l$. Строки матрицы \mathbf{CB} линейно независимы, $\text{rank } \mathbf{CB} = l$.

В теории матричных систем второго порядка вектор \mathbf{x} принято называть перемещением системы, вектор $\dot{\mathbf{x}}$ – скоростью системы.

Пусть задано желаемое значение выхода $\bar{\mathbf{y}}$. Необходимо построить управление в виде обратной связи по скорости и перемещению $\mathbf{u} = -\mathbf{F}_1 \dot{\mathbf{x}} - \mathbf{F}_2 \mathbf{x} + \bar{\mathbf{u}}$, где $\mathbf{F}_1, \mathbf{F}_2 \in \mathbb{R}^{m \times n}$ – матрицы обратной связи, $\bar{\mathbf{u}} \in \mathbb{R}^m$ – некоторое постоянное значение управления такое, что

$$\lim_{t \rightarrow \infty} \mathbf{y}(t) = \bar{\mathbf{y}}$$

при любых начальных значениях перемещения и скорости $\mathbf{x}(0)$, $\dot{\mathbf{x}}(0)$. При этом замкнутая обратная связью система

$$\ddot{\mathbf{x}} + (\mathbf{A}_1 + \mathbf{BF}_1) \dot{\mathbf{x}} + (\mathbf{A}_2 + \mathbf{BF}_2) \mathbf{x} = \mathbf{B} \bar{\mathbf{u}} \quad (2)$$

должна быть асимптотически устойчивой, а переходный процесс по вектору выхода должен точно соответствовать переходному процессу эталонной модели асимптотически устойчивой матричной системы второго порядка

$$\ddot{\mathbf{y}} + \mathbf{D}_1 \dot{\mathbf{y}} + \mathbf{D}_2 (\mathbf{y} - \bar{\mathbf{y}}) = 0, \quad (3)$$

где $\mathbf{D}_1, \mathbf{D}_2 \in \mathbb{R}^{l \times l}$, $\det \mathbf{D}_2 \neq 0$.

2. Предварительные сведения

Матрицу $A(s) = Es^2 + A_1s + A_2$ будем называть матрицей разомкнутой системы, матрицу $A_c(s) = Es^2 + (A_1 + BF_1)s + A_2 + BF_2$ – матрицей замкнутой системы, матрицу $D(s) = Es^2 + D_1s + D_2$ – матрицей эталонной модели. Здесь E – единичная матрица соответствующего порядка, $s \in \mathbb{C}$.

Обозначим через $a(s) = \det A(s)$ характеристический полином разомкнутой системы, через $a_c(s) = \det A_c(s)$ – характеристический полином замкнутой системы, через $d(s) = \det D(s)$ – характеристический полином эталонной модели. Заметим, что $\deg a(s) = \deg a_c(s) = 2n$, $\deg d(s) = 2l$.

Полюсы системы есть корни характеристического полинома. Для асимптотической устойчивости системы необходимо и достаточно, чтобы полюсы системы имели вещественные части меньше нуля.

В работах [11, 12] приводятся необходимые и достаточные условия существования решения задачи о назначении полюсов замкнутой системы и описываются алгоритмы нахождения матриц обратной связи.

Утверждение 1. Все полюсы замкнутой системы можно произвольно задать, выбирая матрицы обратной связи F_1 и F_2 , тогда и только тогда, когда $\text{rank}(A(s), B) = n$ для любого $s \in \mathbb{C}$.

Решить задачу о назначении полюсов замкнутой системы можно, применяя известные методы модального управления. Заметим, что $a_c(s) = \det(sE - \bar{A} + \bar{B}\bar{F})$, где

$$\bar{A} = \begin{pmatrix} 0 & E \\ -A_2 & -A_1 \end{pmatrix}, \quad \bar{B} = \begin{pmatrix} 0 \\ B \end{pmatrix}, \quad \bar{F} = (F_2 \quad F_1).$$

Следовательно, полюсы замкнутой системы есть собственные числа матрицы $\bar{A} - \bar{B}\bar{F}$. Задача о назначении собственных чисел этой матрицы является задачей модального управления для системы с обратной связью по состоянию. Для решения этой задачи разработаны численные методы, реализованные в системах компьютерной математики [13. Р. 343].

Далее будут применяться следующие обозначения для матриц: M^+ – псевдообратная матрица; M_0 – матрица, столбцы которой составляют ортонормированный базис $\ker M = \{x | Mx = 0\}$.

Справедливы равенства [14. С. 31]: $MM^+M = M$, $M^+MM^+ = M^+$, $MM_0 = 0$. Если строки матрицы M линейно независимы, то $M^+ = M^T(MM^T)^{-1}$ и $MM^+ = E$, где E – единичная матрица. Если столбцы матрицы M линейно зависимы, то $M_0 \neq 0$ и $M_0^T M_0 = E$.

3. Алгоритм синтеза обратной связи

Пусть матрицы обратной связи F_1 , F_2 и вектор \bar{u} удовлетворяют соотношениям

$$C(A_1 + BF_1) = D_1C, \quad C(A_2 + BF_2) = D_2C, \quad (4)$$

$$CB\bar{u} = D_2\bar{y}. \quad (5)$$

Тогда выход системы подчиняется уравнению эталонной модели (3). Действительно, из уравнений (1), (4), (5) следует

$$\begin{aligned} C\ddot{x} + CA_1\dot{x} + CA_2x &= CBu, \\ C\ddot{x} + CA_1\dot{x} + CA_2x &= -CBF_1\dot{x} - CBF_2x + CB\bar{u}, \\ \ddot{y} + D_1\dot{y} + D_2(y - \bar{y}) &= 0. \end{aligned}$$

Уравнение (5) имеет бесконечно много решений, поскольку $\text{rank } CB = l < m$. Решение, обладающее минимальной евклидовой нормой [14. С. 34], может быть записано в виде: $\bar{u} = (CB)^+ D_2\bar{y}$.

Уравнения (4) также имеют бесконечно много решений относительно матриц \mathbf{F}_1 и \mathbf{F}_2 . Все множество решений этих уравнений можно записать в виде:

$$\mathbf{F}_1 = \mathbf{G}_1 + \mathbf{G}_0 \mathbf{H}_1, \quad \mathbf{F}_2 = \mathbf{G}_2 + \mathbf{G}_0 \mathbf{H}_2, \quad (6)$$

где $\mathbf{H}_1, \mathbf{H}_2 \in \mathbb{R}^{(m-l) \times n}$ – произвольные матрицы, $\mathbf{G}_0 = (\mathbf{CB})_0 \in \mathbb{R}^{m \times (m-l)}$,

$$\mathbf{G}_1 = (\mathbf{CB})^+ (\mathbf{D}_1 \mathbf{C} - \mathbf{CA}_1) \in \mathbb{R}^{m \times n}, \quad \mathbf{G}_2 = (\mathbf{CB})^+ (\mathbf{D}_2 \mathbf{C} - \mathbf{CA}_2) \in \mathbb{R}^{m \times n}.$$

Матрицы \mathbf{H}_1 и \mathbf{H}_2 будем искать из условия асимптотической устойчивости замкнутой системы (2), уравнение которой с учетом формул (6) принимает вид:

$$\ddot{\mathbf{x}} + (\mathbf{A}_1 + \mathbf{BG}_1 + \mathbf{BG}_0 \mathbf{H}_1) \dot{\mathbf{x}} + (\mathbf{A}_2 + \mathbf{BG}_2 + \mathbf{BG}_0 \mathbf{H}_2) \mathbf{x} = \mathbf{B} \bar{\mathbf{u}}, \quad (7)$$

Полюсы системы (7) есть корни характеристического полинома $a_c(s) = \det(\mathbf{A}_c(s))$, где $\mathbf{A}_c(s) = \mathbf{E}s^2 + (\mathbf{A}_1 + \mathbf{BG}_1 + \mathbf{BG}_0 \mathbf{H}_1)s + \mathbf{A}_2 + \mathbf{BG}_2 + \mathbf{BG}_0 \mathbf{H}_2$ – матрица замкнутой системы.

Составим невырожденную матрицу

$$\mathbf{P} = \begin{pmatrix} \mathbf{C} \\ \mathbf{C}_0^T \end{pmatrix}.$$

Обратная матрица $\mathbf{P}^{-1} = \begin{pmatrix} \mathbf{C}^+ & \mathbf{C}_0 \end{pmatrix}$. Здесь $\mathbf{C}^+ = \mathbf{C}^T (\mathbf{CC}^T)^{-1}$. Выполним невырожденное преобразование матрицы замкнутой системы

$$\mathbf{PA}_c(s)\mathbf{P}^{-1} = \begin{pmatrix} \mathbf{D}(s) & 0 \\ \mathbf{C}_0^T \mathbf{A}_c(s) \mathbf{C}^+ & \mathbf{C}_0^T \mathbf{A}_c(s) \mathbf{C}_0 \end{pmatrix}.$$

Следовательно, характеристический полином замкнутой системы $a_c(s) = d(s)h(s)$, где $h(s) = \det(\mathbf{C}_0^T \mathbf{A}_c(s) \mathbf{C}_0)$.

Замкнутая система асимптотически устойчива тогда и только тогда, когда корни полиномов $d(s)$ и $h(s)$ имеют отрицательные вещественные части. Полином $d(s)$ является характеристическим полиномом асимптотически устойчивой эталонной модели. Следовательно, для асимптотической устойчивости замкнутой системы необходимо и достаточно, чтобы корни полинома $h(s)$ имели отрицательные вещественные части.

Обозначим

$$\begin{aligned} \mathbf{V}(s) &= \mathbf{E}s^2 + \mathbf{V}_1 s + \mathbf{V}_2, & \mathbf{W}_1(s) &= \mathbf{W}_1 s + \mathbf{W}_2, \\ \mathbf{V}_1 &= \mathbf{C}_0^T (\mathbf{A}_1 + \mathbf{BG}_1) \mathbf{C}_0, & \mathbf{V}_2 &= \mathbf{C}_0^T (\mathbf{A}_2 + \mathbf{BG}_2) \mathbf{C}_0, \\ \mathbf{W}_1 &= \mathbf{H}_1 \mathbf{C}_0, & \mathbf{W}_2 &= \mathbf{H}_2 \mathbf{C}_0. \end{aligned} \quad (8)$$

Тогда $\mathbf{C}_0^T \mathbf{A}_c(s) \mathbf{C}_0 = \mathbf{V}(s) + \mathbf{C}_0^T \mathbf{BG}_0 \mathbf{W}(s)$, $h(s) = \det(\mathbf{V}(s) + \mathbf{C}_0^T \mathbf{BG}_0 \mathbf{W}(s))$.

Согласно утверждению 1 все корни полинома $h(s)$ можно произвольно задать, выбирая матрицы \mathbf{W}_1 и \mathbf{W}_2 , тогда и только тогда, когда $\text{rank}(\mathbf{V}(s) + \mathbf{C}_0^T \mathbf{BG}_0 \mathbf{W}(s)) = n - l$ для всех $s \in \mathbb{C}$.

Предположим, что матрицы \mathbf{W}_1 и \mathbf{W}_2 найдены из условия заданных корней полинома $h(s)$. Тогда из соотношений (8) получим искомые значения матриц обратной связи $\mathbf{H}_1 = \mathbf{W}_1 \mathbf{C}_0^T$, $\mathbf{H}_2 = \mathbf{W}_2 \mathbf{C}_0^T$.

В целом можно сформулировать следующее

Утверждение 2. Пусть $\text{rank}(\mathbf{V}(s) + \mathbf{C}_0^T \mathbf{BG}_0 \mathbf{W}(s)) = n - l$ для всех $s \in \mathbb{C}$. Тогда существует управление в виде обратной связи $\mathbf{u} = -\mathbf{F}_1 \dot{\mathbf{x}} - \mathbf{F}_2 \mathbf{x} + \bar{\mathbf{u}}$, при котором выход системы (1) подчиняется уравнению эталонной модели (3), и при этом замкнутая обратной связью система является асимптотически устойчивой.

Алгоритм синтеза обратной связи заключается в следующем:

1. Вычисляем матрицы $\mathbf{G}_0 = (\mathbf{CB})_0$, $\mathbf{G}_1 = (\mathbf{CB})^+(\mathbf{D}_1\mathbf{C} - \mathbf{CA}_1)$, $\mathbf{G}_2 = (\mathbf{CB})^+(\mathbf{D}_2\mathbf{C} - \mathbf{CA}_2)$.
2. Вычисляем матрицы $\mathbf{V}_1 = \mathbf{C}_0^T(\mathbf{A}_1 + \mathbf{BG}_1)\mathbf{C}_0$, $\mathbf{V}_2 = \mathbf{C}_0^T(\mathbf{A}_2 + \mathbf{BG}_2)\mathbf{C}_0$ и находим матрицы \mathbf{W}_1 и \mathbf{W}_2 из условия заданных корней полинома $h(s) = \det(\mathbf{V}(s) + \mathbf{C}_0^T\mathbf{BG}_0\mathbf{W}(s))$.
3. Вычисляем матрицы обратной связи $\mathbf{F}_1 = \mathbf{G}_1 + \mathbf{G}_0\mathbf{W}_1\mathbf{C}_0^T$, $\mathbf{F}_2 = \mathbf{G}_2 + \mathbf{G}_0\mathbf{W}_2\mathbf{C}_0^T$.
4. Находим $\bar{\mathbf{u}} = (\mathbf{CB})^+\mathbf{D}_2\bar{\mathbf{y}}$.

4. Численный пример

Рассмотрим систему (1) с матрицами

$$\mathbf{A}_1 = \begin{pmatrix} -1,2801 & 8,9546 & -3,3933 & -7,601 & 2,4227 \\ -6,2984 & -1,2936 & -6,9115 & -4,0069 & 6,3645 \\ -9,4815 & -0,305 & -5,907 & -0,2965 & 0,5828 \\ 9,6308 & -1,5926 & 3,9773 & -4,6635 & 3,6605 \\ 0,9932 & -3,5893 & 2,3854 & 2,6548 & -7,3084 \end{pmatrix},$$

$$\mathbf{A}_2 = \begin{pmatrix} -0,0288 & 5,7067 & -1,8538 & 0,1049 & -7,6761 \\ 0,2716 & -4,83 & 6,9312 & -4,0596 & -8,0694 \\ 1,7359 & 7,0795 & -6,4603 & -8,6943 & -6,3655 \\ -6,3112 & 0,9241 & -8,4071 & -4,2426 & -7,4568 \\ 4,3951 & -0,1153 & 9,3926 & -1,4376 & -0,1142 \end{pmatrix}, \mathbf{B} = \begin{pmatrix} 1,9349 & 5,3498 & -0,6443 \\ 1,3153 & -5,5939 & -2,6506 \\ -5,4798 & 1,5462 & -5,9651 \\ -5,5633 & -3,0035 & -0,6265 \\ -7,8611 & -6,6435 & 2,8081 \end{pmatrix},$$

$$\mathbf{C} = \begin{pmatrix} -7,6103 & -7,0287 & 7,0456 & 0,1905 & -5,6546 \\ -6,754 & 4,015 & 9,291 & 0,0002 & 7,7904 \end{pmatrix}.$$

Матрицы эталонной модели зададим в виде

$$\mathbf{D}_1 = \begin{pmatrix} 2r & 0 \\ 0 & 2r \end{pmatrix}, \quad \mathbf{D}_2 = \begin{pmatrix} r^2 & 0 \\ 0 & r^2 \end{pmatrix}, \quad r > 0.$$

Характеристический полином эталонной модели $d(s) = (s + r)^4$. Эталонная модель обладает монотонным переходным процессом по переменным выхода. Время переходного процесса определяется параметром r . Пусть $r = 3$.

Вычисляем матрицы

$$\mathbf{G}_0 = (-0,5352 \quad 0,3496 \quad 0,769)^T,$$

$$\mathbf{G}_1 = \begin{pmatrix} -0,1005 & -0,5916 & -0,6181 & 0,6787 & -0,3366 \\ -0,5346 & -0,4264 & -0,072 & -0,6041 & 0,5055 \\ 0,1731 & -0,2178 & -0,3974 & 0,747 & -0,0045 \end{pmatrix},$$

$$\mathbf{G}_2 = \begin{pmatrix} 0,8861 & 0,5039 & -1,3799 & -0,7135 & 0,1987 \\ -0,2365 & -1,3304 & 2,2299 & -0,0467 & -1,7984 \\ 0,7243 & 0,9556 & -1,9743 & -0,4753 & 0,956 \end{pmatrix},$$

$$\mathbf{V}_1 = \begin{pmatrix} -7,233 & 5,3665 & 12,212 \\ 0,6414 & -7,0061 & 12,385 \\ -5,026 & -13,215 & -9,9859 \end{pmatrix}, \quad \mathbf{V}_2 = \begin{pmatrix} 4,2705 & 5,1324 & -0,9356 \\ -2,4274 & 0,0301 & -12,399 \\ -16,317 & -1,852 & 7,8424 \end{pmatrix}.$$

Находим матрицы

$$\mathbf{W}_1 = (9,6517 \quad -8,9177 \quad 17,748), \quad \mathbf{W}_2 = (1,167 \quad 6,8491 \quad -68,704),$$

при которых полином $h(s) = (s + 5)^6$. Вычисляем матрицы обратной связи

$$\mathbf{F}_1 = \begin{pmatrix} -7,2706 & 4,9514 & -5,5369 & 5,4825 & -3,5435 \\ 4,1496 & -4,0476 & 3,1414 & -3,7424 & 1,5896 \\ 10,475 & -8,1819 & 6,6698 & -6,155 & 4,6031 \end{pmatrix},$$

$$\mathbf{F}_2 = \begin{pmatrix} 28,586 & -5,5912 & 21,99 & -4,5629 & -0,5162 \\ -18,332 & 2,6514 & -13,037 & 2,4681 & -1,3314 \\ -39,074 & 9,713 & -35,551 & 5,0555 & 1,9833 \end{pmatrix}.$$

При этих значениях матриц обратной связи характеристический полином замкнутой системы $a_c(s) = (s + 3)^4(s + 5)^6$. Замкнутая система асимптотически устойчива. Заметим, что разомкнутая система не является асимптотически устойчивой.

Пусть заданное значение выхода системы $\bar{\mathbf{y}} = (1 \ 2)^T$. Тогда $\bar{\mathbf{u}} = (-0,1451 \ 0,044 \ -0,121)^T$. Выход системы удовлетворяет уравнению эталонной модели.

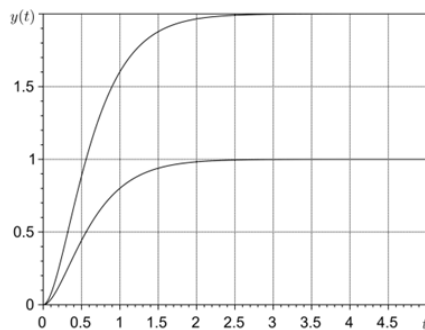


Рис. 1. Переходный процесс по переменным выхода
Fig. 1. Transient process of output variables

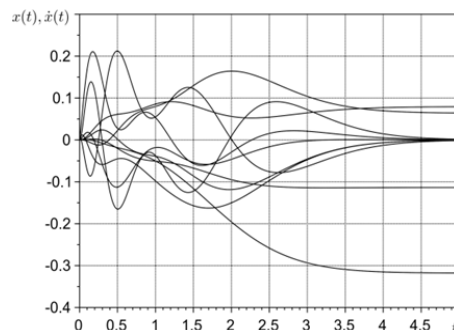


Рис. 2. Переходный процесс по переменным перемещения и скорости
Fig. 2. Transient process of displacement and velocity variables

На рис. 1, 2 показаны результаты моделирования замкнутой системы при нулевых начальных условиях. На рис. 1 показан переходный процесс по переменным выхода, на рис. 2 – переходный процесс по переменным перемещения и скорости.

Все вычисления и моделирование выполнялись в системе компьютерной математики Scilab.

Заключение

В работе решена задача управления качеством переходных процессов по выходу для матричной системы второго порядка. Предложен алгоритм синтеза управления в виде обратной связи по состоянию, при которой выход системы точно соответствует эталонной модели с заданным качеством переходных процессов. При этом замкнутая система асимптотически устойчива. Исследованы условия существования решения задачи. Указаны ограничения, накладываемые на размеры векторов управления и выхода и на матрицы системы. Численный пример и результаты моделирования подтверждают теоретические результаты работы.

Список источников

1. Александров А.Г. Методы построения систем автоматического управления. М. : УРСС, 2008. 232 с.
2. Поляк Б.Т., Хлебников М.В., Рапопорт Л.Б. Математическая теория автоматического управления. М. : УРСС, 2019. 500 с.
3. Gawronski W.K. Dynamics and Control of Structures: a modal approach. New York : Springer, 1998. 231 p. doi: 10.1007/978-0-387-21855-7
4. Du C., Xie L. Modeling and Control of Vibration in Mechanical Systems. Boca Raton, FL : CRC Press, 2010. 336 p. doi: 10.1201/9781315218069
5. Петров Б.Н., Рутковский В.Ю., Земляков С.Д. Адаптивное координатно-параметрическое управление нестационарными объектами. М. : Наука, 1980. 244 с.

6. Мирошник И.В., Никифоров В.О., Фрадков А.Л. Нелинейное и адаптивное управление сложными динамическими системами. М. : Наука, 2000. 549 с.
7. Бронников А.М., Буков В.Н. Условия точного слежения линейной системы за эталонной моделью пониженного порядка // Автоматика и телемеханика. 2008. № 3. С. 60–69.
8. Буков В.Н., Бронников А.М., Сельвесюк Н.И. Децентрализованное управление с модельной координацией составной многосвязной системой // Автоматика и телемеханика. 2009. № 10. С. 3–14.
9. Проскурников А.В., Якубович В.А. Линейные системы с эталонной моделью // Доклады Академии наук. 2007. Т. 415, № 4. С. 441–446.
10. Проскурников А.В., Якубович В.А. Универсальные регуляторы в задачах оптимального управления с эталонной моделью при неизвестных внешних сигналах // Известия РАН. Теория и системы управления. 2012. № 2. С. 49–62.
11. Chu E.K., Datta B.N. Numerically robust pole assignment for second-order systems // International Journal of Control. 1996. V. 64 (4). P. 1113–1127. doi: 10.1080/00207179608921677
12. Chu E.K. Pole assignment for second-order systems // Mechanical Systems and Signal Processing. 2002. V. 16 (1). P. 39–59. doi: 10.1006/mssp.2001.1439
13. Datta B. Numerical Methods for Linear Control Systems. Academic Press, 2004. 695 p. doi: 10.1016/B978-0-12-203590-6.X5000-9
14. Гантмахер Ф.Р. Теория матриц. М. : Физматлит, 2010. 560 с.

References

1. Aleksandrov, A.G. (2008) *Metody postroeniya sistem avtomaticheskogo upravleniya* [Methods for Constructing Automatic Control Systems]. Moscow: URSS.
2. Polyak, B.T., Khlebnikov, M.V. & Rapoport, L.B. (2019) *Matematicheskaya teoriya avtomaticheskogo upravleniya* [Mathematical Theory of Automatic Control]. Moscow: URSS.
3. Gawronski, W.K. (1998) *Dynamics and Control of Structures: A Modal Approach*. New York: Springer. doi: 10.1007/978-0-387-21855-7
4. Du, C. & Xie, L. (2010) *Modeling and Control of Vibration in Mechanical Systems*. Boca Raton, FL: CRC Press. doi: 10.1201/9781315218069
5. Petrov, B.N., Rutkovskiy, V.Yu. & Zemlyakov, S.D. (1980) *Adaptivnoe koordinatno-parametricheskoe upravlenie nestatsionarnymi ob"ektami* [Adaptive Coordinate-Parametric Control of Non-Stationary Objects]. Moscow: Nauka.
6. Miroshnik, I.V., Nikiforov, V.O. & Fradkov, A.L. (2000) *Nelineynoe i adaptivnoe upravlenie slozhnymi dinamicheskimi sistemami* [Nonlinear and Adaptive Control of Complex Dynamic Systems]. Moscow: Nauka.
7. Bronnikov, A.M. & Bukov, V.N. (2008) Usloviya tochnogo sledovaniya lineynoy sistemy za etalonnoy model'yu ponizhennogo poriyadka [Conditions for Exact Tracking of a Linear System After a Reduced-Order Reference Model]. *Avtomatika i telemekhanika*. 3. pp. 60–69.
8. Bukov, V.N., Bronnikov, A.M. & Selvesyuk, N.I. (2009) Detsentralizovannoe upravlenie s model'noy koordinatsiey sostavnoy mnogosvyaznoy sistemoy [Decentralized Control with Model Coordination of a Composite Multivariable System]. *Avtomatika i telemekhanika*. 10. pp. 3–14.
9. Proskurnikov, A.V. & Yakubovich, V.A. (2007) Lineynye sistemy s etalonnoy model'yu [Linear Systems with a Reference Model]. *Doklady Akademii nauk*. 415(4). pp. 441–446.
10. Proskurnikov, A.V. & Yakubovich, V.A. (2012) Universal'nye regulatory v zadachakh optimal'nogo upravleniya s etalonnoy model'yu pri neizvestnykh vnesnykh signalakh [Universal Controllers in Optimal Control Problems with a Reference Model Under Unknown External Signals]. *Izvestiya RAN. Teoriya i sistemy upravleniya*. 2. pp. 49–62.
11. Chu, E.K. & Datta, B.N. (1996) Numerically robust pole assignment for second-order systems. *International Journal of Control*. 64(4). pp. 1113–1127. doi: 10.1080/00207179608921677
12. Chu, E.K. (2002) Pole assignment for second-order systems. *Mechanical Systems and Signal Processing*. 16(1). pp. 39–59. doi: 10.1006/mssp.2001.1439
13. Datta, B. (2004) *Numerical Methods for Linear Control Systems*. Academic Press. doi: 10.1016/B978-0-12-203590-6.X5000-9
14. Gantmacher, F.R. (2010) *Teoriya matrits* [The Theory of Matrices]. Moscow: Fizmatlit.

Информация об авторе:

Перепелкин Евгений Александрович – профессор, доктор технических наук, профессор кафедры прикладной математики Санкт-Петербургского государственного университета аэрокосмического приборостроения (Санкт-Петербург, Россия). E-mail: perepelkin@guap.ru

Автор заявляет об отсутствии конфликта интересов.

Information about the author:

Perepelkin Evgenii A. (Professor, Doctor of Technical Sciences, Saint-Petersburg State University of the Aerospace Instrumentation, Saint-Petersburg, Russian Federation). E-mail: perepelkin@guap.ru

The author declares no conflicts of interests.

Поступила в редакцию 12.09.2025; принята к публикации 02.12.2025

Received 12.09.2025; accepted for publication 02.12.2025

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ

MATHEMATICAL MODELING

Научная статья

УДК 004.8

doi: 10.17223/19988605/73/4

Исследование устойчивости моделей полносверточных нейронных сетей с робастными функциями потерь к импульсным помехам на изображениях**Артем Юрьевич Малкин¹, Николай Григорьевич Марков²**^{1, 2} *Национальный исследовательский Томский политехнический университет, Томск, Россия*¹ *malkin@tpu.ru*² *markovng@tpu.ru*

Аннотация. Исследуется устойчивость моделей полносверточных нейронных сетей mo-u-net, полученных путем использования известных робастных функций потерь (РФП), к импульсным помехам на изображениях деревьев пихты, пораженных уссурийским полиграфом. Для исследования устойчивости применяются метрики точности классификации деревьев IoUc и mIoU. Для исследования моделей создан датасет, фрагменты обучающей выборки которого имеют искусственным образом внесенные импульсные помехи. Анализ результатов исследования каждой из семи моделей показал, что модель с РФП Уэлша является наиболее предпочтительной при работе с зашумленными изображениями.

Ключевые слова: семантическая сегментация (попиксельная классификация) изображения; импульсная помеха; робастная функция потерь; помехоустойчивость сверточной нейронной сети.

Для цитирования: Малкин А.Ю., Марков Н.Г. Исследование устойчивости моделей полносверточных нейронных сетей с робастными функциями потерь к импульсным помехам на изображениях // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2025. № 73. С. 30–40. doi: 10.17223/19988605/73/4

Original article

doi: 10.17223/19988605/73/4

Research of resistance of fully convolutional neural network models with robust loss functions to impulse noise in images**Artem Yu. Malkin¹, Nikolai G. Markov²**^{1, 2} *National Research Tomsk Polytechnic University, Tomsk, Russian Federation*¹ *malkin@tpu.ru*² *markovng@tpu.ru*

Abstract. The resistance of Mo-U-Net full-field neural network models, obtained by using known robust loss functions (RLFs), to impulse noise on images of fir trees affected by the Ussuri polygraph is investigated using the tree classification accuracy metrics IoUc and mIoU. To investigate the models, a dataset was created with fragments of the training sample having artificially introduced impulse noise. The analysis of the results of each of the seven models showed that the model with Welch's RLF is the most preferable when working with noisy images.

Keywords: semantic segmentation (pixel-by-pixel classification) of images; impulse noise; robust loss function; noise immunity of convolutional neural network.

For citation: Malkin, A.Yu., Markov, N.G. (2025) Research of resistance of fully convolutional neural network models with robust loss functions to impulse noise in images. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naja tehnika i informatika – Tomsk State University Journal of Control and Computer Science*. 73. pp. 30–40. doi: 10.17223/19988605/73/4

Введение

Управление лесами, концепция которого соотносит экологическую целостность лесов с потребностями человека, во многом зависит от точной и актуальной информации о состоянии лесов, получаемой при их мониторинге [1, 2]. Технологии мониторинга лесов значительно эволюционировали с годами на пути от трудоемких наземных обследований к более сложным методам дистанционного зондирования Земли (ДЗЗ). Это обусловлено необходимостью использования более оперативных и экономически эффективных методов оценки состояния и управления лесными ресурсами, занимающими большие площади. Сегодня все чаще выполняют оперативный мониторинг лесов с помощью современных систем ДЗЗ. При этом используют высокоточную съемку деревьев с помощью фото- и (или) видеокамер, установленных на космических аппаратах, самолетах (вертолетах) или беспилотных летательных аппаратах (БПЛА), а затем проводят дешифрирование полученных изображений [2–5]. Однако специалисты лесной отрасли постоянно сталкиваются с отсутствием современного инструментария – моделей, методов и информационных систем для автоматического дешифрирования таких изображений. Одним из актуальных направлений, позволяющих решать проблему отсутствия такого инструментария, является создание и использование современных моделей сверточных нейронных сетей (СНС) [5, 6]. Эти модели должны оперативно и с приемлемой точностью решать задачи семантической сегментации (попиксельной классификации) изображений, полученных при дистанционном мониторинге лесов.

Особые требования по точности классификации предъявляются к моделям СНС, используемым для дешифрирования изображений, получаемых при оперативном лесопатологическом мониторинге хвойных лесов. В этих случаях на изображениях необходимо оценить степень поражения (состояние здоровья) каждого дерева хвойной породы, заселенного тем или иным насекомым-вредителем. Для этого, учитывая значительное число состояний здоровья пораженных деревьев, необходимы модели СНС, позволяющие решать с высокой точностью задачу мультиклассификации деревьев на таких изображениях [6, 7]. При этом практически важным для сохранения хвойных лесов является обнаружение деревьев, находящихся на ранней стадии поражения вредителями, поскольку для них своевременно могут быть проведены эффективные фитосанитарные мероприятия.

В последние годы лесопатологический мониторинг лесов ведется с использованием самых разных методов ДЗЗ, но предпочтение чаще всего отдается фотосъемке с использованием БПЛА самолетного или вертолетного типа. При проведении такой съемки в результате сбоя аппаратуры получаемые изображения крон деревьев могут иметь помехи в виде случайных аддитивных шумов и (или) импульсных помех. Более того, если используется передача изображений по радиоканалу с БПЛА в наземный пункт сбора данных, то на изображениях могут появиться дополнительные помехи. В любом случае возникает задача оценки влияния таких помех на точность мультиклассификации пораженных деревьев на изображениях. Иными словами, необходимо выявить устойчивость используемых для мультиклассификации моделей СНС к тем или иным помехам на изображениях.

Анализ результатов ряда исследований [8, 9] показал, что для повышения устойчивости моделей СНС при работе с зашумленными изображениями применяется два основных подхода. В рамках первого из них на основе известных моделей СНС создаются модели со специальными архитектурами, устойчивыми к различным помехам. Так, в работе [8] в архитектуры базовых моделей ResNet-50 и ImageNet встраивается модуль релевантной фокусировки признаков (Relevant Feature Focusing; ReFF), который генерирует аннотации релевантных признаков. Это позволяет новым моделям СНС повысить устойчивость к изменению распределения объектов на изображениях, вызванному помехами. В результате исследований авторами установлено, что включение модуля ReFF в изучаемые модели СНС дает положительный результат даже при наличии существенного объема нерелевантных данных и малом

количестве аннотаций для классов объектов: прирост точности классификации по метрике Accuracy в условиях помех составляет до 15% относительно случаев, когда модуль ReFF не включается в модели. К сожалению, авторы не дают конкретных рекомендаций по проектированию модулей, подобных модулю ReFF, для случаев современных моделей СНС с более сложными архитектурами.

Другим основным подходом к повышению устойчивости моделей СНС к помехам на входных изображениях является применение робастных (устойчивых к шумам) функций потерь при обучении этих моделей. Использование робастных функций потерь (РФП) дает возможность при правильно подобранном коэффициенте потерь этих функций уменьшить значение ошибки в методе ее обратного распространения и, соответственно, повысить точность классификации объектов, в том числе в условиях помех. Автор [9] использовал довольно широкий набор РФП при решении задачи классификации ирисов Фишера на изображениях с помощью нейронных сетей прямого распространения: функции потерь Хьюбера, Тьюки, Рамсея, Коши и др. При этом проводился поиск интервалов значений коэффициента потерь β исследуемых РФП, которые затем можно было рекомендовать для обучения нейронных сетей при решении задачи классификации ирисов Фишера. Отметим, что некоторые РФП помимо обучения нейронных сетей прямого распространения применялись также для обучения рекуррентных сетей [10] и СНС [11–13] при решении задач семантической сегментации и классификации объектов на изображениях. Результаты исследований моделей нейронных сетей с различными РФП, изложенные в работах [9–13], указывают на перспективность робастного подхода к повышению устойчивости моделей глубокого обучения в условиях помех. На него и будем опираться в наших исследованиях.

Цель данной работы – проведение исследований устойчивости моделей полносверточных нейронных сетей класса U-Net [14] с различными РФП при решении задач мультиклассификации в условиях импульсных помех на изображениях. В качестве примера таких изображений используются реальные снимки высокого разрешения с фотокамеры на БПЛА деревьев пихты сибирской *Abies sibirica* (далее – *A. sibirica*), пораженных стволовым вредителем – уссурийским полиграфом *Polygraphus proximus* (далее – *P. proximus*). Перед обучением указанных моделей СНС в эти изображения искусственным образом вводятся импульсные помехи. Исследования устойчивости обученных моделей выполняются при различных амплитудах импульсных помех и площади зашумления изображений.

1. Задача исследования устойчивости моделей полносверточных нейронных сетей

В работе [5] показано, что среди моделей СНС для решения задач мультиклассификации пораженных хвойных деревьев на изображениях по критериям ‘точность классификации – скорость выполнения модели’ предпочтение следует отдать полносверточным нейронным сетям. В последние годы все чаще применяют классическую модель полносверточной нейронной сети U-Net [14] и модели на ее основе, позволяющие решать задачи попиксельной классификации изображений. Так, в [5, 7, 15] приведены обнадеживающие результаты мультиклассификации изображений пораженных уссурийским полиграфом *P. proximus* деревьев пихты *A. sibirica* (пять классов) и пораженных союзным короедом *Ips amitinus* деревьев кедра *Pinus sibirica* (шесть классов), полученные с помощью моделей на основе классической модели U-Net. Однако в случае использования классической модели U-Net наблюдается низкая точность классификации одного (случай деревьев *A. sibirica*) или двух (случай деревьев *Pinus sibirica*) промежуточных классов (состояний здоровья) деревьев [7, 15], что не устраивает специалистов лесной отрасли. Распознавание деревьев в промежуточных состояниях (состояния между здоровым деревом и старым сухостоем) имеет большое практическое значение для их сохранения, поскольку выявленные деревья могут быть своевременно подвергнуты эффективным санитарно-оздоровительным мероприятиям. Поэтому в данной работе предлагается исследовать на устойчивость к импульсным помехам модель Mo-U-Net, детально описанную в [5]. Она является модификацией модели U-Net и показывает, как это следует из [5, 16], практически приемлемую точность классификации хвойных деревьев, находящихся в промежуточных состояниях здоровья. В качестве функции потерь в ней использовалась функция Focal Loss [5].

При обучении модели Mo-U-Net на зашумленных импульсными помехами изображениях будут использоваться следующие семь РФП [9]: Коши, Рамсея, Хьюбера, Эндрюса, Geman-McCluer, Уэлша и

Мешалкина (табл. 1). Из работы [9] следует, что производные от этих робастных функций являются непрерывными на множестве вещественных чисел, и это позволяет использовать такие функции в качестве функций потерь в алгоритме обратного распространения ошибки при обучении модели Mo-U-Net. Заметим, по сути, обучению и исследованию подлежит семь полносверточных моделей нейронных сетей, полученных путем замены у модели Mo-U-Net функции потерь Focal Loss на ту или иную РФП из указанных в табл. 1. Для этого разрабатывается и программно реализуется модификация алгоритма обратного распространения ошибки, учитывающая особенности используемых РФП.

Таблица 1

Робастные функции потерь и их производные

Функция потерь $\rho(z)$	Производная функции потерь $\rho'(z)$
Коши	
$\ln \left(\frac{1}{2} \left(\frac{z}{\beta} \right)^2 + 1 \right)$	$\frac{z}{\frac{1}{2}z^2 + \beta^2}$
Рамсея	
$\frac{1 - (1 + \beta z)\exp(-\beta z)}{\beta^2}$	$z \exp\{-\beta z \}$
Хьюбера	
$\begin{cases} \frac{z^2}{2}, & z \leq \beta, \\ \beta z - \frac{\beta^2}{2}, & z > \beta \end{cases}$	$\begin{cases} z, & z \leq \beta, \\ -\beta, & z < \beta, \\ \beta, & z < \beta \end{cases}$
Эндрюса	
$\begin{cases} \beta \left(1 - \cos \frac{z}{\beta} \right), & z < \pi\beta, \\ 2\beta, & z \geq \beta \end{cases}$	$\begin{cases} \sin \frac{z}{\beta}, & z < \pi\beta, \\ 0, & z \geq \beta \end{cases}$
Geman-McCluer	
$\frac{z^2 / \beta}{1 + z^2 / \beta}$	$\frac{2z}{\beta \left(1 + z^2 / \beta \right)^2}$
Уэлша	
$1 - \exp \left(-\frac{1}{2} \left(\frac{z}{\beta} \right)^2 \right)$	$\frac{1}{\beta^2} z \exp \left(-\frac{1}{2} \left(\frac{z}{\beta} \right)^2 \right)$
Мешалкина	
$\beta^{-1} \left(1 - \exp \left(\frac{-\beta z^2}{2} \right) \right)$	$\exp \left(\frac{-\beta z^2}{2} \right) z$

Для обучения, валидации и исследования устойчивости каждой из семи полносверточных моделей будет использоваться датасет, созданный сотрудниками Института мониторинга климатических и экологических систем СО РАН и Томского политехнического университета [7]. При его формировании использовались RGB-изображения пораженных деревьев *A. sibirica* высокого разрешения (0,1 м), полученные летом 2017 г. в Томской области с использованием фотокамеры, установленной на БПЛА DJI Phantom 3 Standart. При дешифрировании этих изображений экспертами выделялось четыре класса состояния здоровья деревьев: «Живое», «Отмирающее», «Свежий сухостой» и «Старый сухостой», а также пятый класс – «Фон» (деревья других пород и иные объекты земной поверхности), что позволило получить эталонные карты сегментации. Датасет представляет собой совокупность фрагментов исходных изображений размером $256 \times 256 \times 3$ пикселей и соответствующих им фрагментов эталонных карт. Он поделен на три выборки: обучающая – 2 004 фрагмента, валидационная – 672 фраг-

мента, и тестовая – 91 фрагмент. На основе этого датасета должен создаваться новый датасет, фрагменты которого имеют импульсные помехи, вносимые по описанной ниже методике.

Для оценки точности классификации деревьев пихты *A. sibirica* на изображениях с помощью исследуемых моделей Mo-U-Net будет использоваться метрика Intersection over Union (IoU) [14]. Для каждого класса c деревьев она может быть вычислена по формуле:

$$IoU_c = \frac{TP_c}{TP_c + FP_c + FN_c}, \quad (1)$$

где TP_c , FP_c и FN_c – количество, соответственно, истинно положительных, ложно положительных и ложно отрицательных решений для деревьев класса c . В наших исследованиях наряду с выражением (1) удобно использовать интегрированную метрику mean Intersection over Union (mIoU), рассчитываемую как среднее значение IoU_c по всем C классам. Отметим, что значения этих метрик, превышающие 0,5, соответствуют высокой точности классификации деревьев. Будем считать метрики IoU_c и mIoU мерой устойчивости каждой исследуемой модели к импульсным помехам: чем больше значения этих метрик, тем выше устойчивость модели.

2. Методика подготовки и проведения экспериментов

Рассмотрим методику подготовки и проведения экспериментов по исследованию устойчивости полносверточных моделей с различными РФП к импульсным помехам на изображениях. На этапе подготовки к тому или иному запланированному эксперименту на основе описанного выше датасета создается новый датасет. Все фрагменты его обучающей выборки будут иметь искусственным образом внесенные импульсные помехи, фрагменты валидационной и тестовой выборок не зашумляются. При зашумлении фрагментов этой выборки учитывается, какой эксперимент с моделями СНС запланирован, и в зависимости от этого задаются значения параметров зашумления каждого из фрагментов. Первым таким параметром является доля площади зашумления: отношение количества пикселей, значения которых будут изменены помехами, к общему количеству пикселей фрагмента изображения. Вторым задаваемым параметром служит амплитуда A помехи – число, указывающее, во сколько раз будет увеличено значение яркости зашумляемого пикселя для каждого из трех каналов RGB. Выбор координаты очередного пикселя фрагмента для зашумления производится по равномерному случайному закону распределения. Все сказанное позволяет считать, что новый датасет будет включать динамически формируемую в зависимости от предстоящего эксперимента обучающую выборку. При подготовке к экспериментам один параметр фиксируется, а второй изменяется в заданных пределах. Так будет сформировано четыре варианта нового датасета для проведения первой серии экспериментов, когда при фиксированной амплитуде $A = 1,5$ изменяется доля площади зашумления каждого фрагмента выборки: 0,15; 0,25; 0,35; 0,50. Для сравнения результатов по точности классификации деревьев с помощью моделей в условиях импульсных помех с результатами классификации изображений без таких помех должны быть проведены обучение, валидация и исследования моделей в случае фрагментов без помех, т.е. с использованием исходного датасета. Для второй серии экспериментов создается три варианта нового датасета, когда при постоянной доле площади зашумления фрагментов, равной 0,25, изменяется амплитуда помех: 1,5; 2,0; 3,0.

После этапа подготовки варианта нового датасета проводится собственно эксперимент. Он выполняется в два этапа. На первом происходят обучение и валидация модели с выбранной РФП, а на втором этапе – собственно исследование устойчивости обученной модели с использованием тестовой выборки датасета.

При обучении и валидации модели с заданной РФП изменяемым является гиперпараметр – коэффициент β для этой функции потерь, принимающий значения в диапазоне [0,001; 9,000]. Поиск квазиоптимального значения данного гиперпараметра осуществляется при обучении и валидации модели с помощью байесовского алгоритма оптимизации [16], при этом число используемых наборов гиперпараметров равно 100. В соответствии с процедурой из [13] квазиоптимальными являются те значения гиперпараметров, при которых модель показывает максимальное значение точности классификации

по метрике mIoU на валидационной выборке. Полученное таким образом квазиоптимальное значение коэффициента β затем используется в заданной РФП при проведении второго этапа эксперимента по исследованию устойчивости модели. Отметим, что такая процедура определения квазиоптимального значения коэффициента β должна проводиться для каждой РФП и в каждом из запланированных экспериментов.

3. Результаты исследований и их обсуждение

Нетрудно видеть, что проведение всех вышеперечисленных экспериментов с каждой из семи моделей СНС потребует значительных вычислительных ресурсов. Поэтому предлагается сначала провести эксперимент со всеми моделями с использованием нового датасета с наиболее часто встречающимися параметрами зашумления. Полученные результаты исследований на тестовой выборке позволят выбрать модель, наиболее устойчивую к импульсным помехам. Затем с этой моделью проводить все эксперименты при описанных выше в методике вариантах изменения параметров зашумления. Проведенный нами анализ имеющихся изображений хвойных лесов, полученных с помощью фотокамеры, установленной на БПЛА, показал, что весьма вероятным сценарием зашумления изображений при съемке является следующий: площадь каждого фрагмента обучающей выборки зашумлена на 25% (доля импульсных помех от общего числа пикселей фрагмента равна 0,25), а амплитуда A импульсных помех имеет значение 1,5.

Именно с этими значениями параметров были зашумлены фрагменты обучающей выборки нового датасета и были обучены и валидированы все семь моделей СНС. Затем для каждой из РФП путем выявления максимального значения метрики mIoU на валидационной выборке получено квазиоптимальное значение коэффициента β . Полученные значения коэффициента β для всех функций потерь приведены в табл. 2. Здесь же представлены результаты эксперимента в виде значений метрики mIoU на валидационной и тестовой выборках.

Таблица 2

Результаты валидации и тестирования моделей СНС с различными РФП, обученных на фрагментах с долей площади зашумления 0,25 и $A = 1,5$

Функция потерь	Коэффициент β	Выборка	mIoU
Коши	5,6900	Валидационная	0,6705
		Тестовая	0,5651
Рамсея	0,0010	Валидационная	0,7212
		Тестовая	0,6089
Хьюбера	1,7109	Валидационная	0,6980
		Тестовая	0,6023
Эндрюса	3,5703	Валидационная	0,6737
		Тестовая	0,5768
Geman-McCluer	3,9004	Валидационная	0,6938
		Тестовая	0,6149
Уэлша	4,2786	Валидационная	0,7011
		Тестовая	0,6194
Мешалкина	6,6236	Валидационная	0,6869
		Тестовая	0,5884

Из табл. 2 следует, что модель СНС с функцией Уэлша демонстрирует лучшее значение метрики mIoU на тестовой выборке относительно результатов для исследуемых моделей с другими функциями потерь. Поэтому далее эксперименты проводились в соответствии с изложенной методикой только с этой моделью. Для пояснения, как происходит поиск квазиоптимального значения коэффициента β , на рис. 1 показаны результаты поиска такого значения коэффициента β , при котором модель СНС с РФП Уэлша демонстрирует максимальное значение метрики mIoU на валидационной выборке. Для удобства представления и анализа этих кривых значения коэффициента β были нормализованы. Видим, что квазиоптимальное значение коэффициента β , равное 4,2786, было получено на 83-м наборе

гиперпараметров из 100 используемых наборов, когда метрика mIoU на валидационной выборке приняла максимальное значение, равное 0,7011.

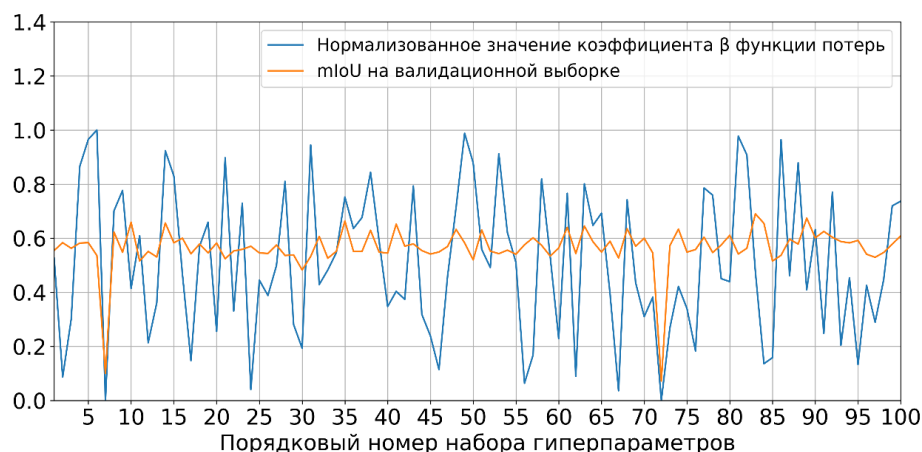


Рис. 1. Результаты поиска значений коэффициента β в случае функции Уэлша

Для исследования влияния площади зашумления импульсными помехами фрагментов обучающей выборки на точность классификации деревьев пихты с помощью модели с РПП Уэлша при постоянной амплитуде помех $A = 1,5$ была проведена первая серия экспериментов. В табл. 3 представлены результаты таких экспериментов. В ней показаны полученные квазиоптимальные значения коэффициента β функции Уэлша и значения метрики IoUc для каждого класса деревьев и метрики mIoU на валидационной и тестовой выборках.

Таблица 3

Точность классификации деревьев по метрикам IoUc и mIoU с помощью модели Mo-U-Net с РПП Уэлша в зависимости от площади зашумления обучающих фрагментов при постоянной амплитуде помех $A = 1,5$

Доля площади зашумления обучающих фрагментов	Коэффициент β	Выборка	IoUc					mIoU
			Живое	Отмирающее	Свеж. сухост.	Стар. сухост.	Фон	
0 (нет помех)	5,4457	Валидация	0,7651	0,6677	0,7914	0,7277	0,9113	0,7726
		Тестовая	0,6538	0,3267	0,7423	0,6202	0,8211	0,6328
0,15	7,9897	Валидация	0,7903	0,6698	0,7669	0,6974	0,9134	0,7676
		Тестовая	0,7023	0,2585	0,7244	0,5882	0,8272	0,6201
0,25	4,2786	Валидация	0,7283	0,5803	0,7031	0,6037	0,8903	0,7011
		Тестовая	0,6723	0,2980	0,7425	0,5694	0,8146	0,6194
0,35	8,2191	Валидация	0,5982	0,2978	0,6593	0,5942	0,8474	0,5994
		Тестовая	0,6183	0,2768	0,6469	0,5369	0,7939	0,5746
0,50	7,3545	Валидация	0,5132	0,2445	0,5978	0,5279	0,8036	0,5374
		Тестовая	0,5425	0,2827	0,7056	0,5235	0,7381	0,5585

Из полученных результатов следует вывод о том, что исследуемая модель по метрике mIoU на тестовой выборке демонстрирует плавное снижение устойчивости к импульсным помехам с ростом доли площади зашумления фрагментов обучающей выборки: устойчивость понижается на 2,0% при доле площади зашумления 0,15 и на 11,7% при максимальной доле площади зашумления, равной 0,50. При этом значение метрики mIoU равно 0,5585 даже при наличии помех на половине площади каждого фрагмента, что превышает порог 0,5 и указывает на довольно высокую точность классификации деревьев пихты. Отсюда следует вывод о возможности практического применения модели в лесной отрасли даже при наличии импульсных помех на половине площади анализируемого изображения. Аналогичные выводы можно сделать из анализа значений метрики IoUc на тестовой выборке для классов деревьев «Живое», «Свежий сухостой» и «Старый сухостой», а также для класса «Фон». Однако для промежуточного класса деревьев «Отмирающее» значения метрики IoUc на тестовой выборке значительно меньше порога 0,5 даже для фрагментов без помех и уменьшаются с ростом доли площади зашумления фрагментов. Такие результаты объясняются малым числом деревьев пихты на изображениях, исполь-

зуемых при формировании обучающей выборки (дисбаланс классов), а функция потерь Уэлша в отличие от известной двухпараметрической функции потерь Focal Loss, используемой при решении подобной задачи в [5], когда значение метрики IoUc в случае деревьев этого класса превышает порог 0,5, не позволяет снизить влияние такого дисбаланса. Это указывает на невозможность практического использования исследуемой модели в случае классификации деревьев данного класса.

Вторая серия экспериментов проводилась для исследования влияния амплитуды импульсных помех на точность классификации деревьев пихты с помощью модели с РФП Уэлша. При этом доля площади зашумления фрагментов не изменялась и была равна 0,25. Полученные результаты валидации и тестирования обученной модели по метрикам IoUc и mIoU представлены в табл. 4. Во втором столбце показаны также вычисленные и используемые в экспериментах квазиоптимальные значения коэффициента β .

Таблица 4

Значения метрик IoUc и mIoU в случаях разных амплитуд импульсных помех при постоянной доле площади зашумления фрагментов, равной 0,25

Амплитуда помех на обучающих фрагментах	Коэффициент β	Выборка	IoUc					mIoU
			Живое	Отмирающее	Свеж. сухост.	Стар. сухост.	Фон	
Без помех	5,4457	Валидационная	0,7651	0,6677	0,7914	0,7277	0,9113	0,7726
		Тестовая	0,6538	0,3267	0,7423	0,6202	0,8211	0,6328
1,5	4,2786	Валидационная	0,7283	0,5803	0,7031	0,6037	0,8903	0,7011
		Тестовая	0,6723	0,2980	0,7425	0,5694	0,8146	0,6194
2,0	2,2407	Валидационная	0,6635	0,4245	0,6698	0,5771	0,8683	0,6406
		Тестовая	0,6347	0,2758	0,7036	0,5129	0,7890	0,5832
3,0	4,0045	Валидационная	0,6574	0,4167	0,6690	0,6089	0,8686	0,6441
		Тестовая	0,6162	0,3744	0,7439	0,5874	0,7990	0,6242

Видим, что при наличии помех имеют место уменьшение значений метрик IoUc и mIoU и, соответственно, снижение устойчивости модели. Например, по метрике mIoU на тестовой выборке снижение на 2,1% для случая $A = 1,5$, на 7,8% для $A = 2,0$ и на 1,4% для $A = 3$. Незначительное снижение устойчивости модели в случае $A = 3$ по сравнению со случаями помех с амплитудами $A = 1,5$ и $A = 2,0$ происходит из-за того, что такой значительный рост яркости зашумленных пикселей позволяет модели лучше выделить детали крон деревьев пихты. По значениям метрики IoUc, превышающим порог 0,5, можно сделать вывод о ее довольно высокой устойчивости и практической применимости в условиях помех со значительными амплитудами в случае деревьев классов «Живое», «Свежий сухой» и «Старый сухой» и класса «Фон». К сожалению, для деревьев класса «Отмирающее» значения метрики IoUc на тестовой выборке для всех вариантов задания значений амплитуд помех значительно меньше порога 0,5, что указывает на низкую устойчивость модели к помехам и поэтому не позволит использовать ее на практике в случае классификации деревьев данного класса.

Заключение

Проведены исследования устойчивости семи полносверточных моделей, полученных на основе модели Mo-U-Net с использованием ряда известных РФП, при решении задачи классификации зашумленных импульсными помехами изображений деревьев пихты. Для этого создан датасет, фрагменты обучающей выборки которого имеют искусственным образом внесенные импульсные помехи. В соответствии с разработанной методикой подготовки и проведения исследований при зашумлении фрагментов учитывается, какой эксперимент с моделями запланирован, и в зависимости от этого задаются значения параметров импульсных помех. При обучении и валидации каждой из моделей при различных значениях амплитуды помех и доли площади зашумления фрагментов обучающей выборки осуществлен поиск квазиоптимальных значений коэффициентов β каждой функции потерь. Полученные квазиоптимальные значения коэффициентов β далее использованы при проведении исследований устойчивости моделей.

В результате исследований показано, что модель с функцией Уэлша демонстрирует лучшее значение метрики mIoU на тестовой выборке среди результатов для моделей с другими функциями потерь. Это позволило провести дальнейшие эксперименты в соответствии с предложенной методикой только с этой моделью. В первую очередь было исследовано влияние площади зашумления импульсными помехами фрагментов обучающей выборки на устойчивость модели с функцией Уэлша при постоянной амплитуде помех $A = 1,5$. При этом выявлено, что по метрике mIoU на тестовой выборке модель демонстрирует плавное снижение устойчивости к помехам с ростом доли площади зашумления фрагментов от 0,15 до 0,5 соответственно на 2,0–11,7%. Значение метрики mIoU равно 0,5585 даже при наличии помех на половине площади каждого фрагмента, что превышает порог 0,5 и указывает на высокую точность классификации деревьев пихты. Для исследования влияния амплитуды импульсных помех на устойчивость модели с функцией Уэлша проведена вторая серия экспериментов при постоянной доле площади зашумления фрагментов, равной 0,25. Из анализа значений метрики mIoU на тестовой выборке следует, что устойчивость модели снижается на 1,4–7,8% в зависимости от амплитуды помех, однако все значения метрики mIoU превышают порог 0,5.

Анализ полученных значений метрики IoUc на тестовой выборке для деревьев классов «Живое», «Свежий сухостой» и «Старый сухостой», а также для класса «Фон» показал, что наблюдается плавное снижение устойчивости модели при увеличении доли площади зашумления фрагментов или амплитуды помех, но все значения этой метрики превышают пороговое значение 0,5, и поэтому для этих классов модель может быть рекомендована для практического применения. Для деревьев класса «Отмирающее» значения этой метрики на тестовой выборке для всех вариантов задания в экспериментах значений амплитуд помех и долей зашумления фрагментов значительно меньше порога 0,5, что говорит о низкой устойчивости модели к помехам и, соответственно, не позволяет использовать ее на практике в случае классификации деревьев данного класса.

Полученные результаты исследования устойчивости семи полносверточных моделей позволили сформировать практически важные рекомендации специалистам лесной отрасли при нейросетевом анализе изображений, полученных при мониторинге хвойных лесов с помощью БПЛА.

Список источников

1. Chang W.Y., Lantz V.A., Hennigar C.R., MacLean D.A. Economic impacts of forest pests: a case study of spruce budworm outbreaks and control in New Brunswick, Canada // Canadian Journal of Forest Research. 2012. V. 42. P. 490–505. doi: 10.1139/x11-190
2. van Lierop P., Lindquist E., Sathyapala S., Franceschini G. Global forest area disturbance from fire, insect pests, diseases and severe weather events // Forest Ecology and Management. 2015. V. 352. P. 78–88. doi: 10.1016/j.foreco.2015.06.010
3. Musolin D., Kirichenko N., Karpun N., Mandelshtam M., Selikhovkin A., Zhuravleva E., Aksenenko E., Golub V., Kerchev I., Vasaitis R., Volkovitch M., Zhuravleva E., Selikhovkin A. Invasive pests of forests and urban trees in Russia: origin pathways, damage, and management // Forests. 2022. V. 13. Art. 521. doi: 10.3390/f13040521
4. Кривец С.А., Бисирова Э.М., Волкова Е.С., Дебков Н.М., Керчев И.А., Мельник М.А., Никифоров А.Н., Чернова Н.А. Технология мониторинга пихтовых лесов в зоне инвазии уссурийского полиграфа в Сибири. Томск : УМИУМ, 2018. 74 с.
5. Марков Н.Г., Мачука К. Модели и методы глубокого обучения для решения задач дистанционного мониторинга лесных ресурсов // Известия Томского политехнического университета. Инжиниринг георесурсов. 2024. Т. 335, № 6. С. 55–74. doi: 10.18799/24131830/2024/6/4600
6. Safonova A., Tabik S., Alcaraz-Segura D., Rubtsov A., Maglinets Y., Herrera F. Detection of fir trees (*Abies sibirica*) damaged by the Bark Beetle in unmanned aerial vehicle images with deep learning // Remote Sensing. 2019. V. 11. P. 643–662. doi: 10.3390/rs11060643
7. Керчев И.А., Маслов К.А., Марков Н.Г., Токарева О.С. Семантическая сегментация поврежденных деревьев пихты на снимках с беспилотных летательных аппаратов // Современные проблемы дистанционного зондирования Земли из космоса. 2021. Т. 18, № 1. С. 116–126. doi: 10.21046/2070-7401-2021-18-1-116-126
8. Adachi K., Yamaguchi S. Learning Robust Convolutional Neural Networks with Relevant Feature Focusing Via Explanations // 2022 IEEE International Conference on Multimedia and Expo (ICME). IEEE, 2022. P. 1–6.
9. Сивак М.А. Робастное обучение нейронных сетей с простой архитектурой для решения задач классификации : дис. ... канд. техн. наук : 05.13.17. Новосибирск, 2022. 117 с.
10. Fanjun L., Ying L. Robust echo state network with Cauchy loss function and hybrid regularization for noisy time series prediction // Applied Soft Computing. 2023. V. 146. Art. 110640.
11. Li X., Lu Q., Dong Y., Tao D. Robust subspace clustering by cauchy loss function // IEEE Transactions on Neural Networks and Learning Systems. 2018. V. 30 (7). P. 2067–2078.

12. Li X., Xia Z., Zhang H. Cauchy activation function and XNet // arXiv preprint arXiv: 2409.19221. 2024. URL: <https://arxiv.org/abs/2409.19221>
13. Малкин А.Ю., Марков Н.Г. Помехоустойчивость модели полносверточной нейронной сети U-Net при семантической сегментации деревьев пихты на зашумленных снимках с БПЛА // Доклады ТУСУР. 2024. Т. 27, № 2. С. 64–70. doi: 10.21293/1818-0442-2024-27-2-64-70
14. Ronneberger O., Fischer P., Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation // Medical Image Computing and Computer-Assisted Intervention. 2015. V. 9351. P. 234–241. doi: 10.1007/978-3-319-24574-4_28
15. Марков Н.Г., Маслов К.А., Керчев И.А., Токарева О.С. Модели U-Net для семантической сегментации поврежденных деревьев сосны сибирской кедровой на снимках с БПЛА // Современные проблемы дистанционного зондирования Земли из космоса. 2022. Т. 19, № 1. С. 65–77. doi: 10.21046/2070-7401-2022-19-1-65-77
16. Wu J., Chen X.Y., Zhang H., Xiong L.D., Lei H., Deng S.H. Hyperparameter optimization for machine learning models based on Bayesian optimization // Journal of Electronic Science and Technology. 2019. Vol. 17 (1). P. 26–40.

References

1. Chang, W.Y., Lantz, V.A., Hennigar, C.R. & MacLean, D.A. (2012) Economic impacts of forest pests: a case study of spruce budworm outbreaks and control in New Brunswick, Canada. *Canadian Journal of Forest Research*. 42. pp. 490–505. doi: 10.1139/x11-190
2. Van Lierop, P., Lindquist, E., Sathyapala, S. & Franceschini, G. (2015) Global forest area disturbance from fire, insect pests, diseases and severe weather events. *Forest Ecology and Management*. 352. pp. 78–88. doi: 10.1016/j.foreco.2015.06.010
3. Musolin, D., Kirichenko, N., Karpun, N., Mandelshtam, M., Selikhovkin, A., Zhuravleva, E. Aksechenko, E., Golub, V., Kerchev, I., Vasaitis, R., Volkovitch, M., Zhuravleva, E. & Selikhovkin, A. (2022) Invasive pests of forests and urban trees in Russia: origin pathways, damage, and management. *Forests*. 13. p. 521. doi: 10.3390/f13040521
4. Krivets, S.A., Bisirova, E.M., Volkova, E.S., Debkov, N.M., Kerchev, I.A., Melnik, M.A. Nikiforov, A.N. & Chernova, N.A. (2018) *Tekhnologiya monitoringa pikhtovykh lesov v zone invazii ussuriyskogo poligrafa v Sibiri* [Monitoring technology of fir forests in the zone of Ussurian polygraph invasion in Siberia]. Tomsk: UMIUM.
5. Markov, N.G. & Machuka, K. (2024) Models and methods of deep learning for remote monitoring of forest resources. *Izvestiya TPU. Inzhiniring georesurov*. 335(6). pp. 55–74. doi: 10.18799/24131830/2024/6/4600
6. Safonova, A., Tabik, S., Alcaraz-Segura, D., Rubtsov, A., Maglinets, Y. & Herrera, F. (2019) Detection of fir trees (*Abies sibirica*) damaged by the bark beetle in UAV images with deep learning. *Remote Sensing*. 11. p. 643. doi: 10.3390/rs11060643
7. Kerchev, I.A., Maslov, K.A., Markov, N.G. & Tokareva, O.S. (2021) Semantic segmentation of damaged fir trees in UAV imagery. *Sovremennye problemy distantsionnogo zondirovaniya Zemli iz kosmosa*. 18(1). pp. 116–126. doi: 10.21046/2070-7401-2021-18-1-116-126
8. Adachi, K. & Yamaguchi, S. (2022) Learning robust convolutional neural networks with relevant feature focusing via explanations. *Proc. IEEE Int. Conf. on Multimedia and Expo (ICME)*. IEEE. pp. 1–6.
9. Sivak, M.A. (2022) *Robustnoe obuchenie neyronnykh setey s prostoy arkhitekturoy dlya resheniya zadach klassifikatsii* [Robust Training of Neural Networks with Simple Architecture for Solving Classification Problems]. Engineering Cand. Diss. Novosibirsk: NSTU.
10. Li, F. & Li, Y. (2023) Robust echo state network with Cauchy loss function and hybrid regularization for noisy time series prediction. *Applied Soft Computing*. 146. Art. 110640.
11. Li, X., Xia, Z. & Zhang, H.C. (2018) Robust subspace clustering by Cauchy loss function. *IEEE Transactions on Neural Networks and Learning Systems*. 30(7). pp. 2067–2078.
12. Li, X., Xia, Z. & Zhang, H.C. (2024) Cauchy activation function and XNet. *arXiv preprint, arXiv:2409.19221*.
13. Malkin, A.Yu. & Markov, N.G. (2024) Pomexoustoychivost' modeli polnosvertochnoy neyronnoy seti U-Net pri semanticheskoy segmentatsii derev'ev pikhty na zashumlennykh snimkakh s BPLA [Noise Immunity of the U-Net Fully Convolutional Neural Network Model in Semantic Segmentation of Fir Trees on Noisy UAV Images]. *Doklady TUSUR*. 27(2). pp. 64–70. doi: 10.21293/1818-0442-2024-27-2-64-70
14. Ronneberger, O., Fischer, P. & Brox, T. (2015) U-Net: Convolutional networks for biomedical image segmentation. *Medical Image Computing and Computer-Assisted Intervention*. Art. 9351. doi: 10.1007/978-3-319-24574-4_28
15. Markov, N.G., Maslov, K.A., Kerchev, I.A. & Tokareva, O.S. (2022) Modeli U-Net dlya semanticheskoy segmentatsii povrezhdennykh derev'ev sosny sibirskoy kedrovoy na snimkakh s BPLA [U-Net Models for Semantic Segmentation of Damaged Siberian Pine Trees on UAV Images]. *Sovremennye problemy distantsionnogo zondirovaniya Zemli iz kosmosa*. 19(1). pp. 65–77. doi: 10.21046/2070-7401-2022-19-1-65-77
16. Wu, J. et al. (2019) Hyperparameter optimization for machine learning models based on Bayesian optimization. *Journal of Electronic Science and Technology*. 17(1). pp. 26–40.

Информация об авторах:

Малкин Артем Юрьевич – аспирант Отделения информационных технологий инженерной школы информационных технологий и робототехники Национального исследовательского Томского политехнического университета (Томск, Россия). E-mail: malkin@tpu.ru

Марков Николай Григорьевич – профессор, доктор технических наук, профессор Отделения информационных технологий инженерной школы информационных технологий и робототехники Национального исследовательского Томского политехнического университета (Томск, Россия). E-mail: markovng@tpu.ru

Вклад авторов: все авторы сделали эквивалентный вклад в подготовку публикации. Авторы заявляют об отсутствии конфликта интересов.

Information about the authors:

Malkin Artem Yu. (Post-Graduate Student, National Research Tomsk Polytechnic University, Tomsk, Russian Federation). E-mail: malkin@tpu.ru

Markov Nikolai G. (Professor, Doctor of Technical Sciences, National Research Tomsk Polytechnic University, Tomsk, Russian Federation). E-mail: markovng@tpu.ru

Contribution of the authors: the authors contributed equally to this article. The authors declare no conflicts of interests.

Поступила в редакцию 02.06.2025; принята к публикации 02.12.2025

Received 02.06.2025; accepted for publication 02.12.2025

Научная статья

УДК 519.248

doi: 10.17223/19988605/73/5

Нестационарные потоки в СеМО без очереди и с детерминированным временем обслуживания

Гурами Шалвович Цициашвили

Институт прикладной математики Дальневосточного отделения РАН, Владивосток, Россия, guram@iam.dvo.ru

Аннотация. Строится математическая модель ациклической сети массового обслуживания с нестационарным пуассоновским входным потоком, без очереди и с детерминированным временем обслуживания. Вычисляются нестационарные интенсивности потоков, проходящих по сети. Доказывается, что если входной поток является пуассоновским, то все остальные потоки, проходящие по сети, также являются пуассоновскими. Причем количество заявок, находящихся в каждом узле сети, также имеет пуассоновское распределение. С помощью специальных интегральных соотношений вычисляются параметры этих пуассоновских распределений.

Ключевые слова: пуассоновский поток; ациклическая сеть; ориентированный граф; теорема о раскрасивании; максимальная длина пути в графе.

Благодарности: Работа выполнена в рамках государственного задания Института прикладной математики ДВО РАН № 075-00459-25-00.

Для цитирования: Цициашвили Г.Ш. Нестационарные потоки в СеМО без очереди и с детерминированным временем обслуживания // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2025. № 73. С. 41–45. doi: 10.17223/19988605/73/5

Original article

doi: 10.17223/19988605/73/5

Non-stationary flows in queuing networks without queue and with deterministic service time

Gurami Sh. Tsitsiashvili

Institute for Applied Mathematics, Far Eastern Branch of RAS, Vladivostok, Russian Federation, guram@iam.dvo.ru

Abstract. The paper builds a mathematical model of an acyclic queuing network with an unsteady Poisson input flow, no queue, and deterministic service time. The non-stationary intensities of the flows passing through the network are calculated. It is proving that if the input flow is Poisson, then all other flows passing through the network are also Poisson. Moreover, the number of customers located in each node of the network also has a Poisson distribution. The parameters of these Poisson distributions are calculated using special integral relations.

Keywords: Poisson flow; acyclic network; directed graph; coloring theorem; maximum path length in graph.

Acknowledgments: The research was carried out within the state assignment for Institute for Applied Mathematics FEB of the RAS N 075-00459-25-00.

For citation: Tsitsiashvili, G.Sh. (2025) Non-stationary flows in queuing networks without queue and with deterministic service time. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naja tehnika i informatika – Tomsk State University Journal of Control and Computer Science*. 73. pp. 41–45. doi: 10.17223/19988605/73/5

Введение

Наряду с классическими мультипликативными теоремами для открытых [1] и замкнутых [2] сетей массового обслуживания (СеМО) в последнее время проведено много их обобщений (см., напр.: [3]). Эти обобщения представляют интерес, если они связаны с новыми практическими приложениями, например при введении в модели СеМО блокирующих вероятностей [4, 5]. Настоящая работа вызвана появлением новых интересных имитационных моделей СеМО в морском транспорте (см., напр.: [6]). Такие модели приводят к новым задачам, требующим для своего решения новых математических приемов.

Однако наряду с определением стационарных характеристик в СеМО требуется рассматривать обслуживание нестационарных входных потоков. Обслуживание нестационарных потоков связано с так называемыми трамповыми перевозками (tramp shipping), т.е. с нерегулярной морской перевозкой грузов. Причем нестационарность входных потоков интересна в том смысле, что позволяет описывать возникновение пробок в СеМО. Поэтому исследование СеМО с нестационарными входными потоками (см., напр.: [7]) является важной и достаточно сложной аналитической и вычислительной задачей. Для получения содержательных решений здесь в первую очередь требуется включение ограничений и дополнительных условий, наиболее часто встречающихся в приложениях и допускающих сравнительно простые решения.

К таким условиям можно отнести отсутствие очереди (при наличии крупных обрабатывающих узлов сети), детерминированное распределение времени обслуживания и ацикличность сети перевозки грузов. С одной стороны, эти условия понятны для эксплуатационников и логистиков. С другой стороны, они достаточно удобны при рассмотрении пуассоновских нестационарных потоков, проходящих через узлы ациклической сети.

Для ациклических сетей с нестационарным входным пуассоновским потоком, детерминированным временем обслуживания и отсутствием очереди строится рекуррентная процедура вычисления интенсивностей пуассоновских потоков, выходящих из узлов сети, и параметров пуассоновских распределений числа заявок в различные моменты времени.

Она основана на выделении наборов узлов U_k сети с заданной максимальной длиной пути k (числом ребер в пути из начальной вершины в другую вершину сети) на каждом шаге. Такая классификация узлов сети приводит к тому, что в каждый узел сети из набора U_k входят ребра только из узлов, содержащихся в наборах $U_{k'}, k' < k$. Это позволяет последовательно по $k \geq 1$ устанавливать пуассоновость и независимость нестационарных потоков, выходящих из узлов набора U_k , и вычислять нестационарную интенсивность этих потоков. Детерминированность времени обслуживания заявок из узлов сети также является источником достаточно простых интегральных формул для определения нестационарных параметров пуассоновских распределений числа заявок в узлах сети.

1. Вычисление максимальных длин путей в ациклических ориентированных графах

Рассмотрим ациклический ориентированный граф (орграф) G с множеством вершин $U = \{1, \dots, n\}$ и множеством ребер V . Полагаем, что в графе G для любой вершины $i \in U$ существует путь из вершины 1 в вершину i . Для каждой вершины i графа G определим максимальную длину пути $l(i)$ из вершины 1 в вершину i , $l(1) = 0$, и положим $S = \max_{i \in U} l(i)$.

Для конструктивного вычисления $l(i)$, $i \in U$, введем матрицу $D^1 = \|d_{ij}^1\|_{i,j=1}^n$, где $d_{ii}^1 = 0$, $i = 1, \dots, n$, $d_{ij}^1 = \infty$, если $(i, j) \notin V$, $d_{ij}^1 = 1$, если $(i, j) \in V$. Тем самым для любой пары вершин, которые не соединяются ребром (путем длины единица), величина $d_{ij}^1 = \infty$.

Построим аналог алгоритма Флойда–Уоршелла [8] для нахождения матрицы максимальных длин всех путей между вершинами ациклического орграфа. Обозначим $D^k = \|d_{ij}^k\|_{i,j=1}^n$, $k = 1, \dots, K$, где

величина $d_{ij}^k = \infty$, если в графе G нет пути, проходящего только через вершины $1, \dots, k$ и соединяющего вершины i, j . Если такие пути существуют, то величина d_{ij}^k равна максимальной длине таких путей. Тогда справедлива следующая теорема [9].

Теорема 1. Матрицы $D^k = \|d_{ij}^k\|_{i,j=1}^n, k = 2, \dots, K$, удовлетворяют соотношениям

$$d_{ij}^k = \max(d_{ij}^{k-1}, d_{ik}^{k-1} + d_{kj}^{k-1}), \text{ если } \max(d_{ij}^{k-1}, d_{ik}^{k-1} + d_{kj}^{k-1}) < \infty, \quad (1)$$

иначе

$$d_{ij}^k = \min(d_{ij}^{k-1}, d_{ik}^{k-1} + d_{kj}^{k-1}). \quad (2)$$

С помощью теоремы 1 можно вычислить максимальные длины путей из вершины 1 в вершины $i \in U$, полагая $l(i) = d_{1i}^n$.

2. Построение нестационарной модели прохождения пуассоновского потока через СеМО

Вычисление переходных интенсивностей. Определим подмножества $L_s = \{i : l(i) = s\}, s = 0, 1, \dots, S$.

Тогда при $r > k, i \in L_k, j \in L_r$ граф G не содержит ребра (i, j) и, значит, в ациклическом орграфе любое ребро $(j, i) \in V, i \in L_k, j \in L_r$ может принадлежать графу G только при $k > r$. Обозначим $U_k = \bigcup_{s=0}^k L_s$

и положим V_k совокупность ребер (i, j) графа G таких, что $i \in U_k, j \notin U_k$. Очевидно, что $\bigcup_{k=0}^{K-1} V_k = V$.

В качестве примера приведем орграф с $U = \{1, 2, 3, 4, 5\}$ (рис 1).

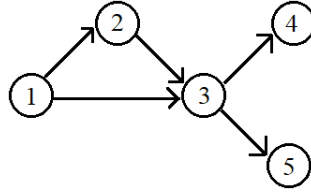


Рис. 1. Пример орграфа G
Fig. 1. Example of digraph G

В этом случае $l(1) = 0, l(2) = 1, l(3) = 2, l(4) = l(5) = 3; L_0 = \{1\}, L_1 = \{2\}, L_2 = \{3\}, L_3 = \{4, 5\}; U_0 = \{1\}, U_1 = \{1, 2\}, U_2 = \{1, 2, 3\}; V_0 = \{(1, 2), (1, 3)\}, V_1 = \{(1, 3), (2, 3)\}, V_2 = \{(3, 4), (3, 5)\}$.

Пусть каждому ребру $(j, i) \in V$ сопоставляется неотрицательная кусочно непрерывная функция $\lambda_{ji}(t)$ от времени t . Эта функция определяет нестационарную интенсивность пуассоновского потока, проходящего по ребру $(j, i) \in V$. Обозначим $\underline{\Lambda}_1(t)$ неотрицательную кусочно непрерывную функцию, являющуюся интенсивностью входного пуассоновского потока в ациклическую сеть. Потребуем, чтобы при некотором $a_i > 0$ выполнялись равенства

$$\underline{\Lambda}_i(t) = \sum_{(j,i) \in V} \lambda_{ji}(t), i \in U, i \neq 1, \bar{\Lambda}_i(t) = \underline{\Lambda}_i(t - a_i), i \in U. \quad (3)$$

Величина $a_i > 0$ определяет сдвиг вправо каждой точки входного потока, т.е. время задержки (обслуживания каждого требования) в вершине i сети. Предположим, что для каждой вершины $i \in U$ задан набор положительных чисел $\{\theta_{ij} : (i, j) \in V\}$: $\sum_{(i,j) \in V} \theta_{ij} = 1$, и выполняются равенства

$$\lambda_{ij}(t) = \theta_{ij} \bar{\Lambda}_i(t), (i, j) \in V. \quad (4)$$

Здесь θ_{ij} определяют вероятности, с которыми каждая заявка, выходящая из вершины $i \in U$, поступает в вершину $j \in U$.

Теорема 2. По изначально заданным функции $\underline{\Lambda}_1(t)$, наборам $\{\theta_{ij} : (i, j) \in V\}$ и числам $a_i, i \in U$, можно с помощью равенств (3), (4) однозначно вычислить функции $\lambda_{ij}(t), (i, j) \in V$.

Доказательство. Доказательство проведем индукцией по ребрам $V_k, k = 1, \dots, S-1$. При $k = 1$ это утверждение вытекает из равенств (3), (4). Предположим, что утверждение теоремы 2 выполняется при некотором $k \geq 1$ и, следовательно, заданы $\lambda_{ij}(t), (i, j) \in V_k$. Тогда, используя равенства (3), (4), можно вычислить сначала $\bar{\Lambda}_i(t)$, а затем $\lambda_{ij}(t), (i, j) \in V_{k+1}$. Корректность такого определения связана с тем, что ребра сети направлены из вершин множества U_k в вершины множества $U_r, k < r$, с большим нижним индексом. Это утверждение иллюстрируется рис. 1.

Определение пуассоновских потоков в сети. Пусть задан пуассоновский поток точек \underline{T}_1 с интенсивностью $\underline{\Lambda}_1(t)$, тогда сдвигом всех точек этого потока вправо на величину a_1 можно получить пуассоновский поток интенсивности \bar{T}_1 . Если считать, что точки пуассоновского потока \underline{T}_1 соответствуют моментам прихода заявок в узел 1 сети, то тогда соответствующие им (с тем же номером) точки потока \bar{T}_1 соответствуют моментам ухода заявок входного потока из узла 1. Далее предположим, что каждая точка выходного потока \bar{T}_1 с вероятностью θ_{1j} поступает в поток T_{1j} , следующий по ребру $(1, j) \in V_1$. Тогда вследствие теоремы о раскрашивании [10] потоки $T_{1j}, (1, j) \in V_1$, являются независимыми и имеют интенсивности $\lambda_{1j}(t)$. Те из потоков $T_{1j}, (1, j) \in V_1$, которые соединяются в узлах $i \in U_2$, образуют пуассоновские потоки $\underline{T}_i, i \in V_2$, и имеют интенсивности $\underline{\Lambda}_i(t), i \in U_2$. Причем пуассоновские потоки $T_{1j}, (1, j) \in V_1, j \notin U_2$, и пуассоновские потоки $\underline{T}_i, i \in V_2$, независимы. Продолжая индукцией по k доказывать пуассоновость и независимость потоков $\underline{T}_i, i \in V_k$, и пуассоновских потоков $T_{ij}, (i, j) \in V_k, j \notin U_k$, можно определить пуассоновские потоки, проходящие по всем ребрам сети G . Интенсивности всех этих потоков совпадают с $\underline{\Lambda}_i(t), i \in V_k, \lambda_{ij}(t), (1, j) \in V_k, j \notin U_k$.

Определение случайного числа точек в узлах сети. Известно, что точки пуассоновского потока \underline{T}_i , входящего в узел $i \in V_k$, образуют моменты прихода заявок в этот узел. Поскольку время пребывания каждой заявки равно a_i , то можно утверждать, что случайное число этих заявок в момент t имеет пуассоновское распределение с параметром $\int_{t-a_i}^t \underline{\Lambda}_i(\tau) d\tau = F_i(t)$. Таким образом, появляется возможность определить не только интенсивности нестационарных пуассоновских потоков, проходящих по ребрам орграфа G , но и зависимость от времени t параметров пуассоновских распределений числа заявок в узлах сети.

Заключение

Таким способом для ациклического орграфа G построена вероятностная модель прохождения входного пуассоновского потока заявок \underline{T}_1 через СеМО. Определены интенсивности нестационарных пуассоновских потоков, входящих в узлы и выходящих из узлов сети. Вычислены параметры пуассоновских распределений числа заявок в узлах сети в нестационарном режиме. Все математические построения основаны на алгоритме вычисления длины максимального пути из начальной вершины ациклического орграфа в остальные вершины и на выделении в орграфе наборов вершин с фиксированной длиной максимального пути. Далее строится методом математической индукции алгоритм вычисления интенсивностей потоков, входящих в узлы и выходящих из узлов сети. Корректность этих вычислений определяется независимостью пуассоновских потоков с вычисленными интенсивностями. Наконец, по интенсивностям входных пуассоновских потоков в узлы сети и по времени пребывания заявок в этих узлах находятся формулы для вычисления параметров пуассоновских распределений числа заявок в узлах сети.

Список источников

1. Jackson J.R. Networks of Waiting Lines // *Oper. Res.* 1957. V. 5 (4). P. 518–521.
2. Gordon K.D., Newell G.F. Closed Queuing Systems with Exponential Servers // *Oper. Research.* 1967. V. 15 (2). P. 254–265.
3. Basharin G.P., Gaidamaka Yu.V., Samouylov K.E. Mathematical Theory of Teletraffic and Its Applications to the Analysis of Multiservice Communication of Next Generation Networks // *Autom. Control. Comput. Sci.* 2013. V. 47. P. 62–69. doi: 10.3103/S0146411613020028
4. Balsamo S., De Nitto V. A survey of Product-form Queuing Networks with Blocking and their Equivalences // *Annals of Operations Research.* 1998. V. 79. P. 97–117.
5. Boucherie R.J., van Dijk N.M. On the arrival theorem for product form queuing networks with blocking // *Performance Evaluation.* 1997. Vol. 29 (3). P. 155–176.
6. Лемперт А.А., Жарков М.Л., Казаков А.Л., Ву Х.З. Моделирование морского контейнерного терминала с использованием сети массового обслуживания // *Управление большими системами.* 2024. Вып. 112. С. 310–337.
7. Ивницкий В.А. Теория произвольного входящего потока с применениями к нестационарным системам и сетям массового обслуживания. Saarbrücken : LAP Lambert, 2012. 640 с.
8. Кормен Т., Лейзерсон Ч., Ривест Р., Штайн К. Алгоритмы: построение и анализ. М. : Вильямс, 2011. 1296 с.
9. Цициашвили Г.Ш., Осипова М.А. Стационарные потоки в ациклических сетях массового обслуживания // *Дальневосточный математический журнал.* 2016. Т. 16, вып. 2. С. 223–228.
10. Кингман Дж. Пуассоновские процессы. М. : МЦНМО, 2007. 136 с.

References

1. Jackson, J.R. (1957) Networks of Waiting Lines. *Oper. Research.* 5(4). pp. 518–521.
2. Gordon, K.D. & Newell, G.F. (1967) Closed Queuing Systems with Exponential Servers. *Oper. Research.* 15(2). pp. 254–265.
3. Basharin, G.P., Gaidamaka Yu.V. & Samouylov K.E. (2013) Mathematical Theory of Teletraffic and Its Applications to the Analysis of Multiservice Communication of Next Generation Networks. *Automation Control and Computer Science.* 47. pp. 62–69. doi: 10.3103/S0146411613020028
4. Balsamo, S. & De Nitto, V. (1998) A survey of Product-form Queuing Networks with Blocking and their Equivalences. *Annals of Operations Research.* 79. pp. 97–117.
5. Boucherie, R.J. & van Dijk, N.M. (1997) On the arrival theorem for product form queuing networks with blocking. *Performance Evaluation.* 29(3). pp. 155–176.
6. Lempert, A.A., Zharkov, M.L., Kazakov, A.L. & Vu, Kh.Z. (2024) Modelirovanie morskogo konteynernogo terminala s ispol'zovaniem seti massovogo obsluzhivaniya [Modeling of a Sea Container Terminal Using a Queuing Network]. *Upravlenie bol'shimi sistemami.* 112. pp. 310–337.
7. Ivitskiy, V.A. (2012) *Teoriya proizvol'nogo vkhodyashchego potoka s primeneniymi k nestatsionarnym sistemam i setyam massovogo obsluzhivaniya* [Theory of an Arbitrary Input Flow with Applications to Non-Stationary Systems and Queuing Networks]. Saarbrücken: LAP Lambert.
8. Cormen, T.H., Leiserson, Ch.E., Rivest, R.L. & Shtayn, K. (2011) *Algoritmy: postroeniye i analiz* [Algorithms: Construction and Analysis]. Translated from English. Moscow: Vil'yams.
9. Tsitsiashvili, G.Sh. & Osipova, M.A. (2016) Statsionarnye potoki v atsiklicheskikh setyakh massovogo obsluzhivaniya [Stationary Flows in Acyclic Queuing Networks]. *Dal'nevostochnyy matematicheskiy zhurnal.* 16(2). pp. 223–228.
10. Kingman, J. (2007) *Puassonovskie protsessy* [Poisson Processes]. Translated from English. Moscow: MTsNMO.

Информация об авторе:

Цициашвили Гурами Шалвович – профессор, доктор физико-математических наук, главный научный сотрудник Института прикладной математики Дальневосточного отделения РАН (Владивосток, Россия). E-mail: guram@iam.dvo.ru

Автор заявляет об отсутствии конфликта интересов.

Information about the author:

Tsitsiashvili Gurami Sh. (Professor, Doctor of Physical and Mathematical Sciences, Main Researcher of the Institute for Applied Mathematics, Far Eastern Branch of RAS, Vladivostok, Russian Federation). E-mail: guram@iam.dvo.ru

The author declares no conflicts of interests.

Поступила в редакцию 23.06.2025; принята к публикации 02.12.2025

Received 23.06.2025; accepted for publication 02.12.2025

ОБРАБОТКА ИНФОРМАЦИИ

DATA PROCESSING

Научная статья

УДК 519.862.6

doi: 10.17223/19988605/73/6

**О связи между многослойной модульной регрессионной моделью
и регрессией в виде производственной функции Леонтьева****Михаил Павлович Базилевский***Иркутский государственный университет путей сообщения, Иркутск, Россия, mik2178@yandex.ru*

Аннотация. Предложена модель многослойной модульной регрессии, отличающаяся от известной спецификации тем, что в ней выходное значение с каждого слоя входит не только в модульную, но и в линейную часть последующего слоя. Задача оценки неизвестных параметров предложенной модели с помощью метода наименьших модулей сведена к задаче частично булевого линейного программирования. Доказано, что для любого порядка переменных качество аппроксимации предложенной модели всегда не хуже, чем качество линейной регрессии и производственной функции Леонтьева. Показано, что любая вложенная кусочно-линейная регрессия является разновидностью многослойной модульной регрессии. Проведены вычислительные эксперименты, подтверждающие корректность доказанной теоремы.

Ключевые слова: регрессионный анализ; многослойная модульная регрессия; производственная функция Леонтьева; метод наименьших модулей; задача частично-булевого линейного программирования.

Для цитирования: Базилевский М.П. О связи между многослойной модульной регрессионной моделью и регрессией в виде производственной функции Леонтьева // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2025. № 73. С. 46–57. doi: 10.17223/19988605/73/6

Original article

doi: 10.17223/19988605/73/6

**On the relationship between the multilayer modular regression model
and regression in the form of Leontief's production function****Mikhail P. Bazilevskiy***Irkutsk State Transport University, Irkutsk, Russian Federation, mik2178@yandex.ru*

Abstract. The article proposes a model of multilayer modular regression, which differs from the known specification in that the output value from each layer is included not only in the modular part, but also in the linear part of the subsequent layer. The problem of estimating unknown parameters of the proposed model using the least absolute deviations method is reduced to the problem of mixed 0-1 integer linear programming. It has been proved that, for any order of variables, the approximation quality of the proposed model is always better than the quality of linear regression or the Leontief production function. Any nested piecewise linear regression can be considered as a type of multilayer modular regression. Computational experiments have been carried out to confirm the correctness of this theorem.

Keywords: regression analysis; multilayer modular regression; Leontief production function; least absolute deviations method; mixed 0-1 integer linear programming problem.

For citation: Bazilevskiy, M.P. (2025) On the relationship between the multilayer modular regression model and regression in the form of Leontief's production function. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naja tehnika i informatika – Tomsk State University Journal of Control and Computer Science*. 73. pp. 46–57. doi: 10.17223/19988605/73/6

Введение

Решение многих научных проблем в настоящее время сопряжено с установлением математической формы связи между выходной переменной и одной или несколькими входными переменными по имеющимся в наличии статистическим данным. Для построения таких зависимостей исследователи часто применяют инструменты регрессионного анализа [1, 2]. Так, например, в [3] с помощью регрессионного анализа установлено влияние на среднемесячный уровень воды озера Ловозеро трех предикторов – среднемесячных расходов воды притока, общей облачности и облачности нижнего яруса. В [4] построена зависимость валовой продукции сельского хозяйства Якутии от поголовья крупного рогатого скота и лошадей, площади пастбищ и сенокосов, количества сельскохозяйственных организаций и численности сельского населения. В [5] получена регрессионная модель удельных энергозатрат на тягу поездов от шести факторов – объемов перевозочной работы, эксплуатационной длины электрифицированных линий, средней массы поезда, среднесуточной производительности локомотива, средней участковой скорости, доли грузовой работы.

Объединяет работы [3–5] то, что в них для проведения регрессионного анализа выбрана модель множественной линейной регрессии, параметры которой довольно просто оценить с помощью метода наименьших квадратов (МНК). Известно, что МНК можно применять и при нелинейных преобразованиях входящих в регрессию переменных, например с помощью элементарных функций. Так, в [6, 7] с помощью МНК оцениваются параметры полиномиальной регрессии, а в [8, 9] – параметры степенной регрессии, также известной как производственная функция (ПФ) Кобба–Дугласа.

Представленные в работах [6–9] нелинейные регрессионные зависимости хорошо изучены и применяются в исследованиях не один десяток лет. Помимо МНК оценка неизвестных параметров регрессионных моделей может осуществляться методом наименьших модулей (МНМ), реализация которого возможна с помощью описанного в [10] алгоритма «итеративного МНК» (метод вариационно-взвешенных квадратичных приближений). В большинстве же исследований МНМ реализуется с использованием аппарата линейного программирования, что стало возможным благодаря работе [11]. А в результате развития аппаратных и программных средств решения задач частично целочисленного линейного программирования [12] на сегодняшний день разработаны и другие нелинейные спецификации регрессионных моделей, с помощью которых можно строить довольно точные математические зависимости с интересными свойствами.

К первому классу таких нелинейных спецификаций относятся модели, фундаментом для которых послужила ПФ Леонтьева [13], содержащая мультиарную операцию \min . В монографии [14] задача МНМ-оценки таких регрессий сведена к задаче частично булевого линейного программирования (ЧБЛП). Позднее эти модели за счет использования мультиарных операций \max и вложенности функций были обобщены и получили название «вложенные кусочно-линейные регрессии» (см., напр.: [15]). Ко второму классу нелинейных спецификаций относятся модели, содержащие операцию «модуль». Впервые задача их МНМ-оценки была сведена к задаче ЧБЛП в работе [16]. Позднее появились многослойные модульные регрессии [17], которые также называют «глубокие модульные регрессии», и «широкие модульные регрессии», представленные в статье [18]. Связь между вложенными кусочно-линейными и многослойными модульными регрессиями до сих пор не исследовалась.

Известно, что для любых чисел a и b бинарные операции \min и \max связаны с операцией модуль следующими соотношениями:

$$\min\{a, b\} = \frac{a + b - |a - b|}{2}, \quad (1)$$

$$\max\{a, b\} = \frac{a + b + |a - b|}{2}. \quad (2)$$

Выражения (1) и (2) позволяют предположить, что класс многослойных модульных регрессий включает в себя класс вложенных кусочно-линейных регрессий. Цель данной статьи состоит в том, чтобы доказать выдвинутое предположение на примере ПФ Леонтьева.

1. Исследование связи между ПФ Леонтьева и многослойной модульной регрессией

Пусть имеется выборка объема n , содержащая значения y_i , $i = \overline{1, n}$, выходной (объясняемой) переменной y и значения x_{ij} , $i = \overline{1, n}$, $j = \overline{1, l}$, l входных (объясняющих) переменных x_1, x_2, \dots, x_l . Рассмотрим модель множественной линейной регрессии с неизвестными параметрами $\alpha_0, \alpha_1, \dots, \alpha_l$:

$$y_i = \alpha_0 + \sum_{j=1}^l \alpha_j x_{ij} + \varepsilon_i, \quad i = \overline{1, n}, \quad (3)$$

где ε_i , $i = \overline{1, n}$, – ошибки регрессии.

Будем считать, что неизвестные параметры линейной регрессии (3) оцениваются с помощью МНМ, который состоит в минимизации суммы модулей ошибок, что означает решение задачи $\sum_{i=1}^n |\varepsilon_i| \rightarrow \min$. Решение этой задачи сводится к решению задачи линейного программирования [11]. Предположим, что с помощью МНМ найдены оценки неизвестных параметров модели (3). Обозначим сумму остатков этой модели как $SAE_{\text{лин}} = \sum_{i=1}^n |e_i|$, где e_i – i -й остаток, равный разнице между фактическим и расчетным значением отклика y в i -м наблюдении.

Рассмотрим ПФ Леонтьева со свободным членом:

$$y_i = \alpha_0 + \min \{ \alpha_1 x_{i1}, \alpha_2 x_{i2}, \dots, \alpha_l x_{il} \} + \varepsilon_i, \quad i = \overline{1, n}, \quad (4)$$

где \min – l -арная операция, возвращающая минимум из перечисленных в фигурных скобках $\{ \}$ l значений. Пусть модель (4) оценена с помощью МНМ в результате решения задачи ЧБЛП [14]. Обозначим сумму остатков этой модели как $SAE_{\text{Леон}}$.

Теорема. Пусть $l \geq 3$, а все индексы входных переменных произвольно распределены по l элементам s_1, s_2, \dots, s_l вектора S . Предположим, что с помощью МНМ оценены неизвестные параметры $\alpha_0, \alpha_1, \dots, \alpha_n, \beta_1, \dots, \beta_n$ многослойной модульной регрессии

$$y_i = \alpha_0 + z_{i,l-2} + \alpha_l x_{i,s_l} - |z_{i,l-2} + \beta_l x_{i,s_l}| + \varepsilon_i, \quad i = \overline{1, n}, \quad (5)$$

$$z_{i,k} = z_{i,k-1} + \alpha_{k+1} x_{i,s_{k+1}} - |z_{i,k-1} + \beta_{k+1} x_{i,s_{k+1}}|, \quad i = \overline{1, n}, \quad k = \overline{2, l-2}, \quad (6)$$

$$z_{i1} = \alpha_1 x_{i,s_1} + \alpha_2 x_{i,s_2} - |\beta_1 x_{i,s_1} + \beta_2 x_{i,s_2}|, \quad i = \overline{1, n}, \quad (7)$$

где $z_{i,k}$ – выходное значение на k -м слое в i -м наблюдении. Обозначим сумму остатков этой регрессии как $SAE_{\text{мод}}$. Тогда будут справедливы неравенства

$$SAE_{\text{мод}} \leq SAE_{\text{лин}}, \quad (8)$$

$$SAE_{\text{мод}} \leq SAE_{\text{Леон}}. \quad (9)$$

Доказательство. Многослойная модульная регрессия (5)–(7) содержит $l - 1$ слоев: 1-й слой отражает равенство (7), промежуточные слои со 2-го до $(l - 2)$ -го – равенства (6), выходной $(l - 1)$ -й слой – равенство (5). Последовательно подставляя выходные значения z с нижних слоев в выражения для верхних слоёв, получим структурную спецификацию

$$y_i = \alpha_0 + \sum_{j=1}^l \alpha_j x_{i,s_j} - F(\alpha, \beta, x) + \varepsilon_i, \quad i = \overline{1, n}, \quad (10)$$

где $F(\alpha, \beta, x)$ – функция нескольких переменных с неизвестными параметрами, содержащая модульные преобразования. Очевидно, что модель (10) есть расширенная версия линейной регрессии (3), откуда следует справедливость неравенства (8) для любого распределения индексов входных переменных по элементам вектора S .

Пусть $\beta_1 = \alpha_1$, $\beta_j = -\alpha_j$, $j = \overline{2, l}$. Многослойная модульная регрессия (5)–(7) будет иметь вид:

$$y_i = \alpha_0 + z_{i,l-2} + \alpha_l x_{i,s_l} - |z_{i,l-2} - \alpha_l x_{i,s_l}| + \varepsilon_i, \quad i = \overline{1, n}, \quad (11)$$

$$z_{i,k} = z_{i,k-1} + \alpha_{k+1}x_{i,s_{k+1}} - |z_{i,k-1} - \alpha_{k+1}x_{i,s_{k+1}}|, \quad i = \overline{1, n}, \quad k = \overline{2, l-2}, \quad (12)$$

$$z_{i1} = \alpha_1x_{i,s_1} + \alpha_2x_{i,s_2} - |\alpha_1x_{i,s_1} - \alpha_2x_{i,s_2}|, \quad i = \overline{1, n}. \quad (13)$$

Модель (11)–(13) есть ограниченная версия модели (5)–(7), поэтому сумма ее остатков не меньше величины $SAE_{\text{мод}}$. Из (1) следует, что $a + b - |a - b| = \min\{2a, 2b\}$. Учитывая это равенство, перепишем регрессию (11)–(13) в виде:

$$y_i = \alpha_0 + \min\{2z_{i,l-2}, 2\alpha_lx_{i,s_l}\} + \varepsilon_i, \quad i = \overline{1, n}, \quad (14)$$

$$z_{i,k} = \min\{2z_{i,k-1}, 2\alpha_{k+1}x_{i,s_{k+1}}\}, \quad i = \overline{1, n}, \quad k = \overline{2, l-2}, \quad (15)$$

$$z_{i1} = \min\{2\alpha_1x_{i,s_1}, 2\alpha_2x_{i,s_2}\}, \quad i = \overline{1, n}. \quad (16)$$

Последовательно подставляя для модели (14)–(16) выходные значения z с нижних слоев в выражения для верхних слоев, получим структурную спецификацию

$$y_i = \alpha_0 + \min\left\{2 \min\left\{2 \min\left\{2\alpha_1x_{i,s_1}, 2\alpha_2x_{i,s_2}\right\}, 2\alpha_3x_{i,s_3}\right\}, \dots\right\} + \varepsilon_i, \quad i = \overline{1, n}. \quad (17)$$

Многослойная конструкция (17), как показано в [19], может быть записана в виде:

$$y_i = \alpha_0 + \min\left\{2^{l-1}\alpha_1x_{i,s_1}, 2^{l-1}\alpha_2x_{i,s_2}, 2^{l-2}\alpha_3x_{i,s_3}, \dots, 2^2\alpha_{l-1}x_{i,s_{l-1}}, 2\alpha_lx_{i,s_l}\right\} + \varepsilon_i, \quad i = \overline{1, n}. \quad (18)$$

Сделав замены $\gamma_1 = 2^{l-1}\alpha_1$, $\gamma_j = 2^{l+1-j}\alpha_j$, $j = \overline{2, l}$, нетрудно видеть, что регрессия (18) есть ПФ Леонтьева (4).

Модель (18) эквивалентна многослойной модульной регрессии (11)–(13), сумма остатков которой не меньше величины $SAE_{\text{мод}}$. Тогда для модели (5)–(7) справедливо неравенство (9) для любого распределения индексов входных переменных по элементам вектора S . **Теорема доказана.**

Сделаем некоторые замечания касаясь доказанной теоремы.

Во-первых, доказанная теорема справедлива и при $l = 2$. Действительно, при $l = 2$ имеем однослойную конструкцию

$$y_i = \alpha_0 + \alpha_1x_{i,s_1} + \alpha_2x_{i,s_2} - |\beta_1x_{i,s_1} + \beta_2x_{i,s_2}| + \varepsilon_i, \quad i = \overline{1, n},$$

для которой по тем же причинам, что и при $l \geq 3$, справедливы неравенства (8) и (9).

Во-вторых, смысл доказанной теоремы состоит в том, что для любых статистических данных и для любого распределения индексов входных переменных по элементам вектора S качество аппроксимации оцененной с помощью МНМ многослойной модульной регрессии (5)–(7) всегда не хуже (на практике чаще лучше), чем качество линейной регрессии (3) либо ПФ Леонтьева (4).

В-третьих, если в многослойной модульной регрессии (5)–(7) заменить знаки «–», стоящие перед операциями модуль, на знаки «+», то аналогичным образом с использованием (2) можно доказать, что качество образованной модели всегда будет не хуже, чем качество линейной регрессии (3) либо регрессии [20] с l -арной операцией \max вида:

$$y_i = \alpha_0 + \max\{\alpha_1x_{i1}, \alpha_2x_{i2}, \dots, \alpha_lx_{il}\} + \varepsilon_i, \quad i = \overline{1, n}.$$

В-четвертых, при $l = 2$ величина $SAE_{\text{мод}}$ многослойной модульной регрессии (5)–(7) не зависит от распределения индексов входных переменных по элементам вектора S , а при $l \geq 3$ – зависит. В любом случае справедливы неравенства (8) и (9).

Таким образом, из теоремы следует, что класс многослойных модульных регрессий включает в себя производственные функции Леонтьева.

Из доказательства теоремы также следует, что любую вложенную кусочно-линейную регрессию с помощью соотношений (1) и (2) всегда можно представить в виде многослойной модульной регрессии. Это можно сделать по следующему алгоритму.

1. Записать вложенную кусочно-линейную регрессию в виде многослойной конструкции только с бинарными операциями \min и \max так, как это показано в [19].

2. С использованием соотношений (1) и (2) перейти от бинарных операций \min и \max к спецификации с модулями.

3. В полученной конструкции обозначить все неизвестные параметры по-разному.

Например, пусть имеется вложенная кусочно-линейная регрессия вида:

$$y_i = \alpha_0 + \min \left\{ \max \left\{ \alpha_1 x_{i1}, \alpha_2 x_{i2} \right\}, \alpha_3 x_{i3}, \alpha_4 x_{i4} \right\} + \varepsilon_i, \quad i = \overline{1, n}. \quad (19)$$

В (19) операция \max – бинарная, \min – тернарная. Запишем модель (19) в виде многослойной конструкции только с бинарными операциями:

$$y_i = \alpha_0 + \min \left\{ \min \left\{ \max \left\{ \alpha_1 x_{i1}, \alpha_2 x_{i2} \right\}, \alpha_3 x_{i3} \right\}, \alpha_4 x_{i4} \right\} + \varepsilon_i, \quad i = \overline{1, n}.$$

С помощью (1) и (2) перейдем от полученной конструкции к многослойной спецификации с модулями:

$$y_i = \alpha_0 + \frac{1}{2} \left(z_{i2} + \alpha_4 x_{i4} - |z_{i2} - \alpha_4 x_{i4}| \right) + \varepsilon_i, \quad i = \overline{1, n}, \quad (20)$$

$$z_{i2} = \frac{1}{2} \left(z_{i1} + \alpha_3 x_{i3} - |z_{i1} - \alpha_3 x_{i3}| \right), \quad i = \overline{1, n}, \quad (21)$$

$$z_{i1} = \frac{1}{2} \left(\alpha_1 x_{i1} + \alpha_2 x_{i2} + |\alpha_1 x_{i1} - \alpha_2 x_{i2}| \right), \quad i = \overline{1, n}. \quad (22)$$

Оцененная с помощью МНМ вложенная кусочно-линейная регрессия (19) и многослойная модульная регрессия (20)–(22) будут иметь одинаковые значения критерия SAE. В (20)–(22) всего 4 неизвестных параметра, которые дублируются. Обозначив же их по-разному, получим в общей сложности 8 неизвестных параметров. Тогда полученная многослойная модульная регрессия при оценивании с помощью МНМ в любом случае будет не хуже по критерию SAE, чем (20)–(22), а, следовательно, и (19).

Таким образом, любой вложенной кусочно-линейной регрессии по указанному алгоритму всегда можно поставить в соответствие многослойную модульную регрессию, которая будет не хуже, а за счет большего числа неизвестных параметров лучше первой. Тем самым многослойная модульная регрессия – гораздо более гибкий инструмент математического моделирования, чем вложенная кусочно-линейная регрессия.

2. Оценивание с помощью МНМ параметров многослойной модульной регрессии (5)–(7)

Многослойная модульная регрессия впервые была предложена в [17], и там же задача нахождения оценок ее параметров с помощью МНМ сведена к задаче ЧБЛП. Предлагаемая в данной статье модель (5)–(7) отличается от описанной ранее в работе [17] регрессии тем, что во второй на текущем слое выходные переменные z с предыдущего слоя используются только под знаком модуля. В многослойной модульной регрессии (5)–(7) выходные переменные с предыдущего слоя используются как под знаком модуля, так и в линейной части текущего слоя.

Используя некоторые математические приемы из [17], получим следующую задачу ЧБЛП для нахождения неизвестных параметров модели (5)–(7):

$$\sum_{i=1}^n (g_i + h_i) \rightarrow \min, \quad (23)$$

$$y_i = \alpha_0 + z_{i,l-2} + \alpha_l x_{i,s_l} - (u_{i,l-1} + v_{i,l-1}) + g_i - h_i, \quad i = \overline{1, n}, \quad (24)$$

$$u_{i,l-1} - v_{i,l-1} = z_{i,l-2} + \beta_l x_{i,s_l}, \quad i = \overline{1, n}, \quad (25)$$

$$z_{i,k} = z_{i,k-1} + \alpha_{k+1} x_{i,s_{k+1}} - (u_{i,k} + v_{i,k}), \quad i = \overline{1, n}, \quad k = \overline{2, l-2}, \quad (26)$$

$$u_{i,k} - v_{i,k} = z_{i,k-1} + \beta_{k+1} x_{i,s_{k+1}}, \quad i = \overline{1, n}, \quad k = \overline{2, l-2}, \quad (27)$$

$$z_{i1} = \alpha_1 x_{i,s_1} + \alpha_2 x_{i,s_2} - (u_{i1} + v_{i1}), \quad i = \overline{1, n}, \quad (28)$$

$$u_{i1} - v_{i1} = \beta_1 x_{i,s_1} + \beta_2 x_{i,s_2}, \quad i = \overline{1, n}, \quad (29)$$

$$u_{i,j} \leq M \cdot \delta_{i,j}, \quad i = \overline{1, n}, \quad j = \overline{1, l-1}, \quad (30)$$

$$v_{i,j} \leq M \cdot (1 - \delta_{i,j}), \quad i = \overline{1, n}, \quad j = \overline{1, l-1}, \quad (31)$$

$$\delta_{i,j} \in \{0, 1\}, \quad i = \overline{1, n}, \quad j = \overline{1, l-1}, \quad (32)$$

$$u_{i,j} \geq 0, \quad v_{i,j} \geq 0, \quad i = \overline{1, n}, \quad j = \overline{1, l-1}, \quad (33)$$

$$g_i \geq 0, \quad h_i \geq 0, \quad i = \overline{1, n}, \quad (34)$$

где M – заданное достаточно большое положительное число; если в i -м наблюдении на j -м слое под знаком модуля содержится положительное число, то оно присваивается переменной $u_{i,j}$, а $v_{i,j} = 0$; если в i -м наблюдении на j -м слое под знаком модуля содержится отрицательное число, то оно присваивается переменной $-v_{i,j}$, а $u_{i,j} = 0$; если в i -м наблюдении на верхнем слое ошибка регрессии ε_i есть положительное число, то оно присваивается переменной g_i , а $h_i = 0$; если в i -м наблюдении на верхнем слое ошибка регрессии ε_i есть отрицательное число, то оно присваивается переменной $-h_i$, а $g_i = 0$; булевы переменные $\delta_{i,j}$, $i = \overline{1, n}$, $j = \overline{1, l-1}$, удовлетворяют правилу

$$\delta_{i,j} = \begin{cases} 1, & \text{если в } i\text{-м наблюдении на } j\text{-м слое под знаком модуля положительное число,} \\ 0, & \text{в противном случае.} \end{cases}$$

Ограничения (24), (25) в сформулированной задаче соответствуют верхнему слою многослойной модульной регрессии, (26), (27) – внутренним слоям, (28), (29) – нижнему слою. Решение задачи ЧБЛП (23)–(34) дает оптимальные МНМ-оценки параметров многослойной модульной регрессии (5)–(7) при $l \geq 3$. При $l = 2$ модель будет состоять только из одного слоя, поэтому ее оценивание будет сводиться к решению более простой задачи ЧБЛП с целевой функцией (23), линейными ограничениями (29)–(34) и

$$y_i = \alpha_0 + \alpha_1 x_{i,s_1} + \alpha_2 x_{i,s_2} - (u_{i1} + v_{i1}) + g_i - h_i, \quad i = \overline{1, n}. \quad (35)$$

3. Вычислительные эксперименты

Для подтверждения корректности полученных в предыдущих разделах математических выводов на персональном компьютере с процессором AMD Ryzen 3 4300U и 16 Гб оперативной памяти проводились вычислительные эксперименты. Чем больше слоев в многослойной модульной регрессии, тем выше вычислительная сложность задачи ЧБЛП, поскольку в ней возрастает число булевых переменных. Поэтому для решения таких задач нужен эффективный решатель (солвер). В этой связи было решено использовать хорошо зарекомендовавший себя и представленный в открытом доступе решатель оптимизационных задач LPSolve [21]. Вместе с тем было решено использовать разработанный в Китае коммерческий продукт Cardinal Optimizer (COPT) [22]. Как заявляют разработчики, этот продукт в некоторых тестах превосходит другие коммерческие аналоги (CPLEX, Gurobi), которые сейчас недоступны в России, а также open-source солверы. Было интересно сравнить скорость работы LPSolve и COPT на примере построения многослойных модульных регрессий. Для управления солвером COPT был использован интерфейс Python.

Для проведения экспериментов использовались две выборки данных.

Первая выборка [23. С. 284] объема $n = 13$ содержит данные о результатах химических экспериментов (данные Хальда). Она включает в себя информацию по следующим переменным: y_1 – выделившееся тепло в калориях на грамм цемента; x_1 – количество трикальций-алюмината (% от веса клинкера); x_2 – количество трикальций-силиката (%); x_3 – количество тетракальций-алюминий-феррита (%); x_4 – количество дикальций-силиката (%).

Вторая выборка объема $n = 23$ включает данные о новых единицах жилья в США и влияющих на них факторах. Эти данные встроены в эконометрический пакет Gretl (файл data4-3.gdt). В них содержится информация по следующим переменным: y_2 – общее количество новых единиц жилья,

введенных в эксплуатацию (тыс.); x_1^* – население США (млн чел); x_2^* – валовой национальный продукт (млрд долл.); x_3^* – уровень безработицы (%); x_4^* – доходность ипотечных кредитов на новое жилье (%).

При проведении экспериментов менялся состав входящих в модель объясняющих переменных. Для этого были назначены следующие подмножества: $\{1, 2\}$, $\{1, 3\}$, $\{1, 4\}$, $\{2, 3\}$, $\{2, 4\}$, $\{3, 4\}$, $\{1, 2, 3\}$, $\{1, 2, 4\}$, $\{1, 3, 4\}$, $\{2, 3, 4\}$ и $\{1, 2, 3, 4\}$. Например, подмножество $\{2, 3\}$ для первой выборки означает, что объясняющими переменными выступают x_2 и x_3 . Для каждой выборки и каждого состава переменных решались следующие задачи:

- 1) в пакете Gretl оценивались параметры линейной регрессии (3) и фиксировалось значение $SAE_{\text{лин}}$;
- 2) с помощью LPSolve оценивались параметры ПФ Леонтьева (4) и фиксировалось значение $SAE_{\text{Леон}}$;
- 3) с помощью LPSolve и COPT оценивались параметры многослойной модульной регрессии (5)–(7) с прямым порядком следования объясняющих переменных, фиксировались значение $SAE_{\text{мод}}$ и время решения задачи ЧБЛП (23)–(34) при $l \geq 3$ либо (23), (29)–(35) при $l = 2$;
- 4) с помощью LPSolve и COPT оценивались параметры многослойной модульной регрессии (5)–(7) с обратным порядком следования объясняющих переменных, фиксировались значение $SAE_{\text{мод}}$ и время решения задачи ЧБЛП (23)–(34) при $l \geq 3$ либо (23), (29)–(35) при $l = 2$.

Прямой порядок следования переменных означает, что если, например, выбран состав $\{1, 2, 3\}$, то первый элемент вектора S равен 1, второй – 2, третий – 3. При обратном порядке первый элемент вектора S равен 3, второй – 2, третий – 1.

Используемые солверы LPSolve и COPT содержат большое количество различных настраиваемых параметров. При решении задач ЧБЛП важно правильно установить такой параметр, как точность целого числа, которую обозначим IntTol. Если установить значение IntTol слишком большим, например 0,1, то значение 0,91 солвер будет идентифицировать как число «1», а значение 0,07 – как число «0». Такая настройка, возможно, увеличит скорость решения задачи, но результат может быть ошибочным. В LpSolve параметр IntTol по умолчанию равен $1e^{-7}$, а в COPT – $1e^{-6}$. Для организации равных условий работы пакетов параметру IntTol в COPT было присвоено значение $1e^{-7}$.

Хотелось бы обратить внимание на ограничения (30) и (31). Если точность целого числа IntTol в солверах выбрана $1e^{-7}$, то при очень большом числе M эти ограничения не будут гарантировать выполнение условий

$$u_{i,j}v_{i,j} = 0, \quad i = \overline{1, n}, \quad j = \overline{1, l-1}, \quad (36)$$

что может привести к ошибочным результатам. Поэтому условия (36) необходимо проверять после решения каждой задачи ЧБЛП. Если они не выполняются, то требуется перерешать задачу, уменьшив параметр IntTol.

Для данных Хальда параметр M был выбран 1 000, а для данных о жилье в США $M = 10\,000$. Вместо переменной y_2 была использована переменная $y_2^* = y_2/M$. Ограничение по времени в солверах на решение одной задачи ЧБЛП составило 1 800 с (30 мин).

Результаты проведенных экспериментов представлены в таблице.

Условия (36) при IntTol = $1e^{-7}$ не выполнялись только при обработке данных о жилье для составов $\{1, 2, 4\}$ и $\{2, 3, 4\}$, поэтому в строках № 19 и 21 таблицы приведены результаты, полученные в LPSolve и COPT при IntTol = $1e^{-8}$.

В таблице символ «*» указывает на субоптимальные решения, найденные в установленный лимит времени 30 мин.

По таблице в результате обработки данных Хальда можно сделать следующие выводы.

1. Как линейная регрессия может превосходить по критерию SAE ПФ Леонтьева (например, строка № 1), так и ПФ может превосходить линейную регрессию (например, строка № 2).
2. При построении многослойной модульной регрессии абсолютно во всех экспериментах для любого порядка переменных оказались справедливы неравенства (8) и (9), что подтверждает корректность доказанной теоремы. Причем, только в 3 случаях из 22 значение $SAE_{\text{мод}}$ оказалось равно $SAE_{\text{лин}}$ или $SAE_{\text{Леон}}$, т.е. в большинстве случаев многослойная модульная регрессия оказалась лучше и линейной регрессии, и ПФ Леонтьева.

3. При изменении порядка следования переменных качество однослойных модульных регрессий по критерию SAE сохранилось, а качество двухслойных и трехслойных модульных регрессий изменилось.

4. В LPSolve для прямого и обратного порядков переменных среднее время построения однослойной модульной регрессии составило 0,06 с, двухслойной – 2,26 с, трехслойной – 25,51 с. Рост времени обусловлен увеличением в задаче ЧБЛП количества булевых переменных при увеличении числа слоев.

5. В COPT были получены точно такие же модели, что и в LPSolve. В COPT для прямого и обратного порядков переменных среднее время построения однослойной модульной регрессии составило 0,33 с, двухслойной – 0,59 с, трехслойной – 1,16 с. Получилось, что однослойные модульные регрессии в LPSolve и COPT оцениваются практически мгновенно, но COPT проигрывает по времени пакету LPSolve. Однако такой результат, вероятнее всего, можно объяснить тем, что при управлении COPT в Python на чтение оптимизационной задачи из программы уходит чуть больше времени, чем в LpSolve. Для двухслойных модульных регрессий время решения задач ЧБЛП в COPT оказалось в 3,84 раза, а для трехслойных – в 22,08 раза меньше, чем в LPSolve.

Результаты экспериментов

№	Состав	SAE _{лин}	SAE _{Леон}	Прямой порядок переменных				Обратный порядок переменных			
				LPSolve		COPT		LPSolve		COPT	
				SAE _{мод}	t, с	SAE _{мод}	t, с	SAE _{мод}	t, с	SAE _{мод}	t, с
Данные Хальда											
1	1, 2	23,126	56,167	23,126	0,048	23,126	0,33	23,126	0,047	23,126	0,36
2	1, 3	107,1	91,6	59,195	0,055	59,195	0,28	59,195	0,036	59,195	0,27
3	1, 4	23,339	107,3	21,972	0,064	21,972	0,33	21,972	0,063	21,972	0,35
4	2, 3	55,057	83,125	54,662	0,084	54,662	0,34	54,662	0,073	54,662	0,41
5	2, 4	82,957	72,421	70,347	0,070	70,347	0,35	70,347	0,085	70,347	0,35
6	3, 4	35,313	165	30,940	0,053	30,940	0,28	30,940	0,043	30,940	0,29
7	1, 2, 3	19,056	56,167	19,049	1,087	19,049	0,50	13,66	0,268	13,66	0,45
8	1, 2, 4	19,663	52	17,017	3,878	17,017	0,59	16,389	1,969	16,389	0,93
9	1, 3, 4	20,203	90,6	20,203	7,424	20,203	0,62	17,882	0,858	17,882	0,58
10	2, 3, 4	22,536	72,421	20,609	1,431	20,609	0,45	20,068	1,130	20,068	0,58
11	1, 2, 3, 4	18,834	52	12,459	40,144	12,459	1,01	13,809	10,876	13,809	1,30
Встроенные в Gretl данные data4-3											
12	1, 2	5,655	5,858	5,655	17,049	5,655	3,01	5,655	16,716	5,655	3,57
13	1, 3	5,896	5,773	5,000	4,503	5,000	0,54	5,000	4,716	5,000	0,49
14	1, 4	4,356	5,902	4,241	38,665	4,241	0,95	4,241	38,982	4,241	0,91
15	2, 3	5,853	5,797	5,164	4,672	5,164	0,60	5,164	4,922	5,164	0,75
16	2, 4	4,514	5,858	4,172	5,480	4,172	0,66	4,172	5,540	4,172	0,52
17	3, 4	6,055	6,053	5,004	5,077	5,004	0,67	5,004	4,921	5,004	0,69
18	1, 2, 3	5,199	5,773	4,764*	1800	4,145	4,77	4,758*	1800	4,308	40,22
19	1, 2, 4	4,346	5,855	3,818	1431,4	3,818	9,69	4,204*	1800	4,013	934,41
20	1, 3, 4	4,323	5,767	3,925*	1800	3,177	5,98	3,540	1785	3,540	5,92
21	2, 3, 4	4,176	5,784	3,343	813,87	3,343	32,03	4,056*	1800	3,713	3,75
22	1, 2, 3, 4	4,074	5,767	4,813*	1800	2,840	261,6	4,139*	1800	3,268	32,23

По таблице в результате обработки данных о жилье можно сделать следующие выводы.

1. Линейная регрессия и ПФ Леонтьева в зависимости от выбранных переменных могут превосходить друг друга по критерию SAE.

2. Не принимая во внимание субоптимальные решения, при построении многослойной модульной регрессии во всех экспериментах для любого порядка переменных определено, что оказались справедливы неравенства (8) и (9), и это вновь подтверждает корректность доказанной теоремы. И снова в большинстве случаев (в 20 из 22) многослойная модульная регрессия оказалась лучше как линейной регрессии, так и ПФ Леонтьева.

3. Изменение порядка следования переменных меняет значение критерия SAE для двухслойных и трехслойных модульных регрессий, но не меняет для однослойных.

4. Вторая выборка данных содержит больше наблюдений, чем первая, поэтому время решения задач ЧБЛП увеличилось. Причем LPSolve в установленный лимит времени 30 мин не справился с 7 задачами из 22. Так, в LPSolve для прямого и обратного порядков переменных среднее время построения однослойной модульной регрессии составило 12,6 с. Для двухслойной и трехслойной модульной регрессии – 1 628,78 и 1 800 с соответственно. Причем при снятии установленного лимита для последних двух спецификаций время будет еще больше.

5. Солвер СОРТ, в отличие от LPSolve, справился абсолютно со всеми задачами. Для прямого и обратного порядков переменных среднее время построения однослойной модульной регрессии составило 1,11 с, двухслойной – 129,59 с, трехслойной – 146,91 с. Среднее время оценки двухслойных регрессий оказалось высоким из-за эксперимента для обратного порядка переменных из множества {1, 2, 4}. Без учета этого эксперимента среднее время составит всего 14,62 с. Таким образом, для однослойных модульных регрессий время решения задач ЧБЛП в СОРТ оказалось в 11,32 раза, для двухслойных – как минимум в 12,57 раза, а для трехслойных – как минимум в 12,25 раза меньше, чем в LPSolve.

Рассмотрим внешний вид моделей с полным составом переменных, построенных по данным о рынке жилья. ПФ имеет вид:

$$\tilde{y}_2^* = 1,00054 + \min\{0,003032x_1^*, 0,000306x_2^*, 0,115717x_3^*, 0,097422x_4^*\}. \quad (37)$$

Как отмечено в строке № 22 таблицы, для регрессии (37) $SAE_{\text{Леон}} = 5,767$. При этом нулевыми оказались остатки в наблюдениях под номерами 3 и 21.

Трехслойная модульная регрессия для прямого порядка переменных имеет вид:

$$\tilde{y}_2^* = 3,46081 + 0,065492x_4^* + z_2 - |z_2 + 0,22326x_4^*|, \quad (38)$$

$$z_2 = z_1 - 0,029006x_3^* - |z_1 + 0,26546x_3^*|, \quad (39)$$

$$z_1 = -0,012955x_1^* + 0,000523x_2^* - |0,0041428x_1^* - 0,00038652x_2^*|. \quad (40)$$

Для регрессии (38)–(40) $SAE_{\text{мод}} = 2,840$. При этом нулевые остатки зафиксированы сразу в девяти наблюдениях – под номерами 1, 2, 3, 7, 11, 13, 15, 17 и 20.

Для наглядности на рис. 1, а изображены графики фактических и расчетных значений зависимой переменной y_2 для модели (37), а на рис. 1, б – для модели (38)–(40).

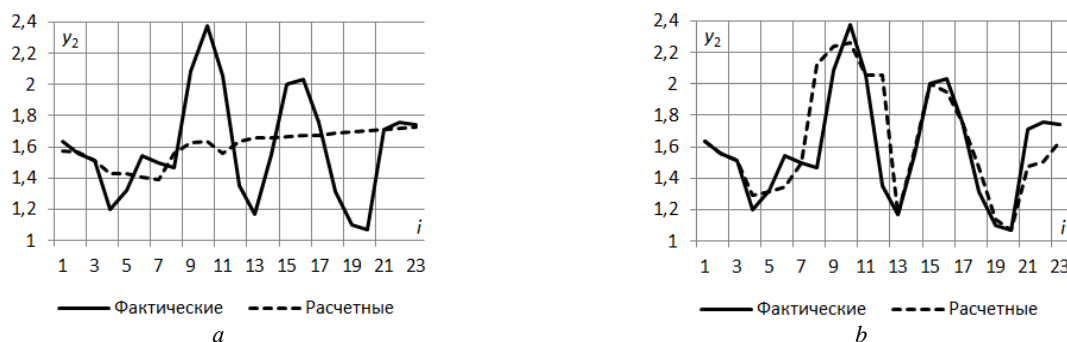


Рис. 1. Фактические и рассчитанные по моделям значения переменной y_2

Fig. 1. Actual and calculated values of the variable y_2 according to the models

По графикам на рис. 1 видно, что точность трехслойной модульной регрессии (38)–(40) существенно выше, чем точность ПФ Леонтьева (37).

Заключение

В статье предложена модель многослойной модульной регрессии (5)–(7), отличающаяся от разработанной в [17] модели тем, что выходное значение с каждого предыдущего слоя входит не только в модульную часть, но и в линейную часть последующего слоя. Доказано, что для любых статистических

данных и для любого распределения индексов входных переменных по элементам вектора S качество аппроксимации оцененной с помощью МНМ многослойной модульной регрессии (5)–(7) всегда не хуже (на практике зачастую лучше), чем качество линейной регрессии (3) и ПФ Леонтьева (4). Показано, что класс многослойных модульных регрессий включает в себя класс вложенных кусочно-линейных регрессий. Разработан алгоритм трансформации любой вложенной кусочно-линейной регрессии в модель многослойной модульной регрессии, оценка которой с помощью МНМ гарантирует не ухудшение качества аппроксимации по критерию SAE. Задача нахождения точных МНМ-оценок неизвестных параметров предложенной многослойной модульной регрессии (5)–(7) сведена к задаче ЧБЛП. На двух разных выборках данных проведены вычислительные эксперименты, доказывающие корректность доказанной теоремы. В ходе экспериментов впервые проведен сравнительный анализ скорости солверов LPSolve и Cardinal Optimizer. Для первой выборки при оценке параметров однослойных модульных регрессий солверы решили задачи ЧБЛП практически мгновенно. При оценке двухслойных и трехслойных модульных регрессий Cardinal Optimizer оказался быстрее в 3,84 и 22,08 раза соответственно, чем LPSolve. Для второй выборки большего объема при оценке однослойных модульных регрессий Cardinal Optimizer оказался в 11,32 раза быстрее, чем LPSolve. При оценке двухслойных и трехслойных модульных регрессий в LPSolve оптимальное решение большинства задач ЧБЛП не было получено в установленный лимит времени 30 мин. При этом Cardinal Optimizer справился абсолютно со всеми задачами, оказавшись быстрее, чем LPSolve, как минимум в 12,57 и 12,25 раза для двухслойных и трехслойных модульных регрессий. Таким образом, выявлено существенное превосходство по скорости солвера Cardinal Optimizer перед LPSolve.

Список источников

1. Montgomery D.C., Peck E.A., Vining G.G. Introduction to linear regression analysis. New York : John Wiley and Sons, 2021. 704 p.
2. Baždarić K., Šverko D., Salarić I., Martinović A., Lucijanić M. The ABC of linear regression analysis: What every author and editor should know // European science editing. 2021. V. 47. doi: 10.3897/ese.2021.e63780
3. Батмазова А.А., Гайдукова Е.В., Дрегваль М.С. Особенности построения регрессионной модели для оценки и прогноза уровня озера Ловозеро // Международный научно-исследовательский журнал. 2024. № 7 (145). doi: 10.60797/IRJ.2024.145.67
4. Бястинова Л.М. Использование многофакторных регрессионных моделей при прогнозировании в сфере сельского хозяйства Якутии // Экономика и природопользование на Севере. 2023. № 4. С. 102–110. doi: 10.25587/2587-8778-2023-4-102-110
5. Каштанов А.Л., Комяков А.А., Никифоров М.М. Прогнозирование и верификация ключевых показателей энергетической эффективности железнодорожного транспорта // Вестник Уральского государственного университета путей сообщения. 2021. № 1. С. 46–54. doi: 10.20291/2079-0392-2021-1-46-54
6. Горшенин А.Ю., Васина Д.И. Сравнение используемых методов при прогнозировании выработки электроэнергии ветроэлектростанциями // Математические структуры и моделирование. 2023. № 3 (67). С. 36–42. doi:10.24147/2222-8772.2023.3.36-42
7. Брачунова У.В., Козловский В.Н., Шакурский М.В., Пантюхин О.В. Основные аспекты разработки математической модели и программного комплекса по оценке зарядного баланса легкового автомобиля // Известия Тульского государственного университета. Технические науки. 2023. № 5. С. 242–249. doi: 10.24412/2071-6168-2023-5-242-243
8. Гичиев Н.С. Анализ двухфакторной модели экономического роста на основе производственной функции Кобба–Дугласа // Региональные проблемы преобразования экономики. 2022. № 3 (137). С. 62–66. doi: 10.26726/1812-7096-2022-3-62-66
9. Ботнарюк М.В., Ксензова Н.Н. Производственная функция Кобба–Дугласа для оценки деятельности морского порта // Научные проблемы водного транспорта. 2022. № 74. С. 85–95. doi: 10.37890/jwt.vi74.348
10. Демиденко Е.З. Линейная и нелинейная регрессии. М. : Финансы и статистика, 1981. 303 с.
11. Wagner H.M. Linear programming techniques for regression analysis // Journal of the American Statistical Association. 1959. V. 54 (285). P. 206–212. doi: 10.1080/01621459.1959.10501506
12. Koch T., Berthold T., Pedersen J., Vanaret C. Progress in mathematical programming solvers from 2001 to 2020 // EURO Journal on Computational Optimization. 2022. V. 10. Art. 100031. doi: 10.1016/j.ejco.2022.100031
13. Клейнер Г.Б. Производственные функции: теория, методы, применение. М. : Финансы и статистика, 1986. 239 с.
14. Носков С.И. Технология моделирования объектов с нестабильным функционированием и неопределенностью в данных. Иркутск : Облформпечать, 1996. 320 с.
15. Носков С.И. Оценивание параметров простой вложенной кусочно-линейной регрессии с линейной составляющей // Вестник Югорского государственного университета. 2024. Т. 20, № 1. С. 19–21. doi: 10.18822/byusu20240119-21
16. Базилевский М.П., Ойдопова А.Б. Оценивание модульных линейных регрессионных моделей с помощью метода наименьших модулей // Вестник Пермского национального исследовательского политехнического университета. Электротехника, информационные технологии, системы управления. 2023. № 45. С. 130–146. doi: 10.15593/2224-9397/2023.1.06

17. Базилевский М.П. Оценивание неизвестных параметров многослойной модульной регрессии методом наименьших модулей // Моделирование, оптимизация и информационные технологии. 2024. Т. 12, № 2 (45). doi:10.26102/2310-6018/2024.45.2.039
18. Базилевский М.П. Оценивание регрессионных моделей с регрессорами в виде модулей линейных комбинаций объясняющих переменных // System Analysis and Mathematical Modeling. 2024. Т. 6, № 3. С. 269–281. doi: 10.17150/2713-1734.2024.6(3).269-281
19. Базилевский М.П. Алгоритм оценивания неизвестных параметров одного вида многослойных неэлементарных линейных регрессий // Ученые записки Комсомольского-на-Амуре государственного технического университета. 2025. № 1 (81). С. 35–44.
20. Носков С.И., Хоняков А.А. Программный комплекс построения некоторых типов кусочно-линейных регрессий // Информационные технологии и математическое моделирование в управлении сложными системами. 2019. № 3 (4). С. 47–55.
21. Mixed Integer Linear Programming (MILP) solver – LPSolve // Sourceforge. URL: <https://sourceforge.net/projects/lpsolve/> (accessed: 07.06.2025).
22. Ge D., Huangfu Q., Wang Z, Wu J., Ye Y. Cardinal Optimizer (COPT) user guide. URL: <https://guide.coap.online/copt/en-doc> (accessed: 07.06.2025).
23. Дрейпер Н., Смит Г. Прикладной регрессионный анализ. М. : Финансы и статистика, 1987. Кн. 2. 353 с.

References

1. Montgomery, D.C., Peck, E.A. & Vining, G.G. (2021) *Introduction to Linear Regression Analysis*. New York: John Wiley and Sons.
2. Baždarić, K., Šverko, D., Salarić, I., Martinović, A. & Lucijanić, M. (2021) The ABC of linear regression analysis: What every author and editor should know. *European Science Editing*. 47. doi: 10.3897/ese.2021.e63780
3. Batmazova, A.A., Gaydukova, E.V. & Dregval, M.S. (2024) Osobennosti postroeniya regressionnoy modeli dlya otsenki i prognoza urovnya ozero Lovozero [Building a Regression Model for Assessing and Forecasting the Water Level of Lovozero Lake]. *Mezhdunarodnyy nauchno-issledovatel'skiy zhurnal*. 7(145). doi: 10.60797/IRJ.2024.145.67
4. Byastinova, L.M. (2023) Ispol'zovanie mnogofaktornykh regressionnykh modeley pri prognozirovanii v sfere sel'skogo khozyaystva Yakuti [Using Multivariate Regression Models for Forecasting in the Agricultural Sector of Yakutia]. *Ekonomika i prirodopol'zovanie na Severe*. 4. pp. 102–110. doi: 10.25587/2587-8778-2023-4-102-110
5. Kashtanov, A.L., Komyakov, A.A. & Nikiforov, M.M. (2021) Prognozirovaniye i verifikatsiya klyuchevykh pokazateley energeticheskoy effektivnosti zheleznodorozhnogo transporta [Forecasting and Verification of Key Energy Efficiency Indicators of Railway Transport]. *Vestnik Ural'skogo gosudarstvennogo universiteta putey soobshcheniya*. 1. pp. 46–54. doi: 10.20291/2079-0392-2021-1-46-54
6. Gorshenin, A.Yu. & Vasina, D.I. (2023) Svravnenie ispol'zuemykh metodov pri prognozirovanii vyrobтки elektroenergii vetroelektrstantsiyami [Comparison of Methods Used in Forecasting Electricity Generation by Wind Power Plants]. *Matematicheskie struktury i modelirovaniye*. 3(67). pp. 36–42. doi:10.24147/2222-8772.2023.3.36-42
7. Brachunova, U.V., Kozlovskiy, V.N., Shakurskiy, M.V. & Pantyukhin, O.V. (2023) Osnovnye aspekty razrabotki matematicheskoy modeli i programmnoy kompleksa po otsenke zaryadnogo balansa legkovogo avtomobilya [Main Aspects of Developing a Mathematical Model and Software Package for Assessing the Charge Balance of a Passenger Car]. *Izvestiya Tul'skogo gosudarstvennogo universiteta. Tekhnicheskie nauki*. 5. pp. 242–249. doi: 10.24412/2071-6168-2023-5-242-243
8. Gichiev, N.S. (2022) Analiz dvukhfaktornoy modeli ekonomicheskogo rosta na osnove proizvodstvennoy funktsii Kobba–Duglasa [Analysis of a Two-Factor Economic Growth Model Based on the Cobb–Douglas Production Function]. *Regional'nye problemy preobrazovaniya ekonomiki*. 3(137). pp. 62–66. doi: 10.26726/1812-7096-2022-3-62-66
9. Botnaryuk, M.V. & Ksenzova, N.N. (2022) Proizvodstvennaya funktsiya Kobba–Duglasa dlya otsenki deyatel'nosti morskogo porta [The Cobb–Douglas Production Function for Assessing the Activities of a Seaport]. *Nauchnye problemy vodnogo transporta*. 74. pp. 85–95. doi: 10.37890/jwt.vi74.348
10. Demidenko, E.Z. (1981) *Lineynaya i nelineynaya regressii* [Linear and Nonlinear Regression]. Moscow: Finansy i statistika.
11. Wagner, H.M. (1959) Linear programming techniques for regression analysis. *Journal of the American Statistical Association*. 54(285). pp. 206–212. doi: 10.1080/01621459.1959.10501506
12. Koch, T., Berthold, T., Pedersen, J. & Vanaret, C. (2022) Progress in mathematical programming solvers from 2001 to 2020. *EURO Journal on Computational Optimization*. 10. Art. 100031. doi: 10.1016/j.ejco.2022.100031
13. Kleynner, G.B. (1986) *Proizvodstvennyye funktsii: teoriya, metody, primeneniye* [Production Functions: Theory, Methods, Application]. Moscow: Finansy i statistika.
14. Noskov, S.I. (1996) *Tekhnologiya modelirovaniya ob"ektov s nestabil'nykh funktsionirovaniem i neopredelennost'yu v dannykh* [Modeling Technology for Objects with Unstable Operation and Data Uncertainty]. Irkutsk: Oblinformatpechat'.
15. Noskov, S.I. (2024) Otsenivaniye parametrov prostoy vlozhennoy kusochno-lineynoy regressii s lineynoy sostavlyayushchey [Estimating Parameters of a Simple Nested Piecewise Linear Regression with a Linear Component]. *Vestnik Yugorskogo gosudarstvennogo universiteta*. 20(1). pp. 19–21. doi: 10.18822/byusu20240119-21
16. Bazilevskiy, M.P. & Oydopova, A.B. (2023) Otsenivaniye modul'nykh lineynykh regressionnykh modeley s pomoshch'yu metoda naimen'shikh moduley [Estimating Modular Linear Regression Models Using the Least Absolute Deviations Method]. *Vestnik Permskogo natsional'nogo issledovatel'skogo politekhnicheskogo universiteta. Elektrotekhnika, informatsionnye tekhnologii, sistemy upravleniya*. 45. pp. 130–146. doi: 10.15593/2224-9397/2023.1.06

17. Bazilevskiy, M.P. (2024) Otsenivanie neizvestnykh parametrov mnogosloynoy modul'noy regressii metodom naimen'shikh moduley [Estimating Unknown Parameters of Multilayer Modular Regression Using the Least Absolute Deviations Method]. *Modelirovanie, optimizatsiya i informatsionnye tekhnologii*. 12(2(45)). doi: 10.26102/2310-6018/2024.45.2.039
18. Bazilevskiy, M.P. (2024) Otsenivanie regressionnykh modeley s regressorami v vide moduley lineynykh kombinatsiy ob"yasnyayushchikh peremennykh [Estimating Regression Models with Regressors in the Form of Modules of Linear Combinations of Explanatory Variables]. *System Analysis and Mathematical Modeling*. 6(3). pp. 269–281. doi: 10.17150/2713-1734.2024.6(3).269-281
19. Bazilevskiy, M.P. (2025) Algoritm otsenivaniya neizvestnykh parametrov odnogo vida mnogosloynnykh neelementarnykh lineynykh regressiy [Algorithm for Estimating Unknown Parameters of One Type of Multilayer Non-elementary Linear Regressions]. *Uchenye zapiski Komsomol'skogo-na-Amure gosudarstvennogo tekhnicheskogo universiteta*. 1(81). pp. 35–44.
20. Noskov, S.I. & Khonyakov, A.A. (2019) Programmy kompleks postroyeniya nekotorykh tipov kusochno-lineynykh regressiy [Software Package for Constructing Certain Types of Piecewise Linear Regressions]. *Informatsionnye tekhnologii i matematicheskoe modelirovanie v upravlenii slozhnymi sistemami*. 3(4). pp. 47–55.
21. *Mixed Integer Linear Programming (MILP) solver – LPSolve*. [Online] Available from: <https://sourceforge.net/projects/lpsolve/> (Accessed: 7th June 2025).
22. Ge, D., Huangfu, Q., Wang, Z., Wu, J. & Ye, Y. (n.d.) *Cardinal Optimizer (COPT) user guide*. [Online] Available from: <https://guide.coap.online/copt/en-doc> (Accessed: 7th June 2025).
23. Draper, N. & Smith, G. (1987) *Prikladnoy regressionnyy analiz* [Applied Regression Analysis]. Vol. 2. Translated from English. Moscow: Finansy i statistika.

Информация об авторе:

Базилевский Михаил Павлович – доцент, кандидат технических наук, доцент кафедры «Математика» Иркутского государственного университета путей сообщения (Иркутск, Россия). E-mail: mik2178@yandex.ru

Автор заявляет об отсутствии конфликта интересов.

Information about the author:

Bazilevskiy Mikhail P. (Associate Professor, Candidate of Technical Sciences, Irkutsk State Transport University, Irkutsk, Russian Federation). E-mail: mik2178@yandex.ru

The author declares no conflicts of interests.

Поступила в редакцию 07.06.2025; принята к публикации 02.12.2025

Received 07.06.2025; accepted for publication 02.12.2025

Научная статья
УДК 550.834, 519.6
doi: 10.17223/19988605/73/7

Вычислительная кинематическая диагностика сред с возможностью линеаризации скорости зондирующего сигнала в условиях томографического эксперимента

Николай Владимирович Гервас¹, Сергей Михайлович Зеркаль²

^{1, 2} Новосибирский государственный технический университет, Новосибирск, Россия

¹ nik.gervas@mail.ru

² zerkal@ngs.ru

Аннотация. Рассматривается решение обратной кинематической задачи одной из основных задач теории распространения волн, постановки, сводящейся к классической компьютерной (вычислительной) томографии. Полнота томографических измерительных (проекционных) данных обеспечивается расположением источников и приемников зондирующего сигнала на окружности, лежащих в плоскости $Z = 0$ евклидова пространства XYZ . Математически такая постановка является «слабо-некорректной» (по отечественной классификации для задач математической физики), что делает решение однозначным и повышает достоверность интерпретации. Излагаются постановка задачи, построение решения и вычислительный алгоритм, а также результаты численного компьютерного моделирования. Применительно к сейсмоакустическим исследованиям определяются возможности и перспективы использования полученных результатов.

Ключевые слова: вычислительный алгоритм; обратная задача; рефрагированные волны; кинематическая диагностика; компьютерная томография.

Для цитирования: Гервас Н.В., Зеркаль С.М. Вычислительная кинематическая диагностика сред с возможностью линеаризации скорости зондирующего сигнала в условиях томографического эксперимента // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2025. № 73. С. 58–64. doi: 10.17223/19988605/73/7

Original article
doi: 10.17223/19988605/73/7

Computational kinematic diagnostics of media with the possibility of linearizing the probing signal velocity in the context of a tomographic experiment

Nikolai V. Gervas¹, Sergey M. Zerkal²

^{1, 2} Novosibirsk State Technical University, Novosibirsk, Russian Federation

¹ nik.gervas@mail.ru

² zerkal@ngs.ru

Abstract. The paper addresses the inverse kinematic problem, which is one of the primary problems of the "Theory of Wave Propagation", formulated in a manner that reduces it to classical computational tomography. The completeness of tomographic (projection) data is ensured by positioning the sources and receivers of the probing signal on a circle lying in the plane $Z = 0$ of the Euclidean space $(X Y Z)$. Mathematically, this formulation is classified as "weakly ill-posed" (according to the domestic classification of mathematical physics problems), which makes the solution unique and enhances the reliability of the interpretation. The paper presents the problem statement, the construction of the solution, the computational algorithm, and the results of numerical simulations. In the context of seismic-acoustic studies, the possibilities and prospects of applying the obtained results are discussed.

Keywords: computational algorithm; inverse problem; refracted waves; kinematic diagnostics; computed tomography.

For citation: Gervas, N.V., Zerkal, S.M. (2025) Computational kinematic diagnostics of media with the possibility of linearizing the probing signal velocity in the context of a tomographic experiment. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naja tehnika i informatika – Tomsk State University Journal of Control and Computer Science*. 73. pp. 58–64. doi: 10.17223/19988605/73/7

Введение

Обратная кинематическая задача (ОКЗ) представляет собой достаточно крупную научную проблему, имеющую математическую основу как задача математической физики, теории распространения волн. В своей численной реализации она относится к вычислительной диагностике. ОКЗ имеет высокое прикладное значение, возникает в основном в условиях оптики, акустики, сейсмологии как задача кинематической диагностики среды распространения соответствующих рефрагированных волн [1–4]. Впервые эта задача была рассмотрена в 1906–1907 гг. Г. Герглотцем и Э. Вихертом, и в последующее время интерес научной общественности к ней только возрастал [5]. К настоящему времени достижения информатики и искусственного интеллекта (ИИ) открывают обширную перспективу для преодоления трудностей, присущих решению ОКЗ.

Геофизические методы, направленные на исследование структуры земных недр, обычно предполагают создание «изображения» изучаемой среды на основе интерпретации полученных данных. Достижение высокой точности в этом процессе невозможно без применения специализированных компьютерных систем, спроектированных с использованием математического моделирования их ключевых элементов и поддерживаемых математическим обеспечением, основанным на разработке соответствующих алгоритмов для решения поставленных задач. Таким образом, геофизика тесно связана с относительно новой областью знаний – вычислительной диагностикой [6], где ключевые понятия включают «вычислительные технологии» и «информационно-вычислительные технологии». Под вычислительной диагностикой понимается совокупность методов и инструментов, направленных на изучение характеристик объектов исследования на основе обработки косвенной информации с использованием вычислительных средств. В отличие от традиционных методов, вычислительная диагностика оперирует огромными объемами информации об исследуемом объекте, которые обрабатываются с применением эффективных алгоритмов на высокопроизводительных вычислительных системах.

Примером применения вычислительной диагностики являются вычислительная томография и ее специализированный вариант – сейсмическая томография, где ключевую роль в исследованиях играет решение обратной кинематической задачи [7, 8].

Технология кинематической диагностики земных недр, связанная с определением скоростного распределения сейсмической волны по географу, начала развиваться с работ немецкого физика Э. Вихерта. Важно отметить, что на практике при выполнении измерений кинематическая информация является наиболее точной в сравнении, например, с динамической [9]. Данное исследование сосредоточено на совершенствовании теоретических основ, позволяющих предложить новые модификации методов диагностических исследований, основанных на теории распространения волн. При этом учитывается необходимость сочетания теоретических разработок с экспериментальной верификацией физических моделей, применимых к задачам сейсмологии [6–9].

1. Постановка задачи и построение решения

Рассмотрим полупространство $Z \geq 0$ в трехмерном пространстве R^3 , заполненное средой с показателем преломления $n(x, y, z) = V^{-1}(x, y, z)$, где $V(x, y, z)$ представляет собой скорость распространения колебаний в среде. Пусть в точке $S_0 = (x_0, y_0, 0)$ генерируется сигнал, а в точке $S_1 = (x_1, y_1, 0)$ фиксируется время прихода рефрагированной волны – $T(S_0, S_1)$. Обратная кинематическая задача состоит в восстановлении функции $V(x, y, z)$ по известной функции $T(S_0, S_1)$. В общем случае эта задача является переопределенной: по функции четырех переменных $T(x_0, y_0, x_1, y_1)$ требуется определить функцию трех переменных $V(x, y, z)$. Исследования по переопределенной постановке обратной кинематической задачи были проведены В.Г. Романовым и Р.Г. Мухометовым [5].

Организация системы наблюдений в форме окружности радиуса r позволяет устранить переопределенность задачи, поскольку функция T теперь зависит от радиуса r и двух углов φ_1, φ_2 (в полярных координатах), указывающих на источник и приемник соответственно. Предполагается, что исследуемая среда является регулярной, т.е. изменения скорости в ней таковы, что каждой паре точек источник–приемник (S_0 и S_1) соответствует одна геодезическая линия (луч) $\Gamma(S_0, S_1)$. Важным условием также является то, что скорость $V(x, y, z)$ представима в виде $V(x, y, z) = V_0(z) + V_1(x, y, z)$, $V_0 = A + Bz$, $V_0 \gg |V_1|$. Соотношение коэффициентов A и B выбрано так, чтобы обеспечивать достаточную глубину проникновения луча для заданной базы наблюдения (расстояние между источником и приемником). Функция $V_0(z)$ считается известной, а коэффициенты A и B заданы; подлежит определению функция $V_1(x, y, z)$.

В дальнейших рассуждениях используется метод линеаризации обратной кинематической задачи для многомерных сред, который последовательно применяется начиная с работ В.Г. Романова [5]. Это позволяет прийти к формуле, являющейся следствием применения линеаризации к функционалу Ферма [10], который связывает кинематические характеристики распространения сигнала:

$$T_1(S_0, S_1) = \int_{\Gamma_0(S_0, S_1)} n_1 dS, \quad (1)$$

где $T_1 = T - T_0$; $n_1 = V^{-1} - V_0^{-1}$; $\Gamma_0(S_0, S_1)$ – геодезическая линия для среды со скоростным распределением V_0 . Поскольку значения T считаются известными (как результат решения прямой задачи, а в практическом применении это вектор измерений), значения T_0 для $V_0 = A + Bz$ можно вычислить явно, что упрощает вычисление требуемых значений T_1 . На этом этапе задача сводится к определению функции n_1 по функции T_1 с использованием интегрального уравнения (1).

Применение системы наблюдений в форме окружности не только устраняет переопределенность задачи, но и задает томографическую постановку для исследуемой задачи. Лучи Γ_0 , «натянутые» на окружность системы наблюдений, формируют поверхность шарового сегмента. Изменяя радиус r , можно получить систему вложенных шаровых сегментов, которые заполняют объем исследуемой области в пространстве R^3 . Определяя n_1 на поверхности таких сегментов, получаем решение трехмерной задачи. Следует отметить, что послойный подход к изучению объекта и методика сбора проекционных данных делают эту задачу сопоставимой с классическими задачами томографии [1].

Пусть функция $T_1(S_0, S_1)$ известна и выполняется равенство (1), где $n_1(x, z) = 1/V(x, z) - 1/V_0(z)$, $x \in R^2$. Точки S_0 и S_1 располагаются на окружности $x_1^2 + x_2^2 = r^2$ ($x_1 \equiv x$, $x_2 \equiv y$). Лучевая траектория $\Gamma_0(S_0, S_1)$ определяется из системы уравнений

$$\begin{cases} |x|^2 + \left(z + \frac{A}{B}\right)^2 = \rho^2, \\ p = \text{const}, \rho = \text{const}, \rho > 0, \\ \langle x, v \rangle - p = 0, v = (\sin \theta, \cos \theta). \end{cases} \quad (2)$$

Последние два уравнения описывают плоскость Q , перпендикулярную к плоскости $Z = 0$ (координатная плоскость x_1, x_2) и пересекающую плоскость по прямой $l(S_0, S_1)$. Геометрический смысл величин p, ρ и θ показан на рис. 1.

Перейдем в выражении (1) от интегрирования по $\Gamma_0(S_0, S_1)$ к интегрированию вдоль $l(S_0, S_1)$. Обозначим

$$z = -\frac{A}{B} \pm \sqrt{\rho^2 - |x|^2} = z^*(x). \quad (3)$$

Будем рассматривать случай $z \geq 0$, тогда перед корнем следует выбрать знак «+». В операторном виде решение задачи дается следующей формулой обращения:

$$n_1(x_1, x_2, z^*) = \left(z^* + \frac{A}{B}\right) R^{-1} f(p, \theta), \quad (4)$$

где R^{-1} – оператор обратного преобразования Радона. Определяя искомую функцию n_1 по формуле обращения (4) в круге $x_1^2 + x_2^2 \leq r^2$, тем самым определяем n_1 на поверхности шарового сегмента, образованного лучами Γ_0 , опирающимися на окружность системы наблюдений [7].

Вывод формулы (4) делает возможным использование численных методов обращения преобразования Радона для решения обратной кинематической задачи. Данная техника, хорошо разработанная для решения других научных задач [6], значительно упрощает работу с данной формулой.

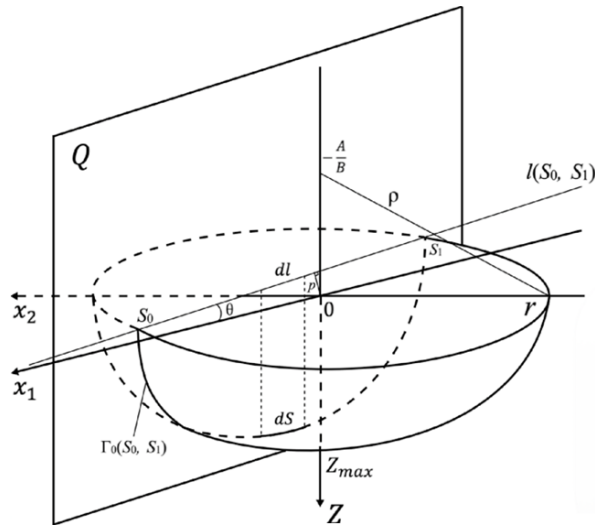


Рис. 1. Иллюстрация геометрических условий линейаризованной постановки обратной задачи
Fig. 1. Illustration of geometric conditions of the linearized formulation of the inverse problem

В результате полного сбора данных формируется проекционная матрица, столбцы которой представляют различные проекции. Таким образом, проекционная матрица имеет размерность $(N \times M)$, где N – количество строк, а M – количество столбцов (элементов в каждой строке). Особенностью данного подхода является то, что первый и последний элементы каждой проекции равны нулю, поскольку в этих случаях источник и приемник совпадают (см. рис. 1). Следовательно, фактически требуется на две пары источник–приемник меньше. В проекционной матрице нулевыми будут первая и последняя строки. Таким образом, формируются проекционные данные $f_{ij} = (p_i, \theta_j)$, $i = 1, N, j = 1, M$, подаваемые на вход программного модуля для решения обратной задачи.

Для решения обратной задачи был применен алгебраический подход, известный в томографии как алгебраическая реконструкция, который реализуется после предварительной алгебраизации преобразования Радона, как указано в формуле (4) [3]. В соответствии с этим подходом были проведены численные эксперименты для восстановления локальной неоднородности скорости в трехмерном пространстве.

Моделировалась локальная неоднородность скорости $V_a(x, y, z) = W \exp(-w_x(x - X_a)^2 - w_y(y - Y_a)^2 - w_z(z - Z_a)^2)$, включенная в среду с линейным законом изменения скорости с глубиной $V = A(1 + \alpha z)$. Здесь $w_x = 3$, $w_y = 3$, $w_z = 20$, $X_a = 0$, $Y_a = 0$, $Z_a = 0,1$, $\alpha = 0,5 \text{ км}^{-1}$, $A = 1 \text{ км/с}$. Система наблюдения представляет собой набор концентрических окружностей с радиусами от 250 до 1 250 м с шагом 250 м. В системе используется 20 источников и приемников, а также 8 проекций. Сетка восстановления имеет размер 20×20 для всех радиусов исследования. Значения восстанавливаемой скорости привязываются к координате погружения луча по глубине в условиях отсутствия неоднородности.

Восстановленная функция локальной неоднородности в проекции на плоскость XY представлена на рис. 2, а, в то время как точные значения скорости на поверхности, составленной лучами при отсутствии неоднородности, показаны на рис. 2, б. Оси координат x и y выполнены в одном масштабе. Пример иллюстрирует случай для радиуса $r = 750 \text{ м}$.

На рис. 3 представлен разрез восстановленной трехмерной аномалии скорости по линии $y = 0$. Глубина z отображается по вертикальной оси, а координата x – по горизонтальной оси. В табл. 1 представлена зависимость скорости зондирующего сигнала от номера изолинии. Точность восстановления можно оценить с помощью среднеквадратичной нормы отклонения Q , которая измеряет различие между найденным решением и известным точным значением (5):

$$Q = \sqrt{\frac{\sum_{i,j} [n_{1r}(x_{1i}, x_{2j}, z^*) - n_{1e}(x_{1i}, x_{2j}, z^*)]^2}{\sum_{i,j} n_{1e}^2(x_{1i}, x_{2j}, z^*)}}, \quad (5)$$

где $z^* = z^*(x_{1i}, x_{2j})$, (i и j – текущие параметры сетки, в узлах которой вычисляются значения n_{1r}).

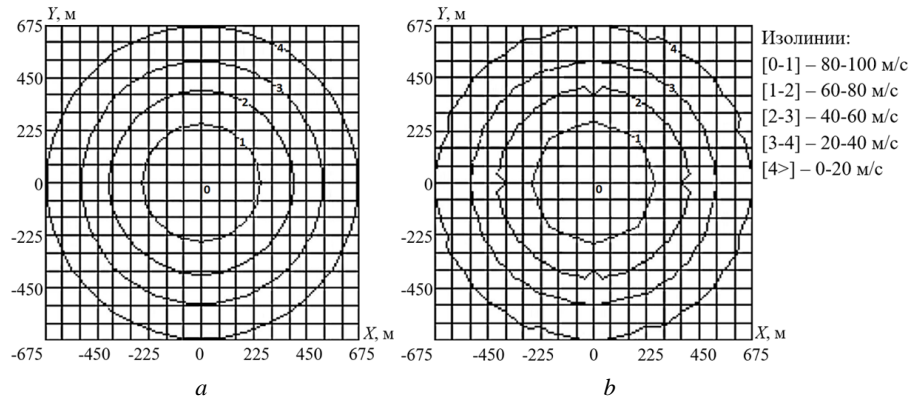


Рис. 2. Восстановленная функция локальной неоднородности в проекции на плоскость XY (a); точные значения скорости на поверхности, составленной лучами при отсутствии неоднородности (b)

Fig. 2. The reconstructed function of local inhomogeneity in projection onto the XY plane – a ; the exact values of velocity on the surface composed of rays in the absence of inhomogeneity – b

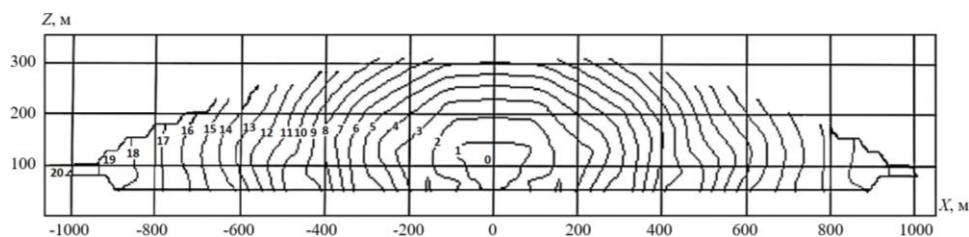


Рис. 3. Разрез восстановленной трехмерной аномалии скорости по линии $y = 0$

Fig. 3. Cross-section of restored three-dimensional speed anomaly along the line $y = 0$

Таблица 1

Зависимость скорости зондирующего сигнала от номера изолинии

Изолинии	0–2	2–4	4–6	6–8	8–10	10–12	12–14	14–16	16–18	18–20
Скорость, м/с	90–100	80–90	70–80	60–70	50–60	40–50	30–40	20–30	10–20	0–10

График зависимости Q от радиуса окружности представлен на рис. 4.

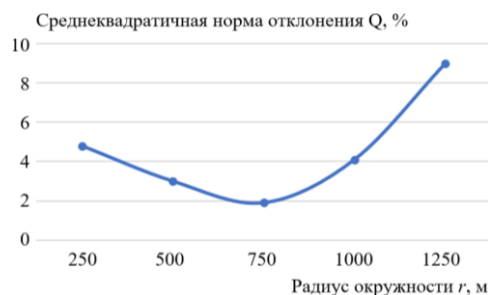


Рис. 4. Зависимость среднеквадратичной нормы отклонения Q от радиуса окружности

Fig. 4. Dependence of the root mean square deviation Q on the radius of the circle

Резкие отклонения при малых и больших радиусах объясняются изменением траектории луча. Моделируемая приповерхностная неоднородность с повышенной скоростью приводит к концентрации лучей в области экстремума, в то время как точное решение вычисляется в точках пространства, свободных от неоднородностей.

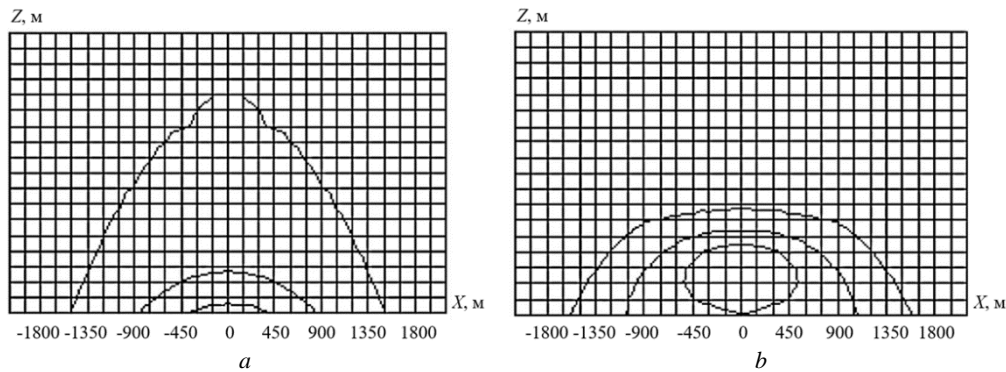


Рис. 5. Относительная насыщенность лучами сетки восстановления, рассчитанная для случая профильных наблюдений при отсутствии неоднородности (a) и при наличии неоднородности (b)

Fig. 5. Relative saturation of the recovery grid rays, calculated for the case of profile observations in the absence of inhomogeneity a) and in the presence of inhomogeneity b)

На рис. 5, a представлена картина относительной насыщенности лучами сетки восстановления, рассчитанная для случая профильных наблюдений при отсутствии неоднородности, а на рис. 5, b – при наличии неоднородности. По горизонтальной оси отображаются координаты профиля, а по вертикальной – координаты глубины сетки разбейения.

Заключение

Излагаемое решение ОКЗ обеспечивается наличием априорной информации об исследуемом объекте, т.е. о скоростном распределении акустического или сейсмического сигнала. Вывод формулы обращения для функционала Ферма требует знания «пути» выполнения интегрирования, эта информация объективно отсутствует в результате измерения времен прихода рефрагированных волн. Очевидно, что отмеченная информация служит основой для алгебраизации функционала Ферма и сведения решения ОКЗ к решению СЛАУ той или иной степени сложности. Таким образом, следует сделать вывод о доминирующем характере используемой априорной информации, что в терминах ИИ определяет разработанный метод решения ОКЗ как «сильный» метод. Естественно, что алгоритмическое обеспечение открыто для потребления дополнительной информации, в том числе возникающей при реализации итерационного подхода, а именно когда выполняется уточнение, «улучшение» выбора V_0 . То есть происходит организация последовательности V_1^i , где i – номер итерации, выполняющей полный цикл решения ОКЗ, построение $V_0^{i+1} = V_0^i + V_1^i$, численное решение прямой кинематической задачи, т.е. вычисление значений T_0^{i+1} и численное построение лучей Γ_0^{i+1} . Алгоритмическое обеспечение этой методики и его программная реализация составляют перспективу дальнейшей разработки данного подхода к решению ОКЗ.

Следует сказать, что в целом алгебраическая реконструкция [11] позволяет решать задачи с нарушением полноты проекционных данных, делающим невозможным использование формул обращения. Сказанное является актуальным для задач сейсмической диагностики сред, содержащих непрозрачные включения для зондирующего сигнала, в том числе имеющие геометрически избирательную чувствительность к этому излучению [12].

Список источников

1. Папкина Ю.И. Звуковое поле в морском волноводе с неоднородной скоростью звука по глубине и трассе // Вестник Московского университета. Сер. 3. Физика. Астрономия. 2021. № 3. С. 27–35.
2. Греков А.Н., Греков Н.А., Сычев Е.Н. Среднечастотные акустические методы и средства для исследования водной среды. Севастополь : Ин-т природно-технических систем, 2020. 126 с.
3. Микушин И.И. Метрологическое обеспечение измерений скорости звука в воде. СПб. : Изд-во СПбГЭТУ, 2023. 205 с.
4. Добринский В.И., Лаврентьев М.М. Способ определения профиля скорости звука по записи акустического поля в одной точке // Тезисы докладов второй Дальневосточной акустической конференции «Человек и океан». Владивосток : ДВПИ, 1978. С. 12–14.
5. Лаврентьев М.М., Романов В.Г., Шишатский С.П. Некорректные задачи математической физики и анализа. М. : Наука, 1980. 286 с.

6. Тихонов А.Н., Арсенин В.Я., Тимонов А.А. Математические задачи компьютерной томографии. М. : Наука, 1987. 159 с.
7. Гервас Н.В., Зеркаль С.М. Вычислительная кинематическая диагностика скорости зондирующего сигнала с использованием априорной информации об исследуемой среде // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2024. № 69. С. 54–63. doi: 10.17223/19988605/69/6
8. Нолет Г., Чепмен К., Слуйс А. Сейсмическая томография. М. : Мир, 1990. 416 с.
9. Пузырев Н.Н. Метод и объекты сейсмических исследований. Введение в общую сейсмологию. Новосибирск: Изд-во СО РАН, НИЦ ОИГТМ, 1997. 301 с.
10. Смирнов В.И. Курс высшей математики. М. : Наука. 1974. Т. 4, ч. 1. 336 с.
11. Бронников А.В., Воскобойников Ю.Е. Комбинированные алгоритмы нелинейной фильтрации защищенных сигналов и изображений // Автотометрия. 1990. № 1. С. 55–66.
12. Peshkov A.V., Zerkal S.M. Computational-heuristic algorithm for tomographic solution of industrial flaw detection problems // Proceedings 16 Int. Scientific and Technical Conference Actual Problems of Electronic Instrument Engineering (APEIE–2023). Novosibirsk, 10–12 Nov. 2023. P. 910–915. doi: 10.1109/APEIE59731.2023.10347879

References

1. Papkova, Yu.I. (2021) Zvukovoe pole v morskoy volnovode s neodnorodnoy skorost'yu zvuka po glubine i trasse [Sound Field in a Marine Waveguide with Inhomogeneous Sound Speed Along Depth and Path]. *Vestnik Moskovskogo universiteta. Ser. 3. Fizika. Astronomiya*. 3. pp. 27–35.
2. Grekov, A.N., Grekov, N.A. & Sychev, E.N. (2020) *Srednechastotnye akusticheskie metody i sredstva dlya issledovaniya vodnoy sredy* [Mid-frequency Acoustic Methods and Means for Studying the Aquatic Environment]. Sevastopol: In-t prirodno-tekhnicheskikh sistem.
3. Mikushin, I.I. (2023) *Metrologicheskoe obespechenie izmereniy skorosti zvuka v vode* [Metrological Support for Measurements of Sound Speed in Water]. St. Petersburg: SPbGETU.
4. Dobrinskiy, V.I. & Lavrentiev, M.M. (1978) Sposob opredeleniya profilya skorosti zvuka po zapisi akusticheskogo polya v odnoy tochke [Method for Determining Sound Speed Profile from Acoustic Field Recording at a Single Point]. *Chelovek i okean* [Human and Ocean]. Proc. of the Conference. Vladivostok: DVPI. pp. 12–14.
5. Lavrentiev, M.M., Romanov, V.G. & Shishatskiy, S.P. (1980) *Nekorrektnye zadachi matematicheskoy fiziki i analiza* [Ill-posed Problems of Mathematical Physics and Analysis]. Moscow: Nauka.
6. Tikhonov, A.N., Arsenin, V.Ya., Timonov, A.A. (1987) *Matematicheskiye zadachi komp'yuternoy tomografii* [Mathematical Problems of Computer Tomography]. Moscow: Nauka.
7. Gervas, N.V. & Zerkal, S.M. (2024) Computational kinematic diagnostics of the probing signal speed using a priori information about the medium under study. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 69. pp. 54–63. doi: 10.17223/19988605/69/6
8. Nolet, G., Chapman, K. & Sluis, A. (1990) *Seismicheskaya tomografiya* [Seismic Tomography]. Translated from English. Moscow: Mir.
9. Puzyrev, N.N. (1997) *Metod i ob"ekty seysmicheskikh issledovaniy. Vvedenie v obshchuyu seysmologiyu* [Method and Objects of Seismic Research. Introduction to General Seismology]. Novosibirsk: SB RAS.
10. Smirnov, V.I. (1974) *Kurs vysshey matematiki* [Course of Higher Mathematics]. Vol. 4(1). Moscow: Nauka.
11. Bronnikov, A.V. & Voskoboynikov, Yu.E. (1990) Kombinirovannyye algoritmy nelineynoy fil'tratsii zashchishchennykh signalov i izobrazheniy [Combined Algorithms for Nonlinear Filtering of Protected Signals and Images]. *Avtometriya*. 1. pp. 55–66.
12. Peshkov, A.V. & Zerkal, S.M. (2023) Computational-heuristic algorithm for tomographic solution of industrial flaw detection problems. In: *Proceedings of the 16th International Scientific and Technical Conference Actual Problems of Electronic Instrument Engineering (APEIE–2023)*. Novosibirsk, November 10–12, 2023. pp. 910–915. doi: 10.1109/APEIE59731.2023.10347879

Информация об авторах:

Гервас Николай Владимирович – аспирант кафедры вычислительной техники Новосибирского государственного технического университета (Новосибирск, Россия). E-mail: nik.gervas@mail.ru

Зеркаль Сергей Михайлович – доктор технических наук, профессор кафедры вычислительной техники Новосибирского государственного технического университета (Новосибирск, Россия). E-mail: zerkal@ngs.ru

Вклад авторов: все авторы сделали эквивалентный вклад в подготовку публикации. Авторы заявляют об отсутствии конфликта интересов.

Information about the authors:

Gervas Nikolai V. (Post-Graduate Student, Novosibirsk State Technical University, Novosibirsk, Russian Federation). E-mail: nik.gervas@mail.ru

Zerkal Sergey M. (Doctor of Technical Sciences, Professor, Novosibirsk State Technical University, Novosibirsk, Russian Federation). E-mail: zerkal@ngs.ru

Contribution of the authors: the authors contributed equally to this article. The authors declare no conflicts of interests.

Поступила в редакцию 28.04.2025; принята к публикации 02.12.2025

Received 28.04.2025; accepted for publication 02.12.2025

Original article

UDK 369:519.2

doi: 10.17223/19988605/73/8

Nonparametric evaluation of continuous r -year deferred m -year term life annuity using information on probabilistic characteristics of lifetime**Yury G. Dmitriev¹, Gennady M. Koshkin²**^{1, 2} National Research Tomsk State University, Tomsk, Russian Federation¹ dmit@mail.tsu.ru² kgm@mail.tsu.ru

Abstract. The theory of pension annuities is closely related to the ideology of net premiums of the life insurance theory. The mathematical theory of insurance is widely used to solve many problems that are determined by the requirements of the market economy. The requirements of practice stimulate the development of insurance theory and the closely related theory of annuities and force researchers to turn to more complex mathematical models in this area. New methods of calculating annuities appear that reduce the time for making optimal decisions in the absence of sufficient information about the markets of new types of pension services. The article considers the problem of estimating continuous r -year deferred m -year term life annuity with making use of information on probabilistic characteristics of lifetime. Insurance companies often offer their clients to conclude contracts of r -year deferred m -year annuities. Nonparametric estimators of life annuities are constructed from data on the lifetimes of individuals. Found the principal terms of the asymptotic mean squared errors (MSEs) of the proposed estimators; their asymptotic normality is proved. It is shown that the use of auxiliary information often leads to a lower MSE of the modified estimator compared to the MSE of the traditional estimator. An adaptive estimator is proposed that can be applied in practice.

Keywords: r -year deferred m -year term life annuity; nonparametric evaluation; auxiliary information; mean squared error; asymptotic normality; adaptive estimator.

For citation: Dmitriev, Yu.G., Koshkin, G.M. (2025) Nonparametric evaluation of continuous r -year deferred m -year term life annuity using information on probabilistic characteristics of lifetime. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naja tehnika i informatika – Tomsk State University Journal of Control and Computer Science*. 73. pp. 65–74. doi: 10.17223/19988605/73/8

Научная статья

doi: 10.17223/19988605/73/8

Непараметрическое оценивание непрерывной m -летней временной ренты, отсроченной на r лет, с учетом информации о вероятностных характеристиках продолжительности жизни**Юрий Глебович Дмитриев¹, Геннадий Михайлович Кошкин²**^{1, 2} Национальный исследовательский Томский государственный университет, Томск, Россия¹ dmit@mail.tsu.ru² kgm@mail.tsu.ru

Аннотация. Теория пенсионных рент тесно связана с идеологией нетто-премий теории страхования жизни. Математическая теория страхования широко используется при решении многих задач, которые определяются требованиями рыночной экономики. Требования практики стимулируют развитие теории страхования и тесно связанную с ней теорию рент и вынуждают исследователей обращаться к более сложным математическим моделям в указанной области. Появляются новые методы расчета рент, которые сокращают время принятия оптимальных решений в условиях отсутствия достаточной информации о рынках новых видов пенсионных услуг. В статье рассматривается проблема оценивания непрерывной m -летней временной ренты, отсроченной на r лет, с учетом информации о вероятностных характеристиках продолжительности жизни. Страховые компании часто предлагают клиентам заключать договоры именно m -летних рент, отсроченных на r лет. Непараметрические

оценки рент строятся по данным продолжительностей жизни индивидуумов. Найдены главные части асимптотических среднеквадратичных ошибок (СКО) предложенных оценок, доказана их асимптотическая нормальность. Показано, что использование дополнительной информации часто приводит к меньшей СКО модифицированной оценки по сравнению с СКО традиционной оценки. Предлагается адаптивная оценка, которая может применяться на практике.

Ключевые слова: m -летняя рента, отсроченная на r лет; непараметрическое оценивание; дополнительная информация; среднеквадратическая ошибка; асимптотическая нормальность; адаптивная оценка.

Для цитирования: Дмитриев Ю.Г., Кошкин Г.М. Непараметрическое оценивание непрерывной m -летней временной ренты, отсроченной на r -лет, с учетом информации о вероятностных характеристиках продолжительности жизни // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2025. № 73. С. 65–74. doi: 10.17223/19988605/73/8

Introduction

The current stage of development of public socio-economic relations requires non-trivial approaches to the ideology of calculating pension annuities [1. P. 13–46, 170–194], which is associated with:

- the impact on the insurance market of such unpredictable phenomena as epidemics, natural disasters, social cataclysms, etc. [2–4],
- the emergence of new types of insurance and pension services [5, 6].

Let X be the age of an individual and at the moment $t = 0$ payments start. The idea of the r -year deferred m -year term life annuity in accordance with [1. P. 150] is this: from the moment $t+r = r$, an individual starts receiving a monetary amount once a year, but payments are made not more than m years. For ease of calculation, such a monetary amount is taken as equal to a conventional unit. So, payments are making in the time interval $(r, r+m]$. It is known that the life annuity is associated with the appropriate type of insurance. Thus, the average total cost of the present continuous r -year deferred m -year term life annuity gives the following formula (see [1. P. 184]):

$${}_r|\bar{a}_{x:\overline{m}|}(\delta) = \frac{1 - {}_r|\bar{A}_{x:\overline{m}|}}{\delta},$$

where ${}_r|\bar{A}_{x:\overline{m}|} = \int_{x+r}^{x+r+m} e^{-\delta t} f_x(t) dt$ is a net premium (the expectation of the present value of an insured unit sum

for an r -year deferred m -year term life insurance at age x), δ is a force of interest, $f_x(t) = \frac{f(x+t)}{S(x)}$ is a prob-

ability density of the future lifetime $T_x = X - x$ of an individual (x) [1. P. 62], $f(x)$ is a probability density of lifetime X of an individual (x), $S(x) = P(X > x)$ is a survival function. The essence of the present continuous r -year deferred m -year term life annuity is as follows: a client of age x who has entered into an agreement transfers to the company the sum of ${}_r|\bar{a}_{x:\overline{m}|}(\delta)$ conventional monetary units; then the company will pay one conventional monetary unit every year throughout the time interval $(r, r+m]$. It is clear that ${}_r|\bar{a}_{x:\overline{m}|}(\delta) > 1$, and the value of the rent ${}_r|\bar{a}_{x:\overline{m}|}^N(\delta)$ increases with the growth of m .

Introduce the random variable

$$z(x) = \frac{1 - e^{-\delta T_x}}{\delta}, \quad r + m \geq T_x > r. \quad (1)$$

Then, averaging the random variable $z(x)$ (1) (see [7–9]), we get the formula of the r -year deferred m -year term life annuity:

$${}_r|\bar{a}_{x:\overline{m}|}(\delta) = E(z) = \frac{1}{\delta} \left(1 - \frac{\Phi(x, r, m, \delta)}{S(x)} \right). \quad (2)$$

Here E is the symbol of the mathematical expectation,

$$\Phi(x, r, m, \delta) = e^{\delta x} \int_{x+r}^{x+r+m} e^{-\delta t} dF(t),$$

$F(x) = P(X \leq x) = 1 - S(x)$ is a distribution function.

Note that the whole life annuity $\bar{a}_x(\delta)$ [7] is the special case of the annuity (2) at $r = 0$ and $m = \infty$.

Let us consider the problem of estimating continuous life annuities based on a sample of individuals' life expectancies [10–12]. The use of classical methods of statistical data processing often does not allow obtaining adequate models on the basis of which the insurance company's development strategy is built. When using classical parametric estimates and models, information about the phenomenon being studied is required with an accuracy of up to unknown parameters. In practice, problems often arise with the selection of suitable parametric estimates and models. Data processing using nonparametric statistical methods allows synthesizing simple and adequate (with known statistical properties) estimates and models in conditions where information about the phenomenon being studied is of a general nature [13].

1. Construction of the r -year deferred m -year term annuity estimator

Assume we have a random sample X_1, \dots, X_N of N individuals' lifetimes. Using the empirical survival function

$$S_N(x) = \frac{1}{N} \sum_{i=1}^N I(X_i > x),$$

where $I(A)$ is the indicator of an event A , we obtain the following estimator of (2):

$${}_r|\bar{a}_{x:\overline{m}|}^N(\delta) = \frac{1}{\delta} \left(1 - \frac{e^{\delta x}}{S_N(x) \cdot N} \sum_{i=1}^N \exp(-\delta X_i) I(x+r+m \geq X_i > x+r) \right) = \frac{1}{\delta} \left(1 - \frac{\Phi_N(x, r, m, \delta)}{S_N(x)} \right), \quad (3)$$

$$\Phi_N(x, r, m, \delta) = \frac{e^{\delta x}}{N} \sum_{i=1}^N \exp(-\delta X_i) I(x+r+m \geq X_i > x+r).$$

2. Bias and mean squared error of estimator (3)

In this section, we will obtain the principal term of the asymptotic MSE and the bias convergence rate of the estimator ${}_r|\bar{a}_{x:\overline{m}|}^N(\delta)$.

Introduce the notation (see [14]): the function $H(t): R^s \rightarrow R^1$, $t = t(x) = (t_1(x), \dots, t_s(x))$, is an s -dimensional bounded function; $H_j(t) = \frac{\partial H(t)}{\partial t_j}$, $j = \overline{1, s}$, $\nabla H(t) = (H_1(t), \dots, H_s(t))$; the symbol T denotes the transpose; $t_N = (t_{1N}, \dots, t_{sN})$ is an s -dimensional statistic, $t_{jN} = t_{jN}(x) = t_{jN}(x; X_1, \dots, X_N)$, $j = \overline{1, s}$; $\|t_N\| = \sqrt{t_{1N}^2 + \dots + t_{sN}^2}$ is the Euclidean norm of t_N ; $\Rightarrow N_s\{\mu, \sigma\}$ is the symbol of weak convergence of sequence of random variables to the s -dimensional normal random variable with a mean $\mu = (\mu_1, \dots, \mu_s)$ and symmetric covariance matrix $\sigma = \|\sigma_{ij}\|$, $0 < \sigma_{jj} = \sigma_{jj}(x) < \infty$, $j = \overline{1, s}$; \Re is the set of the integers.

Definition 1. The function $H(t): R^s \rightarrow R^1$ and the sequence $\{H(t_N)\}$ are said to belong to the class $N_{v,s}(t; \gamma)$, provided that

1) there exists an ε -neighborhood $\{z: |z_i - t_i| < \varepsilon, i = \overline{1, s}\}$, in which the function $H(z)$ and all its partial derivatives $\frac{\partial H(z)}{\partial z_j}$ up to the order v are continuous and bounded;

2) for any values of variables X_1, \dots, X_N the sequence $\{H(t_N)\}$ is dominated by a numerical sequence $C_0 d_N^\gamma$, such that $d_N \uparrow \infty$, as $N \rightarrow \infty$, and $0 \leq \gamma < \infty$.

We present Theorem 1 from [14].

Theorem 1. Let the conditions 1) $H(z)$, $\{H(t_N)\} \in N_{2,s}(t; \gamma)$ and 2) $E\|t_N - t\|^i = O(d_N^{-i/2})$ hold for all $i \in \mathfrak{R}$. Then, for every $k \in \mathfrak{R}$,

$$\left| E[H(t_N) - H(t)]^k - E[\nabla H(t) \cdot (t_N - t)^T]^k \right| = O(d_N^{-(k+1)/2}). \quad (4)$$

If in formula (4) $k=1$, we obtain the principal term $E[\nabla H(t) \cdot (t_N - t)^T]$ of the bias $E[H(t_N) - H(t)]$ for $H(t_N)$, and at $k=2$, we have the principal term $E[\nabla H(t) \cdot (t_N - t)^T]^2$ of the MSE $E[H(t_N) - H(t)]^2$.

Theorem 2. If the survival function $S(x) > 0$ and $S(t)$ is continuous at a point x , then

1) for the bias $b\left({}_r|\bar{a}_{x:\bar{m}}^N(\delta)\right)$ of estimator (3) we have

$$\left| b\left({}_r|\bar{a}_{x:\bar{m}}^N(\delta)\right) \right| = \left| E\left({}_r|\bar{a}_{x:\bar{m}}^N(\delta) - {}_r|\bar{a}_{x:\bar{m}}(\delta)\right) \right| = O(N^{-1});$$

2) the MSE $u^2\left({}_r|\bar{a}_{x:\bar{m}}^N(\delta)\right)$ is given by the formula

$$u^2\left({}_r|\bar{a}_{x:\bar{m}}^N(\delta)\right) = E\left({}_r|\bar{a}_{x:\bar{m}}^N(\delta) - {}_r|\bar{a}_{x:\bar{m}}(\delta)\right)^2 = \frac{\Phi(x, r, m, 2\delta) - \Phi^2(x, r, m, \delta) / S(x)}{N\delta^2 S^2(x)} + O\left(\frac{1}{N^{3/2}}\right).$$

Proof. For the estimator ${}_r|\bar{a}_{x:\bar{m}}^N(\delta)$ (3) in the notation of Theorem 1, we have: $s=2$; $d_N=N$;

$$t_N = (t_{1N}, t_{2N}) = (\Phi_N(x, r, m, \delta), S_N(x)); \quad t = (t_1, t_2) = (\Phi(x, r, m, \delta), S(x));$$

$$H(t_N) = \frac{1}{\delta} \left(1 - \frac{t_{1N}}{t_{2N}} \right) = \frac{1}{\delta} \left(1 - \frac{\Phi_N(x, r, m, \delta)}{S_N(x)} \right) = {}_r|\bar{a}_{x:\bar{m}}^N(\delta);$$

$$H(t) = \frac{1}{\delta} \left(1 - \frac{t_1}{t_2} \right) = \frac{1}{\delta} \left(1 - \frac{\Phi(x, r, m, \delta)}{S(x)} \right) = {}_r|\bar{a}_{x:\bar{m}}(\delta); \quad H_1(t) = -\frac{1}{\delta S(x)}, \quad H_2(t) = \frac{\Phi(x, r, m, \delta)}{\delta S^2(x)},$$

$$\nabla H(t) = (H_1(t), H_2(t)) = \left(-\frac{1}{\delta S(x)}, \frac{\Phi(x, r, m, \delta)}{\delta S^2(x)} \right) \neq 0.$$

The sequence $\{H(t_N)\}$ satisfies the condition 1) of Theorem 1 with $C_0 = \frac{1}{\delta}(1 + e^{-\delta r})$ and $\gamma=0$:

$$\begin{aligned} |H(t_N)| &= \frac{1}{\delta} \left| 1 - \frac{\Phi_N(x, r, m, \delta)}{S_N(x)} \right| \leq \frac{1}{\delta} \left(1 + \frac{\Phi_N(x, r, m, \delta)}{S_N(x)} \right) \leq \frac{1}{\delta} \left(1 + \frac{e^{\delta x} \sum_{i=1}^N \exp(-\delta X_i) \mathbf{I}(x+r+m \geq X_i > x+r)}{\sum_{i=1}^N \mathbf{I}(X_i > x)} \right) \leq \\ &\leq \frac{1}{\delta} \left(1 + \frac{e^{\delta x} e^{-\delta(x+r)} \sum_{i=1}^N \mathbf{I}(x+r+m \geq X_i > x+r)}{\sum_{i=1}^N \mathbf{I}(X_i > x)} \right) \leq \frac{1}{\delta} (1 + e^{-\delta r}). \end{aligned}$$

Further, in view of $t_2 = S(x) > 0$, the function $H(t)$ satisfies the condition 1) of Theorem 1. Also, this function satisfies the condition 2) of Theorem 1 due to Lemma 3.1 [15], as for all $i \in \mathfrak{R}$ such inequalities hold:

$$\begin{aligned} E\left[e^{i\delta x} e^{-i\delta X} \Gamma^i(x+r+m \geq X > x+r) \right] &\leq e^{i\delta x} e^{-i\delta(x+r)} [S(x+r) - S(x+r+m)] = \\ &= e^{-i\delta r} [S(x+r) - S(x+r+m)] \leq 1, \quad E\left[\Gamma^i(X > x) \right] = S(x) \leq 1. \end{aligned}$$

It is well known that $S_N(x)$ is an unbiased and consistent estimator of $S(x)$. Show that $\Phi_N(x, r, m, \delta)$ is an unbiased estimator of $\Phi(x, r, m, \delta)$:

$$E\Phi_N(x, r, m, \delta) = \frac{e^{\delta x}}{N} E \left[\sum_{i=1}^N \exp(-\delta X_i) I(x + r + m \geq X_i > x + r) \right] = \Phi(x, r, m, \delta).$$

The ratio of two unbiased estimators can have a bias. All conditions of Theorem 1 are satisfied and $E[\nabla H(t)(t_N - t)^T] = 0$, therefore, in accordance with (4), we obtain the order of the bias $b\left({}_r\bar{a}_{x:\overline{m}|}^N(\delta)\right)$:

$$\left| E\left({}_r\bar{a}_{x:\overline{m}|}^N(\delta) - {}_r\bar{a}_{x:\overline{m}|}(\delta)\right) - E[\nabla H(t)(t_N - t)^T] \right| = \left| E\left({}_r\bar{a}_{x:\overline{m}|}^N(\delta) - {}_r\bar{a}_{x:\overline{m}|}(\delta)\right) \right| = O(N^{-1}).$$

Now, calculate the variance of estimator $\Phi_N(x, r, m, \delta)$:

$$\begin{aligned} D\Phi_N(x, r, m, \delta) &= D \left[\frac{e^{\delta x}}{N} \sum_{i=1}^N \exp(-\delta X_i) I(x + r + m \geq X_i > x + r) \right] = \\ &= \frac{e^{2\delta x}}{N^2} \sum_{i=1}^N D[\exp(-\delta X_i) I(x + r + m \geq X_i > x)] = \frac{1}{N} (\Phi(x, r, m, 2\delta) - \Phi^2(x, r, m, \delta)). \end{aligned}$$

Similarly, we find the covariance matrix $\Omega_{2 \times 2}(x, r, m, \delta) = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}$ for statistics $S_N(x)$ and

$\Phi_N(x, r, m, \delta)$:

$$\sigma_{11} = ND[\Phi_N(x, r, m, \delta)] = \Phi(x, r, m, 2\delta) - \Phi^2(x, r, m, \delta); \quad \sigma_{22} = ND[S_N(x)] = S(x)(1 - S(x));$$

$$\sigma_{12} = \sigma_{21} = N \text{cov}(S_N(x), \Phi_N(x, r, m, \delta)) =$$

$$= N[E[S_N(x)\Phi_N(x, r, m, \delta)] - E[S_N(x)]E[\Phi_N(x, r, m, \delta)]] = (1 - S(x))\Phi(x, r, m, \delta).$$

Using the previous results on the bias and the covariance matrix, we obtain

$$\begin{aligned} u^2\left({}_r\bar{a}_{x:\overline{m}|}^N(\delta)\right) &= E[\nabla H(t)(t_N - t)^T]^2 + O\left(\frac{1}{N^{3/2}}\right) = \\ &= H_1^2(t)\sigma_{11} + H_2^2(t)\sigma_{22} + 2H_1(t)H_2(t)\sigma_{12} + O\left(\frac{1}{N^{3/2}}\right) = \frac{W(x, r, m, \delta)}{N} + O\left(\frac{1}{N^{3/2}}\right), \end{aligned} \quad (5)$$

$$W(x, r, m, \delta) = \frac{\Phi(x, r, m, 2\delta) - \Phi^2(x, r, m, \delta) / S(x)}{\delta^2 S^2(x)}. \quad (6)$$

The proof of Theorem 2 is completed.

3. Asymptotic normality of estimator (3)

We need Theorem 3 (the central limit theorem from [16]) and Theorem 4 from [14].

Theorem 3. If $\xi_1, \xi_2, \dots, \xi_N, \dots$ is a sequence of independent and identically distributed s -dimensional vectors, $E\{\xi_k\} = 0$, $\sigma = E\{\xi_k^T \xi_k\}$, and $t_N = \frac{1}{N} \sum_{k=1}^N \xi_k$, then, as $N \rightarrow \infty$,

$$\sqrt{N}t_N \Rightarrow N_s\{0, \sigma\}.$$

Theorem 4. If $q_N(t_N - t) \Rightarrow N_s\{\mu, \sigma\}$ for some number sequence $q_N \uparrow \infty$, the function $H(z)$ is differentiable at a point μ , $\nabla H(\mu) \neq 0$, then

$$q_N(H(t_N) - H(\mu)) \Rightarrow N_1\{\mu \nabla H(\mu) \mu^T, \nabla H(\mu) \sigma \nabla H^T(\mu)\}.$$

Theorem 5. Under the conditions of Theorem 2

$$\sqrt{N}\left({}_r\bar{a}_{x:\overline{m}|}^N(\delta) - {}_r\bar{a}_{x:\overline{m}|}(\delta)\right) \Rightarrow N_1\left\{0, \frac{\Phi(x, r, m, 2\delta) - \Phi^2(x, r, m, \delta) / S(x)}{\delta^2 S^2(x)}\right\} = N_1\{0, W(x, r, m, \delta)\}.$$

Proof. In the notation of Theorem 4, we have $s = 2$, $\sigma = \Omega_{2 \times 2}(x, r, m, \delta)$ in accordance with Section 2, and $q_N = \sqrt{N}$. Now, applying Theorem 3, we get

$$\sqrt{N} \left[\Phi_N(x, r, m, \delta), S_N(x) \right] - \left[\Phi(x, r, m, \delta), S(x) \right] \Rightarrow N_2 \left\{ (0, 0), \Omega_{2 \times 2}(x, r, m, \delta) \right\}.$$

The function $H(z)$ is differentiable at the point $t = (\Phi(x, r, m, \delta), S(x))$, $\nabla H(t) \neq 0$. Consequently, all conditions of Theorem 4 hold, and using (5) and (6), we obtain the desired result.

The proof of Theorem 5 is completed.

4. Construction of estimators using information on probabilistic characteristics of lifetime

Let ω be the limiting age, i.e. $S(x) > 0$ for $x < \omega$, and $S(x) = 0$ for $x \geq \omega$. Suppose we know the average of the lifetime functional

$$Eg(X) = \int_0^{\omega} g(x) dF(x) = J, \quad (7)$$

where $g(x)$, $x \in [0, \omega]$, is a known function. If $g(x)$ is the indicator of the set $\{\omega : C_1 \leq g(x) \leq C_2\}$, then one gets the probability $P(C_1 \leq X \leq C_2)$; for $g(x) = x^r$ one comes to the initial moment of order r , and for $g(x) = (x - EX)^r$ – to the central moment of order r , and so on. The estimator using such information can be taken in the following from [17, 18]:

$${}_r \bar{a}_{x:\overline{m}}^N(\delta, \lambda) = \frac{1}{\delta} \left(1 - \frac{\Phi_N(x, r, m, \delta)}{S_N(x)} - \lambda(J_N - J) \right), \quad (8)$$

where $J_N = \frac{1}{N} \sum_{i=1}^N g(X_i)$ is an estimator of J , the parameter λ we will find minimizing the principal term of the asymptotic MSE of ${}_r \bar{a}_{x:\overline{m}}^N(\delta, \lambda)$ (8). Estimator (8) combines the available empirical information containing in (3) and prior information (7).

For estimator ${}_r \bar{a}_{x:\overline{m}}^N(\delta, \lambda)$ in the notation of Theorem 1, we have: $s = 3$, $d_N = N$;

$$t_N = (t_{1N}, t_{2N}, t_{3N}) = (\Phi_N(x, r, m, \delta), S_N(x), J_N); \quad t = (t_1, t_2, t_3) = (\Phi(x, r, m, \delta), S(x), J);$$

$$H(t) = H(t_1, t_2, t_3) = \frac{1}{\delta} \left(1 - \frac{t_1}{t_2} - \lambda(t_3 - J) \right) = \frac{1}{\delta} \left(1 - \frac{\Phi(x, r, m, \delta)}{S(x)} - \lambda(J - J) \right) = {}_r \bar{a}_{x:\overline{m}}^N(\delta, \lambda);$$

$$H(t_N) = \frac{1}{\delta} \left(1 - \frac{\Phi_N(x, r, m, \delta)}{S_N(x)} - \lambda(J_N - J) \right) = {}_r \bar{a}_{x:\overline{m}}^N(\delta, \lambda); \quad (9)$$

$$\nabla H(t) = (H_1(t), H_2(t), H_3(t)) = \left(-\frac{1}{\delta S(x)}, \frac{\Phi(x, r, m, \delta)}{\delta S^2(x)}, -\frac{\lambda}{\delta} \right) \neq 0.$$

5. Bias and MSE of estimators using auxiliary information

Arguing as in the proof of Theorem 2, it is easy to show that the sequence $\{H(t_N)\}$ in (9) satisfies the condition 1) of Theorem 1 with $C_0 = \frac{1}{\delta} (1 + e^{-\delta r} + |\lambda|(K + |J|))$, where $\sup_{x \in [0, \omega]} |g(x)| = K < \infty$, and $\gamma = 0$. Also,

the statistic t_N satisfies the condition 2) due to Lemma 3.1 [15], provided that $Eg^i(X) \leq K^i < \infty$ for all $i \in \mathfrak{R}$.

Hence, given that $E(t_N - t) = 0$, for the bias of (7) we obtain

$$\begin{aligned} & \left| E \left({}_r\bar{a}_{x:\overline{m}|}^N(\delta, \lambda) - {}_r\bar{a}_{x:\overline{m}|}(\delta, \lambda) \right) - E[\nabla H(t)(t_N - t)] \right| = \\ & = \left| E \left({}_r\bar{a}_{x:\overline{m}|}^N(\delta, \lambda) - {}_r\bar{a}_{x:\overline{m}|}(\delta, \lambda) \right) \right| = \left| b \left({}_r\bar{a}_{x:\overline{m}|}^N(\delta, \lambda) \right) \right| = O(N^{-1}). \end{aligned} \quad (10)$$

Now, find the covariance matrix $\Omega_{3 \times 3}(x, r, m, \delta) = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix}$ for statistics $\Phi_N(x, r, m, \delta)$, $S_N(x)$,

and J_N : $\sigma_{33} = NDJ_N = Dg(X)$; $\sigma_{13} = \sigma_{31} = N \text{cov}(J_N, \Phi_N(x, r, m, \delta)) = C_1(x, r, m, \delta) - J\Phi(x, r, m, \delta)$;

$\sigma_{23} = \sigma_{32} = N \text{cov}(S_N(x), J_N) = C_2(x, r, m, \delta) - JS(x)$, where $C_1(x, r, m, \delta) = e^{\delta x} \int_{x+r}^{x+r+m} e^{-\delta u} g(u) dF(u)$,

$C_2(x, r, m) = \int_{x+r}^{x+r+m} g(u) dF(u)$, and σ_{11} , σ_{12} , σ_{21} , σ_{22} are defined in Section 2. Using (4) at $k = 2$, (5), (6),

(9), (10), and covariance matrix $\Omega_{3 \times 3}(x, r, m, \delta)$, we obtain:

$$\begin{aligned} u^2 \left({}_r\bar{a}_{x:\overline{m}|}^N(\delta, \lambda) \right) &= E \left({}_r\bar{a}_{x:\overline{m}|}^N(\delta, \lambda) - {}_r\bar{a}_{x:\overline{m}|}(\delta) \right)^2 = \\ &= E[\nabla H(t)(t_N - t)]^2 + O\left(\frac{1}{N^{3/2}}\right) = \frac{W(x, r, m, \delta, \lambda)}{N} + O\left(\frac{1}{N^{3/2}}\right), \end{aligned} \quad (11)$$

$$\begin{aligned} W(x, r, m, \delta, \lambda) &= \sum_{p=1}^3 \sum_{j=1}^3 H_j(t) \sigma_{jp} H_p(t) = H_1^2(t) \sigma_{11} + H_2^2(t) \sigma_{22} + H_3^2(t) \sigma_{33} + 2H_1(t)H_2(t) \sigma_{12} + \\ &+ 2H_1(t)H_3(t) \sigma_{13} + 2H_2(t)H_3(t) \sigma_{23} = W(x, r, m, \delta) + \frac{\lambda^2 \sigma_{33}}{\delta^2} - \frac{2\lambda H_1 \sigma_{13}}{\delta} - \frac{2\lambda H_2 \sigma_{23}}{\delta} = \\ &= W(x, r, m, \delta) + \lambda^2 Q_1 - 2\lambda Q_2, \end{aligned} \quad (12)$$

where $Q_1 = \frac{\sigma_{33}}{\delta^2} > 0$, $Q_2 = \frac{H_1 \sigma_{13} + H_2 \sigma_{23}}{\delta}$.

Thus, the derived formulas (10)–(12) determine the bias and MSE of the estimate (8) and allow us to formulate the following theorem.

Theorem 6. If the survival function $S(x) > 0$, $S(t)$ is continuous at a point x , $\sup_{x \in [0, \omega]} |g(x)| = K < \infty$, then

1) for the bias of estimator (8) the following relation holds: $\left| b \left({}_r\bar{a}_{x:\overline{m}|}^N(\delta, \lambda) \right) \right| = O(N^{-1})$;

2) the MSE of estimator (8) is given by the formula (11).

The minimum of $W(x, r, m, \delta, \lambda)$ (12) with respect to λ is achieved at $\lambda_0 = Q_2 / Q_1$. Such λ_0 minimizes

the principal term of MSE $u^2 \left({}_r\bar{a}_{x:\overline{m}|}^N(\delta, \lambda) \right)$, and this minimum is as follows:

$$\frac{W(x, r, m, \delta, \lambda_0)}{N} = \frac{1}{N} \left(W(x, r, m, \delta) - \frac{Q_2^2}{Q_1} \right) \leq \frac{W(x, r, m, \delta)}{N}. \quad (13)$$

So, the principal term of MSE (11) at λ_0 is not more than the principal term of MSE (5), and, in accordance with (13), the estimator

$${}_r\bar{a}_{x:\overline{m}|}^N(\delta, \lambda_0) = \frac{1}{\delta} \left(1 - \frac{\Phi_N(x, r, m, \delta)}{S_N(x)} - \lambda_0 (J_N - J) \right) \quad (14)$$

will be called the optimal estimator in the mean square sense. The non-negative quantity Q_2^2 / Q_1 in (13) determines the decrease of the principal term of MSE for the optimal estimator ${}_r\bar{a}_{x:\overline{m}|}^N(\delta, \lambda_0)$ by making use of auxiliary information (7).

Theorem 7. If the conditions of Theorem 6 hold, then

- 1) $\left| b\left({}_r\bar{a}_{x:\bar{m}}^N(\delta, \lambda_0) \right) \right| = O(N^{-1});$
- 2) $u^2\left({}_r\bar{a}_{x:\bar{m}}^N(\delta, \lambda_0) \right) = \frac{W(x, r, m, \delta, \lambda_0)}{N} + O\left(\frac{1}{N^{3/2}} \right),$

where $W(x, r, m, \delta, \lambda_0)$ is defined by the formula (13).

Note that Theorem 7 is a simple consequence of Theorem 6.

6. Asymptotic normality of estimators using auxiliary information

Theorem 8. Under the conditions of Theorem 6

$$\sqrt{N}\left({}_r\bar{a}_{x:\bar{m}}^N(\delta, \lambda) - {}_r\bar{a}_{x:\bar{m}}(\delta) \right) \Rightarrow N_1\{0, W(x, r, m, \delta, \lambda)\}.$$

Proof. In the notation of Theorem 4, we have $s = 3$, $\sigma = \Omega_{3 \times 3}(x, r, m, \delta)$ in accordance with Section 5, and $q_N = \sqrt{N}$. Now, applying Theorem 3, we get

$$\sqrt{N}\left[[\Phi_N(x, r, m, \delta), S_N(x), J_N] - [\Phi(x, r, m, \delta), S(x), J] \right] \Rightarrow N_3\{(0, 0, 0), \Omega_{3 \times 3}(x, r, m, \delta)\}.$$

The function $H(z)$ is differentiable at the point $t = (\Phi(x, r, m, \delta), S(x), J)$, $\nabla H(t) \neq 0$. Consequently, all conditions of Theorem 4 hold, and using (15) and (16), we obtain the desired result.

The proof of Theorem 8 is completed.

The asymptotic normality of the optimal estimator ${}_r\bar{a}_{x:\bar{m}}^N(\delta, \lambda_0)$ is determined by Theorem 9.

Theorem 9. Under the conditions of Theorem 6

$$\sqrt{N}\left({}_r\bar{a}_{x:\bar{m}}^N(\delta, \lambda_0) - {}_r\bar{a}_{x:\bar{m}}(\delta) \right) \Rightarrow N_1\{0, W(x, r, m, \delta, \lambda_0)\}.$$

Theorem 9 is a simple consequence of Theorem 8.

7. Adaptive Estimator

The statistic ${}_r\bar{a}_{x:\bar{m}}^N(\delta, \lambda_0)$ can be used as an estimator for ${}_r\bar{a}_{x:\bar{m}}(\delta)$ if we know λ_0 ; otherwise, it is required to construct an adaptive estimator. We need a more detailed formula for λ_0 :

$$\lambda_0 = \frac{1}{S(x)Dg(X)} \left[\frac{\Phi(x, r, m, \delta)}{S(x)} (C_2(x, r, m) - JS(x)) - C_1(x, r, m, \delta) + J\Phi(x, r, m, \delta) \right]. \quad (15)$$

Using (15), we consider the following adaptive estimator:

$${}_r\bar{a}_{x:\bar{m}}^N(\delta, \hat{\lambda}_0) = \frac{1}{\delta} \left(1 - \frac{\Phi_N(x, r, m, \delta)}{S_N(x)} - \hat{\lambda}_0 (J_N - J) \right) \quad (16)$$

with

$$\hat{\lambda}_0 = \frac{1}{s^2 S_N(x)} \left[\frac{\Phi_N(x, r, m, \delta)}{S_N(x)} (\hat{C}_2(x, r, m) - JS_N(x)) - \hat{C}_1(x, r, m, \delta) + J\Phi_N(x, r, m, \delta) \right], \quad (17)$$

where $s^2 = \frac{1}{N-1} \sum_{i=1}^N (g(X_i) - J_N)^2$ is an unbiased estimator of the variance $Dg(X)$,

$$\hat{C}_2(x, r, m) = N^{-1} \sum_{i=1}^N g(X_i) \mathbf{I}(x + r + m \geq X_i > x + r), \quad \hat{C}_1(x, r, m, \delta) = N^{-1} \sum_{i=1}^N e^{-\delta X_i} g(X_i) \mathbf{I}(x + r + m \geq X_i > x + r).$$

Theorem 10. Under the conditions of Theorem 6

$$\sqrt{N}\left({}_r\bar{a}_{x:\bar{m}}^N(\delta, \hat{\lambda}_0) - {}_r\bar{a}_{x:\bar{m}}(\delta) \right) \Rightarrow N_1\{0, W(x, r, m, \delta, \lambda_0)\}.$$

Proof. The following equality holds:

$$\sqrt{N} \left({}_r|\bar{a}_{x:\overline{m}|}^N(\delta, \hat{\lambda}_0) - {}_r|\bar{a}_{x:\overline{m}|}(\delta) \right) = \sqrt{N} \left({}_r|\bar{a}_{x:\overline{m}|}^N(\delta, \lambda_0) - {}_r|\bar{a}_{x:\overline{m}|}(\delta) \right) + R_N,$$

where $R_N = \delta^{-1}(\lambda_0 - \hat{\lambda}_0)\sqrt{N}(J_N - J)$. All the estimators, used in (17), converge almost surely to their true values according to the strong law of large numbers (the Second Theorem of Kolmogorov [19]). Thus, from the First Continuity Theorem of Borovkov [16], estimator $\hat{\lambda}_0$ converges almost surely to λ_0 . Based on the central limit theorem $\sqrt{N}(J_N - J) \Rightarrow N_1\{0, Dg(X)\}$, we retrieve $R_N \Rightarrow 0$. Now, the statement of Theorem 10 is proved by using Theorem 9.

Conclusion

The paper deals with the problem of estimating the present values of the continuous whole life annuity using auxiliary information about the expectation of life. It is shown that the usage of such auxiliary information can often provide the MSE not more than that of standard estimators. We proved the results on asymptotic properties of the proposed estimators: unbiasedness, consistency and normality. Also, the main parts of the asymptotic MSEs of the estimators were found. An adaptive estimator is constructed; such estimator is equivalent (in the sense of asymptotic distribution) to the estimator with the optimal weight coefficient λ_0 . Note that the improved estimators of life annuities (8) and (12) can be obtained by substituting of empirical survival functions by the smooth empirical survival functions [20].

References

1. Falin, G.I. (2002) *Mathematical foundations of the theory of life insurance and pension schemes*. Moscow: Ankil.
2. Al-Nator, M.S., Al-Nator, S.V. & Soloviev, A.K. (2018) Actuarial and statistical substantiation of the retirement age in the context of socio-economic uncertainty. *Modern Mathematics and Concepts of Innovative Mathematical Education*. 5(1). pp. 144–148.
3. Ovanesyan, N.M. & Midler, E.A. (2016) Long-term life insurance as a factor in the sustainable development of the insurance services market in an inflationary environment. *Financial Research*. 1(50). pp. 128–134.
4. Roik, V.D. (2013) Ways to improve insurance of occupational disease risks. *Occupational Safety and Economics*. 2(23). pp. 4–14.
5. Soloviev, A.K. (2018) Risks of forecasting pension reform in the context of digitalization. *Bulletin of the Department of Statistics of the Plekhanov Russian University of Economics: Statistical studies of the socio-economic development of Russia and prospects for sustainable growth: materials and reports*. Moscow. pp. 257–260.
6. Winter, P. & Planchet, F. (2022) Modern tontines as a pension solution: a practical overview premium. *European Actuarial Journal*. 12. pp. 3–32.
7. Gubina, O.V. & Koshkin, G.M. (2015) Estimation of actuarial present value of the whole continuous life annuity *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 30(1). pp. 38–43.
8. Bowers, N., Gerber, H., Hickman, J., Jones, D. & Nesbitt C. (1986) *Actuarial Mathematics*. Itasca: Society of Actuaries.
9. Gerber, H. (1997) *Life Insurance Mathematics*. 3rd ed. New York: Springer-Verlag.
10. Kokina, E.P. & Tregubova A.A. (2015) Formation of insurance tariff classes: application of statistical methods. *Financial Research*. 2(47). pp. 115–122.
11. Awondo, S., Ramirez, O., Datta, G.S., Colson, G. & Fonsah, E.G. (2018) Estimation of crop yields and insurance premiums using a shrinkage estimator. *North American Actuarial Journal*. 22(2). pp. 289–308.
12. Gubina, O.V. & Koshkin, G.M. (2020) Estimation of the present value of n -year life annuity for endowment insurance. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 50. pp. 39–46.
13. Hu, J. & Hong, L. (2022) A nonparametric sequential learning procedure for estimating the pure premium. *European Actuarial Journal*. 2. pp. 485–502.
14. Koshkin, G.M. (1999) Deviation moments of the substitution estimator and its piecewise smooth approximations. *Siberian Mathematical Journal*. 40(3). pp. 515–527.
15. Ibragimov, I.A. & Khasminskii, R.Z. (1981) *Statistical Estimation: Asymptotic Theory*. Berlin – New York: Springer.
16. Borovkov, A.A. (1998) *Mathematical Statistics*. New York: Gordon & Breach.
17. Dmitriev, Yu.G. & Koshkin, G.M. (1987) Using additional information in nonparametric estimation of density functionals. *Automation and Remote Control*. 48(10). pp. 1307–1316.

18. Dmitriev, Y.G. & Koshkin, G.M. (2018) Nonparametric estimators of probability characteristics using unbiased prior conditions. *Statistical Papers*. 59(4), pp. 1559–1575. doi: 10.1007/s00362-018-1044-7
19. Rao, C.R. (1965) *Linear Statistical Methods and Its Applications*. New York: Wiley.
20. Koshkin, G.M. (2014) Smooth estimators of the reliability functions for non-restorable elements. *Russian Physics Journal*. 57(5). pp. 672–681.

Information about the authors:

Dmitriev Yuri G. (Associate Professor, Doctor of Physical and Mathematical Sciences, Head of the Department of the System Analysis and Mathematical Modeling of the National Research Tomsk State University, Tomsk, Russian Federation). E-mail: dmit@mail.tsu.ru

Koshkin Gennady M. (Professor, Doctor of Physical and Mathematical Sciences, Professor of the Department of the System Analysis and Mathematical Modeling of the National Research Tomsk State University, Tomsk, Russian Federation). E-mail: kgm@mail.tsu.ru

Contribution of the authors: the authors contributed equally to this article. The authors declare no conflicts of interests.

Информация об авторах:

Дмитриев Юрий Глебович – доцент, доктор физико-математических наук, заведующий кафедрой системного анализа и математического моделирования Института прикладной математики и компьютерных наук Национального исследовательского Томского государственного университета (Томск, Россия). E-mail: dmit70@mail.ru

Кошкин Геннадий Михайлович – профессор, доктор физико-математических наук, профессор кафедры системного анализа и математического моделирования Института прикладной математики и компьютерных наук Национального исследовательского Томского государственного университета (Томск, Россия). E-mail: kgm@mail.tsu.ru

Вклад авторов: все авторы сделали эквивалентный вклад в подготовку публикации. Авторы заявляют об отсутствии конфликта интересов.

Received 24.08.2025; accepted for publication 02.12.2025

Поступила в редакцию 24.08.2025; принята к публикации 02.12.2025

Научная статья

УДК 519.24

doi: 10.17223/19988605/73/9

**Численная вероятностная арифметика для обобщенных
кусочно-полиномиальных функций****Борис Станиславович Добронетц¹, Ольга Аркадьевна Попова²**^{1,2} *Сибирский федеральный университет, Красноярск, Россия*¹ *BDobronets@yandex.ru*² *OlgaArc@yandex.ru*

Аннотация. Исследуется важный случай арифметических операций над случайными переменными, заданных своими функциями плотности вероятности. Проводится анализ возможностей существующих подходов и отмечаются их недостатки. Предлагается новый подход, основанный на представлениях законов распределения аргументов в виде обобщенных кусочно-полиномиальных функций. Для этих целей разработана вероятностная арифметика, основу которой составляют модели и методы вычислительного вероятностного анализа. Отмечается, что предложенная техника вычислений позволяет учитывать особенности функций плотности вероятности и рассматривать распределения с «тяжелыми хвостами».

Ключевые слова: вероятностные арифметики; вычислительный вероятностный анализ; обобщенные кусочно-полиномиальные функции; распределения с «тяжелыми хвостами».

Для цитирования: Добронетц Б.С., Попова О.А. Численная вероятностная арифметика для обобщенных кусочно-полиномиальных функций // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2025. № 73. С. 75–80. doi: 10.17223/19988605/73/9

Original article

doi: 10.17223/19988605/73/9

Numerical probabilistic arithmetic for generalized piecewise polynomial functions**Boris S. Dobronets¹, Olga A. Popova²**^{1,2} *Siberian Federal University, Krasnoyarsk, Russia*¹ *BDobronets@yandex.ru*² *OlgaArc@yandex.ru*

Abstract. This paper examines the important case of arithmetic operations on random variables defined by their probability density functions. The capabilities of existing approaches are analyzed and their limitations are highlighted. A new approach is proposed based on representing the distribution laws of arguments as generalized piecewise polynomial functions. For these purposes, probabilistic arithmetic is developed, based on models and methods of computational probabilistic analysis. It is noted that the proposed computational technique allows for the consideration of the properties of probability density functions and the analysis of distributions with "heavy tails."

Keywords: probabilistic arithmetic; computational probabilistic analysis; generalized piecewise polynomial functions; heavy-tailed distributions.

For citation: Dobronets, B.S., Popova, O.A. (2025) Numerical probabilistic arithmetic for generalized piecewise polynomial functions. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naja tehnika i informatika – Tomsk State University Journal of Control and Computer Science*. 73. pp. 75–80. doi: 10.17223/19988605/73/9

Введение

При численном моделировании часто возникают задачи вычисления функциональных зависимостей от случайных аргументов. Важный случай представляют арифметические операции над случайными переменными, заданные своими функциями плотности вероятности. Практическая значимость решения подобных задач определила необходимость создания численных вероятностных арифметик, которые возникли в начале 1980-х гг. как операции над гистограммами.

В 1980 г. была опубликована одна из первых работ по вероятностной арифметике [1], где для реализации операций напрямую использовались свойства гистограмм как кусочно-постоянных функций и тот факт, что вероятность попадания случайной величины в интервал основания столбца гистограммы определяется его высотой. В работе [2] основатель интервального анализа R.E. Moore использовал гистограммную арифметику для оценки рисков. В [3] обсуждается использование гистограммной арифметики для решения практических задач.

Дальнейшее развитие численных вероятностных арифметик шло по различным направлениям и требовало от численных операций определенных свойств: точности, возможности реализации длинных цепочек вычислений, использования бесконечных носителей и учет особенностей, включая «тяжелые хвосты».

Существующие численные арифметики над случайными переменными по способу вычислений и организации вычислительного процесса можно условно разделить на следующие основные группы:

- использующие символьные вычисления;
- на основе замкнутых семейств распределений;
- основанные на методе Монте-Карло;
- операции над функциями плотности вероятности, представленными кусочно-полиномиальными функциями.

Каждый из подходов имеет свои теоретическое обоснование и программную реализацию. Символьный подход опирается на возможности систем компьютерной алгебры и позволяет в явном виде вычислять интегралы, участвующие в выполнении операций на случайных величинах. Однако на практике интегралы часто не имеют представлений в замкнутой форме и (или) не берутся аналитически. Такие случаи, как правило, не подходят для практического использования [4]. Другой класс подходов основан на замкнутых семействах распределений. Самый важный из таких подходов описан в [5] и основан на так называемых H -функциях, которые являются обобщением гипергеометрических функций. Подход, представленный в [5], является в первую очередь аналитическим. В работе [6] представлен пакет RaCAL который можно рассматривать как систему, основанную на работе [5]. Подход, более близкий по духу к символьным вычислениям, представлен в работе [7]. В этом подходе не реализованы операции деления случайных величин, и он не допускает плотностей с разрывами или сингулярности. Группа методов Монте-Карло (МК) – это чисто статистические подходы [8]. Основная слабость подхода Монте-Карло – медленная сходимость. В вычислительном вероятностном анализе (ВВА) [9] функции плотности вероятности представляются в виде кусочно-полиномиальных функций, которые определяются сеткой $\omega = \{x_0 < x_1 < \dots < x_n\}$, на каждом интервале (x_i, x_{i+1}) которой функция задается полиномом.

Одна из существующих проблем вероятностных арифметик связана с особенностями функций плотности вероятности и необходимостью учитывать распределения с «тяжелыми хвостами» [10].

В настоящей работе, в отличие от [9], предлагается расширить представления функций плотности вероятности до обобщенных кусочно-полиномиальных функций. Исследование численных вероятностных арифметик начнем с рассмотрения свойств случайных переменных, включая арифметические операции. Далее рассмотрим применение обобщенных кусочно-полиномиальных функций для реализации численных вероятностных арифметик.

1. Операции над случайными переменными

Пусть x – случайная величина, тогда ее плотность вероятности будем обозначать x . Обозначим через R – множество плотностей вероятности $\{x\}$ случайных величин x ; соответственно, R^n – пространство плотностей вероятности случайных векторов из R^n .

Носителем функции плотности вероятности f будем называть множество

$$\text{supp}(f) = \{x \mid f(x) > 0\}.$$

Известны аналитические формулы для определения плотности вероятности результатов арифметических действий над случайными величинами. Например, для нахождения плотности вероятности $p_{x_1+x_2}$ суммы двух случайных величин $x_1 + x_2$ используется соотношение [5]

$$p_{x_1+x_2}(x) = \int_{-\infty}^{\infty} p(x-v, v) dv = \int_{-\infty}^{\infty} p(v, x-v) dv, \quad (1)$$

где $p(x_1, x_2)$ – совместная функция плотности вероятности случайного вектора (x_1, x_2) . Для нахождения плотности вероятности p_{x_1/x_2} частного двух случайных величин x_1 / x_2 имеет место формула

$$p_{x_1/x_2}(x) = \int_{-\infty}^{\infty} |v| p(xv, v) dv. \quad (2)$$

Плотность вероятности $p_{x_1 x_2}$ произведения двух случайных величин $x_1 x_2$ определяется соотношением

$$p_{x_1 x_2}(x) = \int_{-\infty}^{\infty} p(v, x/v) \frac{1}{|v|} dv. \quad (3)$$

Далее рассмотрим использование аналитических формул для численной реализации вероятностной арифметики.

2. Численная реализация арифметических операций

В этом разделе мы будем рассматривать численные реализации арифметических операций в конечномерных подпространствах пространства распределений.

Пространством распределений будем называть множество неотрицательных интегрируемых функций $f(x)$, $x \in R$, интеграл от которых равен единице:

$$\int_{-\infty}^{\infty} f(x) dx = 1.$$

Отметим, что важным подпространством пространства распределений является конечномерное пространство кусочно-полиномиальных функций, где кусочно-полиномиальная функция f характеризуется сеткой $\omega = \{x_0, x_1, \dots, x_n\}$ и набором полиномов $p_i(x) \in P_m, x \in [x_{i-1}, x_i]$. В специальных случаях будем допускать возможность использования вместо полиномов аналитических функций \ln , \exp и др.

В данной работе, в отличие от [9], предлагается расширить представления функций плотности вероятности до обобщенных кусочно-полиномиальных функций.

Обобщенные кусочно-полиномиальные функции задаются сетками

$$\omega = \{-\infty \leq x_0 < x_1 < \dots < x_n \leq \infty\}.$$

Таким образом, функции плотности вероятности могут иметь бесконечные носители, и на каждом интервале (x_i, x_{i+1}) это либо полином, либо аналитическая функция.

Рассмотрим численную реализацию арифметических операций на примере кусочно-полиномиальных функций. В случае, когда случайные величины x, y являются независимыми, совместную функцию плотности вероятности $p(x, y)$ можно представить в виде произведения $p(x, y) = s_x(x) s_y(y)$. В этом случае на каждом прямоугольнике $(x_i, x_{i+1}) \times (y_j, y_{j+1})$ для вычисления интегралов (1)–(3) можно пользоваться как численными квадратурами, например Гаусса, которые точны на соответствующих полиномах, так и аналитическими вычислениями. В случае когда случайные величины x, y являются

зависимыми, совместную функцию плотности вероятности $p(x, y)$ необходимо вычислять отдельной процедурой.

Однако в случае, когда совместная функция плотности вероятности содержит точку $(0, 0)$, при произведении случайных величин возможны особенности в результирующей функции плотности. В качестве примера рассмотрим вычисление функции плотности вероятности произведения двух независимых случайных величин x_1 и x_2 . Предположим, что x_1 и x_2 распределены по треугольному закону с носителем на отрезке $[-1, 1]$ и вершиной в точке $(0, 1)$. В этом случае $x_1 x_2$ имеет вид

$$(x_1 x_2)(z) = \int_z^1 (1-x)(1-z/x) \frac{1}{|x|} dx + \int_{-1}^{-z} (1+x)(1+z/x) \frac{1}{|x|} dx. \quad (4)$$

В силу того, что исходные функции плотности вероятности представлены кусочно-полиномиальными функциями, интеграл (4) может быть вычислен в явном виде:

$$x_3(z) = (x_1 x_2)(z) = -4 - 4|z| - 2 \ln(|z|) (|z| + 1).$$

В результате получаем x_3 в виде обобщенной кусочно-полиномиальной функции с носителем $(-1, 1)$.

Этот подход можно распространить на общий случай вычисления функции плотности вероятности произведения двух случайных величин, когда носитель совместной функции плотности вероятности содержит точку $(0, 0)$. В этом случае кусочно-полиномиальное представление функции плотности вероятности случайной величины $x_3 = x_1 x_2$ будет содержать отрезки, где к полиномам добавлены аналитические функции. При дальнейших операциях с подобными кусочно-полиномиальными функциями необходимо учитывать эти особенности. Заметим, что особенности при вычислении произведений случайных величин возникают только в случае, когда носитель совместной функции плотности вероятности содержит точку $(0, 0)$.

Важное значение при реализации вероятностных арифметик имеет учет особенностей, представленных «тяжелыми хвостами» (the Fat Tail) [10].

Будем говорить, что функция плотности вероятности имеет «тяжелый хвост», если выполнено условие

$$f(x) \sim x^{1+\alpha}, x \rightarrow \infty, \alpha > 0.$$

Важный представитель подобных распределений – распределение Коши

$$f(x) = \frac{1}{\pi} \left[\frac{a}{a^2 + x^2} \right],$$

где $a > 0$ – параметр масштаба. Известно, что сумма двух распределений Коши – распределение Коши.

Для проверки работы численных вероятностных арифметик была вычислена сумма двух случайных величин с распределениями Коши. Было сделано предположение, что результирующее распределение имеет «тяжелый хвост» $s(t) = ct^{-(1+\alpha(t))}$, где $\alpha(t)$ – парабола. Было сделано преобразование: $\tan(\pi x/2) \leftrightarrow t$. Учитывалось, что $s(t) = 0, s'(t) = 0$, при $t \rightarrow \infty$. Аппроксимация искалась в виде $s(t) = \exp((a \arctan(t) - 1)^2 - 1) - 2 \ln(t) + c$.

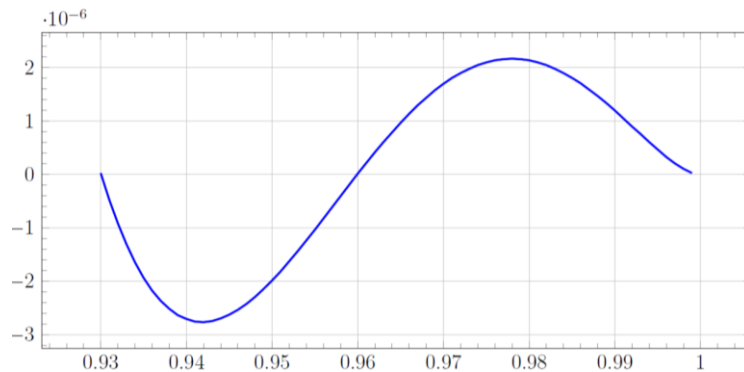


Рис. 1. Ошибки приближения распределения Коши
Fig. 1. Errors in the approximation of the Cauchy distribution

Интегралы (1) брались аналитически, плотность вероятности суммы была вычислена в точках $\xi_1 = 10, \xi_2 = 20$ и были найдены константы a, c . В результате построенного приближения погрешность

на интервале $[10, \infty)$ не превысила величины $3 \cdot 10^{-6}$. На рис. 1 представлена ошибка аппроксимации, горизонтальная шкала x представляет $t = \tan(\pi x / 2)$. Таким образом, точка $x = 1$ соответствует $t = \infty$.

Заключение

Представление функций плотности вероятности в виде обобщенных кусочно-полиномиальных функций позволило при численной реализации вероятностных арифметик учитывать особенности функций плотности вероятности и использовать распределения с «тяжелыми хвостами». Данный подход позволяет повысить качество численного моделирования и дает возможность получать надежные оценки при решении ряда практических задач, например в задачах прогнозной аналитики, а также для оценки и управления финансовыми, инвестиционными и другими видами рисков.

Список источников

1. Colombo A., Jaarsma R. A powerful numerical method to combine random variables // *IEEE Transactions on Reliability*. 1980. V. R-29 (2). P. 126–129.
2. Moore R.E. Risk Analysis without Monte Carlo Methods // *Freiburger Intervall-Berichte*. 1984. V. 84 (1). P. 1–48.
3. Герасимов В.А., Добронец Б.С., Шустров М.Ю. Численные операции гистограммной арифметики и их применения // *Автоматика и телемеханика*. 1991. № 2. С. 83–88.
4. Glen A.G., Evans D.L., Leemis L.M. APPL: A Probability Programming Language // *The American Statistician*, 2001. V. 55 (2). P. 156–166.
5. Springer M.D. *The Algebra of Random Variables*. Hoboken, NJ : John Wiley & Sons, 1979.
6. Korzeń M., Jaroszewicz S. PaCAL: A Python Package for Arithmetic Computations with Random Variables // *Journal of Statistical Software*. 2014. V. 57 (10). P. 1–34. doi: 10.18637/jss.v057.i10
7. Williamson R., Downs T. Probabilistic arithmetic I: numerical methods for calculating convolutions and dependency bounds // *International Journal of Approximate Reasoning*. 1990. V. 4 (2). P. 89–158.
8. Михайлов Г.А., Войтишек А.В. Статистическое моделирование. Методы Монте-Карло : учеб. пособие для бакалавриата и магистратуры. М. : Юрайт, 2018.
9. Добронец Б.С., Попова О.А. Вычислительный вероятностный анализ: модели и методы. Красноярск : Сиб. фед. ун-т, Ин-т космических и информационных технологий, 2020.
10. Clauset A., Shalizi C.R., Newman M.E.J. Power-law distributions in empirical data // *SIAM Rev.* 2009. V. 51. P. 661–703.

References

1. Colombo, A. & Jaarsma, R. (1980) A Powerful Numerical Method to Combine Random Variables. *IEEE Transactions on Reliability*. R-29(2). pp. 126–129
2. Moore, R.E. (1984) Risk Analysis without Monte Carlo Methods. *Freiburger Intervall-Berichte*. 84(1). pp. 1–48.
3. Gerasimov, V.A., Dobronets, B.S. & Shustrov, M.Yu. (1991) Numerical operations of histogram arithmetic and their applications. *Automation and Remote Control*. 2. pp. 83–88.
4. Glen, A.G., Evans, D.L. & Leemis, L.M. (2001) APPL: A Probability Programming Language. *The American Statistician*. 55(2). pp. 156–166.
5. Springer, M.D. (1979) *The Algebra of Random Variables*. Hoboken, New Jersey: John Wiley & Sons.
6. Korzeń, M. & Jaroszewicz, S. (2014) PaCAL: A Python Package for Arithmetic Computations with Random Variables. *Journal of Statistical Software*. 57(10). doi: 10.18637/jss.v057.i10
7. Williamson, R. & Downs, T. (1990) Probabilistic arithmetic I: numerical methods for calculating convolutions and dependency bounds. *International Journal of Approximate Reasoning*. 4(2). pp. 89–158.
8. Mikhailov, G.A. & Voitishchek, A.V. (2018) *Statisticheskoye modelirovanie. Metody Monte-Karlo* [Statistical modeling. Monte-Carlo Methods]. Moscow: Yurait.
9. Dobronets, B.S. & Popova, O.A. (2020) *Vychislitel'nyy veroyatnostnyy analiz: modeli i metody* [Computational Probabilistic Analysis: Models and Methods]. Krasnoyarsk: Siberian Federal University.
10. Clauset, A., Shalizi, C.R. & Newman, M.E.J. (2009) Power-law distributions in empirical data. *SIAM Rev.* 51. pp. 661–703.

Информация об авторах:

Добронец Борис Станиславович – профессор, доктор физико-математических наук, профессор кафедры систем искусственного интеллекта Института космических и информационных технологий Сибирского федерального университета (Красноярск, Россия). E-mail: BDobronets@yandex.ru

Попова Ольга Аркадьевна – доцент, кандидат технических наук, доцент кафедры систем искусственного интеллекта Института космических и информационных технологий Сибирского федерального университета (Красноярск, Россия). E-mail: OlgaArc@yandex.ru

Вклад авторов: все авторы сделали эквивалентный вклад в подготовку публикации. Авторы заявляют об отсутствии конфликта интересов.

Information about the authors:

Dobronets Boris S. (Professor, Doctor of Physics and Mathematics, Siberian Federal University, Institute of Space and Information Technologies, Krasnoyarsk, Russian Federation). E-mail: BDobronets@yandex.ru

Popova Olga A. (Associate Professor, Candidate of Technical Sciences, Siberian Federal University, Institute of Space and Information Technologies, Krasnoyarsk, Russian Federation). E-mail: OlgaArc@yandex.ru

Contribution of the authors: the authors contributed equally to this article. The authors declare no conflicts of interests.

Received 04.09.2025; accepted for publication 02.12.2025

Поступила в редакцию 04.09.2025; принята к публикации 02.12.2025

Научная статья

УДК 519.7

doi: 10.17223/19988605/73/10

Метод идентификации дефектов в датчиках с использованием скользящих наблюдателей высокого порядка

Алексей Нилович Жиравок¹, Александр Валерьевич Зуев²,
Александр Анатольевич Проценко³, Игорь Вячеславович Горностаев⁴,
Анастасия Юрьевна Сердюкова⁵

^{1, 5} Дальневосточный федеральный университет, Владивосток, Россия

^{2, 3, 4} Институт проблем морских технологий Дальневосточного отделения
Российской академии наук, Владивосток, Россия

² Институт автоматизации и процессов управления Дальневосточного отделения
Российской академии наук, Владивосток, Россия

¹ zhirabok@mail.ru

² alvzuev@yandex.ru

³ pro293133@gmail.com

⁴ gornostaev.iv@dvfu.ru

⁵ serdyukova.au@dvfu.ru

Аннотация. Рассматривается задача идентификации (определения величины) дефектов в датчиках технических систем, описываемых линейными уравнениями, в присутствии внешних возмущений. Для решения задачи используются наблюдатели, работающие в скользящем режиме, построенные на основе редуцированной (имеющей меньшую размерность) модели исходной системы, обладающей избирательной чувствительностью по отношению к дефектам и возмущению. Показано, что при выполнении определенных условий существует скользящий режим, который позволяет получить точную оценку величины дефекта.

Ключевые слова: системы; датчики; дефекты; идентификация; скользящие наблюдатели.

Благодарности: Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 25-19-00590, <https://rscf.ru/project/25-19-00590/>

Для цитирования: Жиравок А.Н., Зуев А.В., Проценко А.А., Горностаев И.В., Сердюкова А.Ю. Метод идентификации дефектов в датчиках с использованием скользящих наблюдателей высокого порядка // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2025. № 73. С. 81–89. doi: 10.17223/19988605/73/10

Original article

doi: 10.17223/19988605/73/10

Method of sensor fault identification based on higher-order sliding mode observers

Alexey N. Zhirabok¹, Alexander V. Zuev², Alexander A. Protsenko³,
Igor V. Gornostaev⁴, Anastasia Yu. Serdyukova⁵

^{1, 5} Far Eastern Federal University, Vladivostok, Russian Federation

^{2, 3, 4} Institute of Marine Technology Problems of the Far Eastern Branch
of Russian Academy of Sciences, Russian Federation

² Institute of Automation and Control Processes of the Far Eastern Branch
of Russian Academy of Sciences, Russian Federation

¹ zhirabok@mail.ru

² alvzuev@yandex.ru

³ pro293133@gmail.com

⁴ gornostaev.iv@dvfu.ru⁵ serdyukova.au@dvfu.ru

Abstract. The problem of sensor fault identification in technical systems described by linear equations under the external disturbances is studied. To solve the problem, sliding mode observer is used which is constructed based on the reduced-order model of the initial system. This model is sensitive to the faults and insensitive to the disturbances. It is shown that if some conditions are met, then sliding mode exists which allows obtaining exact estimation of the fault.

Keywords: dynamic systems; sensors; faults; identification; sliding mode observers.

Acknowledgments: The research was carried out with the financial support of the Russian Science Foundation № 25-19-00590, <https://rscf.ru/project/25-19-00590/>

For citation: Zhirabok, A.N., Zuev, A.V., Protsenko, A.A., Gornostaev, I.V., Serduykova, A.Yu. (2025) Method of sensor fault identification based on higher-order sliding mode observers. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naja tehnika i informatika – Tomsk State University Journal of Control and Computer Science*. 73. pp. 81–89. doi: 10.17223/19988605/73/10

Введение

Разнообразные датчики составляют неотъемлемую часть практически любой технической системы. Нередко они являются наименее надежными ее элементами, вследствие чего при появлении в них дефектов датчики могут поставлять искаженную информацию о состоянии системы, что в результате будет приводить к ошибочным реакциям системы управления. Если величину возникших дефектов удастся оценить, эту информацию можно использовать для коррекции искажений и восстановления нормальной работы системы управления.

В настоящее время для оценки (идентификации) величин возникших дефектов активно используются наблюдатели, работающие в скользящем режиме (скользящие наблюдатели) [1–8]. В указанных работах решается задача идентификации для различных классов систем и дефектов, возникающих как в динамике и приводах системы, так и в ее датчиках. Предполагается, что до проведения идентификации решается задача локализации, определяющая, какой датчик неисправен. Для конкретности ниже будем полагать, что искажение показаний датчика описывается неизвестной функцией $d(t)$, которую требуется идентифицировать.

Отметим, что задачи идентификации дефектов в датчиках рассматривались в [2, 7], где было получено только приближенное решение, поскольку итоговое выражение содержало производную $d(t)$. Метод, предложенный в [8], давал точное решение за счет использования специальной системы повышенной размерности, на основе которой строился скользящий наблюдатель. В отличие от этих методов в настоящей работе скользящий наблюдатель строится на основе редуцированной модели исходной системы пониженной размерности, нечувствительной к возмущениям, которая не содержит производной $d(t)$.

Настоящая работа является логическим продолжением статьи [6], где рассматривалась задача идентификации дефектов в датчиках технических систем на основе скользящих наблюдателей. Из [1–6] и аналогичных работ следует, что задача может быть решена при наложении на исходную систему ряда условий, которые далеко не всегда выполняются, что делает невозможным решение задачи идентификации.

Эти условия могут быть в значительной мере ослаблены при использовании так называемых скользящих наблюдателей высокого порядка, рассмотренных в ряде статей [9–12], которые опираются на дифференциатор Леванта [13]. Для реализации этой идеи в статье ставится и решается задача идентификации дефектов в датчиках на основе скользящих наблюдателей высокого порядка. Новизна работы состоит в том, что в отличие от известных работ процедура идентификации нечувствительна к внешним возмущениям и реализуется без наложения условия согласования, что позволяет решить задачу идентификации для более широкого класса систем.

Для этого вначале изложим основные сведения о скользящих наблюдателях высокого порядка на основе работ [9, 10], поскольку подход работ [11, 12] накладывает больше ограничений на исходную систему.

1. Предварительные сведения

Рассмотрим класс систем, описываемых уравнениями

$$\begin{aligned}\dot{x}(t) &= Ax(t) + Bu(t) + L\rho(t), \\ y(t) &= Cx(t),\end{aligned}\tag{1}$$

где $x(t) \in R^n$ и $u(t) \in R^m$ – векторы состояния и управления; $y(t) \in R$ – скалярное измерение; $A \in R^{n \times n}$, $B \in R^{n \times m}$, $L \in R^{n \times 1}$ и $C \in R^{1 \times n}$ – известные постоянные матрицы; $\rho(t) \in R$ – неизвестная скалярная функция времени, описывающая действующие на систему возмущения.

Отметим, что для решения рассматриваемой задачи в работах [1–5] на систему (1) накладывалось условие согласования $\text{rank}(CL) = \text{rank}(L)$, которое снимается в статьях [9–12]. Введем несколько необходимых для дальнейшего изложения понятий [9, 10].

Напомним, что матрицей наблюдаемости системы (1) называется матрица

$$P = \begin{pmatrix} C \\ CA \\ \vdots \\ CA^{n-1} \end{pmatrix}.$$

Предполагается, что пара (C, A) наблюдаема, т.е. $\text{rank}(P) = n$. Известно, что в этом случае существует такая матрица K , что $\tilde{A} = A - KC$ будет устойчивой. Предполагается также, что система (1) является минимально-фазовой, т.е. инвариантные нули тройки (A, C, L) имеют отрицательные вещественные части. Последнее означает, что нули передаточной функции системы (1) являются устойчивыми.

Относительной степенью системы (1) для функции $\rho(t)$ называется число n_1 такое, что

$$CA^j L = 0, \quad j = 1, 2, \dots, n_1 - 2, \quad CA^{n_1-1} L \neq 0.$$

Известно, что $n_1 \leq n$, и соответствующим преобразованием координат систему можно привести к виду:

$$\begin{aligned}\dot{x}_1(t) &= A_{11}x_1(t) + A_{12}x_2(t) + B_1u(t) + L_1\rho(t), \\ \dot{x}_2(t) &= A_{21}x_1(t) + A_{22}x_2(t) + B_2u(t), \\ y(t) &= C_1x_1(t),\end{aligned}\tag{2}$$

где $A_{11} \in R^{n_1 \times n_1}$, $A_{12} \in R^{n_1 \times (n-n_1)}$, $L_1 \in R^{n_1 \times 1}$, $C_1 \in R^{1 \times n_1}$, при этом

$$A_{11} = \begin{pmatrix} 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \\ \beta_1 & \beta_2 & \dots & \beta_{n_1} \end{pmatrix}, \quad A_{12} = \begin{pmatrix} 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \\ \beta_{n_1+1} & \beta_{n_1+2} & \dots & \beta_n \end{pmatrix}, \quad L_1 = (0 \quad \dots \quad 0 \quad q)^T, \quad q \neq 0,$$

где β_1, \dots, β_n – некоторые константы; при $n_1 = n$ подсистема с вектором x_2 отсутствует.

Рассмотрим вначале случай $n_1 = n$, полагая, что неизвестная функция $\rho(t)$ ограничена вместе со своими p производными: $\rho(t) \leq \rho_0$, $\rho^{(i)}(t) \leq \rho_0$, $i = 1, 2, \dots, p$. Кроме того, предполагается, что p -я производная удовлетворяет условию Липшица с константой ρ_1 , т.е. $|\rho^{(p)}(t) - \rho^{(p)}(t')| \leq \rho_1 |t - t'|$.

Для оценки величины функции $\rho(t)$ строятся два наблюдателя, первый из которых представляет собой стандартный наблюдатель Люенбергера полного порядка:

$$\dot{z}(t) = Az(t) + Bu(t) + S(y(t) - Cz(t)), \quad z \in R^n.\tag{3}$$

Второй – это скользящий наблюдатель высокого порядка [13], имеющий следующий вид:

$$\begin{aligned}\dot{v}_1 &= w_1 = -\alpha_{n+p+1} M^{1/(n+p+1)} |v_1 - y + Cz|^{(n+p)/(n+p+1)} \text{sign}(v_1 - y + Cz) + v_2, \\ \dot{v}_2 &= w_2 = -\alpha_{n+p} M^{1/(n+p)} |v_2 - w_1|^{(n+p-1)/(n+p)} \text{sign}(v_2 - w_1) + v_3, \\ &\vdots \\ \dot{v}_n &= w_n = -\alpha_{p+2} M^{1/(p+2)} |v_n - w_{n-1}|^{(p+1)/(p+2)} \text{sign}(v_n - w_{n-1}) + v_{n-1}, \\ &\vdots\end{aligned}\tag{4}$$

$$\begin{aligned}\dot{v}_{n+p} = w_{n+p} &= -\alpha_2 M^{1/2} |v_{n+p-1} - w_{n+p-2}|^{1/2} \text{sign}(v_{n+p-1} - w_{n+p-2}) + v_{n+p}, \\ \dot{v}_{n+p+1} &= -\alpha_1 M \text{sign}(v_{n+p+1} - w_{n+p}),\end{aligned}$$

где M – достаточно большая константа, константы α_i выбираются достаточно большими согласно рекомендациям [11]; в частности, там предложено $\alpha_1 = 1,1$, $\alpha_2 = 1,5$, $\alpha_3 = 2$, $\alpha_4 = 3$, $\alpha_5 = 5$, $\alpha_6 = 8$.

Теорема [10]. Функция $\rho(t)$ может быть оценена в виде:

$$\hat{\rho}(t) = \frac{1}{q} (v_{n+1} + (b_1 v_1 + b_2 v_2 + \dots + b_n v_n)), \quad (5)$$

где b_1, b_2, \dots, b_n – коэффициенты характеристического уравнения матрицы $A - KC$:

$$\det(A - KC - sI) = (-1)^n (s^n - b_n s^{n-1} - \dots - b_1).$$

В [10] доказывается, что при наложенных на систему (1) ограничениях и соответствующем выборе констант M и α_i оценка (5) будет точной после окончания переходного процесса за конечное время. Дополнительно показывается, что если в измерениях присутствует шум с максимальной амплитудой ε , то величина ошибки оценивания функции $\rho(t)$ имеет порядок $\varepsilon^{(p+1)/(n+p+1)}$.

В случае $n_1 < n$ при прежних ограничениях на неизвестную функцию $\rho(t)$ ее оценка может быть получена аналогичным образом путем замены размерности n в формулах (4) и (5) на n_1 .

В простейшем частном случае, когда $n = 1$, система (1) имеет вид:

$$\begin{aligned}\dot{x}(t) &= ax(t) + bu(t) + q\rho(t), \\ y(t) &= x(t),\end{aligned}$$

и неизвестная функция $\rho(t)$ удовлетворяет условию Липшица, т.е. $p = 0$, получаем:

$$\begin{aligned}\dot{z}(t) &= az(t) + bu(t) + K(x(t) - z(t)), \\ \dot{v}_1 = w_1 &= -1,5M^{1/2} |v_1 - x + z|^{1/2} \text{sign}(v_1 - x + z) + v_2, \\ \dot{v}_2 &= -1,1M \text{sign}(v_2 - w_1).\end{aligned}$$

Тогда при $M > |l| \rho_0$ оценка функции $\rho(t)$ имеет вид:

$$\hat{\rho}(t) = \frac{1}{q} (v_2 + (a - K)v_1), \quad K > |a|.$$

2. Построение редуцированной модели

Требование скалярности измерения в модели (1) является недостатком подхода [9, 10], ограничивающим возможности его применения. Этот недостаток, однако, может быть преодолен путем анализа не исходной системы, а ее редуцированной (имеющей меньшую размерность) модели, которая всегда может быть построена так, чтобы быть чувствительной к дефектам, подлежащим оцениванию. Для изложения этой идеи рассмотрим класс технических систем, описываемых линейной моделью

$$\begin{aligned}\dot{x}(t) &= Fx(t) + Gu(t) + L\rho(t), \\ y(t) &= Hx(t) + \sum_{j=1}^l D_j d_j(t).\end{aligned} \quad (6)$$

Здесь $x(t) \in R^n$, $u(t) \in R^m$, $y(t) \in R^l$ – векторы состояния, управления и измерений; $F \in R^{n \times n}$, $G \in R^{n \times m}$, $L \in R^{n \times q}$ и $H \in R^{l \times n}$ – известные постоянные матрицы; $\rho(t) \in R^c$ – неизвестная функция времени, описывающая действующие на систему возмущения; $d_i(t) \in R$ – функция, описывающая дефекты в i -м датчике: при их отсутствии $d_i(t) = 0$, при появлении дефекта $d_i(t)$ становится неизвестной функцией времени, $i = 1, 2, \dots, l$; матрицы D_1, \dots, D_l связывают дефекты с соответствующими компонентами вектора измерений: $D_1 = (1 \ 0 \ \dots \ 0)^T$, \dots , $D_l = (0 \ 0 \ \dots \ 1)^T$. Предполагается, что каждая функция $d_i(t)$ удовлетворяет условию Липшица с некоторой константой.

Для простоты рассмотрим случай, когда дефекты возможны только в одном датчике с соответствующими элементами D и $d(t)$. Требуется оценить функцию $d(t)$ без предположения о минимальной фазовости системы (6).

Редуцированная модель системы (6), нечувствительная к возмущениям, описывается уравнением

$$\begin{aligned}\dot{x}_*(t) &= F_*x_*(t) + G_*u(t) + J_*Hx(t); \\ y_*(t) &= H_*x_*(t),\end{aligned}\quad (7)$$

где $x_* \in \mathbb{R}^k$ – вектор состояния размерности $k < n$, матрицы F_* и H_* размеров $k \times k$ и $1 \times k$ соответственно имеют канонический вид:

$$F_* = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \ddots & \cdots \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix}, \quad H_* = (1 \ 0 \ 0 \ \cdots \ 0). \quad (8)$$

В отличие от системы (6) и строящегося ниже наблюдателя, модель (7) является виртуальным объектом; фактически она представляет собой часть системы (6). Слагаемое $J_*Hx(t)$ использовано вместо $J_*y(t)$ для возможности учета дефектов в датчиках.

Отметим, что для применения методов работ [1–6] к системе (7) требуется выполнение условия $\text{rank}(H_*J_*D) = \text{rank}(J_*D)$ или равенства $PJ_*D = (QH_*)^T$ для некоторой матрицы Q и симметрической положительно определенной матрицы P . Первое, в частности, означает, что дефект должен входить только в первое уравнение системы (7), которое формирует ее выход y_* , второе также носит ограничительный характер, что делает невозможным применение этих методов во многих случаях. Описанный в предыдущем разделе подход не предполагает использования этих ограничений.

Напомним [6], что постоянные матрицы G_* и J_* определяются на основе решения уравнения

$$(N \ -J_{*1} \ \dots \ -J_{*k})(V^{(k)} \ B^{(k)}) = 0, \quad (9)$$

где

$$V^{(k)} = \begin{pmatrix} D^0 H F^k \\ H F^{k-1} \\ \vdots \\ H \end{pmatrix}, \quad B^{(k)} = \begin{pmatrix} D^0 H L & D^0 H F L & D^0 H F^2 L & \cdots & D^0 H F^{k-1} L \\ 0 & H L & H F L & \cdots & H F^{k-2} L \\ 0 & 0 & H L & \cdots & H F^{k-3} L \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix}.$$

Решая уравнение (9) с минимальной размерностью k , начиная с $k = 1$, определяем строку $(N \ -J_{*1} \ \dots \ -J_{*k})$. Далее из соотношений

$$R_* = N D^0, \quad \Phi_1 = R_* H, \quad \Phi_1 F = \Phi_2 + J_{*1} H, \quad \Phi_i F = \Phi_{i+1} + J_{*i} H, \quad i = \overline{2, k-1}, \quad \Phi_k F = J_{*k} H,$$

где D^0 – матрица максимального ранга такая, что $D^0 D = 0$, определяются строки вспомогательной матрицы Φ и находится матрица $G_* = \Phi G$. Поскольку $y(t) = Hx(t) + Dd(t)$, то в модели (7) слагаемое $J_*Hx(t)$ заменяется на $J_*y(t) - J_*Dd(t)$. Для простоты предположим, что вектор J_*D содержит только одну ненулевую компоненту, равную q .

3. Решение задачи

Сравнивая модель (7) с системами (1) и (2), можно сделать вывод, что в качестве системы (1), для которой решается задача оценивания неизвестной функции, может быть использована модель (7) с матрицами $A = F_*$, $B = (G_* \ J_*)$, $L = -J_*D$, $C = H_*$ и переменными $u(t) := \begin{pmatrix} u(t) \\ y(t) \end{pmatrix}$, $y(t) := y_*(t)$ и $\rho(t) := d(t)$.

Нетрудно проверить, что матрица наблюдаемости модели (7) представляет собой единичную матрицу, поэтому модель наблюдаема. Будем предполагать, что матрица J_*D удовлетворяет следующему условию: комплексная частота s , при которой ранг матрицы Розенброка

$$R = \begin{pmatrix} sI - F_* & J_*D \\ H_* & 0 \end{pmatrix}$$

становится меньше $k+1$, имеет отрицательную вещественную часть, т.е. система (7) минимально фазовая.

Наблюдатель (3) в нашем случае принимает вид:

$$\dot{z}(t) = F_*z(t) + G_*u(t) + J_*y(t) + K(R_*y - y_*), \quad z \in R^k. \quad (10)$$

Уравнения (4), описывающие скользящий наблюдатель, сохраняют свой вид с заменой в первом уравнении выражения $v_1 - y + Cz$ на $v_1 - y_* + H_*z$. Значение числа n_1 совпадает с номером компоненты вектора состояния модели (7), в которую входит функция $d(t)$.

С учетом того, что матрица K представляет собой столбец $K = (k_1 \ k_2 \ \dots \ k_k)^T$, а F_* и H_* заданы в канонической форме (8), матрица $A - KC$ принимает вид:

$$F_* - KH_* = \begin{pmatrix} -k_1 & 1 & 0 & \dots & 0 \\ -k_2 & 0 & 1 & \dots & 0 \\ -k_3 & 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \ddots & \dots \\ -k_k & 0 & 0 & \dots & 0 \end{pmatrix}.$$

Ее характеристическое уравнение имеет вид:

$$\det(F_* - KH_* - sI) = (-1)^k (s^k + k_1 s^{k-1} + \dots + k_k).$$

Отсюда по аналогии с (5) получаем формулу для оценки функции $d(t)$:

$$\hat{d}(t) = \frac{1}{q} (v_{n+1} + (k_k v_1 + k_{k-1} v_2 + \dots + k_1 v_n)).$$

4. Пример

Рассмотрим линеаризованную модель трехбакового объекта, приведенную в [6] и описываемую уравнениями

$$\begin{aligned} \dot{x}_1 &= -\gamma_1(x_1 - x_2) + \gamma_2 u_1, \\ \dot{x}_2 &= \gamma_1(x_1 - x_2) - \gamma_3(x_2 - x_3) + \gamma_4 u_2, \\ \dot{x}_3 &= \gamma_3(x_2 - x_3) - \gamma_5 x_3, \\ y_1 &= x_1 + d, \quad y_2 = x_3, \end{aligned}$$

где коэффициенты $\gamma_1 \div \gamma_5$ зависят от конструктивных особенностей объекта, $x_1 \div x_3$ – уровни жидкости в баках. В отличие от работы [6] будем полагать, что измеряются уровни в первом и третьем баках. Для простоты примем $L=0$, а также $\gamma_1 = \dots = \gamma_5 = 1$, что дает следующие матрицы:

$$F = \begin{pmatrix} -1 & 1 & 0 \\ 1 & -2 & 1 \\ 0 & 1 & -2 \end{pmatrix}, \quad G = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad H = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad D = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

Рассмотрим дефект в первом датчике, для которого $D^0 = (0 \ 1)$. Нетрудно проверить, что при $k=1$ уравнение (9) не имеет решения, примем $k=2$:

$$(N \ -J_{*1} \ -J_{*2}) \begin{pmatrix} 1 & -4 & 5 \\ -1 & 1 & 0 \\ 0 & 1 & -2 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} = 0,$$

что дает $N=1$, $J_{*1}=(0 \quad -4)$, $J_{*2}=(1 \quad -3)$. В результате

$$R_*= (0 \quad 1), \quad \Phi = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 2 \end{pmatrix}, \quad G_* = \Phi G = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \quad D_* = -J_* D = \begin{pmatrix} 0 \\ -1 \end{pmatrix}, \quad q = -1.$$

Модель (7) принимает вид:

$$\begin{aligned} \dot{x}_{*1} &= x_{*2} - 4y_2, \\ \dot{x}_{*2} &= y_1 - 3y_2 + u_2 - d, \\ y_* &= x_{*1} = y_2. \end{aligned}$$

Рассмотрим матрицу Розенброка этой модели

$$R = \begin{pmatrix} s & -1 & 0 \\ 0 & s & 1 \\ 1 & 0 & 0 \end{pmatrix}.$$

Нетрудно видеть, что она невырождена, т.е. построенная модель минимально фазовая.

Поскольку $J_* D = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$, то $n_1 = k = 2$; примем $K = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$ и построим наблюдатель Люенбергера (10):

$$\begin{aligned} \dot{z}_1 &= z_2 - 4y_2 + 2(y_2 - y_*) = z_2 - 2y_2 - 2y_*, \\ \dot{z}_2 &= y_1 - 3y_2 + u_2 + (y_2 - y_*) = y_1 - 2y_2 - y_* + u_2, \\ y_* &= z_1. \end{aligned}$$

Так как функция $d(t)$ удовлетворяет условию Липшица, принимаем $p = 0$. Скользящий наблюдатель при $\alpha_1 = 1,1$, $\alpha_2 = 1,5$, $\alpha_3 = 2$ принимает вид:

$$\begin{aligned} \dot{v}_1 &= w_1 = -2M^{1/3} |v_1 - y_2 + z_{*1}|^{2/3} \text{sign}(v_1 - y_2 + z_{*1}) + v_2, \\ \dot{v}_2 &= w_2 = -1,5M^{1/2} |v_2 - w_1|^{1/2} \text{sign}(v_2 - w_1) + v_3, \\ \dot{v}_3 &= -1,1M \text{sign}(v_3 - w_2), \\ M &> 2\|d_1(t)\|. \end{aligned}$$

Найдем коэффициенты характеристического уравнения матрицы $F_* - KH_*$:

$$F_* - KH_* = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} - \begin{pmatrix} 2 \\ 1 \end{pmatrix} (1 \quad 0) = \begin{pmatrix} -2 & 1 \\ -1 & 0 \end{pmatrix}.$$

Тогда

$$\det(sI - (F_* - KH_*)) = \det \begin{pmatrix} s+2 & -1 \\ 1 & s \end{pmatrix} = s^2 + 2s + 1,$$

откуда $b_1 = 1$, $b_2 = 2$. В результате оценка функции $d(t)$ принимает вид:

$$\hat{d}(t) = -(v_3 + v_1 + 2v_2).$$

Отметим, что поскольку функция, описывающая дефект, входит во второе уравнение, методы идентификации, рассмотренные в работах [1–5], в данном случае неприменимы.

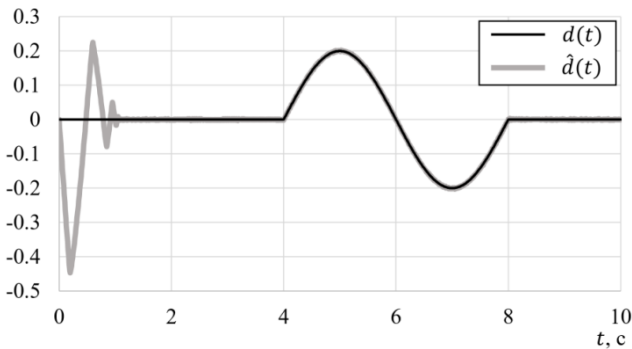


Рис. 1. Результат идентификации функции $d(t)$
Fig. 1. Результат идентификации функции $d(t)$

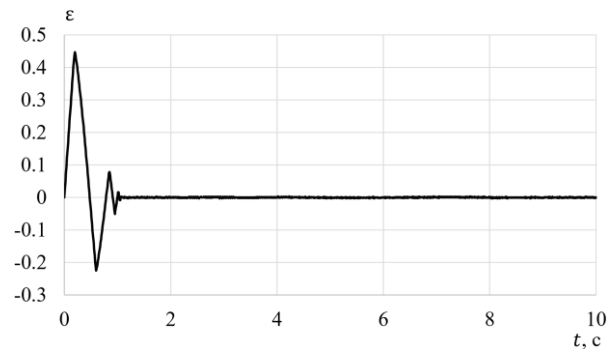


Рис. 2. Ошибка идентификации функции $d(t)$
Fig. 2. Ошибка идентификации функции $d(t)$

В ходе моделирования использовались следующие управляющие воздействия: $u_1(t) = \sin(t)$, $u_2(t) = \sin(0,3t)$. Дефект имитировался появлением сигнала $d(t) = 0,2 \sin(\pi t / 2 - 2\pi)$ на интервале времени 4–8 с. Были приняты следующие начальные условия: $x_1(0) = 0,2$, $x_2(0) = 0,05$, $x_3(0) = 0,02$.

На рис. 1 представлен график оценки функции $d(t)$, на рис. 2 – график ошибки идентификации $\varepsilon(t) = d(t) - \hat{d}(t)$; видно, что она не превышает 1%. Из рисунков видно, что построенные наблюдатели обеспечивают точную оценку величины дефекта после окончания переходного процесса за конечное время.

Заключение

В работе была поставлена и решена задача построения скользящих наблюдателей высокого порядка для идентификации дефектов в датчиках технических систем, описываемых линейными моделями. Задача решается на основе редуцированной (имеющую меньшую размерность) модели исходной системы, нечувствительной к возмущениям. Это позволило уменьшить сложность средств идентификации и ослабить ограничения, накладываемые на исходную систему, для решения поставленной задачи.

Список источников

1. Edwards C., Spurgeon S., Patton R. Sliding mode observers for fault detection and isolation // *Automatica*. 2000. V. 36. P. 541–553.
2. Yan X., Edwards C. Nonlinear robust fault reconstruction and estimation using sliding mode observers // *Automatica*. 2007. V. 43. P. 1605–1614. doi: 10.1016/j.automatica.2007.02.008
3. Keijzer T., Ferrari R. Threshold design for fault detection with first order sliding mode observers // *Automatica*. 2022. V. 146. Art. 110600. doi: 10.1016/j.automatica.2022.110600
4. Hamdi H., Rodrigues M., Rabaout B., Braiek N. A fault estimation and fault tolerant control based sliding mode observer for LPV descriptor systems with time delay // *Int. J. Appl. Math. Comput. Sci.* 2021. V. 31. P. 247–258.
5. Жирабок А.Н., Зуев А.В., Сергиенко О., Шумский А.Е. Идентификация дефектов в нелинейных динамических системах и их датчиках на основе скользящих наблюдателей // *Автоматика и телемеханика*. 2022. № 2. С. 63–89.
6. Жирабок А.Н., Зуев А.В., Шумский А.Е. Идентификация дефектов в датчиках технических систем с использованием скользящих наблюдателей // *Измерительная техника*. 2019. № 10. С. 21–28. doi: 10.1007/s11018-020-01707-1
7. Kalsi K., Hui S., Zak S. Unknown input and sensor fault estimation using sliding-mode observers // *Proceedings 2011 ACC, San Francisco*. 2011. P. 1364–1369.
8. Tan C., Edwards C. Sliding mode observers for robust detection and reconstruction of actuator and sensor faults // *Int. J. Robust and Nonlinear Control*. 2003. V. 13. P. 443–463.
9. Fridman L., Levant A., Davida J. High-order sliding-mode observation and identification for linear systems with unknown inputs // *Proceedings 45th IEEE Conference Decision & Control*. 2006. P. 5567–5572.
10. Fridman L., Levant A., Davida J. Observation of linear systems with unknown inputs via high-order sliding-modes // *Int. J. Systems Science*. 2007. V. 38. P. 773–791.
11. Floquet T., Barbot J.-P. A Canonical form for the design of unknown input sliding mode observer // *Advanced in Variable Structure / eds. C. Edwards et al. Springer Verlag*, 2006. P. 217–292.
12. Floquet T., Edwards C., Spurgeon S. On sliding mode observers for systems with unknown inputs // *Int. J. Adaptive Control and Signal Processing*. 2007. V. 21. P. 638–656.
13. Levant A. Higher-order sliding modes, differentiation and output feedback control // *Int. J. Control*. 2003. V. 76. P. 924–941.

References

1. Edwards, C., Spurgeon, S. & Patton, R. (2000) Sliding mode observers for fault detection and isolation. *Automatica*. 36. pp. 541–553.
2. Yan, X. & Edwards, C. (2007) Nonlinear robust fault reconstruction and estimation using sliding mode observers. *Automatica*. 43. pp. 1605–1614. doi: 10.1016/j.automatica.2007.02.008
3. Keijzer, T. & Ferrari, R. (2022) Threshold design for fault detection with first order sliding mode observers. *Automatica*. 146. Art. 110600. doi: 10.1016/j.automatica.2022.110600
4. Hamdi, H., Rodrigues, M., Rabaout, B. & Braiek, N. (2021) A fault estimation and fault tolerant control based sliding mode observer for LPV descriptor systems with time delay. *International Journal of Applied Mathematics and Computer Science*. 31. pp. 247–258.
5. Zhirabok, A., Zuev, A., Sergienko, O. & Shumsky, A. (2022) Identification of faults in nonlinear dynamical systems and their sensors based on sliding mode observers. *Automation and Remote Control*. 83(2). pp. 214–236.
6. Zhirabok, A., Zuev, A. & Shumsky, A. (2020) Identification of Faults in the Sensors of Technical Systems with the Use of Sliding Mode observers. *Measurement Techniques*. 62. pp. 869–878. doi: 10.1007/s11018-020-01707-1

7. Kalsi, K., Hui, S. & Zak, S. (2011) Unknown input and sensor fault estimation using sliding-mode observers. *Proceedings of the 2011 ACC*, San Francisco. pp. 1364–1369.
8. Tan, C. & Edwards, C. (2003) Sliding mode observers for robust detection and reconstruction of actuator and sensor faults. *International Journal of Robust and Nonlinear Control*. 13. pp. 443–463.
9. Fridman, L., Levant, A. & Davida, J. (2006) High-order sliding-mode observation and identification for linear systems with unknown inputs. *Proceedings of the 45th IEEE Conference Decision & Control*. pp. 5567–5572.
10. Fridman, L., Levant, A. & Davida, J. (2007) Observation of linear systems with unknown inputs via high-order sliding-modes. *International Journal of Systems Science*. 38. pp. 773–791.
11. Floquet, T. & Barbot, J.-P. (2006) A Canonical Form for the design of unknown input sliding mode observer. In: Edwards, C. et al. (eds) *Advanced in Variable Structure*. Springer Verlag. pp. 217–292.
12. Floquet, T., Edwards, C. & Spurgeon, S. (2007) On sliding mode observers for systems with unknown inputs. *International Journal of Adaptive Control and Signal Processing*. 21. pp. 638–656.
13. Levant, A. (2003) Higher-order sliding modes, differentiation and output feedback control. *International Journal of Control*. 76. pp. 924–941.

Информация об авторах:

Жиравок Алексей Нилович – профессор, доктор технических наук, профессор департамента автоматизации и робототехники Дальневосточного федерального университета (Владивосток, Россия). E-mail: zhirabok@mail.ru

Зуев Александр Валерьевич – доктор технических наук, доцент, ведущий научный сотрудник Института проблем морских технологий ДВО РАН; ведущий научный сотрудник Института автоматизации и процессов управления ДВО РАН (Владивосток, Россия). E-mail: alvzuev@yandex.ru

Проценко Александр Анатольевич – кандидат технических наук, старший научный сотрудник Института проблем морских технологий ДВО РАН (Владивосток, Россия). E-mail: pro293133@gmail.com

Горностаев Игорь Вячеславович – кандидат технических наук, старший научный сотрудник Института проблем морских технологий ДВО РАН (Владивосток, Россия). E-mail: gornostaev.iv@dvfu.ru

Сердюкова Анастасия Юрьевна – студент Дальневосточного федерального университета (Владивосток, Россия). E-mail: serdyukova.au@dvfu.ru

Вклад авторов: все авторы сделали эквивалентный вклад в подготовку публикации. Авторы заявляют об отсутствии конфликта интересов.

Information about the authors:

Zhirabok Alexey N. (Professor, Doctor of Technical Sciences, Far Eastern Federal University, Vladivostok, Russian Federation). E-mail: zhirabok@mail.ru.

Zuev Alexander V. (Associate Professor, Doctor of Technical Sciences, Institute of Marine Technology Problems of the Far Eastern Branch of RAS; Institute of Automation and Control Processes of the Far Eastern Branch of RAS, Vladivostok, Russian Federation). E-mail: alvzuev@yandex.ru.

Protsenko Alexander A. (Candidate of Technical Sciences, Institute of Marine Technology Problems of the Far Eastern Branch of RAS, Vladivostok, Russian Federation). E-mail: pro293133@gmail.com.

Gornostaev Igor Vyacheslavovic (Candidate of Technical Sciences, Institute of Marine Technology Problems of the Far Eastern Branch of RAS, Vladivostok, Russian Federation). E-mail: gornostaev.iv@dvfu.ru

Serdyukova Anastasia Yu. (Student of the Far Eastern Federal University, Vladivostok, Russian Federation). E-mail: serdyukova.au@dvfu.ru

Contribution of the authors: the authors contributed equally to this article. The authors declare no conflicts of interests.

Поступила в редакцию 29.08.2025; принята к публикации 02.12.2025

Received 29.08.2025; accepted for publication 02.12.2025

Научная статья

УДК 519.95

doi: 10.17223/19988605/73/11

Регуляризаторы по наборам обобщенных оценок

Николай Александрович Игнатьев

Национальный университет Узбекистана им. Мирзо Улугбека, Ташкент, Узбекистан, n_ignatev@rambler.ru

Аннотация. Предложен новый метод формирования ансамблей алгоритмов распознавания на основе технологии стекинга с использованием регуляризаторов для повышения обобщающей способности моделей. Основное внимание уделено предотвращению переобучения за счет мажорирующих функций, корректирующих отступы объектов от границы между классами. Исследованы условия корректного разделения объектов при обучении базовых алгоритмов и метаалгоритма. Предложен иерархический агломеративный алгоритм группировки признаков, формирующий латентные признаки с учетом внутриклассового сходства и межклассового различия. Показано, что регуляризация отступов и преобразование количественных признаков в номинальные повышают точность распознавания. Установлено, что выбор параметров мажорирующих функций минимизирует расхождение по точности между базовыми и метаалгоритмами.

Ключевые слова: ансамбли алгоритмов; стекинг; регуляризация; мажорирующие функции.

Для цитирования: Игнатьев Н.А. Регуляризаторы по наборам обобщенных оценок // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2025. № 73. С. 90–99. doi: 10.17223/19988605/73/11

Original article

doi: 10.17223/19988605/73/11

Regularizers on sets of generalized estimates

Nikolay A. Ignatev

National University of Uzbekistan named after Mirzo Ulugbek, Tashkent, Uzbekistan, n_ignatev@rambler.ru

Abstract. This paper proposes a novel method for constructing ensembles of recognition algorithms based on stacking technology, incorporating regularizers to enhance the generalization capability of models. The primary focus is on preventing overfitting through the use of majorizing functions that adjust the margins (offsets) of objects from the class boundary. The study investigates the conditions necessary for the correct separation of objects during the training of both base algorithms and the meta-algorithm. A hierarchical agglomerative feature grouping algorithm is proposed, which forms latent features based on intra-class similarity and inter-class differences. It is demonstrated that margin regularization and the transformation of quantitative features into nominal ones improve recognition accuracy. The results show that choosing appropriate parameters for the majorizing functions minimizes the accuracy gap between the base and meta-algorithms. Key advantages of the proposed method:

- flexible feature selection for the meta-algorithm based on a greedy strategy;
- unification of measurement scales through feature transformation;
- robustness to overfitting due to margin regularization.

Keywords: algorithm ensembles; stacking; regularization; majorizing functions.

For citation: Ignatev, N.A. (2025) Regularizers on sets of generalized estimates. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naja tehnika i informatika – Tomsk State University Journal of Control and Computer Science*. 73. pp. 90–99. doi: 10.17223/19988605/73/11

Введение

Повышение обобщающей способности при распознавании является основной целью формирования ансамблей алгоритмов. Как правило, в процессе формирования используют одну из технологий [1] машинного обучения. Чтобы предотвратить переобучение (overfitting), широко используются регуляризаторы. Регуляризаторы являются средством обобщения знания на новые данные, накладывая штрафы на сложность модели.

Смысл термина «обобщенные оценки» [2] ассоциируется с вычислением значений латентных признаков для описания объектов в двухклассовой задаче распознавания. Реализация методов вычисления обобщенных оценок основана на идее противопоставления описаний объектов двух классов как оппозиции друг другу.

Формализация описания многообразия пространств из латентных признаков, формируемых разными методами, сделана в [3]. Методы делятся на линейные и нелинейные, для реализации которых используются или не используются критерии оптимизации.

В статье [2] был предложен алгоритм вычисления обобщенных оценок по комбинации из базовых или элементарных классификаторов. Элементарные классификаторы из одного признака формировали обобщенные оценки объектов для разделения их на два класса. Для каждого признака вне зависимости от шкалы измерений (номинальной или количественной) в представлении прецедентов производилось нелинейное преобразование посредством значений функции принадлежности. Обобщенные оценки, вычисленные по набору элементарных классификаторов, рассматривались как проекции описаний объектов на числовую ось, которая использовалась для отнесения объекта к классу по пороговому решающему правилу.

В [4] описан многокритериальный метод формирования наборов латентных признаков из исходных. Согласно жадной стратегии иерархического агломеративного алгоритма группировки происходит формирование наборов латентных признаков. Множество исходных признаков, входящих в состав латентного с наибольшей дискриминантной способностью для определения принадлежности объектов к классам, считалось информативным набором. Показано, что существует латентный признак, сформированный из подмножества исходных, точность распознавания на котором выше, чем по аналогичному признаку для всего множества. Включение очередного признака в группу при синтезе основано на минимизации отношения внутриклассового сходства и межклассового различия. Если минимизации не происходит, то согласно правила иерархической агломеративной группировки процесс формирования латентного признака завершается. Нет гарантии, что при этом по значениям латентного признака объекты двух классов корректно разделяются на числовой оси.

Согласно технологии стекинга [5], один или несколько латентных признаков, полученных на исходных данных, могут использоваться в качестве входных данных для метаалгоритма, который делает финальное предсказание. Обучение метаалгоритма на входных данных базовых алгоритмов заключается в оптимальном их комбинировании.

Интерес для исследования представляет повышение точности с использованием регуляризаторов. Необходимо определить и обосновать условия, обеспечивающие существование (отсутствие) корректного разделения объектов обучения на классы, не приводящего к переобучению. Обобщающую способность алгоритмов имеет смысл оценивать как частоту ошибок на конечной выборке, так как вероятность ошибки является величиной ненаблюдаемой, которую невозможно вычислить точно [6].

Применение регуляризаторов для повышения точности основывается:

- на увеличении значений отступов между ближайшими объектами из разных классов;
- удалении аномальных объектов (выбросов) из обучающих выборок;
- достижении максимума значения меры компактности классов и выборки в целом.

Отступ – это расстояние между ближайшими объектами двух классов, которые не пересекаются, может представляться и как абсолютная, и как относительная величина. Относительный отступ применялся в регуляризаторах метода «ближайший сосед» (NN) [7]. Классическим пониманием отступа как абсолютной величины пользуются при обучении алгоритмов распознавания с разделяющимися

поверхностями (например, в методе опорных векторов – SVM). Значение отступа является аргументом для мажорирующих функций по технологии бустинга.

Бустинг для задач классификации объектов на два класса K_1 и K_2 строится как серия алгоритмов регрессии $a_1(\cdot), \dots, a_k(\cdot)$, $k \geq 2$, со значениями целевого признака из $\{-1, 1\}$ и правилом:

- если $a_1(S) + \dots + a_k(S) > 0$, то объект S из класса K_1 ;
- если $a_1(S) + \dots + a_k(S) < 0$, то объект S из класса K_2 .

Решение по бустингу заключается в вычислении отступов и выбранной мажорирующей функции. Для практической реализации, как правило, выбор производился из следующего набора функций: квадратичная, кусочно-линейная, сигмоидная, логистическая, экспоненциальная. Итерации бустинга можно повторять. Такой вид бустинга называется градиентным, так как фактически вычисляется градиент функции ошибок (потерь), и новый алгоритм получается, как шаг против градиента функции ошибок. Интерес представляет применимость к значениям латентного признака (обобщенным оценкам объектов) мажорирующих функций для формирования ансамблевых алгоритмов.

1. Предмет исследования

Технология стекинга является методом формирования ансамблей алгоритмов в машинном обучении [7]. Различают базовые модели и модели метауровня, или метамодели. Базовые модели обучаются на исходных данных, затем их результаты используются в качестве входных данных для метамодели. В качестве преимуществ технологии указывалось на возможность комбинирования разных базовых моделей (решающих деревьев, логистической регрессии, градиентного бустинга и т.д.). Комбинирование позволяет использовать сильные и слабые стороны разных моделей с целью повышения точности. Недостатки технологии выражались в более высокой вычислительной сложности и рисках переобучения.

Для устранения недостатков вводились ограничения на применение технологии стекинга. Запрещалось обучать метаалгоритм на данных, по которым обучались базовые алгоритмы. Считалось, что игнорирование запрета приводит к переобучению и недостоверным результатам на новых данных.

Интерес для исследования представляет процесс формирования латентного признакового пространства в качестве входных данных для метаалгоритма. Реализация процесса происходит по результатам иерархической агломеративной группировки с учетом следующих условий:

- вхождение всех исходных признаков в состав латентных не является обязательным;
- максимальная точность распознавания при обучении достигается по базовым алгоритмам;
- базовый алгоритм необходим для оценки объекта лишь по части признакового пространства.

Пусть $a_1(\cdot), \dots, a_k(\cdot)$ – набор из последовательности базовых алгоритмов. Действия базового алгоритма:

- выбор и включение исходного признака в состав латентного;
- корректировка значений латентного признака (оценок) объектов по отступу от границы между классами с помощью мажорирующих функций;
- включение оценок объектов в обучающую выборку как дополнительного признака для обучения метаалгоритма.

Каждый базовый алгоритм, включаемый в набор, не уменьшает точность распознавания предыдущего. Эффективность распознавания при обучении растет, так как известны отступы объектов от границы между классами, а также в каком направлении от нее нужно производить коррекцию (регуляризацию) оценок с помощью мажорирующих функций. Высокая точность распознавания, полученная на последнем базовом алгоритме из набора, не является основанием для выводов о высокой обобщающей способности для метаалгоритма.

Требуется изучение влияния мажорирующих функций на результаты метаалгоритма. Проблема качества обучения имеет скорее комбинаторную, нежели вероятностную природу [5]. Применительно к рассматриваемым ансамблевым алгоритмам речь идет о комбинациях латентных и исходных признаков в зависимости от параметров мажорирующих функций.

Эффект от применения мажорирующих функций проявляется в повышении точности распознавания на обучающей выборке по базовым алгоритмам. Процедуры обучения на основе жадных стратегий являются причиной порождения эффекта переобучения. Включение дополнительных признаков в описание объектов обучающей выборки проводится с целью сохранения в них частичной информации о скрытых закономерностях в данных. На использование этих закономерностей адаптирован метаалгоритм при разделении объектов на классы.

С целью унификации шкал измерений производится преобразование значений количественных признаков в градации номинальных. Для преобразования используется разбиение на непересекающиеся интервалы, оптимальные значения границ которых определяются по специальному критерию.

Есть предположение, что переобучение по технологии стекинга связано с выбором параметров базовых алгоритмов. Как правило, число базовых алгоритмов изначально не фиксировано, так как оно зависит от использования правил иерархической агломеративной группировки и настраиваемых параметров мажорирующей функции. Правила группировки реализованы на основе выбора первого исходного признака для организации ансамбля, результатов анализа отношения внутриклассового сходства и межклассового различия, условия останова процесса формирования значений латентного признака.

Для выбора параметров мажорирующих функций необходим анализ сходимости процесса обучения к состоянию корректного разделения обучающей выборки на классы и максимальному показателю обобщающей способности по метаалгоритму. Показателем для контроля сходимости к корректному разделению является отношение внутриклассового сходства и межклассового различия. Анализ вариативности отношений востребован для исследования устойчивости метаалгоритма от переобучения.

Для вычисления выходных данных (значений дополнительных признаков) по базовым алгоритмам помимо отступов объектов от границы требовался идентификатор класса. Вычисление оценок для произвольного объекта по набору базовых алгоритмов связывалось с необходимостью использования функции потерь. Вид функции и множество ее допустимых значений в общем-то неизвестны. Для базового алгоритма нет информации, в каком направлении от границы между классами производить корректировку значений отступа.

Утверждается, что для произвольного объекта:

- базовые алгоритмы не вычисляют значения оценок;

- метаалгоритм может производить распознавание лишь с учетом привнесения в состав таблицы обучения дополнительных признаков с помощью базовых алгоритмов.

Следовало обосновать применимость предлагаемого метода на данных, которые не участвовали в процессе обучения. Обоснование строится на доказательстве утверждения, что для распознавания объекта не требуется вычислять значения его оценок по базовым алгоритмам. Метаалгоритм для распознавания использует исходные данные объекта и значения дополнительных признаков, полученных на объектах обучающей выборки.

Информацию о вариациях отношений значений внутриклассового сходства к межклассовому различию при добавлении исходного признака в латентный можно получать в процессе реализации метода агломеративной иерархической группировки [4]. Ценность (востребованность) информации заключается в ее использовании в процессе принятия решений как при включении, так и при отказе от включения очередного исходного признака в качестве кандидата в состав латентного.

2. Постановка задачи

Рассматривается стандартная постановка задачи распознавания для объектов из двух непересекающихся классов K_1 и K_2 . Описание объектов в обучающей выборке $E_0 = \{S_1, \dots, S_m\}$ представлено набором разнотипных признаков $X(n) = (x_1, \dots, x_n)$, σ из которых являются номинальными, $(n - \sigma)$ – количественными.

Считается, что набор базовых алгоритмов формируется с использованием параметров мажорирующих функций по значениям отступов объектов от границы между классами. Для предобработки данных используются процедуры:

- преобразования значений количественных признаков в градации номинальных;
- вычисления значения весов признаков и их вкладов в распознавание объектов классов;
- оптимизации критериев для синтеза латентных признаков из исходных.

Требуется:

- определить параметры для вычисления величин штрафов по мажорирующим функциям;
- построить набор базовых алгоритмов с использованием мажорирующих функций;
- разработать метаалгоритм для распознавания принадлежности объектов к классам с использованием дополнительных признаков, определяемых по базовым алгоритмам.

Пусть для значений признака $x_c \in X(n)$ в описании объектов $E_0 = K_1 \cup K_2$ построена упорядоченная по неубыванию последовательность

$$r_1, \dots, r_j, \dots, r_m. \quad (1)$$

В качестве границ двух непересекающихся интервалов $[\pi_1; \pi_2]$, $(\pi_2; \pi_3]$, определяемых по (1), используются $\pi_1 = r_1$, $\pi_2 = r_j$, $1 < j < m$, $\pi_3 = r_m$. Интервалы $[\pi_1; \pi_2]$ и $(\pi_2; \pi_3]$ идентифицируются соответственно как первый и второй. Вес признака у объектов классов по (1) вычисляется как максимум произведения внутриклассового сходства и межклассового различия по критерию из [8]:

$$\left(\frac{\sum_{d=1}^2 \sum_{i=1}^2 (u_i^d - 1) u_i^d}{\sum_{i=1}^2 |K_i| (|K_i| - 1)} \right) \left(\frac{\sum_{d=1}^2 \sum_{i=1}^2 u_i^d (|K_{3-i}| - u_{3-i}^d)}{2|K_1||K_2|} \right) \rightarrow \max_{\pi_1 < \pi_2 < \pi_3}, \quad (2)$$

где $u_i^d(u_{3-i}^d)$ – количество значений признака x_c у объектов из класса K_i (K_{3-i}) в d -м интервале. Множество допустимых значений критерия (2) принадлежит $(0; 1]$ и используется для оценки объектов классов на числовой оси. Если в каждом интервале содержатся все значения признака объектов из одного класса, то его вес равен 1.

Граница между классами (порог) для количественного признака x_c вычисляется как

$$\Gamma_c = \frac{\pi_2 + b}{2}, \quad (3)$$

где b – ближайшее к π_2 значение из интервала $(\pi_2; \pi_3]$, определяемого по (2). В данной работе граница (3) используется для преобразования значений признака $x_c \in X(n)$ в виде двух градаций (по факту принадлежности к одному из интервалов) в номинальной шкале измерений. Далее будем считать, что выборка E_0 представлена значениями номинальных признаков.

Обозначим через g_{1c}^j, g_{2c}^j – количество значений градации $j \in \{1, \dots, \mu\}$ признака $x_c \in X(n)$ в описании объектов соответственно класса K_1 и K_2 . Межклассовое различие по признаку x_c определяется как величина

$$\lambda_c = 1 - \frac{\sum_{t=1}^2 \sum_{d=1}^{\mu} g_{1c}^d g_{2c}^d}{2|K_1||K_2|}.$$

Степень однородности (мера внутриклассового сходства) β_c значений градаций признака при $\mu \geq 2$ по классам K_1, K_2 вычисляется по формулам

$$D_{dc} = \begin{cases} (|K_d| - l_{dc} + 1)(|K_d| - l_{dc}), & p_c > 2, \\ |K_d|(|K_d| - 1), & p_c \leq 2, \end{cases}$$

$$\beta_c = \begin{cases} \frac{\sum_{j=1}^{p_c} g_{1c}^j (g_{1c}^j - 1) + g_{2c}^j (g_{2c}^j - 1)}{D_{1c} + D_{2c}}, & D_{1c} + D_{2c} > 0, \\ 0, & D_{1c} + D_{2c} = 0, \end{cases}$$

где l_{dc} – число градаций признака x_c в описании объектов из K_d , $d = 1, 2$.

Вес признаку $x_c \in X(n)$ определяется как

$$\omega_c = \beta_c \lambda_c. \quad (4)$$

Множество допустимых значений весов признаков, вычисленных по (4), лежит в интервале $[0; 1]$. Для получения обобщенных оценок объектов [2] на E_0 используются вклады градаций признаков. Вклад градации $j \in \{1, \dots, \mu\}$ признака $x_c \in X(n)$ вычисляется как

$$\eta_c(j) = \omega_c \left(\frac{\alpha_{cj}^1}{|K_1|} - \frac{\alpha_{cj}^2}{|K_2|} \right), \quad (5)$$

где $\alpha_{cj}^1, \alpha_{cj}^2$ – количество значений градации j признака x_c соответственно в классах K_1 и K_2 ; ω_c – вес признака x_c по (4). Обобщенная оценка объекта $S_r \in E_0$, $S_r = \{x_{ri}\}$, по описанию на наборе $X(d) \subset X(n)$ без использования мажорирующих функций вычисляется как

$$R(S_r) = \sum_{x_i \in X(d)} \eta_i(x_{ri}).$$

2.1. Формирование дополнительных признаков для метаалгоритма

Согласно технологии стекинга вычисление каждого дополнительного признака реализуется отдельным базовым алгоритмом. Базовый алгоритм явно не участвует в вычислении оценок произвольного допустимого объекта. Считается, что для каждого $x_i \in X(n)$ определены вес ω_i по (4) и значения вкладов $\eta_i(j)$, $j \in \{1, \dots, \mu\}$, по (5). Особенности реализации вычисления дополнительных признаков в качестве входных данных для метаалгоритма заключаются:

- в выборе первого признака из $X(n)$ для вычисления обобщенных оценок (значений дополнительных признаков) объектов E_0 базовыми алгоритмами;
- наборе правил для включения (не включения) признака в группу;
- вычислении значений обобщенных оценок объектов E_0 по вкладам признаков (5) и отступу между классами по мажорирующей функции.

Обозначим через P , $TUPLAM$ – множество индексов признаков соответственно из $X(n)$ и формируемых алгоритмом группировки, $f(\cdot)$ – мажорирующая функция, α – параметр для регуляризации отступа, $0 < \alpha < 1$, δ ($0 < \delta < 0,5$) – порог для отношения внутриклассового сходства θ и межклассового различия по латентному признаку γ , κ – максимальное число дополнительных признаков, $\kappa \leq n - 1$. Реализация алгоритма по шагам будет следующей:

Шаг 1. $P = \{i \mid x_i \in X(n)\}$.

Шаг 2. Вычислить $u = \arg \max_{j \in P} \omega_j$. $TUPLAM = \{u\}$.

Цикл по $t \in \{1, \dots, m\}$ $R(S_t) = \eta_u(a_{tu})$. **Конец цикла**; $cr1 = 10$. $P = P/\{u\}$.

Шаг 3. **Цикл** по $u \in P$. **Цикл** по $t \in \{1, \dots, m\}$. $b_t = R(S_t) + \eta_u(a_{tu})$.

Если $S_t \in K_1$, то $b_t = b_t + \alpha f(-b_t)$ иначе $b_t = b_t - \alpha f(-b_t)$. **Конец цикла**;

$$M_1 = \sum_{S_t \in K_1} b_t \cdot M_2 = \sum_{S_t \in K_2} b_t \cdot M_1 = M_1 / |K_1| \cdot M_2 = M_2 / |K_2|. \Theta = 0. \gamma = 0.$$

Цикл по $t \in \{1, \dots, m\}$. Если $S_t \in K_1$, то $\theta = \theta + |b_t - M_1|$, $\gamma = \gamma + |b_t - M_2|$. Иначе $\theta = \theta + |b_t - M_2|$, $\gamma = \gamma + |b_t - M_1|$. **Конец цикла**;

Если $\theta/\gamma < cr1$, то $cr1 = \theta/\gamma$, $q = u$. **Конец цикла**;

Шаг 4. $crit = cr1$. $P = P/\{q\}$. $TUPLAM = TUPLAM \cup \{q\}$. $cr1 = 10$.

Цикл по $t \in \{1, \dots, m\}$. $R(S_t) = R(S_t) + \eta_q(a_{tq})$.

Если $S_t \in K_1$, то $R(S_t) = R(S_t) + \alpha f(-R(S_t))$. Иначе $R(S_t) = R(S_t) - \alpha f(-R(S_t))$.

Конец цикла; Вывод $\{R(S_t)\}_{t \in \{1, \dots, m\}}$. Если $|TUPLAM| < \kappa$ and $crit > \delta$, то идти 3.

Вывод $TUPLAM$.

Шаг 5. **Конец**.

Множество значений $\{R(S_t)\}_{t \in \{1, \dots, m\}}$, полученное на шаге 4 алгоритма, формируют дополнительные (латентные) признаки в описания объектов K_1 и K_2 .

Выводы, которые можно сделать по результатам иерархического агломеративного алгоритма:

- мощность множества исходных признаков для обучения метаалгоритма $|TUPLAM| \leq n$;
- число дополнительных (латентных) признаков $p = |TUPLAM| - 1$.

Поставим в соответствие каждому исходному признаку индексы согласно порядку их следования в $TUPLAM$. Обозначим набор исходных и дополнительных признаков объектов E_0 для реализации метаалгоритма как $Y(2p - 1) = (y_0, \dots, y_p, r_1, \dots, r_p)$. С учетом такого обозначения произвольный допустимый объект S будет представлен признаками из $TUPLAM$ как $S = (a_0, \dots, a_p)$. В описание по $Y(2p - 1)$ объекта обучающей выборки $S_i \in E_0$, $S_i = (a_{i0}, \dots, a_{ip}, d_{i1}, \dots, d_{ip})$ включены дополнительные признаки d_{i1}, \dots, d_{ip} . Реализация метаалгоритма по шагам для распознавания объекта S будет такой:

Шаг 1. $B1(a_0) = \{S_i \in K_1 | a_0 = a_{i0}\}$, $B2(a_0) = \{S_i \in K_2 | a_0 = a_{i0}\}$, $j = 0$.

Шаг 2. $j = j + 1$. $B1(a_j) = \{S_i \in B1(a_{j-1}) | a_j = a_{ij}, d_{ij} > 0\}$, $B2(a_j) = \{S_i \in B2(a_{j-1}) | a_j = a_{ij}, d_{ij} < 0\}$.

Шаг 3. Если $j < p$, то идти 2.

Шаг 4.
$$\begin{cases} S \in K_1, |B1(a_j)| / |K_1| > |B2(a_j)| / |K_2|, \\ S \in K_2, |B2(a_j)| / |K_2| > |B1(a_j)| / |K_1|, \\ 0, |B1(a_j)| / |K_1| = |B2(a_j)| / |K_2|. \end{cases}$$

Шаг 5. Конец.

2.2. О точности ансамблевых алгоритмов

Для вычисления точности ансамблевых алгоритмов распознавания методы кросс-валидации неприменимы. Доказательство этого утверждения относительно ансамблевых алгоритмов, формируемых по методу вычисления обобщенных оценок, приводится в [2]. Возникла необходимость поиска альтернативных способов оценки точности. Предлагается исследование связи между точностью на обучении по базовым алгоритмам и метаалгоритму.

Для обозначения отношения внутриклассового сходства к межклассовому различию, так же как в описании алгоритма иерархической агломеративной группировки, будем использовать θ/γ . Значение данного отношения рассматривается как средство для проверки истинности гипотезы о компактности объектов классов на многообразии латентных признаков. Интерес для исследования при формировании ансамблевых алгоритмов представляют:

- размеры отступов между классами в зависимости от параметров мажорирующей функции;
- наличие (отсутствие) равномерной сходимости значений $\theta/\gamma \rightarrow \min$ при иерархической агломеративной группировке по последовательности латентных признаков базовых алгоритмов;
- условия отсутствия корректного разделения объектов обучающей выборки по базовому и метаалгоритму.

Результаты исследования свойств сходимости $\theta/\gamma \rightarrow \min$ могут быть использованы при:

- выборе параметров мажорирующей функции;
- обосновании сходства точности по базовому и метаалгоритму.

Преобразование значений количественных признаков в градации номинальных увеличивает вероятность появления совпадающих описаний объектов обучения из двух классов. Граница между классами по (3), используемая для такого преобразования, определяется на основе реальной (не гипотетической) плотности распределения значений признака.

Проблема появления объектов из разных классов с совпадающими описаниями решается за счет использования мажорирующих функций. С помощью этих функций формируются несовпадающие значения латентных или дополнительных признаков для метаалгоритма. Наличие набора дополнительных признаков уменьшает вероятность отказа от распознавания по метаалгоритму.

3. Вычислительный эксперимент

Формирование ансамблей алгоритмов зависит от свойств распределений признаков в описании прецедентов для обучения. Учет этих свойств позволяет гибко подстраивать систему выбора как исходных, так и дополнительных признаков для метаалгоритма. В качестве мажорирующей используется сигмоидная функция с параметром $\alpha \in (0; 1)$. Применение других мажорирующих функций и их влияние на отступы с целью сравнительного анализа в данном исследовании не рассматриваются. Считается, что для количественных признаков произведено преобразование в градации номинальных по (1). В табл. 1 представлен пример построения ансамбля при наличии равномерной сходимости $\theta/\gamma \rightarrow \min$ на данных Molecular-biology [9]. Значение отношения θ/γ получено на последнем в последовательности дополнительном признаке, сформированным базовым алгоритмом. Условием останова процесса формирования было $\theta/\gamma < 0,1$. В скобках указана точность по метаалгоритму.

Таблица 1

Результаты распознавания при наличии равномерной сходимости

α	Число дополнительных признаков	Значение отношения θ/γ	Точность, %
0,10	9	0,0982	100 (100)
0,20	6	0,0885	100 (100)
0,30	5	0,0818	100 (100)

Сходимость по θ/γ при разных значениях параметра α (см. табл. 1) к относительно малому значению δ ($\delta = 0,1$) при корректном разделении объектов на классы по базовому и метаалгоритму увеличивает возможности выбора набора из исходных и дополнительных признаков. При $\alpha = 0,3$ набор был представлен шестью из 56 исходных и пятью дополнительными признаками.

На данных Heart + Disease [10] показана важность отбора значения параметров α и κ для установления максимального соответствия результатов между базовыми алгоритмами и метаалгоритмом при отсутствии равномерной сходимости по отношению θ/γ (табл. 2).

Таблица 2

Результаты распознавания при отсутствии равномерной сходимости

α	Число дополнительных признаков κ	
	5	12
0,05	91,11 (87,04)	92,96 (92,96)
0,10	94,44 (86,67)	98,15 (98,15)
0,20	99,63 (97,41)	99,63 (99,26)
0,30	100,00 (97,41)	100,00 (99,26)

Компромисс или малое расхождение при высокой точности распознавания между результатами по базовым и метаалгоритму достигнут на $\alpha = 0,2$ и $\kappa = 12$. Максимум по базовому алгоритму составил 99,63%, и 99,26% – по метаалгоритму (см. табл. 2).

Для демонстрации влияния коэффициентов регуляризации на значение меры компактности обучающей выборки и связи коэффициентов с обобщающей способностью метрического алгоритма «ближайший сосед» в [6] были использованы данные Spambase [11]. Так же, как и в [6], для тестирования предлагаемого ансамбля алгоритмов было произведено разбиение 4 204 объектов Spambase на две равные по мощности выборки. При этом использован порядок следования четных и нечетных номеров индексов объектов в каждом классе. Каждая выборка (Chet и Nechet) применялась для обучения и тестирования.

По причине отсутствия на данных Spambase равномерной сходимости $\theta/\gamma \rightarrow \min$ для анализа состава исходных и дополнительных признаков использовалось значение параметра κ . Результаты распознавания по составам исходных и дополнительных признаков при коэффициенте $\alpha = 0,2$ приводятся в табл. 3.

Таблица 3

Результаты распознавания на данных Chet и Nechet

Выборка	Число дополнительных признаков κ		
	10	20	56
Chet	100,00 (94,96)	100,00 (97,95)	100,00 (99,24)
Nechet	100,00 (93,34)	100,00 (96,34)	100,00 (99,48)

Как видно из табл. 3, точность распознавания по метаалгоритму не превосходит точность по базовому алгоритму.

Для проверки обобщающей способности ансамбля алгоритмов распознавания в качестве прецедентов использованы выборки Chet и Nechet, результаты которых приведены в табл. 4. В скобках указана точность, полученная на регуляризаторах для метрического алгоритма NN из [7].

Таблица 4

Точность распознавания на тестовых выборках, %

Обучающая выборка	Тестовая выборка	
	Chet	Nechet
Chet	–	99,48 (88,20)
Nechet	99,24(88,73)	–

В табл. 4 демонстрируется превосходство по точности при использовании регуляризаторов по обобщенным оценкам относительно регуляризаторов для метрического алгоритма NN.

Заключение

Разработан новый метод формирования ансамблей алгоритмов распознавания по технологии стекинга. Отметим изменения, которые привнесены в эту технологию:

- число базовых алгоритмов связано с выбором параметров модели;
- есть два ограничения на число признаков, определяемых явно или по специальному условию; выполнение условия зависит от значения отношения внутриклассового сходства и межклассового различия;
- роль базовых алгоритмов сводится к вычислению дополнительных признаков для обучающей выборки;
- высокая обобщающая способность на обучающих и тестовых выборках по метаалгоритму объясняется размерами отступов между классами.

Метод рекомендуется для использования в моделях, основанных на знаниях. Совершенствование метода связано с решением проблемы выбора прецедентов для обучающей выборки и разработкой новых способов формирования множества дополнительных признаков.

Список источников

1. Zhou Z.H. Ensemble learning: foundations and algorithms. Chapman & Hall/CRC, 2021. 394 p.
2. Ignatev N.A. On Nonlinear Transformations of Features Based on the Functions of Objects Belonging to Classes // Pattern Recognition and Image Analysis. 2021. V. 31 (2). P. 197–204.
3. Игнатьев Н.А., Акбаров Б.Х. Оценка близости структур отношений объектов обучающей выборки на многообразиях наборов латентных признаков // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2023. № 65. С. 69–78. doi: 10.17223/19988605/65/7
4. Ignatev N.A., Rahimova M.A. Formation and analysis of sets of informative features of objects by pairs of classes // Scientific and Technical Information Processing. 2022. V. 49 (6). P. 439–445.
5. Hastie T., Tibshirani R., Friedman J. The elements of statistical learning: data mining, inference and prediction. 2nd ed. Springer, 2009. 767 p. (Springer Series in Statistics).
6. Воронцов К.В. Комбинаторный подход к оценке качества обучаемых алгоритмов // Математические вопросы кибернетики / под ред. О.Б. Лупанов. М. : Физматлит, 2004. Т. 13. С. 5–36.

7. Игнатьев Н.А., Турсунмуротов Д.Х. Цензурирование обучающих выборок с использованием регуляризации отношений связности объектов классов // Научно-технический вестник информационных технологий, механики и оптики. 2024. Т. 24 (2). С. 2226–1494. doi: 10.17586/2226-1494-2024-24-2-322-329
8. Згуральская Е.Н. Алгоритм выбора оптимальных границ интервалов разбиения значений признаков при классификации // Известия Самарского научного центра Российской академии наук. 2012. № 4-3. С. 826–829.
9. UCI repository of machine learning databases/molecular-biology/promoter-gene-sequences. URL: <https://archive.ics.uci.edu/dataset/67/molecular+biology+promoter+gene+sequences> (accessed: 02.07.2025).
10. UCI repository of machine learning databases. Ionosphere. URL: <http://archive.ics.uci.edu/ml/datasets/Heart+Disease> (accessed: 02.07.2025).
11. UCI repository of machine learning databases. spambase. URL: <https://archive.ics.uci.edu/dataset/94/spambase> (accessed: 02.07.2025).

References

1. Zhou, Z.H. (2021) *Ensemble Learning: Foundations and Algorithms*. Chapman & Hall/CRC.
2. Ignatev, N.A. (2021) On Nonlinear Transformations of Features Based on the Functions of Objects Belonging to Classes. *Pattern Recognition and Image Analysis*. 31(2). pp. 197–204.
3. Ignatev, N.A. & Akbarov, B.Kh. (2023) Estimation of the proximity of structures of relations of objects of the training sample on manifolds of sets of latent features. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 65. pp. 69–78. doi: 10.17223/19988605/65/7
4. Ignatev, N.A. & Rahimova, M.A. (2022) Formation and Analysis of Sets of Informative Features of Objects by Pairs of Classes. *Scientific and Technical Information Processing*. 49(6). pp. 439–445.
5. Hastie, T., Tibshirani, R. & Friedman, J. (2009) *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd ed. Springer.
6. Vorontsov, K.V. (2004) Combinatorial approach to assessing the quality of learning algorithms. *Matematicheskie voprosy kibernetiki*. 13. pp. 5–36.
7. Ignatev, N.A. & Tursunmurotov, D.Kh. (2024) Censoring training samples using regularization of relatedness relations of class objects. *Nauchno-tekhnicheskij vestnik informatsionnykh tekhnologiy, mekhaniki i optiki*. 24(2). pp. 322–329. doi: 10.17586/2226-1494-2024-24-2-322-329
8. Zguralskaya, E.N. (2012) Algorithm for selecting optimal boundaries of intervals for partitioning feature values during classification. *Izvestiya Samarskogo nauchnogo tsentra Rossiyskoy akademii nauk*. 4–3. pp. 826–829.
9. UCI Repository of Machine Learning Databases. Molecular Biology. Promoter Gene Sequences. [Online] Available from: <https://archive.ics.uci.edu/dataset/67/molecular+biology+promoter+gene+sequences> (Accessed: 2nd July 2025).
10. UCI Repository of Machine Learning Databases. Ionosphere. [Online] Available from: <http://archive.ics.uci.edu/ml/datasets/Heart+Disease> (Accessed: 2nd July 2025).
11. UCI Repository of Machine Learning Databases. Spambase. [Online] Available from: <https://archive.ics.uci.edu/dataset/94/spambase> (Accessed: 2nd July 2025).

Информация об авторе:

Игнатьев Николай Александрович – доктор физико-математических наук, профессор кафедры программного инжиниринга и искусственного интеллекта Национального университета Узбекистана им. Мирзо Улугбека (Ташкент, Узбекистан). E-mail: n_ignatev@rambler.ru

Автор заявляет об отсутствии конфликта интересов.

Information about the author:

Ignatev Nikolay A. (Doctor of Physical and Mathematical Sciences, Professor of the National University of Uzbekistan named after Mirzo Ulugbek, Tashkent, Uzbekistan). E-mail: n_ignatev@rambler.ru

The author declares no conflicts of interests.

Поступила в редакцию 19.07.2025; принята к публикации 02.12.2025

Received 19.07.2025; accepted for publication 02.12.2025

Научная статья

УДК 004.738

doi: 10.17223/19988605/73/12

Подход к выделению значимых признаков сетевой активности устройств Интернета вещей**Ольга Сергеевна Исаева¹, Сергей Владиславович Исаев², Никита Владимирович Кулясов³**^{1, 2, 3} *Институт вычислительного моделирования Сибирского отделения Российской академии наук, Красноярск, Россия*¹ *isaeva@icm.krasn.ru*² *si@icm.krasn.ru*³ *razor@icm.krasn.ru*

Аннотация. Исследуются признаки сетевой активности устройств Интернета вещей и предлагается метод сокращения размерности признакового пространства для повышения эффективности анализа данных. Предложенный подход устраняет мультиколлинеарность, нелинейную зависимость и избыточность признаков, сохраняя их семантическую интерпретируемость. В его основе лежит комбинированное использование статистических характеристик взаимной информации, корреляции, критериев стабильности и значимости для фильтрации признаков. Применение подхода позволило существенно сократить признаковое пространство и улучшить его свойства: численную устойчивость данных, обобщающую способность моделей, качество кластеризации.

Ключевые слова: Интернет вещей; датасет сетевых угроз; статистический анализ; устойчивость признаков; веса признаков в методе главных компонент.

Благодарности: Работа поддержана Красноярским математическим центром, финансируемым Минобрнауки РФ в рамках мероприятий по созданию и развитию региональных НОМЦ (соглашение 075-02-2025-1606).

Для цитирования: Исаева О.С., Исаев С.В., Кулясов Н.В. Подход к выделению значимых признаков сетевой активности устройств Интернета вещей // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2025. № 73. С. 100–109. doi: 10.17223/19988605/73/12

Original article

doi: 10.17223/19988605/73/12

Approach to identifying significant features of IoT device network activity**Olga S. Isaeva¹, Sergey V. Isaev², Nikita V. Kulyasov³**^{1, 2, 3} *Institute of Computational Modelling SB RAS, Krasnoyarsk, Russian Federation*¹ *isaeva@icm.krasn.ru*² *si@icm.krasn.ru*³ *razor@icm.krasn.ru*

Abstract. The study investigates the features of network activity in Internet of Things (IoT) devices and proposes a method for reducing the dimensionality of the feature space to enhance data analysis efficiency. The proposed approach eliminates multicollinearity, nonlinear dependencies, and feature redundancy while preserving their semantic interpretability. It is based on the combined use of statistical characteristics such as mutual information, correlation, stability criteria, and significance measures for feature filtering. Applying this approach significantly reduced the feature space and improved its properties: numerical stability of the data, generalization ability of models, clustering quality.

Keywords: Internet of Things; threat datasets; statistical analysis; feature stability; feature weights in principal component analysis.

Acknowledgments: This work is supported by the Krasnoyarsk Mathematical Center and financed by the Ministry of Science and Higher Education of the Russian Federation in the framework of the establishment and development of regional Centers for Mathematics Research and Education (Agreement No. 075-02-2025-1606).

For citation: Isaeva, O.S., Isaev, S.V., Kulyasov, N.V. (2025) Approach to identifying significant features of IoT device network activity. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naja tehnika i informatika – Tomsk State University Journal of Control and Computer Science*. 73. pp. 100–109. doi: 10.17223/19988605/73/12

Введение

Среди ключевых трендов развития современного общества широкую популярность набирает концепция Интернета вещей (Internet of Things, IoT), направленная на построение киберфизических систем, объединяющих физические и цифровые объекты на основе информационно-коммуникационных технологий [1]. Его применение становится актуальным в различных сферах жизни: от контроля качества продукции до прогнозирования отказов технических систем и мониторинга состояния окружающей среды. Однако значительные преимущества, предоставляемые технологиями Интернета вещей, требуют усиленного внимания к вопросам безопасности и надежности связанных с обеспечением их функционирования процессов, которые являются привлекательной целью для кибератак из-за разнообразия и уязвимости устройств и протоколов их подключения [2]. Особенностью сетевого трафика Интернета вещей является изменчивость, вызванная множественностью происходящих событий, сценариев атак и их последствий, зависящих от развития средств уклонения. Сложность диагностирования аномалий увеличивается по мере роста числа взаимосвязанных систем и разнообразия типов входных данных [3]. Постоянное изменение характеристик нормального поведения затрудняет автоматическое обнаружение аномалий, которые могут возникать из-за различных факторов, таких как отказ используемых устройств или внешняя атака, что требует не только их выявления, но и классификации видов и причин деструктивных воздействий. Методы машинного обучения позволяют получать дополнительную информацию о происходящих в сетевой инфраструктуре событиях безопасности, а также предотвращать нежелательные инциденты. Но для их эффективной работы необходимы специализированные датасеты, охватывающие длительные периоды наблюдения и содержащие сведения о характерных видах атак, типичном поведении пользователей и свойствах реальных сетей, в рамках которых строится система обнаружения вторжений.

Актуальность применения методов машинного обучения для обеспечения сетевой безопасности широко освещается в современной научной литературе. В работах [4–5] представлен детальный обзор популярных наборов данных, используемых для выявления сетевых вторжений, а также предложены критерии оценки их применимости для решения практических задач. В [6] рассматривается проблема несбалансированности данных публичных датасетов. Предложен метод синтеза и сжатия выборок для уменьшения дисбаланса классов. Важным аспектом анализа сетевых данных являются извлечение признаков из исходных файлов и их последующее сопоставление с размеченными данными [7]. При этом исследования отмечают наличие ряда ограничений в публичных датасетах, таких как использование искусственных имитационных сред, неоднородность данных, ошибки в расчетах значений признаков, их дублирование и некорректное разбиение на сессии. Эти факторы существенно влияют на качество моделей машинного обучения. В [8] отмечается важность учета контекста, включающего не только архитектурные особенности сети, но и свойства протоколов, используемых на отдельных уровнях межмашинного взаимодействия.

Для преодоления ограничений, присущих публичным датасетам, авторами настоящего исследования была разработана и внедрена инфраструктура сбора данных и имитации угроз безопасности для сети Интернета вещей Красноярского научного центра [9]. После организации сбора данных встала задача исследования признакового пространства для сокращения его размерности, исключения избыточности, повышения обобщающей способности моделей, улучшения их точности, устойчивости и интерпретируемости результатов. Существующие подходы к решению этой задачи можно разделить на четыре группы: встраиваемые (Embedded), обертывающие (Wrapper), фильтрующие (Filtered) и действующие на основе экспертных оценок. Встраиваемые методы интегрируют выбор признаков в процесс обучения, что позволяет использовать скрытую структуру данных. Однако такой подход потребует

обработки не только множественности признаков, но и данных, генерируемых IoT-устройствами. Обертывающие методы используют алгоритмы роевого интеллекта для эвристического обхода комбинаций признаков и выбора их подмножества, соответствующего целевой функции [10]. Такой способ поиска является NP-сложной задачей и также не подходит для крупномасштабных данных Интернета вещей. Процесс фильтрации признаков заключается в выявлении наиболее информативных и значимых из них. Они конструируют критерии ранжирования признаков с помощью статистических характеристик, корреляции, взаимной информации или условной энтропии [11]. Методы экспертных оценок выполняют отбор признаков на основе знаний о предметной области, но при большом размере признакового пространства их необходимо интегрировать с методами фильтрации.

Цель данной работы – анализ признаков сетевой активности IoT-устройств для уменьшения размерности признакового пространства за счет применения фильтрации, основанной на критериях стабильности и информационной значимости. Подход, представленный в исследовании, учитывает особенности функционирования реальной сети Интернета вещей, но может быть обобщен на другие типы сетевого трафика. Работа является важным шагом к построению системы информационной безопасности, специализированной для исследуемой корпоративной сети, учитывающей ее архитектурные и технические особенности.

1. Особенности сбора данных сетевой активности

В рамках исследований, проводимых в Красноярском научном центре, технология Интернета вещей была внедрена для мониторинга микроклимата в помещениях, где размещено сетевое оборудование [12]. Для анализа безопасности IoT-сети в реальных условиях эксплуатации созданы инструменты сбора данных и имитации угроз. Межмашинное взаимодействие строится на основе протокола прикладного уровня MQTT (Message Queuing Telemetry Transport). На рис. 1 приведена схема размещения основных узлов инфраструктуры сбора данных сетевой активности устройств Интернета вещей.

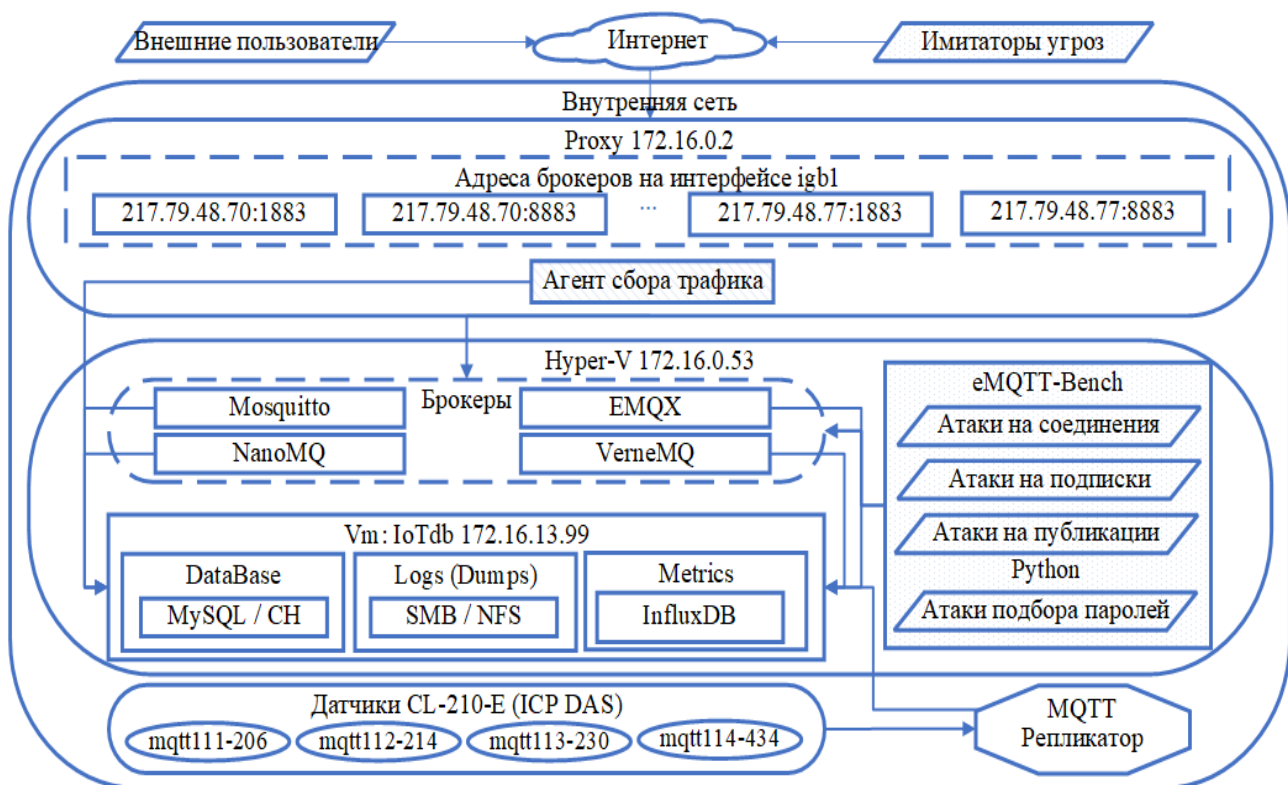


Рис. 1. Инфраструктура сбора данных устройств Интернета вещей
Fig. 1. Infrastructure for collecting data from IoT devices

Основные элементы сети: издатель (Publisher), производящий данные, подписчик (Subscriber), их использующий, и брокер (Broker), выполняющий распределение данных от издателей подписчикам. Использование технологии IoT в сочетании с протоколом MQTT позволяет создать эффективную систему мониторинга, способную оперативно обрабатывать большие потоки данных в темпе их поступления. В рамках исследования были развернуты брокеры на платформах Eclipse Mosquitto, EMQX, NanoMQ и VerneMQ. Сбор данных осуществлялся как на самих брокерах, так и на проху-сервере, через который передается внутренний и внешний сетевой трафик по портам протокола MQTT, включая легитимные сессии, а также несанкционированные попытки соединения и сканирования различных сервисов.

Собранные сетевые данные являются неструктурированными, и для их эффективного использования требуется этап предобработки, включающий разделение трафика на сессии и вычисление параметров полученных сессий. Для этого мы выбрали программное обеспечение с открытым кодом NTLFlowLyzer, которое использовано [13] при построении публичного датасета BCCC-CIC-IDS201. Данный инструмент является усовершенствованной версией сетевого анализатора CICFlowMeter, лежащего в основе популярного датасета CICIDS2017. Мы выполнили проверку параметров выделяемых сессий для сетевых журналов, собранных в рамках нашего исследования. Для решения специфических задач, связанных с Интернетом вещей, в наших данных содержатся метрики брокеров и флаги используемых протоколов [14]. В сформированный датасет вошло более 300 признаков, охватывающих шесть ключевых категорий: временные характеристики, флаги протоколов TCP и MQTT, параметры скорости соединений, статистические данные по заголовкам пакетов, свойства полезной нагрузки и объемные характеристики при массовой передаче данных.

Дальнейшее исследование выявило ряд проблем, существенно затрудняющих применение собранных данных для анализа аспектов безопасности Интернета вещей. К таким проблемам относятся разноплановость представленной информации для типичных сессий и высокоразмерность признакового пространства, которая приводит к экспоненциальному росту объема данных, необходимого для получения надежных результатов моделей машинного обучения, а также наличие избыточных и слабоинформативных признаков, которое снижает эффективность алгоритмов кластеризации и увеличивает их временную сложность. Еще одной значимой проблемой является высокая корреляция признаков, которая искажает структуру данных и приводит к несогласованным результатам кластеризации.

Использование методов снижения размерности, заключающееся в формировании новых признаков, ограничивает возможность семантической интерпретации взаимосвязей между объектами в многомерных данных. Традиционные подходы к сокращению признакового пространства, такие как рекурсивное исключение признаков (RFE) [15], адаптируют набор признаков под конкретный класс моделей, однако не учитывают сложные линейные и нелинейные зависимости между признаками, что сказывается на качестве получаемых подпространств. Стандартные метрики оценки качества моделей (например, MSE, MAE, точность, F1-мера и др.) не могут быть применены в нашем случае, поскольку они предполагают сравнение предсказаний модели с известными целевыми значениями. В нашей задаче требуется первоначально построить целевой показатель путем кластеризации данных, что затруднительно ввиду большого признакового пространства. Таким образом, необходим механизм направленной фильтрации признаков, обеспечивающий сохранение обобщающих свойств и удовлетворяющий выбранным критериям качества данных.

2. Сокращение признакового пространства

Для решения вышеописанных проблем предложен метод сокращения размерности признакового пространства, выделяющий сложные зависимости между признаками с сохранением его информативности. Введем обозначения, необходимые для описания подхода:

$$X = [x_{ij}] \in \mathbb{R}^{n \times m}, \quad (1)$$

где X – матрица наблюдений размерности $n \times m$, n – количество объектов наблюдения, m – количество признаков, x_{ij} – значение j -го признака для i -го объекта наблюдения, $I = \{1, 2, \dots, n\}$ – множество иден-

тификаторов, соответствующих каждому объекту наблюдения, $J = \{1, 2, \dots, m\}$ – множество идентификаторов, соответствующих каждому признаку. Каждый объект наблюдения $i \in I$ описывается вектором $X_i = (x_{i1}, x_{i2}, \dots, x_{im}) \in \mathbb{R}^m$ значений всех признаков для i -го объекта. Признак $j \in J$ представлен вектором $P_j = (x_{1j}, x_{2j}, \dots, x_{nj}) \in \mathbb{R}^n$ значений j -го признака для всех объектов.

$$X = [X_1 \ X_2 \ \dots \ X_n] = [P_1 \ P_2 \ \dots \ P_m]^T. \quad (2)$$

Для центрированной матрицы X_c можно построить квадратную матрицу $\Sigma_X = X_c^T X_c$ размерности $m \times m$ (здесь и далее мы не умножаем на нормировочный коэффициент, поскольку данные предварительно стандартизованы) и вычислить $\lambda_1, \lambda_2, \dots, \lambda_m$ – собственные числа, v_1, v_2, \dots, v_m – собственные векторы, $\delta_1, \delta_2, \dots, \delta_m$, $\delta_i = \sqrt{\lambda_i}$ – сингулярные числа.

Введем целевую функцию, описывающую требования к признаковому пространству:

$$L(X) = \alpha_1 \cdot \kappa(X) + \alpha_2 \cdot M(X) + \sum_{l=1}^L \alpha_{3,l} \cdot R(F_l, X) \quad (3)$$

Весовые коэффициенты α позволяют изменять влияние каждого компонента на результат целевой функции, $\kappa(X) = \delta_{\max} / \delta_{\min}$ – число обусловленности, δ_{\max} – наибольшее, а $\delta_{\min} \neq 0$ – наименьшее сингулярные числа, показывают наличие линейной зависимости и мультиколлинеарности данных, $M(X)$ выполняет оценку меры нелинейной зависимости признаков:

$$M(X) = \frac{2}{m(m-1)} \sum_{j=1}^m \sum_{k=j+1}^m M_{jk}, \quad M_{jk} = \sum_{x_{ij} \in P_j} \sum_{x_{ik} \in P_k} \frac{p(x_{ij}, x_{ik})}{n} \cdot \ln \frac{n \cdot p(x_{ij}, x_{ik})}{p(x_{ij}) \cdot p(x_{ik})}, \quad (4)$$

где M_{jk} определяет величину взаимной информации между P_j и P_k [16], $p(x_{ij}, x_{ik})$ – частота совместного появления значений признаков P_j и P_k , $p(x_{ij})$ и $p(x_{ik})$ – частота появления каждого значения признака в отдельности.

Значение $R(F_l, X)$ – мера сложности Радемахера [17] для семейства функций F_l относительно выборки X :

$$R^{(l)}(F_l, X) = \sup_{f \in F_l} \frac{1}{n} \sum_{i=1}^n \varepsilon_i f(X_i), \quad (5)$$

где ε_i – случайная величина, F_l – семейство функций вида f . Это теоретическая мера, которая показывает способность аппроксимировать данные и избегать переобучения.

Пусть A – множество действий по изменению признакового пространства P . На каждом шаге $h \in S^h = \{0, 1, \dots, H\}$ действие $A^h \in A$ применяется к множеству признаков так, что $A^h(P^{h-1}) = P^h$. Тогда итоговое множество признаков P^H получается следующим образом:

$$P^H = \arg \min_{h \in S^h} \left(L(P^h) \mid A^h \in A, P^h = A^h(P^{h-1}), L(P^h) \leq L(P^{h-1}) \right), \quad (6)$$

при условии: $P_j \in \mathbb{R}^m, m \leq m_{\max} \wedge \sum_{i=1}^n \sum_j I(x_{ij} = \emptyset) = 0 \wedge \forall j \ \sigma_j \geq \tau_\sigma$, где m_{\max} – ограничение на размерность пространства, $I(\cdot)$ – индикаторная функция, σ_j – стандартное отклонение, τ_σ – пороговое значение. Действия из множества A заключаются в фильтрации признакового пространства P^h размерности m^h и удалении из него признаков, удовлетворяющих следующим условиям:

I. Подмножество коррелирующих признаков:

$$C^I = \{C^{j\{k\}}\}_{j,k} = \bigcup_{1 \leq j \leq k \leq m^h} \{(P_j, P_k) \in P^h : |\rho_{jk}| \geq \tau_\rho\}, \quad (7)$$

где $P_j, P_k \in P^h$, m^h – число признаков в P^h , ρ_{jk} – коэффициент корреляции между признаками, τ_ρ – пороговое значение корреляции.

II. Подмножество мультиколлинеарных признаков:

$$C^{II} = \left\{ C^{j\{k\}} \right\}_{j,k} = \bigcup_{k \neq j} \left(P_j \cup \left\{ P_k \in P^h : |\beta_k| > \tau_\beta, V_k \geq \tau_V \right\} \right), \quad (8)$$

$P_j \in P^h = \arg \max_k (V_k \geq \tau_V)$, τ_V – граница мультиколлинеарности, V_k – коэффициент инфляции дисперсии (Variance Inflation Factor; VIF), который показывает, насколько сильно взаимная корреляция между признаками увеличивает дисперсию:

$$V_k = (1 - D^2)^{-1}, \quad D^2 = \frac{\sum_{i=1}^n (x_{ij} - \bar{x}_j)^2}{\sum_{i=1}^n (x_{ij} - x_{ij})^2}, \quad (9)$$

где D^2 – коэффициент детерминации, \bar{x}_j – среднее значение, $\tilde{x}_{ij} = \beta_0 + \sum_{k \neq j} x_{ik} \beta_k$ – предсказанное значение, вычисленное через разложение $P_j = \beta_0 + \beta_1 P_1 + \beta_2 P_2 + \dots + \beta_m P_m$, $\beta_k = \rho_{jk} (\sigma_j / \sigma_k)$, ρ_{jk} – коэффициент корреляции, σ_j – стандартное отклонение.

III. Подмножество взаимозависимых признаков:

$$C^{III} = \left\{ C^{j\{k\}} \right\}_{j,k} = \bigcup_{1 \leq j \leq k \leq m^h} \left\{ (P_j, P_k) \in P^h : |M_{jk}| \geq \tau_M \right\}, \quad (10)$$

где M_{jk} вычисляется по (5), τ_M – порог сильной зависимости.

Для множества $C = C^I \cup C^{II} \cup C^{III}$ вводятся критерии исключения признаков. Признак $P \in C^{j\{k\}} \subseteq C$ исключается из P^h и $C = C \setminus C^{j\{k\}}$, если выполняется условие

$$\forall Q \in C^{j\{k\}} \setminus P : (S(P) < S(Q)) \vee (W(P) < W(Q)), \quad (11)$$

где $S(P) < S_{stab}$ – коэффициент стабильности, S_{stab} – порог, $W(P)$ – агрегированный вес признака.

Коэффициент стабильности вычисляется как

$$S(P) = \frac{1}{T-1} \sum_{t=1}^{T-1} M(P^{(t)}, P^{(t+1)}), \quad (12)$$

где T – количество случайных выборок, полученных из X , $M(P^{(t)}, P^{(t+1)})$ – взаимная информация (4) между признаком P на t -й и $(t+1)$ -й выборках.

Агрегированный вес признака определяются как

$$W(P) = \sum_{i=1}^k |v_i(P)| \cdot \frac{\lambda_i}{\sum_{m=1}^k \lambda_m}, \quad (13)$$

где $k = \min \left\{ j \mid \left(\sum_{i=1}^j \lambda_i / \sum_{i=1}^m \lambda_i \right) \geq \zeta \right\}$ – глубина редуцированного пространства, содержащего ζ дисперсии исходного набора данных, λ_i – собственное значение, $v_i(P)$ – компонента собственного вектора V_i , соответствующая признаку P . Действия по выбору и фильтрации признаков выполняются до тех пор, пока не останется групп для рассмотрения или признаков по критерию (11).

3. Применение подхода к данным сетевой активности

Предложенный подход был апробирован на данных сетевой активности устройств Интернета вещей, собранных в корпоративной сети научного центра. В рамках анализа данных были построены диаграммы рассеивания, демонстрирующие зависимости между парами признаков. На рис. 2 представлен пример визуализации, выполненный на логарифмической шкале, для признаков «Длительность потока (сек)» и «Объем переданных данных (байт)». Для каждой пары признаков добавлена тепловая карта плотности, показывающая распределение точек на графике, и построена аппроксимация данных с использованием полиномиальных или линейных моделей. Визуализация позволила оценить характер взаимосвязей между признаками, а также выделить аномалии в данных.

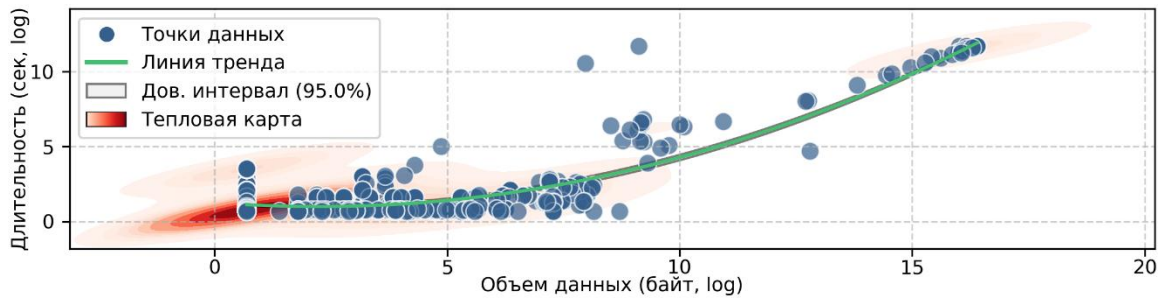


Рис. 2. Пример зависимых признаков
Fig. 2. Example of dependent features

Сформировано множество $C = C^I \cup C^II \cup C^III$ зависимых признаков по формулам (7)–(10). Для каждого признака из полученного множества рассчитан коэффициент стабильности по формуле (12) и исключены признаки, удовлетворяющие критерию (11), не сохраняющие стабильность своих характеристик на различных подвыборках данных и демонстрирующие существенную зависимость от более устойчивых признаков.

Для определения весовых коэффициентов признаков проведен анализ собственных значений и накопительной объясненной дисперсии (рис. 3). Установлено минимальное количество главных компонент, достаточное для описания данных (95% объясненной дисперсии). Вычислены весовые коэффициенты признаков в этих компонентах. Признаки, которые не вносят значимый вклад в формирование главных компонент или имеют меньший вес по сравнению с другими зависимыми от них признаками, согласно критерию (13), были удалены из рассмотрения. На каждом шаге вычисляется значение целевой функции (3), и результаты изменения признакового пространства, не удовлетворяющие условию (6), не принимаются.

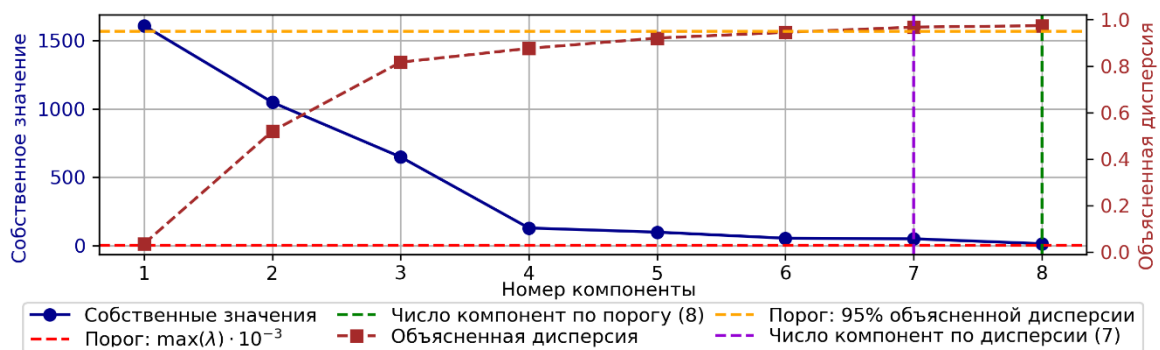


Рис. 3. Определение числа компонент для расчёта весов признаков
Fig. 3. Determining the number of components for calculating feature weights

В таблице приведены пошаговые результаты применения предложенного подхода и их сравнение с методом рекурсивного исключения признаков (RFE) для модели случайного леса.

Пошаговые результаты сокращения размерности

Выполненные действия	Признаков	Число обусл.	Сложность по Радермахеру		
			Лог. регресс.	Случ. лес	Лин. мод., случ. вес
0. Предобработка	275	768,26	0,264	0,199	0,417
1. Корр., нестабильные	223	450,6	0,253	0,196	0,415
2. Корр., незначимые	79	90,4	0,181	0,191	0,233
3. Мультиколл., нестабильные	25	17,08	0,108	0,163	0,145
4. Взаимные, нестабильные	22	12,5	0,099	0,143	0,117
5. Взаимные, незначимые	17	2,23	0,053	0,079	0,087
Время выполнения: 6,72 с (~ 1 000 строк), 58,46 с (~ 10 200 строк). Корр.: 0. Взаимозависимые: 0					
RFE (случ. лес)	17	75,17	0,097	0,182	0,083
Время выполнения: 41,42 с (~ 1 000 строк), 147,42 с (~ 10 200 строк). Корр.: 8. Взаимозависимые: 56					

Применение подхода к сокращению признакового пространства демонстрирует не только значительное сокращение размерности, но и улучшение ключевых характеристик данных. Снижение числа обусловленности свидетельствует о повышении численной устойчивости данных, что может быть дополнительно усилено за счет регуляризации сокращенной матрицы. Заметное уменьшение сложности моделей, описывающих данные, способствует снижению риска переобучения и улучшению их обобщающей способности.

Выполнено сравнение результатов кластеризации для полного и обработанного признакового пространства. На рис. 4 приведены графики силуэтного коэффициента, которые показывают повышение компактности кластеров и их лучшую разделенность.

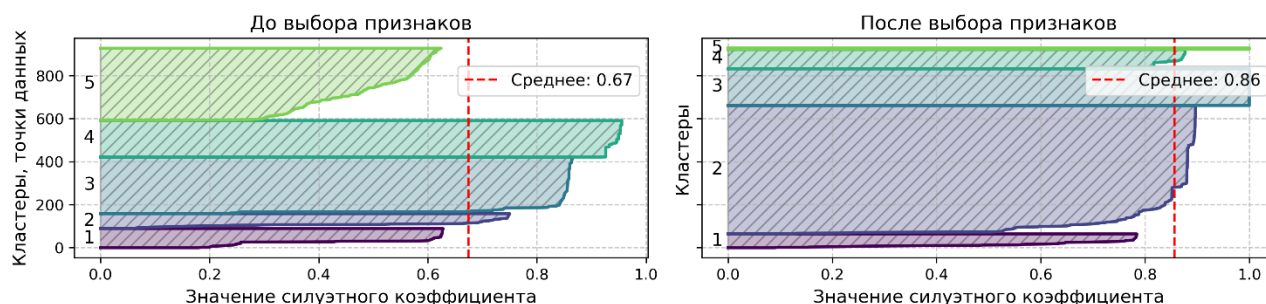


Рис. 4. Сравнение силуэтного коэффициента
Fig. 4 Comparison of silhouette coefficient

В результате анализа объектов, отнесенных к выделенным кластерам, было установлено, что сокращение размерности признакового пространства обеспечило возможность описать характеристики сформированных групп и предложить объяснения для наблюдаемого разбиения. Первый кластер объединяет попытки установления соединений без получения ответа от принимающих устройств. Второй содержит легитимные сессии, сопровождаемые передачей данных. Третий также включает легитимные сессии, но значительной длительности, характеризующие взаимодействие с устройствами интернета вещей. Четвёртый кластер состоит из запросов на соединение и коротких сессий, соответствующих сканированию сети на транспортном уровне без детализации по прикладному протоколу. Пятый включает короткие соединения, представляющие собой более глубокое сканирование сети на уровне прикладного протокола, в отличие от поверхностного анализа, характерного для предыдущего кластера. Таким образом, проведенное исследование позволило повысить степень интерпретируемости данных.

Заключение

Предложенный в исследовании подход направлен на решение ключевых проблем, возникающих при работе с высокоразмерными данными, таких как избыточность признаков, наличие корреляций и сложных нелинейных зависимостей. Он позволяет эффективно сокращать размерность признакового пространства, одновременно улучшая численные характеристики данных и снижая сложность моделей машинного обучения. Основные преимущества подхода включают:

1. Эффективное устранение избыточных зависимостей, достигаемое за счет комбинированного использования взаимной информации и анализа корреляций.
2. Обеспечение семантической интерпретируемости данных за счет сохранения исходных признаков и связей между объектами.
3. Улучшение характеристик кластеризации и обобщающей способности моделей благодаря уменьшению нестабильности признаков и снижению сложности.

Существенное сокращение размерности признакового пространства без потери информативности делает данный подход применимым для широкого спектра задач – от предварительной обработки данных до повышения эффективности алгоритмов классификации, регрессии и кластеризации.

Список источников

1. Курбатов В.И., Папа О.М. Интернет вещей: основные концепции и тренды // Гуманитарные, социально-экономические и общественные науки. 2023. № 1. С. 48–54.
2. Минаев В.А., Швырев Б.А., Ромашкин Т.Р. Безопасность интернета вещей: основные решения // Информация и безопасность. 2023. Т. 26 (2). С. 163–168.
3. Chatterjee A., Ahmed B.S. IoT anomaly detection methods and applications: A survey // Internet of Things. 2022. V. 19. Art. 100568. doi: 10.1016/j.iot.2022.100568
4. Ring M., Wunderlich S., Scheuring D., Landes D., Hotho A. A survey of network-based intrusion detection data sets // Computers & Security. 2019. V. 86. P. 147–167. doi: 10.1016/j.cose.2019.06.005
5. Xinpeng C. CICIDS2017 and UNBSW-NB15 // IEEE Dataport. 2023. doi: 10.21227/ykpn-jx78. URL: <https://iee-dataport.org/documents/cicids2017-and-unbsw-nb15>
6. Liu L., Wang P., Lin J., Liu L. Intrusion detection of imbalanced network traffic based on machine learning and deep learning // IEEE Access. 2021. V. 9. P. 7550–7563. doi: 10.1109/ACCESS.2020.3048198
7. Moustafa N. A new distributed architecture for evaluating AI-based security systems at the edge: Network TON_IoT datasets // Sustainable Cities and Society. 2021. V. 72. Art. 102994.
8. Горюнов М.Н., Мацкевич А.Г., Рыболовлев Д.А. Синтез модели машинного обучения для обнаружения компьютерных атак на основе набора данных CICIDS2017 // Труды Института системного программирования РАН. 2020. Т. 32 (5). С. 81–94. doi: 10.15514/ISPRAS–2020–32(5)–6.
9. Исаева О.С., Кулясов Н.В., Исаев С.В. Инфраструктура сбора данных и имитации угроз безопасности сети интернета вещей // Сибирский аэрокосмический журнал. 2025. Т. 26, № 1. С. 8–20. doi: 10.31772/2712-8970-2025-26-1-8-20
10. Altarabichi M.G., Nowaczyk S., Pashami S., Mashhadi P.S. Surrogate-assisted genetic algorithm for wrapper feature selection // IEEE Congress on Evolutionary Computation. 2021. P. 776–785. doi: 10.1109/CEC45853.2021.9504718
11. Liu S., Motani M. Improving mutual information based feature selection by boosting unique relevance // Journal of Artificial Intelligence Research. 2025. V. 82. P. 1267–1292. doi: 10.1613/jair.1.17219
12. Исаева О.С. Построение цифрового профиля устройств Интернета вещей // Информационные и математические технологии в науке и управлении. 2023. № 2 (30). С. 36–44. doi: 10.25729/ESI.2023.30.2.004
13. Shafi M., Lashkari A., Roudsari A. NTLFlowLyzer: Towards generating an intrusion detection dataset and intruders behavior profiling through network and transport layers traffic analysis and pattern extraction // Computers & Security. 2025. V. 148. Art. 104160. doi: 10.1016/j.cose.2024.104160
14. Исаева О.С. Построение онтологии для систематизации характеристик сети Интернета вещей // Онтология проектирования. 2024. Т. 14, № 2 (52). С. 243–255. doi: 10.18287/2223-9537-2024-14-2-243-255
15. Priyatno A.M., Widiyaningtyas T. A systematic literature review: recursive feature elimination algorithms // JITK. 2024. V. 9 (2). P. 196–207. doi: 10.33480/jitk.v9i2.5015
16. Цурко В.В., Михальский А.И. Оценка статистической связи случайных величин через взаимную информацию // Автоматика и телемеханика. 2022. Вып. 5. С. 76–86. doi: 10.31857/S0005231022050063
17. Koltchinskii V. Rademacher complexities and bounding the excess risk // Journal of Machine Learning Research. 2010. V. 11. P. 2457–2485.

References

1. Kurbatov, V.I. & Papa, O.M. (2023) Internet of things: main concepts and trends. *Gumanitarnye, sotsial'no-ekonomicheskie i obshchestvennye nauki*. 1. pp. 48–54.
2. Minaev, V.A., Shvyrev, B.A. & Romashkin, T.R. (2023) Internet of things security: main solutions. *Informatsiya i bezopasnost' – Information and Security*. 26(2). pp. 163–168.
3. Chatterjee, A. & Ahmed, B. S. (2022) IoT anomaly detection methods and applications: a survey. *Internet of Things*. 19. Art. 100568. doi: 10.1016/j.iot.2022.100568.
4. Ring, M., Wunderlich, S., Scheuring, D., Landes, D. & Hotho A. (2019) A survey of network-based intrusion detection data sets. *Computers & Security*. 86. pp. 147–167. doi: 10.1016/j.cose.2019.06.005
5. Xinpeng, C. (2023) CICIDS2017 and UNBSW-NB15. *IEEE Dataport*. doi: 10.21227/ykpn-jx78.
6. Liu, L., Wang, P. & Lin, J. (2021) Intrusion detection of imbalanced network traffic based on machine learning and deep learning. *IEEE Access*. 9. pp. 7550–7563. doi: 10.1109/ACCESS.2020.3048198
7. Moustafa, N. (2021) A new distributed architecture for evaluating AI-based security systems at the edge: Network TON_IoT datasets, *Sustainable Cities and Society*. 72. Art. 102994.
8. Goryunov, M.N., Matskevich, A.G. & Rybolovlev, D.A. (2020) Synthesis of a machine learning model for detecting computer attacks based on the CICIDS2017 dataset. *Trudy ISP RAN*. 32(5). pp. 81–94. doi: 10.15514/ISPRAS–2020–32(5)–6
9. Isaeva, O.S., Kulyasov, N.V. & Isaev, S.V. (2025) Infrastructure for collecting data and simulating security threats to the internet of things. *Sibirskiy aerokosmicheskii zhurnal – Siberian Aerospace Journal*. 26(1). pp. 8–20. doi: 10.31772/2712-8970-2025-26-1-8-20

10. Altarabichi, M.G., Nowaczyk, S., Pashami, S. & Mashhadi, P.S. (2021) Surrogate-assisted genetic algorithm for wrapper feature selection. *IEEE Congress on Evolutionary Computation*. pp. 776–785. DOI: 10.1109/CEC45853.2021.9504718.
11. Liu, S. & Motani, M. (2025) Improving mutual information-based feature selection by boosting unique relevance. *Journal of Artificial Intelligence Research*. 82. pp. 1267–1292. doi: 10.1613/jair.1.17219.
12. Isaeva, O.S. (2023) Construction of a digital profile of internet of things devices. *Informatsionnye i matematicheskie tekhnologii v nauke i upravlenii – Information and Mathematical Technologies in Science and Management*. 2(30). pp. 36–44. doi: 10.25729/ESI.2023.30.2.004
13. Shafi, M., Lashkari, A. & Roudsari, A. (2025) NTLFlowLyzer: Towards generating an intrusion detection dataset and intruders' behavior profiling through network and transport layers traffic analysis and pattern extraction. *Computers & Security*. 148. Art. 104160. doi: 10.1016/j.cose.2024.104160
14. Isaeva, O.S. (2024) Construction of an ontology for systematization of characteristics of the Internet of Things network. *Ontologiya proektirovaniya – Ontology of Design*. 14(2(52)). pp. 243–255. doi: 10.18287/2223-9537-2024-14-2-243-255
15. Priyatno, A.M. & Widiyaningtyas, T. (2024) A systematic literature review: recursive feature elimination algorithms. *JITK*. 9(2). pp. 196–207. doi: 10.33480/jitk.v9i2.5015
16. Tsurko, V.V. & Mikhalskii, A.I. (2022) Estimation of statistical connection of random variables through mutual information. *Automation and Telemekhanics*. 5. pp. 76–86. doi: 10.31857/S0005231022050063
17. Koltchinskii, V. (2010) Rademacher complexities and bounding the excess risk. *Journal of Machine Learning Research*. 11. pp. 2457–2485.

Информация об авторах:

Исаева Ольга Сергеевна – доктор технических наук, старший научный сотрудник Регионального научно-образовательного математического центра «Красноярский математический центр» Института вычислительного моделирования СО РАН – обособленного подразделения ФИЦ КНЦ СО РАН (Красноярск, Россия). E-mail: isaeva@icm.krasn.ru

Исаев Сергей Владиславович – кандидат технических наук, заместитель директора Института вычислительного моделирования СО РАН – обособленного подразделения ФИЦ КНЦ СО РАН (Красноярск, Россия). E-mail: si@icm.krasn.ru

Кулясов Никита Владимирович – программист Регионального научно-образовательного математического центра «Красноярский математический центр» Института вычислительного моделирования СО РАН – обособленного подразделения ФИЦ КНЦ СО РАН (Красноярск, Россия). E-mail: razor@icm.krasn.ru

Вклад авторов: все авторы сделали эквивалентный вклад в подготовку публикации. Авторы заявляют об отсутствии конфликта интересов.

Information about the authors:

Isaeva Olga S. (Doctor of Technical Sciences, Senior Researcher, Regional Mathematical Center “Krasnoyarsk Mathematical Center”, Institute of Computational Modelling SB RAS – subdivision of the FRC KSC SB RAS, Krasnoyarsk, Russia). E-mail: isaeva@icm.krasn.ru

Isaev Sergey V. (Candidate of Technical Sciences, Deputy Director, Institute of Computational Modelling SB RAS – subdivision of the FRC KSC SB RAS, Krasnoyarsk, Russia). E-mail: si@icm.krasn.ru

Kulyasov Nikita V. (Programmer, Regional Scientific and Educational Mathematical Center “Krasnoyarsk Mathematical Center”, Institute of Computational Modelling SB RAS – subdivision of the FRC KSC SB RAS, Krasnoyarsk, Russia). E-mail: razor@icm.krasn.ru

Contribution of the authors: the authors contributed equally to this article. The authors declare no conflicts of interests.

Поступила в редакцию 27.04.2025; принята к публикации 02.12.2025

Received 27.04.2025; accepted for publication 02.12.2025

Научная статья

УДК 519.2

doi: 10.17223/19988605/73/13

**Рекуррентный обобщенный асинхронный поток событий
с продлевающимся мертвым временем в особом случае****Людмила Алексеевна Нежелская¹, Валентина Денисовна Пономаренко²**^{1,2} *Национальный исследовательский Томский государственный университет, Томск, Россия*¹ *ludne@mail.ru*² *valya.ponomarenko.00@mail.ru*

Аннотация. Рассматривается дважды стохастический рекуррентный обобщенный асинхронный поток событий с двумя состояниями, функционирующий в стационарном режиме в условиях продлевающегося мертвого времени фиксированной длительности. Методом моментов решается задача оценивания длительности мертвого времени в рекуррентном обобщенном асинхронном потоке в особом случае соотношения параметров. Приводятся численные результаты статистических экспериментов, поставленных на имитационной модели потока, и анализ полученных оценок длительности мертвого времени.

Ключевые слова: рекуррентный обобщенный асинхронный поток событий; особый случай соотношения параметров потока; инфинитезимальные характеристики; продлевающееся мертвое время; совместная плотность вероятности; условия рекуррентности; преобразование Лапласа; метод моментов; статистические эксперименты.

Для цитирования: Нежелская Л.А., Пономаренко В.Д. Рекуррентный обобщенный асинхронный поток событий с продлевающимся мертвым временем в особом случае // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2025. № 73. С. 110–121. doi: 10.17223/19988605/73/13

Original article

doi: 10.17223/19988605/73/13

**A recurrent generalized asynchronous event flow
with extended dead time in a special case****Lyudmila A. Nezhel'skaya¹, Valentina D. Ponomarenko²**^{1,2} *National Research Tomsk State University, Tomsk, Russian Federation*¹ *ludne@mail.ru*² *valya.ponomarenko.00@mail.ru*

Abstract. We consider a double stochastic recurrent generalized asynchronous flow of events with two states, operating in a stationary mode under conditions of an extended dead time of a fixed duration. The problem of estimating the duration of dead time in a recurrent generalized asynchronous flow of events has been solved for a specific parameter relationship. Numerical results of statistical experiments performed on a simulation model are presented.

Keywords: double stochastic flow of events; generalized asynchronous flow of events; infinitesimal characteristics; extending dead time; joint probability density; low recurrence conditions; Laplace transform; method of moments; statistical experiments.

For citation: Nezhel'skaya, L.A., Ponomarenko, V.D. (2025) A recurrent generalized asynchronous event flow with extended dead time in a special case. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naja tehnika i informatika – Tomsk State University Journal of Control and Computer Science*. 73. pp. 110–121. doi: 10.17223/19988605/73/13

Введение

Среди математических моделей, описывающих потоки заявок в компьютерных и телекоммуникационных сетях, наиболее приближенной к реальности считается модель дважды стохастического потока [1]. Для нее характерно следующее: сопровождающий случайный процесс является принципиально ненаблюдаемым, и моменты наступления событий случайны. Дважды стохастические потоки делятся на два класса: к первому классу относятся потоки, интенсивность которых есть непрерывный случайный процесс [2, 3], ко второму – потоки с интенсивностью в виде кусочно-постоянного случайного процесса с конечным числом состояний [4–9].

Многие исследования в теории массового обслуживания предполагают, что все события входящего потока в систему доступны для наблюдения. Однако это условие на практике часто не выполняется, поскольку поступившее событие может вызывать период мертвого времени регистрирующего прибора, в течение которого последующие события потока становятся ненаблюдаемыми (теряются) [10]. Наличие мертвого времени осложняет решение задачи оценивания состояний потока или его параметров по наблюдаемым моментам времени наступления событий.

В работах [11–14] решены задачи оценивания длительности мертвого времени, когда наступившие в течение периода мертвого времени события не вызывают его продления (непродлевающееся мертвое время).

Поток событий, рассматриваемый в данной работе, относится ко второму классу дважды стохастических потоков. В отличие от упомянутых исследований здесь предполагается, что событие, наступившее в период мертвого времени, хотя и не наблюдается, способно продлить общий период ненаблюдаемости (продлевающееся мертвое время). При этом длительность мертвого времени считается фиксированной, а общий период ненаблюдаемости потока является случайной величиной.

В работе [15] решена задача оценивания длительности продлевающегося мертвого времени в обобщенном асинхронном потоке событий в случае соотношения параметров $z_2 - z_1 \neq 0$ (общий случай). Однако интерес представляет решение задачи оценивания длительности продлевающегося мертвого времени в рекуррентном обобщенном асинхронном потоке событий в случае $z_2 - z_1 = 0$, называемом в дальнейшем **особым** случаем. Аналитически выводится уравнение моментов, решение которого возможно численно; проводятся статистические эксперименты с целью установления качества получаемых оценок.

1. Постановка задачи

Рассматривается обобщенный асинхронный поток событий, обладающий следующими свойствами: стационарностью, ординарностью, наличием последействия. В общем случае исследуемый поток является коррелированным.

Сопровождающий процесс данного потока $\lambda(t)$ является принципиально ненаблюдаемым кусочно-постоянным случайным процессом с двумя состояниями S_1 и S_2 ; будем говорить, что имеет место состояние S_i процесса $\lambda(t)$, если $\lambda(t) = \lambda_i$, $i = 1, 2$, $\lambda_1 > \lambda_2 \geq 0$. Длительность пребывания процесса $\lambda(t)$ в состоянии S_i является случайной величиной с функцией распределения $F_i(t) = 1 - e^{-\lambda_i t}$, $t \geq 0$, $i = 1, 2$. В течение времени пребывания процесса $\lambda(t)$ в состоянии S_1 имеет место пуассоновский поток событий с параметром λ_1 ; в течение времени пребывания процесса $\lambda(t)$ в состоянии S_2 – пуассоновский поток событий с параметром λ_2 .

В момент перехода сопровождающего случайного процесса $\lambda(t)$ из состояния S_1 в состояние S_2 с вероятностью p ($0 < p \leq 1$) инициируется дополнительное событие потока в состоянии S_2 (сначала переход из S_1 в S_2 , затем наступление дополнительного события в S_2). Аналогично при переходе сопровождающего случайного процесса $\lambda(t)$ из состояния S_2 в состояние S_1 с вероятностью q ($0 < q \leq 1$)

инициируется дополнительное событие потока в состоянии S_1 (сначала переход из S_2 в S_1 , затем наступление дополнительного события в S_1).

В момент наступления каждого события потока наступает период ненаблюдаемости фиксированной длительности T (мертвое время), так что другие события, наступившие в течение времени T , недоступны наблюдению. Каждое событие, ненаблюдаемое в течение мертвого времени, вызывает продление периода ненаблюдаемости на величину T ; наблюдаться будет лишь то событие, которое наступило после окончания последнего периода ненаблюдаемости.

На рис. 1 приведены схема формирования наблюдаемого потока событий и одна из реализаций сопровождающего процесса $\lambda(t)$, где $S_i, i=1,2$, – состояния процесса $\lambda(t)$; t_1, t_2, \dots – моменты времени наступления событий в наблюдаемом потоке. Наблюдаемые события обозначены незакрашенными кругами, а ненаблюдаемые события, т.е. недоступные наблюдению из-за наличия мертвого времени, обозначены закрашенными кругами. Штриховкой обозначен период ненаблюдаемости. Длительность общего периода ненаблюдаемости ξ – случайная величина.

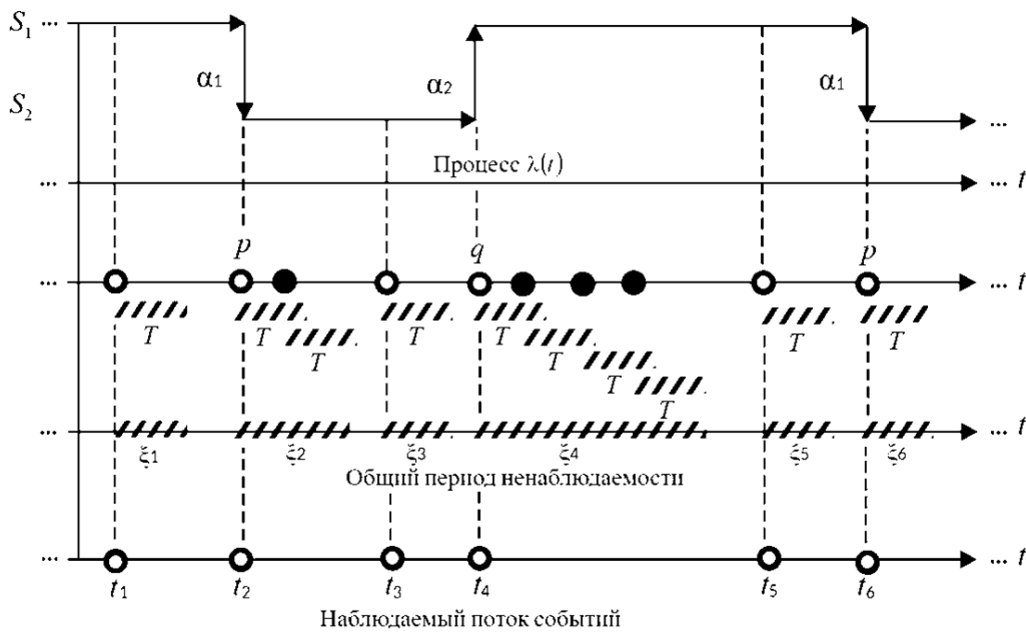


Рис. 1. Формирование наблюдаемого потока событий
Fig. 1. Formation of the observed event flow

Матрицы инфинитезимальных характеристик сопровождающего процесса $\lambda(t)$ имеют вид [16]:

$$\mathbf{D}_0 = \begin{bmatrix} -(\lambda_1 + \alpha_1) & (1-p)\alpha_1 \\ (1-q)\alpha_2 & -(\lambda_2 + \alpha_2) \end{bmatrix}, \quad \mathbf{D}_1 = \begin{bmatrix} \lambda_1 & p\alpha_1 \\ q\alpha_2 & \lambda_2 \end{bmatrix}.$$

Элементами матрицы \mathbf{D}_1 являются интенсивности переходов процесса $\lambda(t)$ из состояния S_i в состояние $S_j, i, j=1,2$, с наступлением события потока. Недиagonальные элементы матрицы \mathbf{D}_0 – интенсивности переходов процесса $\lambda(t)$ из состояния S_i в состояние $S_j, i, j=1,2$, без наступления события. Диагональные элементы матрицы \mathbf{D}_0 – интенсивности выхода процесса $\lambda(t)$ из своих состояний, взятые с противоположным знаком.

В работе [15] проведено исследование в случае $z_2 - z_1 \neq 0$, где

$$z_1 = \frac{1}{2} \left(\lambda_1 + \alpha_1 + \lambda_2 + \alpha_2 - \sqrt{(\lambda_1 + \alpha_1 - \lambda_2 - \alpha_2)^2 + 4\alpha_1\alpha_2(1-p)(1-q)} \right),$$

$$z_2 = \frac{1}{2} \left(\lambda_1 + \alpha_1 + \lambda_2 + \alpha_2 + \sqrt{(\lambda_1 + \alpha_1 - \lambda_2 - \alpha_2)^2 + 4\alpha_1\alpha_2(1-p)(1-q)} \right),$$

$$z_2 > z_1.$$

Отметим, что при получении одномерной плотности в [15] имеет место деление на величину $z_2 - z_1$. Из явного вида параметров z_1 и z_2 следует, что ситуация $z_2 - z_1 = 0$ возможна, если одновременно выполняются условия $\lambda_1 + \alpha_1 - \lambda_2 - \alpha_2 = 0$, $(1-p)(1-q) = 0$.

Решение системы уравнений приводит к трем вариантам соотношения параметров потока:

- 1) $\lambda_1 + \alpha_1 - \lambda_2 - \alpha_2 = 0$, $p = 1, 0 < q < 1$;
- 2) $\lambda_1 + \alpha_1 - \lambda_2 - \alpha_2 = 0$, $q = 1, 0 < p < 1$;
- 3) $\lambda_1 + \alpha_1 - \lambda_2 - \alpha_2 = 0$, $p = q = 1$.

Цель настоящего исследования – вывод и решение уравнения моментов для получения оценки длительности продлевающегося мертвого времени в исследуемом потоке в особом случае соотношения параметров потока, а также проведение на имитационной модели потока серии статистических экспериментов и анализ качества полученных оценок.

2. Условия рекуррентности обобщенного асинхронного потока в особом случае

Будем рассматривать первый вариант задания параметров потока. Предполагается, что обобщенный асинхронный поток событий функционирует в условиях полной наблюдаемости ($T = 0$) в стационарном режиме.

Лемма 1. Для обобщенного асинхронного потока событий, функционирующего в условиях полной наблюдаемости, сопровождающий кусочно-постоянный случайный процесс $\lambda(t)$ является марковским процессом [16].

Пусть $t_1, t_2, \dots, t_k, \dots$ – моменты времени наступления событий в потоке.

Лемма 2. Последовательность $\{\lambda(t_k)\}$, порождаемая совокупностью моментов наступления событий $t_1, t_2, \dots, t_k, \dots$, является вложенной цепью Маркова [16].

Обозначим $\tau_k = t_{k+1} - t_k$, $\tau_k \geq 0$, – значение длительности k -го интервала между соседними событиями t_k и t_{k+1} , $k = 1, 2, \dots$, наблюдаемого потока. В силу стационарного режима функционирования потока для плотности вероятности значений τ_k справедливо $p(\tau_k) = p(\tau)$, $\tau_k \geq 0$, $\forall k$.

Это позволяет без ограничения общности считать момент наступления события t_k равным нулю, или, что то же самое, момент наступления события есть $\tau = 0$.

В общем случае исследуемый поток в условиях полной наблюдаемости является коррелированным потоком. Для достижения поставленной цели необходимо выписать условия рекуррентности потока и определить явный вид плотности вероятности значений длительности интервала между соседними событиями потока.

Воспользуемся следующей теоремой.

Теорема 1. Плотность вероятности значений длительности интервала между соседними событиями в обобщенном асинхронном потоке в особом случае имеет вид [17]:

$$p(\tau) = [(\lambda_1 + \alpha_1) - \pi_2(0)\alpha_2(1-q)(1-(\lambda_1 + \alpha_1)\tau)]e^{-(\lambda_1 + \alpha_1)\tau}, \tau \geq 0; \quad \pi_2(0) = \frac{\alpha_1(\lambda_1 + \alpha_1)}{\alpha_1(\lambda_1 + \alpha_1) + \alpha_2(\lambda_1 + \alpha_1 q)}. \quad (1)$$

Обозначим $p(\tau_1, \tau_2)$ – совместная плотность вероятности длительностей двух соседних интервалов (t_1, t_2) , (t_2, t_3) ; $\tau_1 = t_2 - t_1$, $\tau_2 = t_3 - t_2$ – значения длительностей, $\tau_1 \geq 0$, $\tau_2 \geq 0$. В силу стационарного режима расположение одного интервала (t_k, t_{k+1}) либо двух соседних интервалов (t_k, t_{k+1}) , (t_{k+1}, t_{k+2}) между моментами наступления событий $t_1, t_2, \dots, t_k, \dots$, $k = 1, 2, \dots$, на временной оси произвольно.

Справедлива следующая теорема.

Теорема 2. Совместная плотность вероятности $p(\tau_1, \tau_2)$ длительностей двух соседних интервалов для обобщенного асинхронного потока событий в особом случае имеет вид [17]:

$$p(\tau_1, \tau_2) = p(\tau_1)p(\tau_2) - \frac{[\alpha_1\alpha_2(1-q)]^2(\lambda_1\lambda_2 - \alpha_1\alpha_2q)}{[(\lambda_1 + \alpha_1)^2 - (\lambda_1\lambda_2 - \alpha_1\alpha_2q)]^2} \times \\ \times (1 - (\lambda_1 + \alpha_1)\tau_1)(1 - (\lambda_1 + \alpha_1)\tau_2)e^{-(\lambda_1 + \alpha_1)(\tau_1 + \tau_2)}, \\ \tau_1 \geq 0, \tau_2 \geq 0, \quad (2)$$

где $p(\tau_k)$ определены в (1) для $\tau = \tau_k, k = 1, 2$.

Из анализа (2) вытекает условие рекуррентности $\lambda_1\lambda_2 - q\alpha_1\alpha_2 = 0$, с использованием которого находится величина $\pi_2(0)\alpha_2(1-q) = \alpha_2 - \lambda_1$, входящая в (1). Тогда плотность вероятности значений длительности интервала между соседними событиями в рекуррентном обобщенном асинхронном потоке событий принимает вид:

$$p(\tau) = [\lambda_1 + \lambda_2 + (\alpha_1\alpha_2 - \lambda_1\lambda_2)\tau]e^{-(\lambda_1 + \alpha_1)\tau}, \tau \geq 0. \quad (3)$$

3. Преобразование Лапласа плотности вероятности общего периода ненаблюдаемости рекуррентного потока

Пусть ξ – длительность общего периода ненаблюдаемости в рекуррентном обобщенном асинхронном потоке, функционирующем в условиях продлевающегося мертвого времени фиксированной длительности $T \neq 0$. Последовательность t_1, t_2, \dots моментов наступления событий в наблюдаемом потоке образует вложенную цепь Маркова, и рекуррентность наблюдаемого потока сохраняется.

Рассмотрим функцию Пальма $\varphi_0(T) = \int_T^\infty \tilde{p}(x)dx$ – вероятность того, что на интервале $(0, T)$ событий рекуррентного потока не наступит при условии, что в начальный момент времени интервала $(0, T)$ событие наступило [18]; здесь $\tilde{p}(x)$ – плотность вероятности длительности интервала между соседними событиями в дважды стохастическом рекуррентном потоке.

Теорема 3. Преобразование Лапласа плотности вероятности значений длительности общего периода ненаблюдаемости в рекуррентном обобщенном асинхронном потоке с продлевающимся мертвым временем в особом случае имеет вид:

$$g_\xi(s) = \frac{\varphi_0(T)}{e^{sT}} \left\{ 1 - \frac{1 - e^{-(\lambda_1 + \alpha_1 + s)T}}{(\lambda_1 + \alpha_1 + s)^2} [(\lambda_1 + \alpha_1)^2 + s(\lambda_1 + \lambda_2)] + (\alpha_2 - \lambda_1)(\lambda_1 + \alpha_1) \frac{T}{\lambda_1 + \alpha_1 + s} e^{-(\lambda_1 + \alpha_1 + s)T} \right\}^{-1}, \quad (4)$$

где $\varphi_0(T) = [1 + (\alpha_2 - \lambda_1)T]e^{-(\lambda_1 + \alpha_1)T}$.

Доказательство. Воспользуемся результатом, полученным в [17]. Преобразование Лапласа плотности вероятности значений длительности общего периода ненаблюдаемости в рекуррентном дважды стохастическом потоке событий, функционирующем в условиях продлевающегося мертвого времени, имеет вид:

$$g_\xi(s) = \frac{\varphi_0(T)}{e^{sT}} \left[1 - \int_0^T e^{-sx} \tilde{p}(x)dx \right]^{-1}. \quad (5)$$

Переобозначая плотность $p(\tau)$, определенную в (3), на $\tilde{p}(x)$, подставляя в (5) и выполняя необходимые преобразования, приходим к (4). Теорема доказана.

Лемма 3. Математическое ожидание длительности общего периода ненаблюдаемости ξ в рекуррентном обобщенном асинхронном потоке событий с продлевающимся мертвым временем в особом случае имеет вид:

$$M\xi = \frac{1}{\varphi_0(T)(\lambda_1 + \alpha_1)^2} \left[(1 - \varphi_0(T))(\lambda_1 + \alpha_1) + (\alpha_2 - \lambda_1)(1 - e^{-(\lambda_1 + \alpha_1)T}) \right], \quad (6)$$

где $\varphi_0(T)$ определена в (4).

Доказательство. Нетрудно показать, что с использованием вида $g_\xi(s)$, определенного в (4), математическое ожидание общего периода ненаблюдаемости ξ запишется в виде:

$$M\xi = -g'_\xi(s)\big|_{s=0} = T + \frac{1}{\Phi_0(T)} \int_0^T x\tilde{p}(x)dx. \quad (7)$$

Подставляя $\tilde{p}(x)$, определенную в (3), в (7) и выполняя необходимые преобразования, находим (6). Лемма доказана.

4. Преобразование Лапласа плотности вероятности длительности интервала между соседними событиями в наблюдаемом потоке

Рассмотрим интервал времени (t_k, t_{k+1}) , значение длительности которого есть $\tau_k = t_{k+1} - t_k$. В то же время длительность этого интервала равна $\tau = \xi + \eta$, где η – длительность интервала между моментом окончания общего периода ненаблюдаемости и моментом t_{k+1} . В силу произвольного расположения интервала на временной оси индекс k опущен. Случайные величины η и ξ являются зависимыми. Тогда плотность вероятности $p(\tau)$ длительности интервала между событиями в наблюдаемом потоке запишется в виде:

$$p(\tau) = \int_0^\tau p(\xi) p(\eta|\xi) d\xi = \int_0^\tau p(\xi) p(\tau - \xi|\xi) d\xi. \quad (8)$$

Теорема 4. Преобразование Лапласа плотности вероятности значений длительности интервала между соседними событиями в рекуррентном обобщенном асинхронном потоке с продлевающимся мертвым временем в особом случае имеет вид:

$$g_\tau(s) = \frac{1}{(\lambda_1 + \alpha_1 + s)^2 (\alpha_1 + \alpha_2)} \left\{ (\lambda_1 + \alpha_1)^2 (\alpha_1 + \alpha_2 + s) g_\xi(s) - (\lambda_1 - \alpha_2)^2 s g_\xi(\alpha_1 + \alpha_2 + s) \right\}, \quad (9)$$

где функция $g_\xi(s)$ определена формулой (4); $g_\xi(\alpha_1 + \alpha_2 + s)$ определена формулой (4), в которой нужно заменить аргумент s на $\alpha_1 + \alpha_2 + s$.

Доказательство. Найдем выражение для $p(\tau - \xi|\xi)$. Пусть момент наступления события в наблюдаемом потоке есть $\tau = 0$. Рассмотрим временной интервал $(0, \tau) = (0, \xi + \eta)$ и зафиксируем ξ . Пусть $p_{ij}(\tau - \xi)$ – условные вероятности того, что в интервале длительности $\eta = \tau - \xi$ не наступит событий наблюдаемого потока и $\lambda(\xi + \eta) = \lambda_j$ при условии, что $\lambda(\xi) = \lambda_i$, $i, j = 1, 2$. По смыслу введенные условные вероятности не отличаются от вероятностей $p_{ij}(\tau)$ для обобщенного асинхронного потока в особом случае, полученных в [17]:

$$\begin{aligned} p_{11}(\tau) &= e^{-(\lambda_1 + \alpha_1)\tau}, \quad p_{12}(\tau) = 0, \\ p_{21}(\tau) &= (1 - q)\alpha_2 \tau e^{-(\lambda_1 + \alpha_1)\tau}, \quad p_{22}(\tau) = e^{-(\lambda_1 + \alpha_1)\tau}, \quad \tau \geq 0. \end{aligned} \quad (10)$$

Тогда условные вероятности $p_{ij}(\tau - \xi)$, $i, j = 1, 2$, определяются формулами (10), в которых аргумент τ заменен на $\tau - \xi$.

Введем $P_i(\tau - \xi) = p_{i1}(\tau - \xi) + p_{i2}(\tau - \xi)$ – условная вероятность того, что в интервале (ξ, τ) событий наблюдаемого потока не произойдет при условии, что $\lambda(\xi) = \lambda_i$, $i = 1, 2$. Тогда условная плотность вероятности длительности интервала (ξ, τ) по определению есть $p_i(\tau - \xi) = -P'_i(\tau - \xi)$, $i = 1, 2$. Учитывая (10), находим

$$\begin{aligned} p_1(\tau - \xi) &= (\lambda_1 + \alpha_1) e^{-(\lambda_1 + \alpha_1)(\tau - \xi)}, \\ p_2(\tau - \xi) &= -\left\{ \alpha_2(1 - q) - (\lambda_1 + \alpha_1) [\alpha_2(1 - q)(\tau - \xi) + 1] \right\} e^{-(\lambda_1 + \alpha_1)(\tau - \xi)}, \\ &\quad \tau \geq \xi. \end{aligned} \quad (11)$$

Введем вероятность $\pi_i(\tau|\xi)$ – условная вероятность того, что $\lambda(\tau) = \lambda_i$, $i = 1, 2$, при условии, что в момент $\tau = 0$ событие наступило и наступило мертвое время длительности ξ . Тогда

$$p(\tau - \xi|\xi) = \pi_1(\tau - \xi|\xi)p_1(\tau - \xi) + \pi_2(\tau - \xi|\xi)p_2(\tau - \xi). \quad (12)$$

Вероятности $\pi_i(\tau - \xi|\xi)$ по смыслу совпадают с вероятностями $\pi_i(T)$, $i = 1, 2$, определенными в [17], в которых вместо периода ненаблюдаемости T следует рассматривать ξ :

$$\pi_2(\tau - \xi|\xi) = \pi_2(\xi) = \pi_2 - [\pi_2 - \pi_2(0|\xi)]e^{-(\alpha_1 + \alpha_2)\xi}, \quad \pi_2(0|\xi) = \frac{p_{12} + \delta\pi_2[1 - e^{-(\alpha_1 + \alpha_2)\xi}]}{1 - \delta e^{-(\alpha_1 + \alpha_2)\xi}}, \quad (13)$$

$$p_{12} = \frac{\alpha_1}{\lambda_1 + \alpha_1}, \quad \pi_2 = \frac{\alpha_1}{\alpha_1 + \alpha_2}, \quad \delta = \frac{\lambda_1\lambda_2 - q\alpha_1\alpha_2}{(\lambda_1 + \alpha_1)^2}; \quad \pi_1(\tau - \xi|\xi) = 1 - \pi_2(\tau - \xi|\xi).$$

Принимая во внимание условие рекуррентности $\lambda_1\lambda_2 - q\alpha_1\alpha_2 = 0$, запишем формулы (13) в виде:

$$\pi_2(0|\xi) = \frac{\alpha_1}{\lambda_1 + \alpha_1}, \quad \pi_2(\xi) = \frac{\alpha_1}{(\lambda_1 + \alpha_1)(\alpha_1 + \alpha_2)}(\lambda_1 + \alpha_1 + (\alpha_2 - \lambda_1)e^{-(\alpha_1 + \alpha_2)\xi}), \quad \delta = 0. \quad (14)$$

Подставляя (11), (14) в (12), находим

$$p(\tau - \xi|\xi) = \begin{cases} 0, & 0 \leq \tau < \xi, \\ \{(\lambda_1 + \alpha_1) - \pi_2(\xi)\alpha_2(1 - q)[1 - (\lambda_1 + \alpha_1)(\tau - \xi)]\} e^{-(\lambda_1 + \alpha_1)(\tau - \xi)}, & \tau \geq \xi. \end{cases} \quad (15)$$

Найдем преобразование Лапласа плотности $p(\tau)$, используя (8) и (15). Имеем

$$g_\tau(s) = \int_0^\infty e^{-s\tau} p(\tau) d\tau =$$

$$= \int_0^\infty e^{-s\tau} \left[\int_0^\tau p(\xi) p(\tau - \xi|\xi) d\xi \right] d\tau = \int_0^\infty p(\xi) \left[\int_\xi^\infty e^{-s\tau} p(\tau - \xi|\xi) d\tau \right] d\xi =$$

$$= \int_0^\infty p(\xi) \left[\int_\xi^\infty e^{-s\tau} \{(\lambda_1 + \alpha_1) - \alpha_2(1 - q)[1 - (\lambda_1 + \alpha_1)(\tau - \xi)]\} \pi_2(\xi) e^{-(\lambda_1 + \alpha_1)(\tau - \xi)} d\tau \right] d\xi.$$

В выражении для $g_\tau(s)$ выполним замену переменных: $\tau - \xi = \eta$, $\tau = \xi + \eta$. Тогда

$$g_\tau(s) = \int_0^\infty p(\xi) \left[\int_0^\infty e^{-s(\xi + \eta)} \{(\lambda_1 + \alpha_1) - \alpha_2(1 - q)[1 - (\lambda_1 + \alpha_1)\eta]\} \pi_2(\xi) e^{-(\lambda_1 + \alpha_1)\eta} d\eta \right] d\xi.$$

Выполняя достаточно трудоемкие преобразования, приходим к (9). Теорема доказана.

Замечание 1. Для второго варианта задания параметров потока имеет место условие рекуррентности $\lambda_1\lambda_2 - \alpha_1\alpha_2p = 0$, с учетом которого все выкладки аналогичны предыдущему случаю и приводят к преобразованию Лапласа $g_\tau(s)$, определенному в (9). Третий вариант задания параметров потока не вызывает интереса, поскольку в указанном случае поток вырождается в простейший.

5. Оценка длительности мертвого времени

Преобразование Лапласа (9) позволяет получить начальные моменты

$$M(\tau^l) = (-1)^l g_\tau^{(l)}(s) \Big|_{s=0}, \quad l = 1, 2, \dots$$

Будем решать задачу оценивания длительности мертвого времени T методом моментов. Введем статистики $C_l = \frac{1}{n} \sum_{k=1}^n \tau_k^l$, $l = 1, 2, \dots$, где $\tau_k = t_{k+1} - t_k$ – значение длительности интервала между моментами t_k и t_{k+1} наступления событий в рекуррентном обобщенном асинхронном потоке с продлевающимся мертвым временем.

Предположим, что параметры потока $\alpha_1, \alpha_2, \lambda_1, \lambda_2, p, q$ являются известными. При количестве наблюдений $n \rightarrow \infty$ выборочный момент C_l стремится к теоретическому моменту $M(\tau)$ [19]. Тогда для оценки длительности мертвого времени T имеем уравнение моментов $M(\tau) = C_1$.

Вычислив производную от (9) по s в точке $s = 0$ и взяв ее со знаком минус, получим уравнение для нахождения значения \hat{T} оценки $\hat{\mathbf{T}}$:

$$\begin{aligned} & \frac{2}{\lambda_1 + \alpha_1} - \frac{1}{\alpha_1 + \alpha_2} + \frac{1}{(\lambda_1 + \alpha_1)^2 \varphi_0(T)} \left((\lambda_1 + \alpha_1) [1 - \varphi_0(T)] - (\lambda_1 - \alpha_2) [1 - e^{-(\lambda_1 + \alpha_1)T}] \right) + \\ & + \frac{(\lambda_1 - \alpha_2)^2 \varphi_0(T) e^{-(\alpha_1 + \alpha_2)T}}{(\lambda_1 + \alpha_1)^2 (\alpha_1 + \alpha_2)} \left(1 - \frac{(\lambda_1 + \alpha_1)^2 + (\alpha_1 + \alpha_2)(\lambda_1 + \lambda_2)}{(\lambda_1 + 2\alpha_1 + \alpha_2)^2} [1 - e^{-(\lambda_1 + 2\alpha_1 + \alpha_2)T}] - \right. \\ & \left. - \frac{(\lambda_1 + \alpha_1)(\lambda_1 - \alpha_2)}{\lambda_1 + 2\alpha_1 + \alpha_2} T e^{-(\lambda_1 + 2\alpha_1 + \alpha_2)T} \right)^{-1} = C_1, \end{aligned} \quad (16)$$

где $\varphi_0(T)$ определена в (4).

Обозначим левую часть уравнения (16) через $f(T)$. Можно показать, что функция $f(T)$, $T \geq 0$, является возрастающей. Уравнение $f(T) = C_1$ решается численно на интервале $0 \leq T \leq \tau_{\min}$, $\tau_{\min} = \min\{\tau_k\}$, $k = \overline{1, n}$. Возможные ситуации:

- 1) $f(T=0) \geq C_1$; в качестве значения оценки $\hat{\mathbf{T}}$ выбирается $\hat{T}^* = 0$;
- 2) $f(T=0) < C_1$, тогда возможны случаи: а) корень уравнения моментов (16) попадает в полуинтервал $(0, \tau_{\min}]$, тогда он и выбирается в качестве значения оценки $\hat{T} = \hat{T}^*$; б) корень уравнения (16) больше τ_{\min} , тогда в качестве значения оценки $\hat{\mathbf{T}}$ выбирается $\hat{T}^* = \tau_{\min}$.

6. Результаты статистических экспериментов

Для получения статистики C_1 построена имитационная модель, выходом которой является последовательность t_1, t_2, \dots моментов наступления событий.

Имитационная модель реализована на языке программирования C# с использованием технологии WPF и библиотеки Math.NET Numerics. Результат работы программы имитационного моделирования для длительности мертвого времени $T = 0,3$ и параметров потока $\lambda_1 = 2$, $\lambda_2 = 1$, $\alpha_1 = 4$, $\alpha_2 = 5$, $p = 1$, $q = 0,1$ приведен на рис. 2.

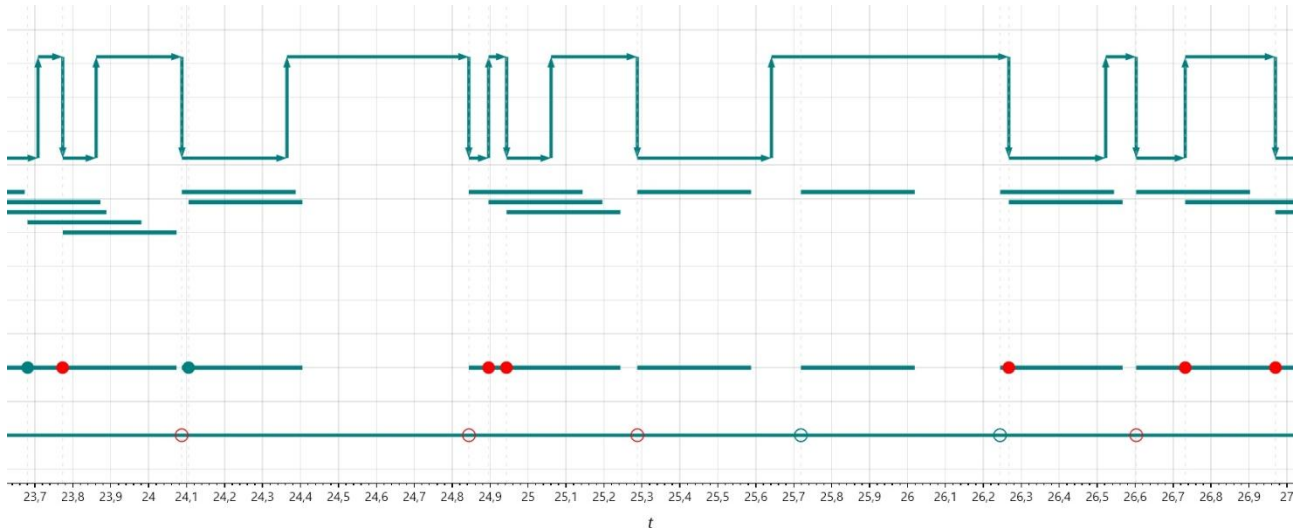


Рис. 2. Результат работы программы имитационного моделирования

Fig. 2. The result of the simulation program

В верхней части рис. 2 показана реализация сопровождающего процесса $\lambda(t)$; ниже короткие горизонтальные линии обозначают периоды мертвого времени фиксированной длительности T ; далее ниже показана реализация случайной величины – общего периода ненаблюдаемости, где круги зеле-

ного цвета обозначают ненаблюдаемые события пуассоновского потока, круги красного цвета обозначают ненаблюдаемые дополнительные события потока; наконец, на нижней временной оси отмечены наблюдаемые события – дополнительные события обозначены окружностями с красной границей, события пуассоновского потока – окружностями с границей зеленого цвета.

Для получения значения \hat{T} оценки \hat{T} длительности мертвого времени T разработан следующий алгоритм оценивания:

1) задаются параметры потока с учетом условия особого случая $\lambda_1 + \alpha_1 = \lambda_2 + \alpha_2$ и условий рекуррентности $\lambda_1 > \lambda_2, \lambda_1 \lambda_2 = p \alpha_1 \alpha_2, 0 < p < 1, q = 1$ или $\lambda_1 > \lambda_2, \lambda_1 \lambda_2 = q \alpha_1 \alpha_2, 0 < q < 1, p = 1$;

2) табулируется функция $f(T)$ и численно устанавливается, что $f(T)$ является возрастающей функцией;

3) в течение T_m ед. времени работает имитационная модель потока для получения выборки $\tau_1, \tau_2, \dots, \tau_n$, находится $\tau_{\min} = \min\{\tau_k\}$, вычисляется статистика C_1 ;

4) численно решается уравнение (15) методом Ньютона; корень уравнения моментов (16) является единственным;

5) в результате шагов 1–4 вычисляется \hat{T}_1^* – значение оценки \hat{T} в первом опыте;

6) шаги 2–4 повторяются для N опытов и находятся значения оценок $\hat{T}_1^*, \hat{T}_2^*, \dots, \hat{T}_N^*$.

Выборочное среднее оценки и выборочная вариация оценки вычисляются по формулам $\hat{M}(\hat{T}) = \frac{1}{N} \sum_{j=1}^N \hat{T}_j^*, \hat{V}(\hat{T}) = \frac{1}{N} \sum_{j=1}^N (\hat{T}_j^* - T)^2$, где T – значение длительности мертвого времени, известное из имитационной модели потока.

Выборочное среднее количество событий исходного потока и выборочное среднее количество событий наблюдаемого потока вычисляются по формулам $\bar{n}_{исх} = \frac{1}{N} \sum_{j=1}^N n_{исх}^{(j)}, \bar{n}_{набл} = \frac{1}{N} \sum_{j=1}^N n_{набл}^{(j)}$, где $n_{исх}^{(j)}$ – количество событий исходного потока в j -м опыте, $n_{набл}^{(j)}$ – количество событий наблюдаемого потока в j -м опыте.

На рис. 3 представлен график функции $f(T)$ при значениях параметров $\lambda_1 = 2, \lambda_2 = 1, \alpha_1 = 4, \alpha_2 = 5, p = 1, q = 0,1$. Статистика C_1 является результатом работы программы имитационного моделирования для $T = 0,5$ и времени моделирования $T_m = 2\,000$ ед. времени.

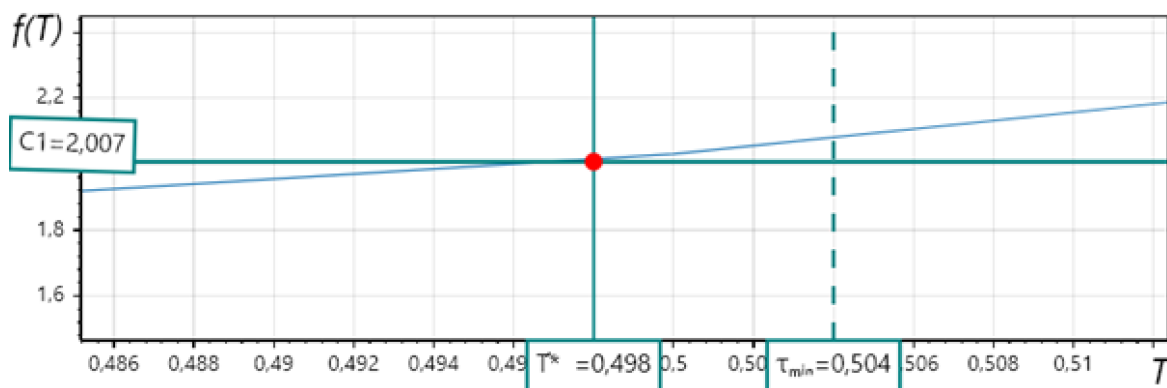


Рис. 3. График функции $f(T)$
Fig. 3. Graph of the function $f(T)$

Пересечение графика функции $f(T)$ и статистики C_1 отмечено на рис. 3 красной точкой. Абсцисса точки пересечения есть значение оценки \hat{T}^* – решение уравнения моментов (16) в одном опыте. Возрастание функции $f(T)$ в области определения $0 \leq T \leq 1$ свидетельствует о единственности решения уравнения моментов для заданного набора параметров.

Для установления качества полученной оценки выполнен эксперимент, который заключается в том, что при одних и тех же параметрах потока $\lambda_1 = 2$, $\lambda_2 = 1$, $\alpha_1 = 4$, $\alpha_2 = 5$, $p = 1$, $q = 0,1$ меняется значение длительности мертвого времени T от $T = 0,1$ до $T = 1$ ед. времени; количество опытов полагается равным $N = 1\,000$ и время моделирования $T_m = 2\,000$ ед. времени.

Численные результаты работы алгоритма оценивания приведены в таблице. В первой строке задается значение длительности мертвого времени T ; во второй и третьей строках – выборочное среднее оценки $\hat{M}(\hat{T})$ и выборочная вариация $\hat{V}(\hat{T})$ соответственно; в четвертой и пятой строках – выборочное среднее количество событий исходного потока $\bar{n}_{исх}$ и выборочное среднее количество событий наблюдаемого потока $\bar{n}_{набл}$; в шестой строке – процент наблюдаемых событий $p = (\bar{n}_{набл} / \bar{n}_{исх}) \cdot 100\%$.

Зависимость $\hat{M}(\hat{T}), \hat{V}(\hat{T}), \bar{n}_{исх}, \bar{n}_{набл}, p$ от значения длительности мертвого времени T

T	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1
$\hat{M}(\hat{T})$	0,0986	0,1970	0,2963	0,3967	0,4970	0,5975	0,6928	0,7985	0,8933	0,8668
$\hat{V}(\hat{T})$	$1,0 \cdot 10^{-6}$	$2,0 \cdot 10^{-6}$	$3,0 \cdot 10^{-6}$	$5,0 \cdot 10^{-6}$	$9,0 \cdot 10^{-6}$	$1,4 \cdot 10^{-5}$	$3,0 \cdot 10^{-5}$	$5,6 \cdot 10^{-5}$	$2,7 \cdot 10^{-3}$	$6,8 \cdot 10^{-2}$
$\bar{n}_{исх}$	16 002	16 010	16 004	16 002	16 005	16 001	16 002	16 000	15 997	15 999
$\bar{n}_{набл}$	11 415	7 710	5 024	3 193	1 992	1 224	744	448	268	160
p	71,33	48,16	31,39	19,95	12,45	7,65	4,65	2,80	1,68	1,00

Анализ результатов оценивания, представленных в таблице, показывает, что с ростом значения длительности мертвого времени число событий в наблюдаемом потоке уменьшается, выборочная вариация $\hat{V}(\hat{T})$ растет, что является естественным.

Замечание 2. Поскольку: а) длительности интервала между соседними событиями наблюдаемого потока $\tau_1, \tau_2, \dots, \tau_k$ являются независимыми и одинаково распределенными случайными величинами; б) теоретический момент $M(\tau)$ существует; в) численно показано, что уравнение моментов (16) имеет единственное решение, оценки \hat{T} являются состоятельными [19].

Заключение

В данной работе рассмотрен дважды стохастический обобщенный асинхронный поток событий с двумя состояниями, функционирующий в стационарном режиме в условиях продлевающегося мертвого времени, в особом случае соотношения параметров потока. Полученные результаты показывают возможность оценивания длительности продлевающегося мертвого времени, выступающего искажающим фактором, по результатам текущих наблюдений за потоком событий. Предложенный подход, основанный на применении преобразования Лапласа и метода моментов, позволяет получить приемлемые в смысле выборочной вариации оценки длительности мертвого времени при достаточно больших выборках наблюдений.

Список источников

1. Вишневский В.М., Дудин А.Н., Клименко В.И. Стохастические системы с коррелированными потоками. Теория и применение в телекоммуникационных сетях. М.: Техносфера, 2018. 564 с.
2. Cox D.R. The analysis of non-Markovian stochastic processes by the inclusion of supplementary variables // Proceedings of the Cambridge Philosophical Society. 1955. V. 51 (3). P. 433–441.
3. Kingman Y.F.C. On doubly stochastic Poisson process // Proceedings of the Cambridge Philosophical Society. 1964. V. 60 (4). P. 923–930.
4. Башарин Г.П., Кокотушкин В.А., Наумов В.А. О методе эквивалентных замен расчета фрагментов сетей связи. Ч. 1 // Известия Академии наук СССР. Техническая кибернетика. 1979. № 6. С. 92–99.
5. Башарин Г.П., Кокотушкин В.А., Наумов В.А. О методе эквивалентных замен расчета фрагментов сетей связи. Ч. 2 // Известия Академии наук СССР. Техническая кибернетика. 1980. № 1. С. 55–61.
6. Neuts M.F. A versatile Markov point process // Journal of Applied Probability. 1979. V. 16. P. 764–779.

7. Lucantoni D.M. New results on the single server queue with a bath markovian arrival process // *Communication in Statistics Stochastic Models*. 1991. V. 7. P. 1–46.
8. Nezhelskaya L., Tumashkina D. Optimal state estimation of semi-synchronous event flow of the second order under its complete observability // *Communications in Computer and Information Science*. 2018. V. 912. P. 93–105.
9. Горцев А.М., Нежелская Л.А. Оценивание параметров асинхронного потока с иницированием лишних событий методом моментов // *Вестник Томского государственного университета*. 2006. № S18. С. 267–273.
10. Апанасович В.В., Коляда А.А., Чернявский А.Ф. Статистический анализ случайных потоков в физическом эксперименте. Минск : Университетское, 1988. 256 с.
11. Василевская Т.П., Горцев А.М., Нежелская Л.А. Оценивание длительности мертвого времени и параметров синхронного альтернирующего потока с проявлением либо не проявлением событий // *Вестник Томского государственного университета*. 2004. № S9-2. С. 129–138.
12. Калягин А.А., Нежелская Л.А. Сравнение МП- и ММ-оценок длительности мертвого времени в обобщенном полусинхронном потоке событий // *Вестник Томского государственного университета. Управление, вычислительная техника и информатика*. 2015. № 3 (32). С. 23–32.
13. Горцев А.М., Ниссенбаум О.В. Оценивание длительности мертвого времени и параметров асинхронного альтернирующего потока событий с иницированием лишнего события // *Вестник Томского государственного университета*. 2004. № 284. С. 137–145.
14. Горцев А.М., Веткина А.В. Оценивание методом моментов параметра равномерного распределения длительности непродлевающегося случайного мертвого времени в рекуррентном полусинхронном потоке событий в общем и особом случаях // *Вестник Томского государственного университета. Управление, вычислительная техника и информатика*. 2022. № 61. С. 47–60.
15. Нежелская Л.А., Пономаренко В.Д. Обобщенный асинхронный поток событий с продлевающимся мертвым временем // *Вестник Томского государственного университета. Управление, вычислительная техника и информатика*. 2024. № 69. С. 82–94.
16. Нежелская Л.А. Оценка состояний дважды стохастических потоков событий : учеб. пособие. Томск : Изд-во Том. гос. ун-та, 2020. 210 с.
17. Нежелская Л.А. Оценка состояний и параметров дважды стохастических потоков событий : дис. ... д-ра физ.-мат. наук. Томск, 2016. 341 с.
18. Хинчин А.Я. Работы по математической теории массового обслуживания. М. : Физматгиз, 1963. 236 с.
19. Малинковский Ю.В. Теория вероятностей и математическая статистика. Гомель : ГГУ им. Ф. Скорины, 2004. Ч. 2: Математическая статистика. 146 с.

References

1. Vishnevsky, V.M., Dudin, A.N. & Klimenok, V.I. (2018) *Stokhasticheskie sistemy s korrelirovannymi potokami. Teoriya i primeneniye v telekommunikatsionnykh setyakh* [Stochastic Systems with Correlated Flows. Theory and Application in Telecommunication Networks]. Moscow: Tekhnosfera.
2. Cox, D.P. (1955) The analysis of non-Markovian stochastic process by the inclusion of supplementary variables. *Proceeding of the Cambridge Philosophical Society*. 51(3). pp. 433–441.
3. Kingman, Y.F.C. (1964) On double stochastic Poisson process. *Proceeding of the Cambridge Philosophical Society*. 60(4). pp. 923–930.
4. Basharin, G.P., Kokotushkin, V.A. & Naumov, V.A. (1979) O metode ekvivalentnykh zamen rascheta fragmentov setey svyazi. Ch. 1 [On the equivalent substitutions method for computing fragments of communication networks. Part 1]. *Izvestiya AN SSSR. Tekhnicheskaya kibernetika*. 6. pp. 92–99.
5. Basharin, G.P., Kokotushkin, V.A. & Naumov, V.A. (1980) O metode ekvivalentnykh zamen rascheta fragmentov setey svyazi. [On the equivalent substitution method for computing fragments of communication networks]. *Izvestiya AN SSSR. Tekhnicheskaya Kibernetika*. 1. pp. 55–61.
6. Neuts, M.F. (1979) A versatile Markov point process. *Journal of Applied Probability*. 16. pp. 764–779.
7. Lucantoni, D.M. (1991) New results on the single server queue with a bath Markovian arrival process. *Communication in Statistics Stochastic Models*. 7. pp. 1–46.
8. Nezhelskaya, L. & Tumashkina, D. (2018). Optimal state estimation of semi-synchronous event flow of the second order under its complete observability. *Communications in Computer and Information Science*. 912. pp. 93–105.
9. Gortsev, A.M. & Nezhelskaya, L.A. (2006) Otsenivanie parametrov asinkhronnogo potoka s initsirovaniem lishnikh sobyitiy metodom momentov [Estimation of parameters of an asynchronous flow with the initiation of extra events by the method of moments]. *Vestnik Tomskogo gosudarstvennogo universiteta – Tomsk State University Journal*. S18. pp. 267–273.
10. Apanasovich, V.V., Kolyada, A.A. & Chernyavsky, A.F. (1988) *Statisticheskii analiz sluchaynykh potokov v fizicheskoy eksperimente* [Statistical Analysis of Random Flows in Physical Experiment]. Minsk: Universitetskoye.
11. Vasilevskaya, T.P., Gortsev, A.M. & Nezhelskaya, L.A. (2004) Otsenivanie dlitel'nosti mertvogo vremeni i parametrov sinkhronnogo al'terniruyushchego potoka s proyavleniem libo neproyavleniem sobyitiy [Estimation of dead time duration and parameters of synchronous alternating flow with manifestation or non-manifestation of events]. *Vestnik Tomskogo gosudarstvennogo universiteta*.

- Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science.* S9–2. pp. 129–138.
12. Kalyagin, A.A. & Nezhelskaya, L.A. (2015) The comparison of maximum likelihood estimation and method of moments estimation of dead time value in a generalized semysynchronous flow of events. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science.* 3(32). pp. 23–32.
 13. Gortsev, A.M. & Nissenbaum, O.V. (2004) Otsenivanie dlitel'nosti mertvogo vremeni i parametrov asinkhronnogo al'terniruyushchego potoka sobytiy s initsirovaniem lishnego sobytiya [Estimation of the dead time duration and parameters of the asynchronous alternating stream of events with initialization of an extra event]. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science.* 284. pp. 137–145.
 14. Gortsev, A.M. & Vetkina, A.V. (2022) Estimation by the method of moments of the parameter of the uniform distribution of the duration of non-transient random dead time in the recurrent semi-synchronous flow of events in general and special cases. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science.* 61. pp. 47–60.
 15. Nezhelskaya L.A. & Ponomarenko V.D. (2024) Generalized asynchronous flow of events with extending dead time. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science.* 69. pp. 82–94.
 16. Nezhelskaya, L.A. (2020) *Otsenka sostoyaniy dvazhdy stokhasticheskikh potokov sobytiy* [Estimation of States of Doubly Stochastic Flows of Events]. Tomsk: Tomsk State University Press.
 17. Nezhelskaya, L.A. (2016) *Otsenka sostoyaniy i parametrov dvazhdy stokhasticheskikh potokov sobytiy* [Evaluation of states and parameters of doubly stochastic flows of events]. Physics and Mathematics Dr. Diss. Tomsk.
 18. Khinchin, A.Ya. (1963) *Raboty po matematicheskoy teorii massovogo obsluzhivaniya* [Works on the Mathematical Theory of Queuing]. Moscow: Fizmatgiz.
 19. Malinkovsky, Y.V. (2004) *Teoriya veroyatnostey i matematicheskaya statistika* [Theory of Probabilities and Mathematical Statics]. Vol. 2. Gomel: GSU.

Информация об авторах:

Нежелская Людмила Алексеевна – профессор, доктор физико-математических наук, заведующая кафедрой прикладной математики Института прикладной математики и компьютерных наук Национального исследовательского Томского государственного университета (Томск, Россия). E-mail: ludne@mail.ru

Пonomarenko Валентина Денисовна – магистрант кафедры прикладной математики Института прикладной математики и компьютерных наук Национального исследовательского Томского государственного университета (Томск, Россия). E-mail: valya.ponomarenko.00@mail.ru

Вклад авторов: все авторы сделали эквивалентный вклад в подготовку публикации. Авторы заявляют об отсутствии конфликта интересов.

Information about the authors:

Nezhel'skaya Lyudmila A. (Professor, Doctor of Physical and Mathematical Sciences, Institute of Applied Mathematics and Computer Science, National Research Tomsk State University, Tomsk, Russia Federation). E-mail: ludne@mail.ru

Ponomarenko Valentina D. (Master's Student, Department of Applied Mathematics, Institute of Applied Mathematics and Computer Science, National Research Tomsk State University, Tomsk, Russia Federation). E-mail: valya.ponomarenko.00@mail.ru

Contribution of the authors: the authors contributed equally to this article. The authors declare no conflicts of interests.

Поступила в редакцию 14.09.2025; принята к публикации 02.12.2025

Received 14.09.2025; accepted for publication 02.12.2025

ИНФОРМАТИКА И ПРОГРАММИРОВАНИЕ

INFORMATICS AND PROGRAMMING

Научная статья

УДК 004.048

doi: 10.17223/19988605/73/14

**Нейронные сети в диагностике злокачественных новообразований
по выдыхаемому воздуху**

**Артём Викторович Обходский¹, Денис Евгеньевич Кульбакин²,
Елена Владимировна Обходская³, Владислав Сергеевич Лаконкин⁴,
Евгений Олегович Родионов⁵, Виктор Иванович Сачков⁶,
Владимир Иванович Чернов⁷, Евгений Лхамацыренович Чойнзонов⁸**

^{1, 4} Национальный исследовательский Томский политехнический университет, Томск, Россия

^{2, 4, 5, 7, 8} Научно-исследовательский институт онкологии Томского национального исследовательского
медицинского центра Российской академии наук, Томск, Россия

^{3, 6} Национальный исследовательский Томский государственный университет, Томск, Россия

⁵ Сибирский государственный медицинский университет, Томск, Россия

¹ art707@yandex.ru

² kulbakin_d@mail.ru

³ fil330a@yandex.ru

⁴ vs113@tpu.ru

⁵ rodionov_eo@oncology.tomsk.ru

⁶ vicsachkov@gmail.com

⁷ chernov@tnimc.ru

⁸ choynzonov@tnimc.ru

Аннотация. Рассматриваются результаты исследования набора архитектур нейронных сетей с изменяемыми параметрами обучения для обработки данных с газоаналитических медицинских приборов, предназначенных для неинвазивной диагностики рака легких и верхних дыхательных путей. Алгоритм обеспечивает структурирование формата входных паттернов для нейронной сети с учетом обеспечения критерия максимума информации во входных данных. Диагностические данные с газоаналитических медицинских приборов представляют собой массивы целочисленных значений кодов с аналого-цифровых преобразователей. Алгоритм нейросетевой обработки данных реализован на языке программирования Python. В исследовании использовались оцифрованные пробы выдыхаемого воздуха от 154 человек. Для случаев отдельной дифференциации здоровых добровольцев, пациентов с раком легких и верхних дыхательных путей алгоритм нейросетевой обработки данных показал точность, в среднем превышающую 87%.

Ключевые слова: обработка данных; классификатор; искусственная нейронная сеть; архитектура нейронной сети; формат входных данных; оптимизация входных данных; параметры обучения; признак дифференцирования; эффективность классификатора.

Благодарности: Исследование выполнено при финансовой поддержке гранта РНФ в рамках научного проекта № 23-15-00177 «Теоретическое и клиническое обоснование молекулярной оценки состава выдыхаемого воздуха для диагностики онкологических заболеваний», <https://rscf.ru/project/23-15-00177/>

Для цитирования: Обходский А.В., Кульбакин Д.Е., Обходская Е.В., Лаконкин В.С., Родионов Е.О., Сачков В.И., Чернов В.И., Чойнзонов Е.Л. Нейронные сети в диагностике злокачественных новообразований по выдыхаемому воздуху // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2025. № 73. С. 122–133. doi: 10.17223/19988605/73/14

Neural networks in the diagnosis of malignant neoplasms by exhaled air

Artem V. Obkhodskiy¹, Denis E. Kulbakin², Elena V. Obkhodskaya³, Vladislav S. Lakonkin⁴,
Evgeniy O. Rodionov⁵, Victor I. Sachkov⁶, Vladimir I. Chernov⁷, Evgeny L. Choyazonov⁸

^{1, 4} National Research Tomsk Polytechnic University, Tomsk, Russian Federation

^{2, 4, 5, 7, 8} Cancer Research Institute, Tomsk National Research Medical Center
of the Russian Academy of Sciences, Tomsk, Russian Federation

^{3, 6} National Research Tomsk State University, Tomsk, Russian Federation

⁵ Siberian State Medical University, Tomsk, Russian Federation

¹ art707@yandex.ru

² kulbakin_d@mail.ru

³ fil330a@yandex.ru

⁴ vsl13@tpu.ru

⁵ rodionov_eo@oncology.tomsk.ru

⁶ vicsachkov@gmail.com

⁷ chernov@tnimc.ru

⁸ choyazonov@tnimc.ru

Abstract. The results of a study the set of neural network architectures with variable learning parameters for processing data from gas-analytical medical devices designed for noninvasive diagnosis of lung and upper respiratory tract cancer are considered. The algorithm provides structuring of the input pattern format for the neural network, taking into account the criterion of maximum information in the input data. Diagnostic data from gas-analytical medical devices are arrays of integer values of codes from analog-to-digital converters. The neural network data processing algorithm is implemented in the Python programming language. The study used digitized exhaled air samples from 154 people. For cases of separate differentiation of healthy volunteers, patients with lung and upper respiratory tract cancers, the neural network data processing algorithm showed an average accuracy exceeding 86%.

Keywords: data processing; classifier; artificial neural network; neural network architecture; input data format; input data optimization; learning parameters; differentiation feature; classifier efficiency.

Acknowledgments: The study was supported by the Russian Science Foundation grant № 23-15-00177 “Theoretical and clinical substantiation of the molecular assessment of the composition of exhaled air for the diagnosis of oncological diseases”, <https://rscf.ru/en/project/23-15-00177>

For citation: Obkhodskiy, A.V., Kulbakin, D.E., Obkhodskaya, E.V., Lakonkin, V.S., Rodionov, E.O., Sachkov, V.I., Chernov, V.I., Choyazonov, E.L. (2025) Neural networks in the diagnosis of malignant neoplasms by exhaled air. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 73. pp. 122–133. doi: 10.17223/19988605/73/14

Введение

Интенсивное внедрение технологий искусственного интеллекта и нейронных сетей в медицинскую практику, в том числе в скрининг, позволяет обеспечить своевременный доступ людей к медицинским услугам в условиях значительного увеличения объема биомедицинских данных, необходимых для постановки диагноза, а также для назначения эффективного лечения с учетом индивидуальных особенностей пациентов. Автоматизация процессов диагностики и назначения терапии с применением программного обеспечения на алгоритмах искусственного интеллекта становится все более востребованной при отсутствии возможности личного посещения медицинских учреждений, пиковом увеличении количества обследуемых людей и интенсивном развитии технологий персонализированной медицины [1].

Повышение эффективности исследований и развитие инструментов анализа медицинских данных предполагает разные направления для применения технологий искусственного интеллекта. В работе [2] применялись нейросетевые алгоритмы многоклассовой сегментации изображений и метрики оценки качества mIoU и mDis для решения задачи различения нечетких локальных визуальных

признаков в биологических тканях во время сегментации изображений лапароскопической хирургии. Были достигнуты сравнительно высокие показатели эффективности нейросетевого алгоритма в среднем на уровне mIoU и mDis 78,19 и 83,84% соответственно. Данный подход позволяет автоматизировать процесс сегментации изображений на практике.

Алгоритмы искусственного интеллекта применяются при разработке лекарств, в том числе при выявлении и оптимизации противовоспалительных агентов для терапии рака. В работе [3] были выделены основные провоспалительные механизмы, участвующие в прогрессировании опухоли, такие как NF- κ B, STAT3, COX-2 и ось JAK/IL-6, и продемонстрировано, как инструменты искусственного интеллекта, включая глубокое обучение и языковые модели, могут ускорить разработку и проверку терапевтических кандидатов. Появление индивидуализированной терапии рака, основанной на глубоком анализе микроокружения опухоли, тоже представляет собой революционную трансформацию в онкологии [4, 5].

В последние годы алгоритмы искусственного интеллекта все чаще входят в состав технологических инструментов цифровой ортопедии, в том числе в области виртуализированной дополненной реальности, 3D-печати и хирургических роботов [5, 6]. В медицинской практике становится частым применение дистанционной роботизированной хирургии, что позволяет решать проблему неравномерного распределения медицинских специалистов высшей категории на местах.

Визуализация в медицине является одним из основных инструментов для диагностики и лечения разных заболеваний. Интеграция технологий машинного зрения с искусственным интеллектом применяется в гепатологии [7], диагностике и лечении кожных заболеваний [8], онкологии [9, 10] и др.

Современные большие языковые модели (LLM) искусственного интеллекта в медицине обещают решать задачи от поддержки принятия клинических решений до обучения пациентов [11]. Достижения LLM в виде агентов еще больше расширяют их полезность, обеспечивая мультимодальную обработку и многозадачную обработку в сложных диагностических и клинических процессах [12]. Большие языковые модели, такие как GPT-4o, DeepSeek-R1, Gemini 2.0, Command-R, Claude 3, Qwen и Grok 3, демонстрируют различные частные характеристики, но в целом общую возможность в формировании сложной логики. Они быстро расширяются, ускоряя переход медицинского сектора от парадигмы Интернет+ к парадигме ИИ+ [13, 14].

Современные технологии искусственного интеллекта дают возможность для освоения и применения новых принципов работы медицинских изделий, которые ранее не могли быть оценены как эффективные в силу того, что обработка данных в этих медицинских приборах выполнялась оператором, а программная реализация алгоритмов обработки данных представлялась нерешаемой задачей [15]. Учитывая этот факт, технологии выявления летучих органических соединений в выдыхаемом воздухе все чаще рассматриваются в качестве перспективного инструмента для ранней диагностики рака. Это обусловлено неинвазивностью метода и возможностью проведения обследования человека за короткое время, что особенно важно при скрининге. Благодаря низкой растворимости в крови, летучие органические соединения легко проникают в альвеолярный воздух и выводятся из организма с выдыхаемым воздухом, что делает их доступными для анализа и диагностики заболеваний.

Общим результатом рассмотренных выше работ является повышение эффективности процессов за счет внедрения автоматизированных алгоритмов интеллектуальной обработки данных. В нашей работе представлены результаты создания базовой нейронной сети и алгоритма обработки данных, предназначенных для нейросетевой классификации данных с газоаналитических медицинских приборов диагностики рака легких и верхних дыхательных путей. Применение технологий искусственного интеллекта напрямую может влиять на увеличение средней продолжительности и улучшение качества жизни для большого числа людей, поскольку в случаях отдельных заболеваний раннее выявление практически определяет исход выживаемости человека в борьбе с заболеванием или способствует снижению уровня инвалидности.

1. Данные с газоаналитического медицинского прибора

В исследовании измерение концентрации компонентов выдыхаемого воздуха проводится с помощью газоаналитических медицинских приборов. Во всех приборах реализуется единый порядок взятия

проб выдыхаемого воздуха у всех участников исследования в контролируемых нормальных внешних условиях. Во время работы прибора полупроводниковые неселективные датчики функционируют в режиме термоциклирования, т.е. периодически нагреваются и охлаждаются. В процессе сбора данных регистрируется 10 полных циклов – от начальной точки нагрева до завершающей стадии остывания процесса термоциклирования [15, 16]. Всего в приборе применяется 24 датчика, а выходной сигнал с каждого датчика преобразуется в цифровой десятичный код в диапазоне от 0 до 1 023 с помощью аналого-цифрового преобразователя (АЦП). Пример формы сигнала с датчика 12 для групп пациентов с раком легких и здоровых добровольцев представлен на рис. 1.

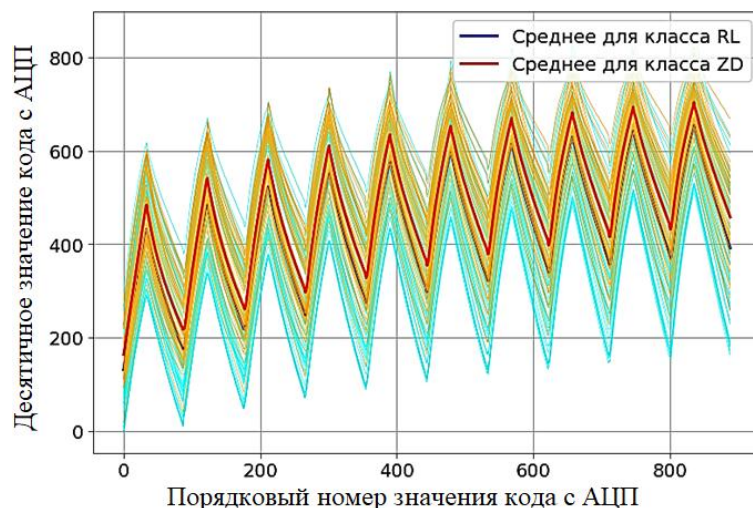


Рис. 1. Форма сигналов с датчика 12 для групп пациентов с раком легких (класс RL) и здоровых добровольцев (класс ZD)

Fig. 1. Waveform from sensor 12 for groups of lung cancer patients (class RL) and healthy volunteers (class ZD)

В качестве критерия для возможной последующей качественной дифференциации двух классов с помощью нейронной сети на начальном этапе исследования бралось во внимание превышение более чем на 1% относительного отклонения средних значений этих классов. Такое отличие демонстрирует, например, датчик 12 на последнем импульсе термоциклирования. Средние значения сигналов для двух классов датчика 12 представлены на рис. 2. На последнем – десятом – импульсе сигналы с датчиков выходят на уставку, и возникает статическое различие между классами. Для различных патологий наборы датчиков, сигналы которых имеют большие отклонения средних значений, отличаются. В частности, для выборки здоровых пациентов и больных с раком головы и шеи разница средних значений для датчика № 12 меньше в 2–3 раза, а значит, извлекаемые паттерны будет сложнее различить нейронной сетью по этому датчику.

Выходные XML-файлы с газоаналитических медицинских приборов содержат избыточное количество данных в виде кодов АЦП, которые нецелесообразно в полном объеме применять в нейросетевой обработке. В связи с этим они преобразуются в один входной для нейронной сети txt-файл, в котором содержится информация о количестве объектов обучающей выборки, размеры входного и выходного слоев нейронной сети, массивы данных обучающей выборки, содержащие значения кодов АЦП последнего импульса термоциклирования. При чтении массива данных десятого импульса термоциклирования из XML-файлов применяется прореживание данных, что позволяет уменьшить размер входного слоя в пять раз без значительной потери качества нейросетевого классификатора. Каждый из 18 подряд выстроенных паттернов (признаков, значений) во входном массиве данных нейронной сети принадлежит 24 датчикам. Таким образом, размер входного слоя составляет 432, размер выходного слоя – 2. В выходном слое возможны два случая: [1, 0] – здоровый доброволец (состояние 1, класс ZD); [0, 1] – пациент с раком легких (состояние 2, класс RL). На рис. 3 представлена сигнатура по 24 датчикам для одного паттерна (признака, значения, точки) в пике десятого импульса термоциклирования. Данная сигнатура обладает такой специфичностью, что для человека задача классификации по ней представляется невозможной, но не для нейронной сети.

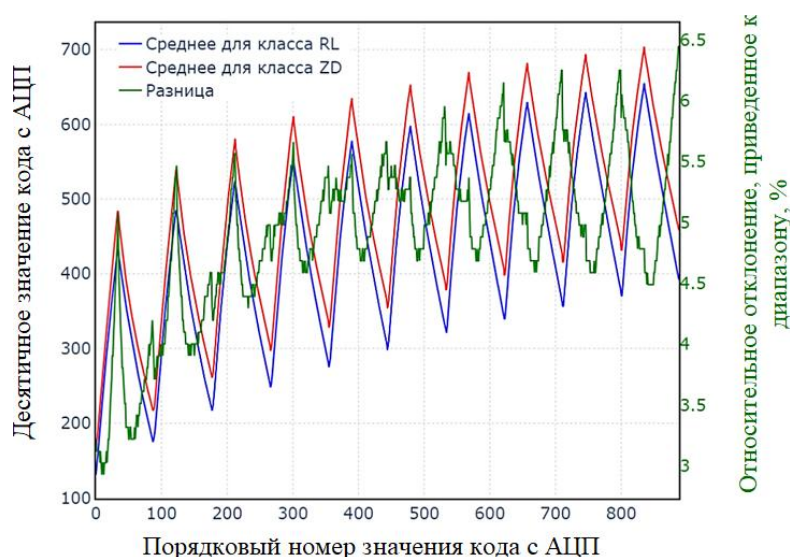


Рис. 2. Средние значения сигналов с датчика 12 для групп пациентов с раком легких (класс RL) и здоровых добровольцев (класс ZD) и относительное отклонение между ними
 Fig. 2. The average values of the signals from sensor 12 for groups of patients with lung cancer (class RL) and healthy volunteers (class ZD) and the relative deviation between them

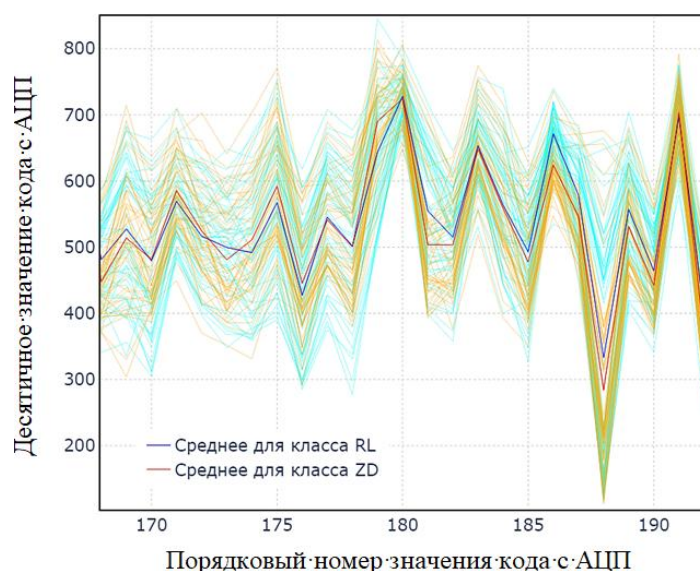


Рис. 3. Паттерн для групп пациентов с раком легких (класс RL) и здоровых добровольцев (класс ZD) по 24 датчикам для восьмой (пиковой) точки десятого импульса термоциклирования
 Fig. 3. A pattern for groups of lung cancer patients (class RL) and healthy volunteers (class ZD) with 24 sensors for the eighth (peak) point of the tenth thermal cycling pulse

Не менее важным вопросом является изучение влияния сущности паттернов на точность нейросетевого классификатора. С учетом определенной далее в работе архитектуры и гиперпараметров нейронной сети исследования показали, что использование паттернов в виде разницы последней волны и первой или отношение последней волны на первую не вносят особого вклада как с точки зрения математики, так и с точки зрения физики газовых процессов. Основной вклад вносится последней волной, когда происходит выход на уставку.

2. Архитектура нейронной сети

Задача поиска оптимальной архитектуры нейронной сети для обработки данных является одной из ключевых в области машинного обучения и искусственного интеллекта. Это связано с тем, что

архитектура нейронной сети напрямую влияет на ее способность решать поставленные задачи: качество работы модели, ее производительность, вычислительную эффективность и возможность масштабирования [17].

Исходя из последующего применения в предварительной диагностике заболеваний, наиболее подходящими кандидатами для дифференциации здоровых добровольцев и пациентов с патологией являются нейронная сеть прямого распространения типа многослойный персептрон (полносвязная нейронная сеть) и сверточная нейронная сеть. В отличие от полносвязной нейронной сети сверточная выявляет не столько вклад каждого отдельного паттерна в результат, сколько вклад совокупностей этих паттернов. На рис. 4 приведена оптимальная архитектура нейронной сети для классификации имеющегося набора экспериментальных данных с газоаналитических медицинских приборов.

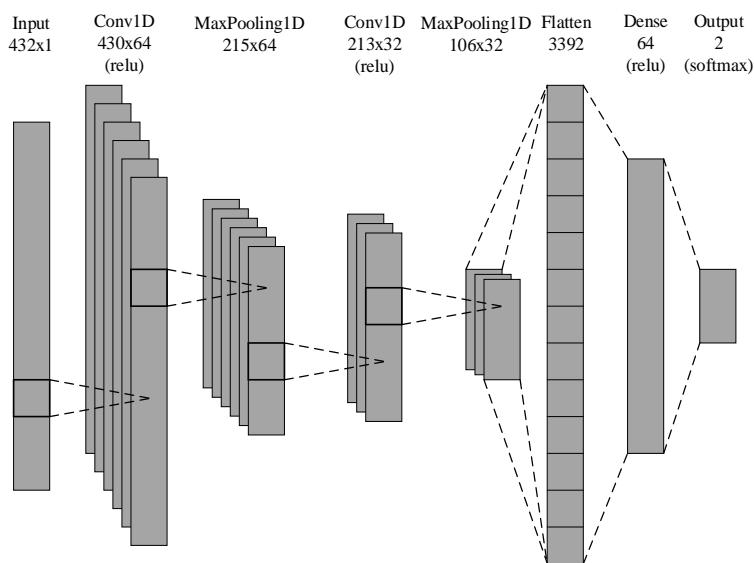


Рис. 4. Архитектура нейронной сети для диагностики рака легких и верхних дыхательных путей по выдыхаемому воздуху
Fig. 4. Neural network architecture for lung and upper respiratory tract cancer diagnosis by exhaled air

Для поиска оптимальной архитектуры нейронной сети использовались такие стратегии, как случайный поиск [16] и байесовская оптимизация. Суть обеих методик заключается в автоматизации процесса подбора параметров нейронной сети. При случайном поиске параметры подбирались случайным образом, а при байесовской оптимизации использовалась вероятностная модель для представления неизвестной целевой функции (функции потерь). Приоритет в работе был отдан байесовской оптимизации.

При использовании этой стратегии определялось количество и содержание скрытых слоев, в каждом слое задавались диапазон изменения количества нейронов и шаг, с которым это количество менялось. Задавались варианты функций активации, например linear, tanh, sigmoid или relu. Для компиляции модели выбирался оптимизатор, например Adam или SGD, задавался диапазон изменения скорости обучения, выбиралась категория потерь. Для решаемой задачи это всегда categorical_crossentropy. Определялась метрика, например accuracy, precision, recall или f1-score. В нашей работе оценка качества нейронной сети проводилась по метрике accuracy. Далее определялись количество комбинаций гиперпараметров для проверки, количество запусков для каждой комбинации, количество эпох и размер партии. В результате реализации байесовской оптимизации получена архитектура нейронной сети с параметрами, обеспечивающими лучшие показатели по выбранной метрике, точность на первой перекрестной группе для рака легких составила 90%. Во всех скрытых слоях нейронной сети использовалась функция активации relu. Для используемых наборов данных нейронные сети с тремя скрытыми слоями показали себя эффективными: обучение проходит относительно быстро, а качество получается сравнительно высоким. В дальнейшем в нейросетевом классификаторе предполагается использование трехслойной архитектуры сверточной нейронной сети.

Нейронная сеть в работе обучалась методом обратного распространения сигнала ошибки с учетом сверточной архитектуры. Завершение процесса обучения нейронной сети происходило по достижении

минимума ошибки, получаемого на верифицируемом множестве данных образцов выдыхаемого воздуха. Верифицирующее множество наборов данных выделялось из исходного набора данных проб выдыхаемого воздуха, из которого также отбирались пробы для обучения без пересечения обучающих и верифицирующих наборов.

Полная размерность входного слоя нейронной сети соответствует количеству значений с аналого-цифровых преобразователей по 24 датчикам прореженного десятого импульса термоциклирования и составляет 432 значения. Выходной слой – два нейрона, принимающие значения в диапазоне от 0 до 1. Исходный порог разделения проб составляет 0,5, но в процессе обучения истинное значение порога уточняется с учетом разного количества проб в классах обучающего набора нейронной сети. Значение на втором нейроне, близкое к 1, выше порога разделения означает вероятностную оценку принадлежности пробы выдыхаемого воздуха к классу пациентов с раком легких или верхних дыхательных путей, ниже порога до 0 – к классу здоровых добровольцев.

При определении оптимальных параметров обучения нейронной сети на наборах данных, полученных с газоаналитических медицинских приборов, проводился подбор гиперпараметров: количества эпох обучения, размера партии и скорости обучения. В результате оптимальное значение скорости обучения `learning_rate` составило 0,001. Исследование зависимости эффективности классификатора от количества скрытых слоев показало, что чем их больше, тем точнее классификатор, однако увеличение количества скрытых слоев более трех не способствует значительному увеличению точности нейронной сети, но способствует существенному увеличению времени ее обучения. Размер партии влияет на скорость обучения нейронной сети и ее эффективность, этот гиперпараметр всегда должен быть как можно больше, но он ограничен ресурсами вычислительной машины. В нашей работе размер партии `batch_size` составил 128. Количество эпох обучения, как правило, выбирается с учетом необходимости научить нейронную сеть обобщать, а не запоминать. Исследования показали, что для нашего набора данных этот процесс происходит за 20 эпох.

5. Экспериментальная часть

Программная реализация алгоритма нейросетевой обработки данных осуществлялась в среде Jupyter Notebook на языке программирования Python под управлением операционной системы Linux Ubuntu. Для этого использовались встраиваемые библиотеки TensorFlow и Keras.

При запуске алгоритма выполняются предварительная обработка исходных данных, обучение нейронной сети и перекрестная проверка. После обработки исходных данных осуществляются их анализ и визуализация. По результатам обучения нейронной сети формируются зависимости потерь и точности от эпох, визуализация t-SNE и выполняется решение обратной задачи нейронной сети. По результатам перекрестной проверки строятся зависимости ROC и PR [18], вычисляются их показатель AUC [19] и пороги разделения проб, формируется диаграмма распределения проб и матрицы неточностей.

В первом эксперименте по нейросетевой обработке данных с газоаналитических медицинских приборов использовался набор проб от 47 здоровых добровольцев и 53 пациентов с раком легких. После обучения нейронной сети были сформированы метрики оценки ее эффективности, в том числе ROC-характеристика, и определено значение параметра AUC. График ROC-характеристики представлен на рис. 5. Точность предварительной диагностики с классификатором на основе сверточной нейронной сети составила в среднем 90% при средних показателях чувствительности 86,79% и специфичности 93,62%.

Порог разделения положительных и отрицательных проб определен путем анализа ROC-характеристики. Его значение составило 0,534, и оно учитывается при проведении перекрестной проверки и для определения показателей специфичности и чувствительности нейросетевого классификатора. Величина AUC-ROC составила 0,950, что указывает на высокое качество классификации проб выдыхаемого воздуха с помощью сверточной нейронной сети и имеющегося обучающего набора экспериментальных данных.

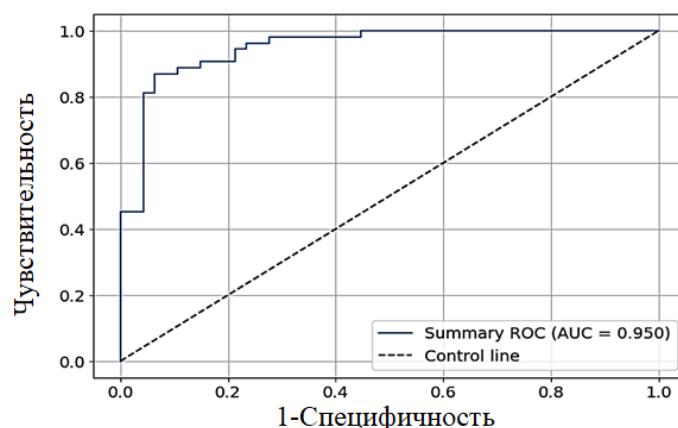


Рис. 5. ROC-характеристика качества обучения нейронной сети для классификации здоровых добровольцев и пациентов с раком легких

Fig. 5. ROC-characteristic of the neural network training quality for classifying healthy volunteers and patients with lung cancer

Во втором эксперименте были проведены аналогичные исследования с использованием набора данных, включающего 47 здоровых добровольцев и 53 пациента с раком верхних дыхательных путей, в том числе головы и шеи. Этот эксперимент позволил расширить исследование и оценить воспроизводимость полученных результатов в другой клинической группе. Анализ данных пациентов с раком верхних дыхательных путей проводился с использованием аналогичных методологических подходов, как и в первом эксперименте, что обеспечило сопоставимость результатов. Полученные результаты подтвердили статистическую значимость выявленных закономерностей и продемонстрировали потенциал применяемого подхода для диагностики онкологических заболеваний разных локализаций. Результаты формирования ROC-характеристики для второго эксперимента приведены на рис. 6.

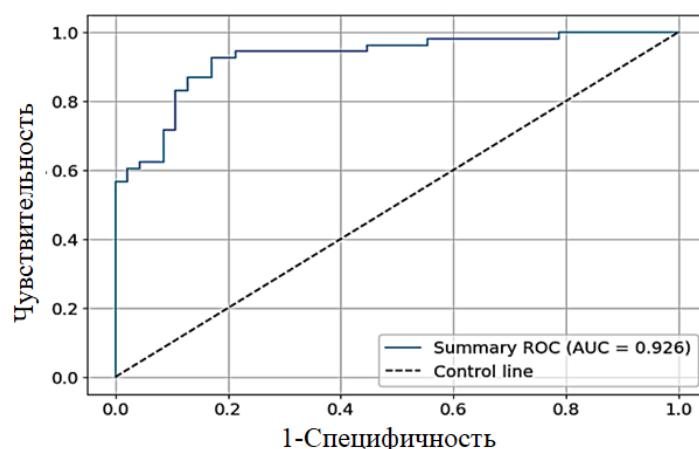


Рис. 6. ROC-характеристика качества обучения нейронной сети для классификации здоровых добровольцев и пациентов с раком верхних дыхательных путей

Fig. 6. ROC-characteristic of the neural network training quality for classifying healthy volunteers and patients with upper respiratory tract cancer

В результате второго эксперимента точность предварительной диагностики составила в среднем 87% при показателях чувствительности 92,45% и специфичности 82,98%. Порог разделения положительных и отрицательных проб выдыхаемого воздуха был найден с помощью анализа ROC-характеристики и составил 0,448. Величина AUC составила 0,926, что ниже, чем в первом эксперименте, но определяющее высокое качество работы классификатора.

Результаты исследований, представленные в работе, подтверждают наличие обобщенного признака дифференциации выдыхаемого воздуха от здоровых добровольцев, пациентов с раком легких и пациентов с раком верхних дыхательных путей. На имеющемся сравнительно небольшом наборе экспериментальных данных с учетом применяемых метрик качества обучения нейронной сети этот

признак дифференциации классов определяется с высокой достоверностью ($P < 0,05$). Решение обратной задачи позволяет выявить датчики и, как следствие, группы летучих органических соединений, характерные для отдельной рассматриваемой патологии в сравнении с группой здоровых добровольцев.

На текущем этапе исследования изучались возможность и эффективность использования цифровых паттернов выдыхаемого воздуха с разным форматом и разными архитектурами нейронных сетей для дифференциации здоровых добровольцев и пациентов с патологией. С учетом полученных результатов можно также заключить, что большой перспективой обладает технология оцифровки не только выдыхаемого воздуха человека, но и всех возможных прямых и косвенных данных, характеризующих состояние здоровья человека, включая анамнестические показатели, биохимические и др., с формированием медицинского информационного цифрового двойника человека. Применение цифровых двойников человека совместно с алгоритмами нейросетевой обработки данных, учет дополнительных сведений, например таких, как возраст, имеющиеся хронические заболевания и др., позволит выявлять закономерности в изменении состояния здоровья и с высокой точностью определять рекомендации для его улучшения.

Заключение

В работе предложена архитектура нейронной сети для проведения процедуры предварительной диагностики злокачественных новообразований легких и верхних дыхательных путей по анализу выдыхаемого воздуха. В исследовании применялись полносвязные и сверточные нейронные сети для классификации сигналов, поступающих в виде цифровых кодов с неселективных полупроводниковых газовых датчиков. Сравнительный анализ архитектур показал, что сверточная нейронная сеть (CNN) демонстрирует большую эффективность по сравнению с полносвязной (DNN) на данных с газоаналитических медицинских проборов. Это обусловлено способностью CNN выявлять общие сигнатуры в сигнале за счет использования ядер свертки, что особенно актуально при анализе временных рядов термоциклирования датчиков. Полносвязные сети, хотя и способны обучаться сложным зависимостям, страдают от избыточности параметров, что приводит к более быстрому переобучению, особенно при увеличении объема входных данных.

Исследование показало, что использование полного набора данных сигнала с одного цикла измерения не является необходимым для достижения высокой точности классификации. Наибольшую информативность несет последняя волна сигнала – финальный этап термоциклирования, когда датчики достигают максимальной чувствительности к отдельным летучим органическим соединениям малой концентрации в выдыхаемом воздухе. Прореживание данных в пределах последнего импульса термоциклирования не оказывает существенного влияния на качество классификатора, что позволяет сократить объем входных данных без потери точности. В то же время подача на вход всей последовательности сигнала увеличивает время обучения, требует больше вычислительных ресурсов и ускоряет процесс перехода в состояние переобучения, особенно в случае полносвязной архитектуры нейронной сети.

Сверточная архитектура оказалась более устойчивой к вариативности экспериментальных данных и обеспечила лучшую обобщающую способность. Характеристики точности CNN достигли уровня 86,79% по показателю чувствительности и 93,62% по специфичности для 47 проб выдыхаемого воздуха от здоровых добровольцев и 53 – от пациентов с раком легких; 92,45% по показателю чувствительности и 82,98% по специфичности для 47 проб здоровых добровольцев и 53 – пациентов с раком верхних дыхательных путей, что сопоставимо с эффективностью современных рентгенологических методов визуальной диагностики рассматриваемых патологий. Эти показатели были получены в условиях обеспечения баланса между сложностью модели и объемом входных данных, что обусловлено необходимостью как выбора оптимальной архитектуры нейронной сети, так и оптимизации формата входных данных и информационного пространства входных признаков дифференциации. Таким образом, сверточная нейронная сеть с использованием только последнего импульса сигнала термоциклирования представляет собой достаточное и рациональное решение с точки зрения производительности, точности и устойчивости к переобучению, что делает ее предпочтительной для интеграции в газо-

аналитические медицинские приборы скрининга злокачественных новообразований легких и верхних дыхательных путей.

Список источников

1. Kumar A. et al. Revolutionizing healthcare with 5G and AI: Integrating emerging technologies for personalized care and cancer management // *Intelligent Hospital*. 2025. V. 1 (1). Art. 100005. P. 1–10. doi: 10.1016/j.inhs.2025.100005
2. Peng K. et al. Image Segmentation Network for Laparoscopic Surgery // *Biomimetic Intelligence and Robotics*. 2025. V. 5. P. 1–12. doi: 10.1016/j.birob.2025.100236
3. Paithankar V., Devnani D., Nimburkar T.A. A Review Article on "AI-Guided Discovery of Novel Antiinflammatory Agents for Cancer Therapy: A New Era in Drug Development" // *Intelligent Hospital*. 2025. V. 1 (1). Art. 100007. doi: 10.1016/j.inhs.2025.100007
4. Feng S., Yin X., Shen Y. Artificial intelligence powered precision: Unveiling the tumor microenvironment for a new frontier in personalized cancer therapy // *Intelligent Medicine*. 2025. V. 5. P. 95–98. doi: 10.1016/j.imed.2025.02.001
5. Zhang L. et al. Osteosarcoma knowledge graph question answering system: deep learning-based knowledge graph and large language model fusion // *Intelligent Medicine*. 2025. V. 5. P. 99–110. doi: 10.1016/j.imed.2024.12.001
6. Asadi O., Hajhosseini M., Shirzadi S. et al. Improvement of classification accuracy of functional near-infrared spectroscopy signals for hand motion and motor imagery using a common spatial pattern algorithm // *Intelligent Medicine*. 2025. V. 5. P. 123–131. doi: 10.1016/j.imed.2024.05.004
7. Feng Y. et al. Application of artificial intelligence-based computer vision methods in liver diseases: a bibliometric analysis // *Intelligent Medicine*. 2025. V. 5. P. 111–122. doi: 10.1016/j.imed.2024.09.008
8. Li T. et al. Application of multimodal deep learning in the auxiliary diagnosis and treatment of dermatological diseases // *Intelligent Medicine*. 2025. V. 5. P. 132–140. doi: 10.1016/j.imed.2024.10.002
9. Ram U.S., Pogue J.A., Soike M. et al. Assessing quantitative performance and expert review of multiple deep learning-based frameworks for computed tomography-based abdominal organ auto-segmentation // *Intelligent Oncology*. 2025. V. 1. P. 160–171. doi: 10.1016/j.intonc.2025.03.003
10. Wang S., Guo X., Ma J. et al. AWCDL: Automatic weight calibration deep learning for detecting HER2 status in whole-slide breast cancer image // *Intelligent Oncology*. 2025. V. 1. P. 128–138. doi: 10.1016/j.intonc.2025.03.008
11. Adeoye J., Su Y. Deep learning with data transformation improves cancer risk prediction in oral precancerous conditions // *Intelligent Medicine*. 2025. V. 5. P. 141–150. doi: 10.1016/j.imed.2024.11.003
12. Chen X., Xiang J., Lu S. et al. Evaluating large language models and agents in healthcare: key challenges in clinical applications // *Intelligent Medicine*. 2025. V. 5. P. 151–163. doi: 10.1016/j.imed.2025.03.002
13. Wei M., Teh B.S., Xu B. The new paradigm for global oncologists in the era of generative AI // *Intelligent Oncology*. 2025. V. 1. P. 189–192. doi: 10.1016/j.intonc.2025.06.002
14. Cai W. DeepSeek AI: Transforming medical AI with cost-efficiency, transparency, and privacy preservation // *Intelligent Oncology*. 2025. V. 1. P. 172–175. doi: 10.1016/j.intonc.2025.03.006
15. Chernov V.I., Choyzonov E.L., Kulbakin D.E., Obkhodskaya E.V., Obkhodskiy A.V., Popov A.S., Sachkova A.S., Sachkov V.I. Cancer Diagnosis by Neural Network Analysis of Data from Semiconductor Sensors // *Diagnostics*. 2020. V. 10 (9). Art. 677. doi: 10.3390/diagnostics10090677
16. Обходский А.В., Обходская Е.В., Кульбакин Д.Е., Чойнзонов Е.Л., Родионов Е.О., Сачков В.И., Чернов В.И., Попов А.С. Программный комплекс нейросетевой классификации данных газоаналитического обследования дыхательной системы // *Вестник Томского государственного университета. Управление, вычислительная техника и информатика*. 2024. № 69. С. 112–123. doi: 10.17223/19988605/69/12
17. Сорокоумова Д.А., Корелин О.Н., Сорокоумов А.В. Построение и обучение нейронной сети для решения задачи распознавания речи // *Труды Нижегородского государственного технического университета им. П. Е. Алексеева*. 2015. № 3 (110). С. 77–84.
18. Maaten van der L.J.P., Hinton G.E. Visualizing High-Dimensional Data Using t-SNE // *J. of Machine Learning Research*. 2008. V. 9. P. 2579–2605.
19. Sokolova M., Japkowicz N., Szpakowicz S. Beyond Accuracy, F-Score and ROC: A Family of Discriminant Measures for Performance Evaluation // *Advances in Artificial Intelligence*. 2006. V. 4304. P. 1015–1021.

References

1. Kumar, A. et al. (2025) Revolutionizing healthcare with 5G and AI: Integrating emerging technologies for personalized care and cancer management. *Intelligent Hospital*. 1(1). Art. 100005. pp. 1–10. doi: 10.1016/j.inhs.2025.100005
2. Peng, K. et al. (2025) Image Segmentation Network for Laparoscopic Surgery. *Biomimetic Intelligence and Robotics*. 5. pp. 1–12. doi: 10.1016/j.birob.2025.100236
3. Paithankar, V., Devnani, D. & Nimburkar, T.A. (2025) A Review Article on "AI-Guided Discovery of Novel Antiinflammatory Agents for Cancer Therapy: A New Era in Drug Development." *Intelligent Hospital*. 1(1). Art. 100007. doi: 10.1016/j.inhs.2025.100007
4. Feng, S., Yin, X. & Shen, Y. (2025) Artificial intelligence powered precision: Unveiling the tumor microenvironment for a new frontier in personalized cancer therapy. *Intelligent Medicine*. 5. pp. 95–98. doi: 10.1016/j.imed.2025.02.001
5. Zhang, L. et al. (2025) Osteosarcoma knowledge graph question answering system: deep learning-based knowledge graph and large language model fusion. *Intelligent Medicine*. 5. pp. 99–110. doi: 10.1016/j.imed.2024.12.001

6. Asadi, O., Hajhosseini, M., Shirzadi, S. et al. (2025) Improvement of classification accuracy of functional near-infrared spectroscopy signals for hand motion and motor imagery using a common spatial pattern algorithm. *Intelligent Medicine*. 5. pp. 123–131. doi: 10.1016/j.imed.2024.05.004
7. Feng, Y. et al. (2025) Application of artificial intelligence-based computer vision methods in liver diseases: a bibliometric analysis. *Intelligent Medicine*. 5. pp. 111–122. doi: 10.1016/j.imed.2024.09.008.
8. Li, T. et al. (2025) Application of multimodal deep learning in the auxiliary diagnosis and treatment of dermatological diseases. *Intelligent Medicine*. 5. pp. 132–140. doi: 10.1016/j.imed.2024.10.002.
9. Ram, U.S., Pogue, J.A., Soike, M. et al. (2025) Assessing quantitative performance and expert review of multiple deep learning-based frameworks for computed tomography-based abdominal organ auto-segmentation. *Intelligent Oncology*. 1. pp. 160–171. doi: 10.1016/j.intonc.2025.03.003
10. Wang, S., Guo, X., Ma, J. et al. (2025) AWC DL: Automatic weight calibration deep learning for detecting HER2 status in whole-slide breast cancer image. *Intelligent Oncology*. 1. pp. 128–138. doi: 10.1016/j.intonc.2025.03.008
11. Adeoye, J. & Su, Y. (2025) Deep learning with data transformation improves cancer risk prediction in oral precancerous conditions. *Intelligent Medicine*. 5. pp. 141–150. doi: 10.1016/j.imed.2024.11.003
12. Chen, X., Xiang, J., Lu, S. et al. (2025) Evaluating large language models and agents in healthcare: key challenges in clinical applications. *Intelligent Medicine*. 5. pp. 151–163. doi: 10.1016/j.imed.2025.03.002
13. Wei, M., The, B.S. & Xu, B. (2025) The new paradigm for global oncologists in the era of generative AI. *Intelligent Oncology*. 1. pp. 189–192. doi: 10.1016/j.intonc.2025.06.002
14. Cai, W. (2025) DeepSeek AI: Transforming medical AI with cost-efficiency, transparency, and privacy preservation. *Intelligent Oncology*. 1. pp. 172–175. doi: 10.1016/j.intonc.2025.03.006
15. Chernov, V.I., Choyzonov, E.L., Kulbakin, D.E., Obkhodskaya, E.V., Obkhodskiy, A.V., Popov, A.S., Sachkova, A.S. & Sachkov, V.I. (2020) Cancer Diagnosis by Neural Network Analysis of Data from Semiconductor Sensors. *Diagnostics*. 10(9). Art. 677. doi: 10.3390/diagnostics10090677
16. Obkhodskiy, A.V., Kulbakin, D.E., Obkhodskaya, E.V., Popov, A.S., Rodionov, E.O., Sachkov, V.I., Chernov, V.I. & Choyzonov, E.L. (2024) Neural network classification software for the gas analytical survey data of respiratory system. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 69. pp. 112–123. doi: 10.17223/19988605/69/12
17. Sorokoumova, D.A., Korelin, O.N. & Sorokoumov, A.V. (2015) Postroeniye i obucheniye neyronnoy seti dlya resheniya zadachi raspoznavaniya rechi [Construction and Training of a Neural Network for Solving Speech Recognition Tasks]. *Trudy Nizhegorodskogo gosudarstvennogo tekhnicheskogo universiteta im. R. Ye. Alekseeva*. 3(110). pp. 77–84.
18. Maaten, van der L.J.P. & Hinton, G.E. (2008) Visualizing High-Dimensional Data Using t-SNE. *Journal of Machine Learning Research*. 9. pp. 2579–2605.
19. Sokolova, M., Japkowicz, N. & Szpakowicz, S. (2006) Beyond Accuracy, F-Score and ROC: A Family of Discriminant Measures for Performance Evaluation. *Advances in Artificial Intelligence. AI 2006. Lecture Notes in Computer Science*. 4304. pp. 1015–1021.

Информация об авторах:

Обходский Артём Викторович – кандидат технических наук, доцент инженерной школы ядерных технологий Национального исследовательского Томского политехнического университета (Томск, Россия). E-mail: art707@yandex.ru

Кульбакин Денис Евгеньевич – доктор медицинских наук, заведующий отделением Научно-исследовательского института онкологии Томского национального исследовательского медицинского центра РАН (Томск, Россия). E-mail: kulbakin_d@mail.ru

Обходская Елена Владимировна – кандидат технических наук, старший научный сотрудник лаборатории химических технологий Национального исследовательского Томского государственного университета (Томск, Россия). E-mail: fil330a@yandex.ru

Лаконкин Владислав Сергеевич – лаборант-исследователь Научно-исследовательского института онкологии Томского национального исследовательского медицинского центра Российской академии наук; студент инженерной школы ядерных технологий Национального исследовательского Томского политехнического университета (Томск, Россия). E-mail: vsl13@tpu.ru

Родионов Евгений Олегович – кандидат медицинских наук, старший научный сотрудник Научно-исследовательского института онкологии Томского национального исследовательского медицинского центра Российской академии наук; ассистент кафедры онкологии Сибирского государственного медицинского университета (Томск, Россия). E-mail: rodionov_eo@oncology.tomsk.ru

Сачков Виктор Иванович – доктор химических наук, заведующий лабораторией химических технологий Национального исследовательского Томского государственного университета (Томск, Россия). E-mail: vicsachkov@gmail.com

Чернов Владимир Иванович – профессор, доктор медицинских наук, член-корреспондент РАН, заместитель директора по научной и инновационной работе, заведующий отделением радионуклидной терапии и диагностики Научно-исследовательского института онкологии Томского национального исследовательского медицинского центра РАН (Томск, Россия). E-mail: chernov@tnimc.ru

Чойзонзов Евгений Лхамцыренович – профессор, доктор медицинских наук, академик РАН, директор Научно-исследовательского института онкологии Томского национального исследовательского медицинского центра РАН (Томск, Россия). E-mail: choynzonov@tnimc.ru

Вклад авторов: все авторы сделали эквивалентный вклад в подготовку публикации. Авторы заявляют об отсутствии конфликта интересов.

Information about the authors:

Obkhodskiy Artem V. (Candidate of Technical Sciences, National Research Tomsk Polytechnic University, Tomsk, Russian Federation), E-mail: art707@yandex.ru

Kulbakin Denis E. (Doctor of Medical Sciences, Head of Department of Head and Neck Tumors, Cancer Research Institute, Tomsk National Research Medical Center of the RAS, Tomsk, Russian Federation). E-mail: kulbakin_d@mail.ru

Obkhodskaya Elena V. (Candidate of Technical Sciences, National Research Tomsk State University, Tomsk, Russian Federation). E-mail: fil330a@yandex.ru

Lakonkin Vladislav S. (Laboratory Research Assistant, Tomsk National Research Medical Center of the Russian Academy of Sciences; Student, School of Nuclear Technology, National Research Tomsk Polytechnic University, Tomsk, Russian Federation). E-mail: vsl13@tpu.ru

Rodionov Evgeniy O. (Candidate of Medical Sciences, Cancer Research Institute, Tomsk National Research Medical Center of the RAS; Siberian State Medical University, Tomsk, Russian Federation). E-mail: rodionov_eo@oncology.tomsk.ru

Sachkov Victor I. (Doctor of Chemistry Sciences, National Research Tomsk State University, Tomsk, Russian Federation). E-mail: vicsachkov@gmail.com

Chernov Vladimir I. (Professor, Doctor of Medical Sciences, Head of Nuclear Medicine Department, Cancer Research Institute; Tomsk National Research Medical Center of the RAS, Tomsk, Russian Federation). E-mail: chernov@tnimc.ru

Choynzonov Evgeny L. (Professor, Doctor of Medical Sciences, Member of the RAS, Director of the Cancer Research Institute, Tomsk National Research Medical Center; Head of Oncology Department of Siberian State Medical University, Tomsk, Russian Federation). E-mail: choynzonov@tnimc.ru

Contribution of the authors: the authors contributed equally to this article. The authors declare no conflicts of interests.

Поступила в редакцию 27.07.2025; принята к публикации 02.12.2025

Received 27.07.2025; accepted for publication 02.12.2025

Научная статья

УДК 519.688

doi: 10.17223/19988605/73/15

Концепция архитектуры приложения для построения аналитической аппроксимации распределения вероятностей числа заявок в СМО на основе имитационного моделирования**Прокудина Юлия Андреевна¹, Моисеев Александр Николаевич²**^{1, 2} Национальный исследовательский Томский государственный университет, Томск, Россия¹ prokudina_ua@mail.ru² moiseev.tsu@gmail.com

Аннотация. Представлена концепция общей архитектуры приложения для построения аналитической аппроксимации распределения вероятностей числа заявок в системах массового обслуживания. Главная идея, заложенная в основу данного приложения, – построение аналитических аппроксимаций распределения вероятностей на основе численных результатов, полученных в серии экспериментов имитационного моделирования соответствующих систем массового обслуживания. Качество генерируемых аппроксимаций оценивается с использованием метрик близости распределений вероятностей. Предложены общий алгоритм процесса построения аналитических аппроксимаций, необходимые для его реализации классы, пакеты и структуры данных, а также детализация их взаимодействий для решения поставленной задачи.

Ключевые слова: имитационное моделирование; система массового обслуживания; аппроксимация распределения вероятностей.

Для цитирования: Прокудина Ю.А., Моисеев А.Н. Концепция архитектуры приложения для построения аналитической аппроксимации распределения вероятностей числа заявок в СМО на основе имитационного моделирования // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2025. № 73. С. 134–142. doi: 10.17223/19988605/73/15

Original article

doi: 10.17223/19988605/73/15

Concept of application architecture for constructing analytical approximation of the probability distribution of the number of customers in queuing systems based on simulation modeling**Yuliya A. Prokudina¹, Alexander N. Moiseev²**^{1, 2} National Research Tomsk State University, Tomsk, Russian Federation¹ prokudina_ua@mail.ru² moiseev.tsu@gmail.com

Abstract. The article presents the idea and concept of the general architecture of the application for constructing an analytical approximation of the probability distribution of the number of customers in queuing systems. The main idea underlying this application is to construct analytical approximations of the probability distribution based on numerical results obtained in a series of simulation experiments of the corresponding queuing systems. The quality of the generated approximations will be assessed using metrics of the proximity of probability distributions. The article proposes a general algorithm for the process of constructing analytical approximations, classes, packages and data structures required for its implementation, as well as a detailing of their interactions to solve the problem.

Keywords: simulations; queuing system; approximation of probability distribution.

For citation: Prokudina, Y.A., Moiseev, A.N. (2025) Concept of application architecture for constructing analytical approximation of the probability distribution of the number of customers in queuing systems based on simulation modeling. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 73. pp. 134–142. doi: 10.17223/19988605/73/15

Введение

При исследовании систем массового обслуживания (СМО) [1] одной из главных задач является получение в аналитическом виде распределения вероятностей числа заявок в системе, обычно в стационарном режиме работы. При этом из-за трудностей математического характера только в редких случаях возможно получить точный аналитический результат. В некоторых случаях удастся аналитически получить аппроксимации распределений, используя асимптотические методы [2, 3] или аппроксимации исходных распределений [4]. В остальных случаях возможно получение рекуррентных алгоритмов [5] либо – чаще всего – проведение численного или имитационного моделирования [6].

Имитационное моделирование позволяет получить численный результат практически для любой конфигурации СМО. Поэтому возникает идея – провести серию экспериментов по имитационному моделированию одной и той же СМО с разными значениями параметров, получить наиболее подходящие в каждом случае распределения вероятностей числа заявок в системе, а затем систематизировать эти результаты и оценить влияние каждого входного параметра на результирующую функцию распределения. Таким образом, в качестве результата будет предложена аппроксимация функции распределения вероятностей числа заявок в системе в виде какой-то известной функции распределения, параметры которой определяются в терминах исходных параметров модели. В приложении реализован банк распределений, среди которых производится поиск.

В настоящее время в ИПМКН ТГУ ведется разработка программного комплекса SimQ [7], предназначенного для имитационного моделирования систем массового обслуживания. Предполагается использовать ядро этой системы для выполнения непосредственно процесса имитационного моделирования и получения эмпирических распределений, а затем с помощью надстроенного каскада модулей построить аналитическую аппроксимацию числа заявок в системе. В данной работе предложена концепция архитектуры программного комплекса, который будет выполнять эту задачу, описаны задачи каждого модуля, основные идеи по реализации алгоритмов.

1. Концепция архитектуры приложения

Чтобы реализовать описанный процесс, предлагается следующая концепция архитектуры разрабатываемого программного комплекса (рис. 1; на рисунках используется нотация UML [8]). Мы выделяем пять основных частей системы: Интерфейс пользователя (UI), Управление (AF_Controller), Модуль оценки параметров (Estimator), Ядро моделирования SimQ и Модуль подбора аналитических выражений (Analytic).

Пакет UI предназначен для элементов, формирующих интерфейс пользователя. С помощью этих элементов пользователь выбирает конфигурацию системы массового обслуживания, задает параметры модели и параметры процесса построения аналитической аппроксимации. Здесь же находятся объекты, с помощью которых производится вывод результатов по окончании процесса построения аналитической аппроксимации.

Пакет AF_Controller реализует функции управления процессом построения аппроксимации, многократно запускает имитационное моделирование выбранной системы (конфигурации СМО) для различных значений варьируемых параметров модели, получает от модуля Estimator соответствующие оценки распределений и передает их в модуль Analytic.

Работа пакета Estimator подробно описана в [9]. Его задачами являются поиск наиболее близкого распределения вероятностей из имеющейся коллекции распределений и оценка его параметров. На самом деле оценки строятся для каждого распределения из коллекции, если это возможно, при этом вычисляется метрика близости построенной оценки к эмпирическому распределению (в настоящей статье для этого используется расстояние Колмогорова).

Ядро моделирования SimQ [7] занимается непосредственно имитационным моделированием заданной СМО и выдает в качестве результата эмпирическое распределение вероятностей числа заявок в системе.

Наконец, задачей пакета Analytic является построение аппроксимации распределения вероятностей числа заявок в выбранной СМО в аналитической форме. Для этого используются выражения для распределений из имеющейся коллекции, а также построенные с помощью Estimator оценки распределений и их метрики близости к результатам имитационного моделирования.

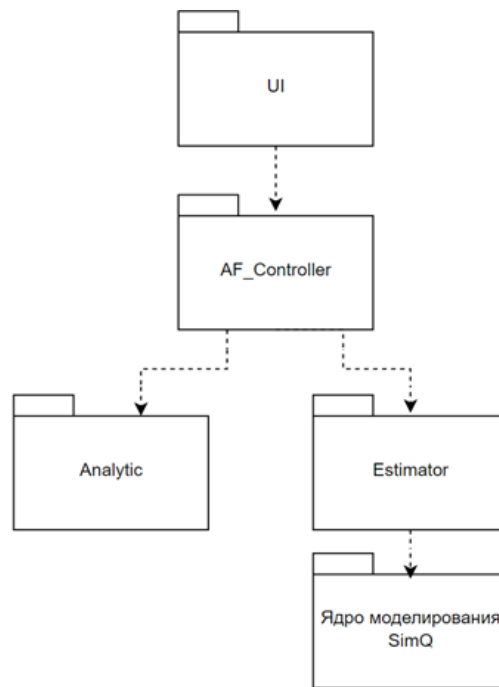


Рис. 1. Концепция архитектуры приложения для построения аналитической аппроксимации

Fig. 1. Concept of the architecture of the application for constructing analytical approximation

В результате пользователь получает в аналитическом виде предполагаемое распределение вероятностей числа заявок в системе для выбранной модели СМО. Это распределение в рамках данной работы будем называть аналитической аппроксимацией. В принципе, приложение может выдавать пользователю все аналитические аппроксимации, полученные для каждого элемента коллекции распределений, сортированные по значению некоторой агрегированной метрики близости, чтобы пользователь наглядно видел точность каждого вида аппроксимации и мог выбрать подходящую, даже если значение метрики близости (погрешность) было не минимальным. Это полезно, например, в тех случаях, когда погрешность отличается незначительно, но одно из распределений является более удобным для использования на практике.

2. Общий контур процесса построения аппроксимации

Рассмотрим общий ход предлагаемого процесса построения аналитической аппроксимации распределения вероятностей в разрабатываемой системе (рис. 2).

Пользователь выбирает конфигурацию (тип) СМО, в некоторых случаях задает значения неизменяемых параметров модели, а также параметры самого процесса построения. Эти данные поступают в AF_Controller, который в цикле перебирает все значения варьируемых параметров модели, необходимые для построения аналитических выражений, запускает для них имитационное моделирование соответствующей СМО в Ядре моделирования SimQ, передает получаемые при этом эмпирические распределения модулю Estimator, который строит для них оценки распределения и вычисляет значения метрики качества для каждой оценки. Эти оценки и соответствующие значения метрик записываются в Структуру ПРО («Параметры–Распределение–Оценки»), которая индексируется по выбранным значениям параметров СМО. Этот процесс продолжается до тех пор, пока не будут выбраны все необходимые значения варьируемых параметров модели.

Далее сформированная Структура ПРО передается модулю Analytic, который на ее основе строит аналитические аппроксимации распределения вероятностей числа заявок в СМО выбранной пользователем конфигурации, а также для каждой такой аппроксимации вычисляет некоторую агрегированную метрику качества, например, используя значения метрики качества оценок распределений, полученных ранее и хранящихся в Структуре ПРО.

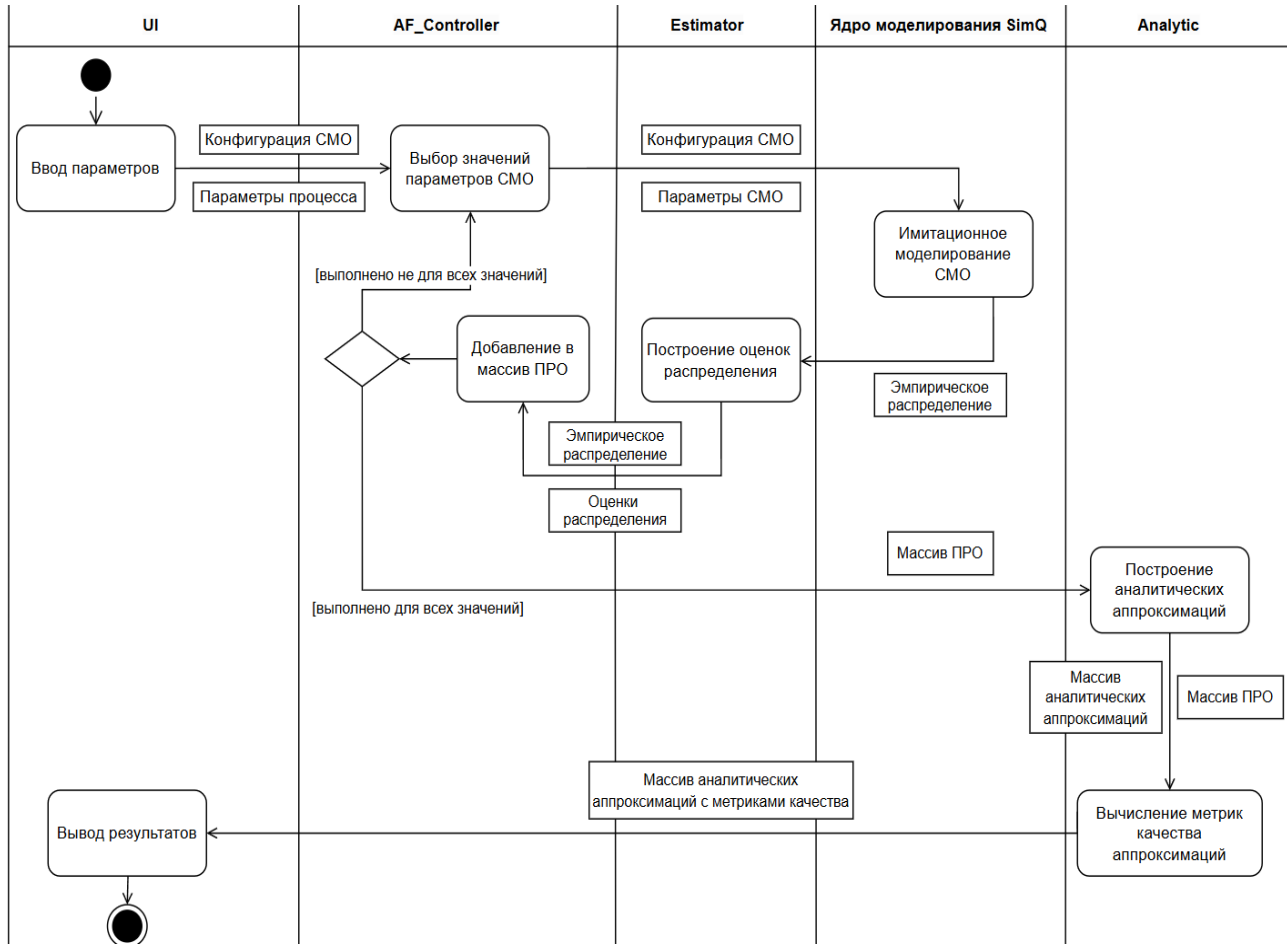


Рис. 2. Процесс построения аналитической аппроксимации
Fig. 2. Process of constructing of the analytical approximation

Результат выдается пользователю в виде аналитических выражений аппроксимаций для распределения вероятностей, выполненных в терминах исходных (варьируемых) параметров СМО.

3. Общая архитектура системы, связанная с построением аналитических аппроксимаций

На рис. 3 показана часть элементов пакетов системы, которые задействованы в процессе построения аналитических аппроксимаций.

В пакете **AF_Controller** содержатся следующие важные для текущего рассмотрения классы:

CollectAFs – класс, который запускает вычисление аналитических функций параметров оцененных распределений (аппроксимаций) [9]. Его операции:

AFs_Collect(init: String): AF[]* – метод, который запускает вычисление аналитических аппроксимаций параметров и выдает результаты в виде массива структур AF (см. ниже). Сначала вызывается метод формирования массива структур ПРО (массив содержит формулы для всех возможных параметров всех необходимых оценок) класса CalcPDE, и на основе этого массива запускается формирование массива структур AF, обращаясь к пакету Analytic. *init* – атрибут, который хранит конфигурацию и значения параметров СМО и параметры симуляции.

PDEs_Collect(): PDE[]* – метод, который в цикле перебирает все значения варьируемых параметров модели, запускает формирование структур ППО (класс PDE, см. ниже), обращаясь к методам класса CalcPDE, и собирает из этих структур массив.

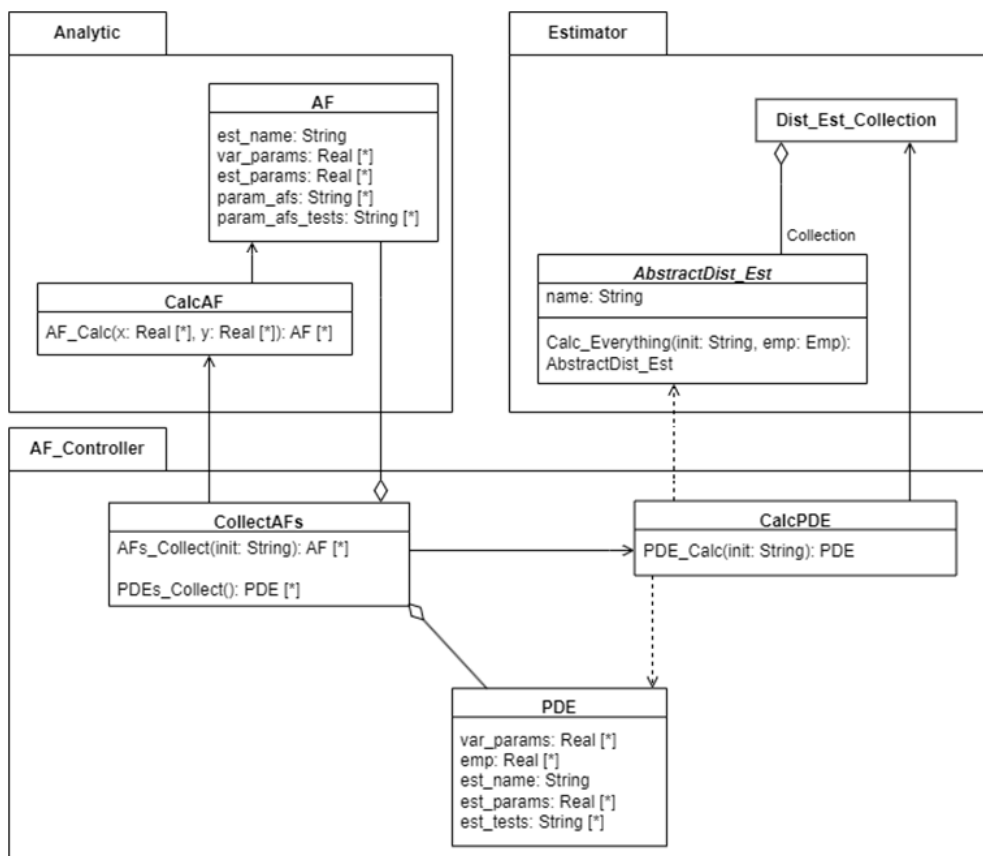


Рис. 3. Структура пакетов Analytic, Estimator, AF_Controller
Fig. 3. Structure of the Analytic, Estimator, AF_Controller packages

CalcPDE – класс, который запускает процесс имитационного моделирования, чтобы получить эмпирическое распределение и его оценки (аппроксимации распределениями из банка распределений с оценками их параметров), а после формирует массив ППО.

PDE_Calc(init: String): PDE – вызывает методы пакетов Ядро SimQ для генерации эмпирических распределений и Estimator для получения оценок и соответствующих значений метрик качества, а также вызывает свои методы для записи этих оценок и метрик в Структуру ППО. Возвращает Структуру ППО.

PDE – класс, который представляет в приложении структуру ППО, содержит следующие атрибуты:

- var_params: Real[*]* – массив исходных параметров системы и их значений;
- emp: Real[*]* – массив значений эмпирического распределения;
- est_name: String* – название оценки (распределения);
- est_params: Real[*]* – массив оценок параметров;
- est_tests: String[*]* – массив метрик качества оценок параметров.

Структура PDE и пример заполнения данными приведены на рис. 4. Массив исходных параметров системы и их значений *var_params* содержит значения параметров исходной СМО, в том числе заданные и изменяемые в классе *CollectAFs*; массив значений эмпирического распределения *emp* состоит из пар «значение–вероятность», представляющих эмпирическое распределение вероятностей, полученное в результате имитационного моделирования; оценки распределения: названия оценок *est_name* – ключевые слова, определяющие распределение вероятностей из банка распределений,

массив оценок параметров *est_params* – массив пар «параметр–значение», полученный на основе построения оценок параметров для одного *est_name* на основе текущего *emp*, массив метрик качества *est_tests* содержит набор значений метрик качества (точности) соответствующей оценки распределения. Обозначения λ , γ и β приведены в качестве примера варьируемых параметров, для конкретной конфигурации, рассматриваемой СМО, они могут означать, например, соответственно интенсивность входящего потока, интенсивность обслуживания заявок и интенсивность ухода заявок из очереди. Другие обозначения, используемые на рисунке: *Kolm.Dist*, ω^2 – метрики качества (расстояние Колмогорова и метрика «омега-квадрат» соответственно), p – параметр геометрического распределения, μ , σ – параметры дискретизированного нормального распределения (математическое ожидание и среднеквадратическое отклонение).

Варьируемые параметры <i>var_params</i>	Эмп. распр. <i>emp</i>	Назв. оценки (распр.) <i>est_name</i>	Оценки параметров <i>est_params</i>	Метрики качества <i>est_tests</i>
$\lambda = 1, \gamma = 3, \beta = 2$	$[0 = 0.001, 1 = 0.0003, \dots]$	Дис. нор.	$\mu = 3, \sigma = 6$	$[Kolm.Dist. = 0.05, w2 = 0.9, \dots]$
		Геомет.	$p = 0.8$	$[Kolm.Dist. = 0.04, w2 = 0.8, \dots]$
	

Рис. 4. Структура типа PDE (ППО) и пример ее заполнения
Fig. 4. Structure of PDE type and an example of its filling

Пакет **Estimator**, представленный впервые в [9], модифицирован и дополнен. Он содержит следующие классы, важные для текущей работы:

AbstractDist_Est – абстрактный класс, представляющий одно распределение вероятностей из банка распределений и включающий в себя всю информацию об оценке его параметров (в том числе метрики качества). Его атрибуты и операции:

name: String – имя распределения (оценки распределения);

Calc_Everything(init: String, emp: Emp): AbstractDist_Est – для заданного эмпирического распределения *emp* строит оценку и вычисляет для нее значения метрик качества.

Dist_Est_Collection – класс, позволяющий осуществлять операции над классами-потомками *AbstractDist_Est*: создавать экземпляры этих классов, вызывать их методы и т.п. Фактически представляет собой точку доступа к коллекции распределений вероятностей из банка распределений. Доступ осуществляется через следующий атрибут:

Collection: AbstractDist_Est[]* – в нем содержатся экземпляры каждого класса-потомка *AbstractDist_Est* (по одному на класс) из банка распределений. Этот массив формируется на старте приложения и позволяет другим классам работать с банком распределений для оценок.

Пакет **Analytic** содержит следующие важные для рассмотрения классы:

CalcAF – его задачей является построение (аппроксимация) аналитических зависимостей значений параметров оценок распределений от исходных параметров конфигурации СМО. Основным интерфейсом этого класса:

AF_Calc(x: Real[], y: Real[*]): AF[*]* – метод, который формирует аналитические аппроксимации параметров оценок распределения в виде функций исходных параметров (конфигурации) СМО. На вход подается массив значений x оцениваемого параметра; массив значений y параметра конфигурации СМО. Возвращает массив аналитических оценок выражений для параметра x через параметр y .

AF – класс, который представляет в приложении структуру, содержащую необходимую информацию по аналитическим оценкам параметров распределения. Имеет следующие атрибуты:

est_name: String – название распределения;

var_params: Real[]* – массив варьируемых параметров и их значений;

est_params: Real[]* – массив параметров распределения;

param_afs: String[]* – массив аналитических функций оценок параметров;

param_afs_tests: String[]* – массив метрик качества для аналитических функций оценок параметров.

Этот тип данных предназначен для сохранения аналитических выражений оценок параметров распределения, полученных в методе *AF_Calc(...)* класса *CalcAF*. AF состоит из названия распределения-оценки, значений варьируемых параметров, параметров оценки, массивов полученных аналитических функций оценок параметров, массивов метрик качества для каждой аналитической оценки. Структура данного класса и пример заполнения атрибутов представлены на рис. 5.

Назв. оценки (расп.) <i>est_name</i>	Вариир. парам-ы и значения <i>var_params</i>	Параметры оценки <i>est_params</i>	Аналит. фу-ии оценок параметров <i>param_afs</i>	Метрики качества <i>param_afs_tests</i>
Дис. нор. расп.	$\lambda = [1, 2, 3, 4]$ $\gamma = [1, 2, 3, 4]$ $\beta = [1, 2, 3, 4]$	μ	$[3\lambda + \gamma - 2^\beta - 0.124,$ $2.4^\lambda + \gamma/\beta - 0.62,$ $-4^\beta - 0.4]$	$[Kolm. Dist. = 0.012, w2 = 0.91, \dots]$
		σ	$[0.1\lambda + 0.4,$ $0.3^\gamma + 0.43\beta^4 - 0.23]$	

Рис. 5. Структура класса AF и пример ее заполнения
Fig. 5. Structure of AF class and an example of its filling

4. Реализация общего контура процесса построения аналитических аппроксимаций

На рис. 6 представлена общая реализация поведения прототипа в виде взаимодействия между элементами системы.

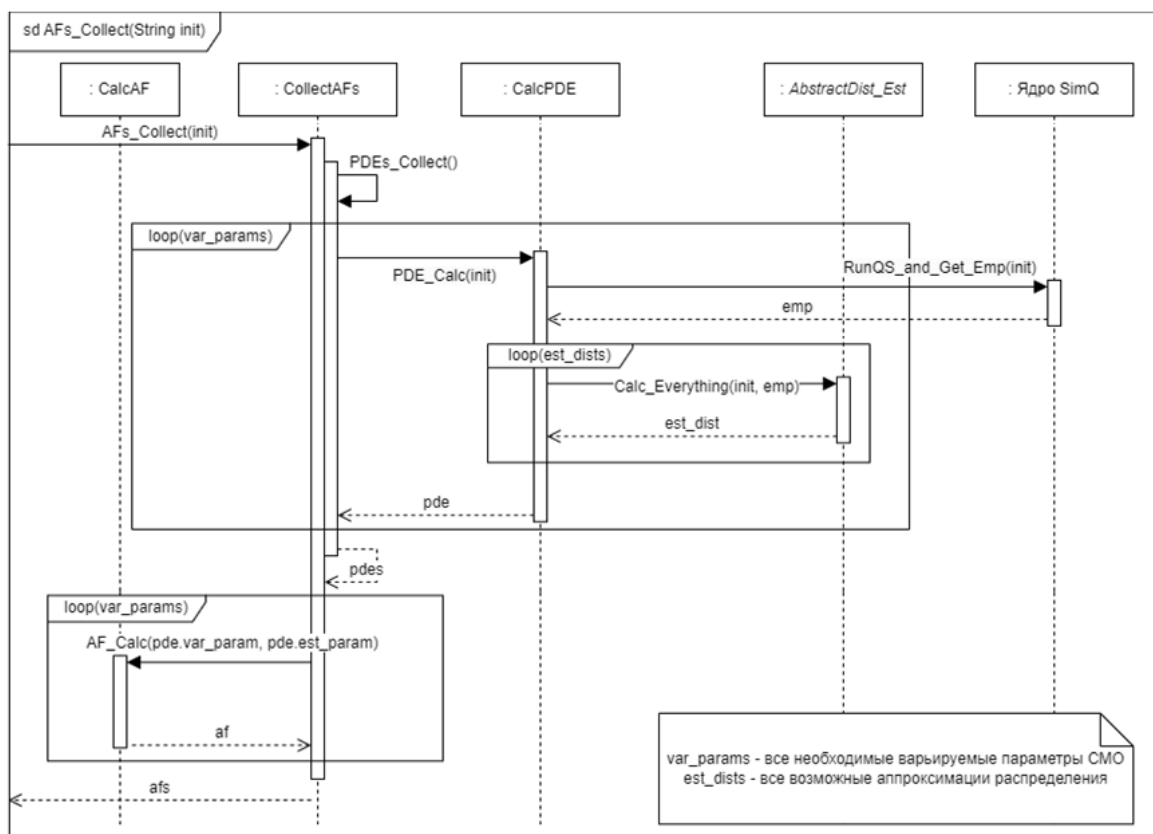


Рис. 6. Реализация общего контура процесса построения аналитических аппроксимаций
Fig. 6. Implementation of the general outline of the process of constructing analytical approximations

Вычисление аналитических функций начинается с запуска метода *AFs_Collect(...)* объекта класса *CollectAFs*, который сначала формирует массив Структур ПРО (коллекцию объектов типа PDE). Для этого в цикле с разными значениями входных параметров он обращается к классу *CalcPDE*, вызывая его метод *PDE_Calc(...)* для генерации массива Структур ПРО. Метод *PDE_Calc(...)*, в свою очередь, обращается к Ядру SimQ, запуская процесс имитационного моделирования заданной СМО с соответствующими значениями параметров. В результате получаем эмпирическое распределение целевой случайной величины. Далее в цикле вычисляются оценки этого распределения (аппроксимации и метрики их точности) через обращение к объектам из коллекции *Collection* класса *Dist_Est_Collection* пакета *Estimator*, и возвращается Структура ПРО (объект PDE).

После этого на основе массива сформированных Структур ПРО класс *CollectAFs* в цикле вызывает метод класса *CalcAF* для построения аналитических функций параметров оценок.

Основная теоретическая проблема, решение которой является критически важным для реализации модуля *Analytic* и работы всей системы, – это разработка методов построения самих аналитических аппроксимаций параметров оценок по имеющимся наборам данных. Здесь авторы видят возможность использования как прямых методов (построение уравнений и их решение), так и методов регрессионного анализа [10, 11] и даже, возможно, методов машинного обучения [12–14]. Выполнена опытная апробация работоспособности предложенной концепции архитектуры с прототипом модуля *Analytic*.

Заключение

Таким образом, в результате работы сформулированы общая концепция архитектуры приложения и основной контур построения аналитических аппроксимаций распределения числа заявок в СМО, конфигурацию которой определяет пользователь. Представленная концепция предлагает общую архитектуру, ключевые элементы и описание основных контуров взаимодействия программного комплекса, предназначенного для поиска аппроксимаций распределений вероятностей числа заявок и других вероятностных характеристик функционирования СМО в аналитическом виде. Реализация программного комплекса позволит получать аналитические выражения или их аппроксимации для указанных характеристик для СМО различных конфигураций, в том числе и таких, где получение аналитических результатов не представляется возможным.

Список источников

1. Назаров А.А., Терпугов А.Ф. Теория массового обслуживания: учебное пособие. 2-е изд., испр. Томск : Изд-во НТЛ, 2010. 228 с.
2. Назаров А.А., Моисеева С.П. Метод асимптотического анализа в теории массового обслуживания. Томск : Изд-во НТЛ, 2006. 112 с.
3. Моисеев А.Н., Назаров А.А. Бесконечнолинейные системы и сети массового обслуживания. Томск : Изд-во НТЛ, 2015. 240 с.
4. Moiseev A.N., Nazarov A. A., Paul S.V. Asymptotic Diffusion Analysis of Multi-Server Retrial Queue with Hyper-Exponential Service // *Mathematics*. 2020. V. 8 (4). Art. 531. P. 1–16.
5. Полховская А.В., Данилюк Е.Ю., Моисеева С.П., Бобкова О.С. Вероятностная модель системы совместного доступа с коллизиями, *H*-настойчивостью и отказами // *Вестник Томского государственного университета. Управление, вычислительная техника и информатика*. 2022. № 58. С. 35–46.
6. Приступа П.В., Сущенко С.П., Михеев П.А., Моисеев А.Н. Анализ эффективности прямой коррекции ошибок на уровне транспортного протокола // *Автоматика и телемеханика*. 2022. № 7. С. 59–81. doi: 10.31857/S0005231022070042
7. Прокудина Ю.А., Оруджов Э.А., Моисеев А.Н. Концепция архитектуры программного комплекса SimQ // *Труды Томского государственного университета. Сер. физико-математическая*. 2023. Т. 308. С. 155–159.
8. Буч Г., Рамбо Дж., Якобсон И. Язык UML : руководство пользователя / пер. с англ. Н. Мухина. 3-е изд. М. : Академия АйТи ; ДМК Пресс, 2022. 495 с.
9. Прокудина Ю.А., Моисеев А.Н. Прототип приложения для поиска распределения, наиболее близкого к заданному эмпирическому // *Труды Томского государственного университета. Сер. физико-математическая*. 2024. Т. 309. С. 199–205.
10. Гельман Э., Хилл Д., Вехтари А. Регрессия: теория и практика. М. : ДМК Пресс, 2022. 748 с.
11. Воскобойников Ю.Е. Регрессионный анализ данных в пакете MATHCAD : учеб. пособие. СПб. : Лань, 2022. 224 с.
12. Вишневский В.М., Ефросинин Д.В. Теория очередей и машинное обучение. М. : ИНФРА-М, 2024. 370 с.
13. Sherzer E., Senderovich A., Baron O., Krass D. Can machines solve general queueing systems? 2022. URL: <https://arxiv.org/abs/2202.01729> (accessed: 17.02.2025).

14. Raheema M., Abdullah A. Function approximation using neuraland fuzzy methods // Communications on Applied Electronics 2016. V. 6 (3). P. 35–42.

References

1. Nazarov, A.A. & Terpuhov, A.F. (2010) *Teoriya massovogo obsluzhivaniya* [Queuing System Theory]. 2nd ed. Tomsk: NTL.
2. Nazarov, A.A. & Moiseeva, S.P. (2006) *Metod asimptoticheskogo analiza v teorii massovogo obsluzhivaniya* [Method of asymptotic analysis in queueing theory]. Tomsk: NTL.
3. Moiseev, A.N. & Nazarov, A.A. (2015) *Beskonechno lineynye sistemy i seti massovogo obsluzhivaniya* [Infinitely linear systems and queueing networks]. Tomsk: NTL.
4. Moiseev A.N., Nazarov A.A. & Paul, S.V. (2020) Asymptotic Diffusion Analysis of Multi-Server Retrial Queue with Hyper-Exponential Service. *Mathematics*. 8(4). Art. 531. pp. 1–16. doi: 10.3390/math8040531
5. Polkhovskaya, A.V., Danilyuk, E.Yu., Moiseeva, S.P. & Bobkova O.S. (2022) The probabilistic model of sharing system with collisions, H-persistence and rejections. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 58. pp. 35–46. doi: 10.17223/19988605/58/4
6. Pristupa, P.V., Suschenko, S.P., Mikheev, P.A. & Moiseev, A.N. (2022) Analysis of the effectiveness of forward error correction at the transport protocol level. *Avtomatika i Telemekhanika*. 7. pp. 59–81. doi: 10.31857/S0005231022070042
7. Prokudina, Yu.A., Oruzhov, E.A. & Moiseev, A.N. (2023) Kontseptsiya arkhitektury programmno kompleksa SimQ [Concept of the SimQ Software Complex Architecture]. *Trudy Tomskogo gosudarstvennogo universiteta. Ser. fiziko-matematicheskaya*. 308. pp. 155–159.
8. Booch, G., Rambo, J. & Yakobson, I. (2022) *Yazyk UML: rukovodstvo pol'zovatelya* [The Unified Modeling Language User Guide]. Translated from English by N. Mukhin. 3rd ed. Moscow: Akademiya AyTi; DMK Press.
9. Prokudina, Yu.A. & Moiseev, A.N. (2024) Prototip prilozheniya dlya poiska raspredeleniya, naiboleye blizkogo k zadannomu empiricheskomu [Application Prototype for Finding the Distribution Closest to a Given Empirical One]. *Trudy Tomskogo gosudarstvennogo universiteta. Ser. fiziko-matematicheskaya*. 309. pp. 199–205.
10. Gelman, E., Hill, D. & Vekhtari, A. (2022) *Regressiya: teoriya i praktika* [Regression and Other Stories]. Moscow: DMK Press.
11. Voskoboynikov, Yu.E. (2022) *Regressiionnyy analiz dannykh v pakete MATHCAD* [Regression Data Analysis in the MATHCAD Package]. St. Petersburg: Lan'.
12. Vishnevskiy, V.M. & Efrosinin, D.V. (2024) *Teoriya ocheredey i mashinnoe obuchenie* [Queueing Theory and Machine Learning]. Moscow: INFRA-M.
13. Sherzer, E., Senderovich, A., Baron, O. & Krass, D. (2022) *Can machines solve general queueing systems?* [Online] Available from: <https://arxiv.org/abs/2202.01729> (Accessed: 26th February 2025).
14. Raheema, M. & Abdullah, A. (2016) Function Approximation using Neuraland Fuzzy Methods. *Communications on Applied Electronics*. 6(3). pp. 35–42.

Информация об авторах:

Прокудина Юлия Андреевна – аспирант, ассистент кафедры программной инженерии Института прикладной математики и компьютерных наук Национального исследовательского Томского государственного университета (Томск, Россия). E-mail: prokudina_ua@mail.ru

Моисеев Александр Николаевич – доцент, доктор физико-математических наук, заведующий кафедрой программной инженерии Института прикладной математики и компьютерных наук Национального исследовательского Томского государственного университета (Томск, Россия). E-mail: moiseev.tsu@gmail.com

Вклад авторов: все авторы сделали эквивалентный вклад в подготовку публикации. Авторы заявляют об отсутствии конфликта интересов.

Information about the authors:

Prokudina Yuliya A. (Post-Graduate Student, Assistant, National Research Tomsk State University, Tomsk, Russian Federation). E-mail: prokudina_ua@mail.ru

Moiseev Alexander N. (Associate Professor, Doctor of Physical and Mathematical Sciences, National Research Tomsk State University, Tomsk, Russian Federation). E-mail: moiseev.tsu@gmail.com

Contribution of the authors: the authors contributed equally to this article. The authors declare no conflicts of interests.

Поступила в редакцию 12.03.2025; принята к публикации 02.12.2025

Received 12.03.2025; accepted for publication 02.12.2025

Научный журнал

**ВЕСТНИК
ТОМСКОГО
ГОСУДАРСТВЕННОГО
УНИВЕРСИТЕТА**

**УПРАВЛЕНИЕ,
ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА
И ИНФОРМАТИКА**

**TOMSK STATE UNIVERSITY
JOURNAL OF CONTROL AND COMPUTER SCIENCE**

2025. № 73

Редактор Е.Г. Шумская
Оригинал-макет Е.Г. Шумской
Редакторы-переводчики: Г.М. Кошкин; В.Н. Горенинцева
Дизайн обложки Л.Д. Кривцовой

Подписано к печати 23.12.2025 г. Формат 60x84¹/₈.
Гарнитура Times. Усл. печ. л. 16,7.
Тираж 250 экз. Заказ № 6601. Цена свободная.

Дата выхода в свет 28.01.2026 г.

Журнал отпечатан на полиграфическом оборудовании
Издательства Томского государственного университета
634050, г. Томск, Ленина, 36
Тел. 8(382-2)–52-98-49; 8(382-2)–52-96-75
Сайт: <http://publish.tsu.ru>; E-mail: rio.tsu@mail.ru