

ВЕСТНИК ТОМСКОГО ГОСУДАРСТВЕННОГО УНИВЕРСИТЕТА

УПРАВЛЕНИЕ,
ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА
И ИНФОРМАТИКА

TOMSK STATE UNIVERSITY
JOURNAL OF CONTROL AND COMPUTER SCIENCE

Научный журнал

2023

№ 64

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

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
Нежелская Людмила Алексеевна – ответственный секретарь, д-р физ.-мат. наук, НИ ТГУ. Тел. +73822529485
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)
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
Пашинская Татьяна Юрьевна – 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 Манс 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 Орзингер (Enco Orzinger) – проф., д-р. философии (PhD), University of Rome, Италия
Паоло Принетто (Paolo Prinetto) – проф. Politechnic Institute, Torino, Италия
Раймонд Убар (Raimund Ubar) – проф., д-р. (PhD), University of Technology Tallinn, Эстония
Нина Владимировна Евтушенко – проф., д-р. техн. наук, ИСП РАН им. Иванникова 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

Ахмедова Ж.Б. О необходимых условиях оптимальности первого порядка в задаче управления, описываемой системой интегро-дифференциальных уравнений дробного порядка	4
Мансимов К.Б., Нагиева И.Ф. Необходимые условия оптимальности первого и второго порядков в одной задаче оптимального управления с нетиповым критерием качества	11
Паршуков А.Н. Метод синтеза модального регулятора в условиях структурных возмущений в объекте управления	21

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ

Дмитриев В.М., Кочергин М.И., Ганджа Т.В. Аспектно-предикатное представление компонентных цепей для адаптивного анализа	30
Паньков А.А., Писарев П.В., Баяндин С.Р. Математическая модель и модальный анализ термоиндикаторного полимерного покрытия	41

ОБРАБОТКА ИНФОРМАЦИИ

Боровской И.Г., Шельмина Е.А., Матолыгин А.А., Ильин Е.П. Нейросетевой анализ данных ультразвуковой дефектоскопии магистральных газопроводов с использованием программно-аппаратной системы ДВУ	50
Воробьев А.В., Воробьева Г.Р. Геоинформационная система динамической пространственной кластеризации распределенных источников данных	61
Лемешко Б.Ю. О проблемах и ошибках применения критериев согласия	74
Кеба А.В., Нежелская Л.А. Сравнение оценок длительности мертвого времени, полученных методом моментов и методом максимального правдоподобия, в обобщенном рекуррентном МАР-потоке событий с двумя состояниями	91
Подкур П.Н., Смоленцев Н.К. Вейвлеты Мейера с коэффициентом масштабирования $N = 12$ для классификации сигналов по высокочастотным признакам	102
Смагин В.И., Ким К.С. Робастная фильтрация для систем с неизвестным входом и интервальной неопределенностью	113

ИНФОРМАТИКА И ПРОГРАММИРОВАНИЕ

Костелей Я.В., Жданов Д.С., Голобокова Е.В., Буреви А.Ш. Автоматизированная система планирования и проведения экспериментов для сбора и анализа медицинских данных в контексте открытых исследований	120
Никитенкова С.П. Применение закона Бенфорда в обнаружении deepfake-изображений	128
Фархадов М.П., Панкратова Е.В., Блинова О.В., Смирнов В.А. Принципы построения формального контекстно-логического конструктора запросов для точного поиска событий в больших массивах гетерогенной информации	138

CONTENTS

CONTROL OF DYNAMICAL SYSTEMS

Ahmadova Z.B. On necessary first-order optimality conditions in a control problem described by a system of fractional-order integro-differential equations	4
Mansimov K.B., Nagiyeva I.F. Necessary conditions for the optimality of the first and second orders in one optimal control problem with atypical quality criteria	11
Parshukov A.N. Method of synthesis of a modal regulator under conditions of structural disturbances in a control object	21

MATHEMATICAL MODELING

Dmitriev V.M., Kochergin M.I., Gandzha T.V. Aspect-predicate representation of component circuits for adaptive analysis	30
Pan'kov A.A., Pisarev P.V., Bayandin S.R. Mathematical model and modal analysis of thermoindicator polymer coating	41

DATA PROCESSING

Borovskoy I.G., Shelmina E.A., Matolygin A.A., Ilin E.P. Neural network analysis of ultrasonic flaw detection data of main gas pipelines using a hardware and software DVU system	50
Vorobev A.V., Vorobeva G.R. Geoinformation system for dynamic spatial clustering of distributed data sources	61
Lemeshko B.Yu. About the problems and errors in applying the goodness of fit tests	74
Keba A.V., Nezhel'skaya L.A. Comparison of dead time estimates obtained by the method of moments and the maximum likelihood method in generalized MAP with two states	91
Podkur P.N., Smolentsev N.R. Meyer wavelets with scaling factor $N = 12$ for classifying signals by high-frequency features	102
Smagin V.I., Kim K.S. Robust filtering in discrete systems with unknown input and interval uncertainty	113

INFORMATICS AND PROGRAMMING

Kosteley Y.V., Zhdanov D.S., Golobokova E.V., Bureev A.Sh. Automated system for planning and conducting experiments for the medical data collection and analysis in the open research context	120
Nikitenkova S.P. Application of Benford's Law in Deepfake Image Detection	128
Farkhadov M.P., Pankratova E.V., Blinova O.V., Smirnov V.A. Principles of constructing a formal context-logical query constructor for accurate search of events in large arrays of heterogeneous information	138

УПРАВЛЕНИЕ ДИНАМИЧЕСКИМИ СИСТЕМАМИ

CONTROL OF DYNAMICAL SYSTEMS

Научная статья

УДК 517.977

doi: 10.17223/19988605/64/1

**О необходимых условиях оптимальности первого порядка в задаче управления,
описываемой системой интегро-дифференциальных уравнений
дробного порядка**

Жаля Билал кызы Ахмедова

*Бакинский государственный университет, Баку, Азербайджан, Институт систем управления
Министерства науки и образования Азербайджана, Баку, Азербайджан, akja@rambler.ru*

Аннотация. Рассматривается задача оптимального управления для систем интегро-дифференциальных уравнений типа Вольтерра с дробной производной Капуто. Функционал качества является функционалом терминального типа. Применяя аналог модифицированного метода приращений, установлено необходимое условие оптимальности первого порядка в форме принципа максимума Л.С. Понтрягина.

Ключевые слова: задача оптимального управления; функция Гамильтона–Понтрягина; принцип максимума; необходимое условие оптимальности; дробная производная Капуто.

Для цитирования: Ахмедова Ж.Б. О необходимых условиях оптимальности первого порядка в задаче управления, описываемой системой интегро-дифференциальных уравнений дробного порядка // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2023. № 64. С. 4–10. doi: 10.17223/19988605/64/1

Original article

doi: 10.17223/19988605/64/1

On necessary first-order optimality conditions in a control problem described by a system of fractional-order integro-differential equations

Zhala B. Ahmadova

*Baku State University, Baku, Azerbaijan, Institute of Control Systems of the Ministry of Science
and Education of Azerbaijan, Baku, Azerbaijan, akja@rambler.ru*

Abstract. The optimal control problem for systems of Volterra-type integro-differential equations with Caputo fractional derivative is considered. The quality functional is a functional of terminal type. Using an analogue of the modified increment method, a necessary first-order optimality condition is established in the form of the L.S. Pontryagin.

Keywords: optimal control problem; Hamilton-Pontryagin function; maximum principle; necessary optimality condition; fractional Caputo derivative.

For citation: Ahmadova, Z.B. (2023) On necessary first-order optimality conditions in a control problem described by a system of fractional-order integro-differential equations. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naja tehnika i informatika – Tomsk State University Journal of Control and Computer Science*. 64. pp. 4–10. doi: 10.17223/19988605/64/1

Введение

Принцип максимума Л.С. Понтрягина, являясь одним из основных результатов теории необходимых условий оптимальности первого порядка, доказан для различных задач оптимального управления, описываемых обыкновенными дифференциальными уравнениями (см., напр.: [1–3]). В последние годы стали интенсивно изучаться задачи оптимального управления объектами, описываемые различными дифференциальными уравнениями с дробными производными (см., напр., [4–8]).

В [4] исследуется задача оптимального управления динамическими системами дробного порядка с использованием метода моментов.

В [5], применяя классическую теорию управления к дробно-дифференциальной системе в ограниченной области, рассмотрена дробная задача оптимального управления для дифференциальной системы с запаздыванием. Дробная производная по времени рассматривается в смысле Римана–Лиувилля. Изучены существование и единственность решения системы дробных дифференциалов с запаздыванием в гильбертовом пространстве. В работе также показано, что рассматриваемая задача оптимального управления имеет единственное решение.

Точное представление многих динамических систем приводит к набору дробных дифференциальных уравнений (ДДУ). В статье [6] представлены общая формулировка и схема решения класса дробных задач оптимального управления (ДЗОУ). Дробная производная используется в смысле Римана–Лиувилля. Показатель эффективности ДДУ рассматривается как функция состояния и управляющих переменных, а динамические ограничения определяются набором ДДУ. Вариационное исчисление, множители Лагранжа и формулы дробного интегрирования по частям используются в задачах ДДУ для получения уравнений Эйлера–Лагранжа.

В работе [7] рассматривается задача оптимального управления для линейной стационарной динамической системы дробного порядка. Исследованы такие условия, при которых проблема моментов может быть поставлена и является разрешимой. Рассмотрены частные случаи (одномерная линейная стационарная система, двойной интегратор и маятник), для которых получены решения задачи и исследованы вопросы качественной динамики. Продемонстрирована возможность постановки и исследования задачи оптимального управления в форме обобщенной проблемы моментов для системы, описываемой уравнением переноса с производной дробного порядка по времени.

В [8] рассматривается новая общая формулировка дробных задач оптимального управления, показатель качества которых представлен в форме дробного интеграла, а динамика задается системой дробных дифференциальных уравнений в смысле Капуто. Использован новый подход для доказательства необходимых условий оптимальности в виде принципа максимума Понтрягина для дробно нелинейных задач оптимального управления.

В [9] исследована задача оптимального управления, описываемая интегро-дифференциальными уравнениями типа Вольтерра с дробной производной Капуто. В рассматриваемой задаче функционал качества является функционалом терминального типа. Установлен аналог принципа максимума Понтрягина.

В [10] рассматривается класс нелинейных фрактально-дробных задач оптимального управления в смысле Атанганы–Римана–Лиувилля с невырожденным ядром Миттаг-Леффлера. Предложен численный метод, основанный на обобщенных вейвлетах Лукаса и методе Ритца. В работе также показано преимущество предлагаемого метода на численных примерах.

Отметим также, что в работе [11] в отличие от нашей работы рассмотрена дискретная задача оптимального управления, в которой доказан дискретный аналог принципа максимума Л.С. Понтрягина.

В настоящей работе рассматриваются процессы, описываемые обыкновенными интегро-дифференциальным уравнением типа Вольтерра дробного порядка. С помощью схемы, основанной на методе приращений, доказан аналог принципа максимума Л.С. Понтрягина.

1. Постановка задачи

Пусть управляемый процесс описывается дробной системой интегро-дифференциальных уравнений типа Вольтерра

$${}^C D_t^\alpha x(t) = \int_{t_0}^t f(t, \tau, x(\tau), u(\tau)) d\tau, \quad t \in T = [t_0, t_1] \quad (1)$$

с начальным условием

$$x(t_0) = x_0, \quad (2)$$

где

$${}^C D_t^\alpha x(t) = \frac{1}{\Gamma(1-\alpha)} \int_{t_0}^t \frac{x'(\tau)}{(t-\tau)^\alpha} d\tau, \quad \alpha \in [0, 1],$$

левая дробная производная Капуто [8, 12, 13], α – показывает степень дробной производной, C – означает дробную производную Капуто, t_0 – начальная точка заданного отрезка, на котором задан управляемый непрерывный процесс, $u(t)$ – r -мерная кусочно-непрерывная (с конечным числом точек разрыва первого рода) вектор-функция управляющих воздействий со значениями из заданного непустого и ограниченного множества, т.е.

$$u(t) \in U \subset R^r, \quad t \in [t_0, t_1] \quad (3)$$

(допустимое управление), x_0 – заданный постоянный вектор, $f(t, \tau, x, u)$ – заданная n -мерная вектор-функция, непрерывная по совокупности переменных вместе с частными производными по x .

Рассмотрим задачу о минимизации терминального функционала

$$S(u) = \varphi(x(t_1)) \quad (4)$$

при ограничениях (1)–(3). Здесь $\varphi(x)$ – заданная непрерывно-дифференцируемая скалярная функция.

Допустимое управление $u(t)$, доставляющее минимальное значение функционалу (4) при ограничениях (1)–(3), назовем оптимальным управлением, а соответствующий процесс $(u(t), x(t))$ – оптимальным процессом.

2. Формула приращения функционала качества

Пусть $(u(t), x(t))$ и $(\bar{u}(t) = u(t) + \Delta u(t), \bar{x}(t) = x(t) + \Delta x(t))$ – два допустимых процесса.

Запишем приращение функционала качества

$$\Delta S(u) = S(\bar{u}) - S(u) = \varphi(\bar{x}(t_1)) - \varphi(x(t_1)). \quad (5)$$

Ясно, что приращение $\Delta x(t)$ траектории $x(t)$ является решением следующей задачи:

$${}^C D_t^\alpha \Delta x(t) = \int_{t_0}^t [f(t, \tau, \bar{x}(\tau), \bar{u}(\tau)) - f(t, \tau, x(\tau), u(\tau))] d\tau, \quad (6)$$

$$\Delta x(t_0) = 0. \quad (7)$$

Пусть $\psi = \psi(t)$ – пока произвольная n -мерная вектор-функция. Умножая обе стороны уравнения (6) скалярно на $\psi(t)$ и интегрируя полученное соотношение по t на интервале от t_0 до t_1 , получим

$$\int_{t_0}^{t_1} \psi'(t) {}^C D_t^\alpha \Delta x(t) dt = \int_{t_0}^{t_1} \psi'(t) \left[\int_{t_0}^t [f(t, \tau, \bar{x}(\tau), \bar{u}(\tau)) - f(t, \tau, x(\tau), u(\tau))] d\tau \right] dt. \quad (8)$$

По формуле интегрирования по частям для дробных производных (см., напр.: [12, 13]) и при начальном условии (7) имеем

$$\int_{t_0}^{t_1} \psi'(t) {}^C D_t^\alpha \Delta x(t) dt = \int_{t_0}^{t_1} {}^C D_t^\alpha \psi'(t) \Delta x(t) dt + {}_t I_{t_1}^{1-\alpha} \psi(t) \Delta x(t) \Big|_{t_0}^{t_1} =$$

$$\begin{aligned}
 &= \int_{t_0}^{t_1} {}^C D_{t_1}^\alpha \psi'(t) \Delta x(t) dt + {}_t I_{t_1}^{1-\alpha} \psi'(t_1) \Delta x(t_1) + {}_t I_{t_1}^{1-\alpha} \psi'(t_0) \Delta x(t_0) = \\
 &= \int_{t_0}^{t_1} {}^C D_{t_1}^\alpha \psi'(t) \Delta x(t) dt + {}_t I_{t_1}^{1-\alpha} \psi'(t_1) \Delta x(t_1),
 \end{aligned} \tag{9}$$

где ${}_t I_{t_1}^{1-\alpha}$ – правый интеграл Римана-Лиувилля [12, 13].

Используя формулу Фубини, можно записать

$$\begin{aligned}
 &\int_{t_0}^{t_1} \psi'(t) \left[\int_{t_0}^t [f(t, \tau, \bar{x}(\tau), \bar{u}(\tau)) - f(t, \tau, x(\tau), u(\tau))] d\tau \right] dt = \\
 &= \int_{t_0}^{t_1} \left[\int_t^{t_1} \psi'(\tau) [f(\tau, t, \bar{x}(t), \bar{u}(t)) - f(\tau, t, x(t), u(t))] d\tau \right] dt.
 \end{aligned} \tag{10}$$

Поэтому формулу приращения (5) функционала качества (4) можем записать следующим образом:

$$\begin{aligned}
 \Delta S(u) &= \varphi(\bar{x}(t_1)) - \varphi(x(t_1)) + \int_{t_0}^{t_1} {}^C D_{t_1}^\alpha \psi'(t) \Delta x(t) dt + {}_t I_{t_1}^{1-\alpha} \psi'(t_1) \Delta x(t_1) - \\
 &- \int_{t_0}^{t_1} \left[\int_t^{t_1} \psi'(\tau) [f(\tau, t, \bar{x}(t), \bar{u}(t)) - f(\tau, t, x(t), u(t))] d\tau \right] dt.
 \end{aligned} \tag{11}$$

Введем аналог функции Гамильтона–Понтрягина:

$$H(t, x(t), u(t), \psi(t)) = \int_t^{t_1} \psi'(\tau) f(\tau, t, x(t), u(t)) d\tau.$$

Тогда формулу (11) можно написать в виде:

$$\begin{aligned}
 \Delta S(u) &= \varphi(\bar{x}(t_1)) - \varphi(x(t_1)) + \int_{t_0}^{t_1} {}^C D_{t_1}^\alpha \psi'(t) \Delta x(t) dt + {}_t I_{t_1}^{1-\alpha} \psi'(t_1) \Delta x(t_1) - \\
 &- \int_{t_0}^{t_1} [H(t, \bar{x}(t), \bar{u}(t), \psi(t)) - H(t, x(t), u(t), \psi(t))] dt.
 \end{aligned} \tag{12}$$

Используя формулу Тейлора (см. напр.: [13, 14]), из (12) получим

$$\begin{aligned}
 \Delta S(u) &= \frac{\partial \varphi'(x(t_1))}{\partial x} \Delta x(t_1) + o_1(\|\Delta x(t_1)\|) + {}_t I_{t_1}^{1-\alpha} \psi'(t_1) \Delta x(t_1) + \int_{t_0}^{t_1} {}^C D_{t_1}^\alpha \psi'(t) \Delta x(t) dt - \\
 &- \int_{t_0}^{t_1} [H(t, x(t), \bar{u}(t), \psi(t)) - H(t, x(t), u(t), \psi(t))] dt - \int_{t_0}^{t_1} \frac{\partial H(t, x(t), u(t), \psi(t))}{\partial x} \Delta x(t) dt - \\
 &- \int_{t_0}^{t_1} \left[\frac{\partial H(t, x(t), \bar{u}(t), \psi(t))}{\partial x} - \frac{\partial H(t, x(t), u(t), \psi(t))}{\partial x} \right] \Delta x(t) dt - \int_{t_0}^{t_1} o_2(\|\Delta x(t)\|) dt.
 \end{aligned} \tag{13}$$

Здесь $\|\alpha\|$ есть норма вектора $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_n)'$, определяемая формулой $\|\alpha\| = \sum_{i=1}^n |\alpha_i|$, а $o(\alpha)$ – вели-

чина, имеющая более высокий порядок малости, чем α , т.е. $\frac{o(\alpha)}{\alpha} \rightarrow 0$ при $\alpha \rightarrow 0$.

Пусть вектор-функция $\psi(t)$ является решением следующей дробной системы интегродифференциальных уравнений:

$${}_t^C D_{t_1}^\alpha \psi(t) = \frac{\partial H(t, x(t), u(t), \psi(t))}{\partial x}, \tag{14}$$

с начальным условием

$${}_t I_{t_1}^{1-\alpha} \psi'(t_1) = -\frac{\partial \varphi(x(t_1))}{\partial x}. \tag{15}$$

Задачу (14), (15) назовем сопряженной системой для рассматриваемой задачи. Учитывая формулы (14) и (15), из (13) получим, что

$$\begin{aligned} \Delta S(u) = & - \int_{t_0}^{t_1} [H(t, x(t), \bar{u}(t), \psi(t)) - H(t, x(t), u(t), \psi(t))] dt + o_1(\|\Delta x(t_1)\|) - \\ & - \int_{t_0}^{t_1} \left[\frac{\partial H(t, x(t), \bar{u}(t), \psi(t))}{\partial x} - \frac{\partial H(t, x(t), u(t), \psi(t))}{\partial x} \right] \Delta x(t) dt - \int_{t_0}^{t_1} o_2(\|\Delta x(t)\|) dt. \end{aligned} \quad (16)$$

3. Оценка нормы приращения траектории

Полученная формула приращения (16) позволяет доказать необходимое условие оптимальности. Для этого нам понадобится оценка нормы приращения $\Delta x(t)$ траектории $x(t)$.

Из (6), (7), используя определение дробной производной Капуто, получаем, что

$$\begin{aligned} \Delta x(t) = & \frac{1}{\Gamma(\alpha)} \int_{t_0}^t \int_{t_0}^{\tau} \frac{f(\tau, s, \bar{x}(s), \bar{u}(s)) - f(\tau, s, x(s), u(s))}{(s - \gamma)^{1-\alpha}} ds d\tau = \\ = & \frac{1}{\Gamma(\alpha)} \int_{t_0}^t \int_{t_0}^{\tau} (s - \gamma)^{\alpha-1} [f(\tau, s, \bar{x}(s), \bar{u}(s)) - f(\tau, s, x(s), \bar{u}(s)) + \\ & + f(\tau, s, x(s), \bar{u}(s)) - f(\tau, s, x(s), u(s))] ds d\tau = \\ = & \frac{1}{\Gamma(\alpha)} \int_{t_0}^t \int_{t_0}^{\tau} (s - \gamma)^{\alpha-1} [f(\tau, s, \bar{x}(s), \bar{u}(s)) - f(\tau, s, x(s), \bar{u}(s))] ds d\tau + \\ & + \frac{1}{\Gamma(\alpha)} \int_{t_0}^t \int_{t_0}^{\tau} (s - \gamma)^{\alpha-1} \Delta_{\bar{u}} f(\tau, s, x(s), u(s)) ds d\tau, \end{aligned} \quad (17)$$

где по определению

$$\Delta_{\bar{u}} f(\tau, s, x(s), u(s)) = f(\tau, s, x(s), \bar{u}(s)) - f(\tau, s, x(s), u(s)).$$

Так как в силу сделанных предположений функция f удовлетворяет условию Липшица по x , то, переходя к норме и используя условие Липшица, получаем, что

$$\|\Delta x(t)\| \leq L_1 \frac{1}{\Gamma(\alpha)} \int_{t_0}^t \int_{t_0}^{\tau} (\tau - \gamma)^{\alpha-1} \|\Delta x(s)\| ds d\tau + \frac{1}{\Gamma(\alpha)} \int_{t_0}^t \int_{t_0}^{\tau} (\tau - \gamma)^{\alpha-1} \Delta_{\bar{u}} f(\tau, s, x(s), u(s)) ds d\tau, \quad (18)$$

где $L_1 = \text{const} > 0$ – некоторая постоянная.

Применяя аналог формулы Гронуолла–Беллмана [15], получим

$$\|\Delta x(t)\| \leq L_2 \frac{1}{\Gamma(\alpha)} \int_{t_0}^t \int_{t_0}^{\tau} (\tau - \gamma)^{\alpha-1} \Delta_{\bar{u}} f(\tau, x(\tau), u(\tau)) ds d\tau. \quad (19)$$

где $L_2 = \text{const} > 0$ – некоторая постоянная.

Пусть $\theta \in [t_0, t_1)$ – произвольная точка непрерывности управляющей функции $u(t)$, а $\varepsilon > 0$ – произвольное малое число, такое что $\theta + \varepsilon < t_1$, $v \in U$ – произвольный вектор.

Специальное приращение управления $u(t)$ определим по формуле

$$\Delta u_{\varepsilon}(t) = \begin{cases} v - u(t), & t \in [\theta, \theta + \varepsilon), \\ 0, & t \in [t_0, t_1] \setminus [\theta, \theta + \varepsilon). \end{cases} \quad (20)$$

Через $\Delta x_{\varepsilon}(t)$ обозначим специальное приращение траектории $x(t)$, соответствующее специальному приращению игольчатой вариации (20) управления $u(t)$.

Из неравенства (19) следует, что

$$\|\Delta x_{\varepsilon}(t)\| \leq L_3 \varepsilon, \quad (21)$$

где $L_3 = \text{const} > 0$ – некоторая постоянная.

Принимая во внимание оценку (21) и формулу (20), из формулы приращения (16) на основании теоремы о среднем получаем справедливость разложения

$$\begin{aligned}\Delta S_{\varepsilon}(u) &= S(u + \Delta u_{\varepsilon}) - S(u) = - \int_{t_0}^{t_1} [H(t, x(t), u(t) + \Delta u_{\varepsilon}(t), \psi(t)) - H(t, x(t), u(t), \psi(t))] dt = \\ &= - \int_{\theta}^{\theta + \varepsilon} [H(t, x(t), v, \psi(t)) - H(t, x(t), u(t), \psi(t))] dt + o(\varepsilon) = \\ &= -\varepsilon [H(\theta, x(\theta), v, \psi(\theta)) - H(\theta, x(\theta), u(\theta), \psi(\theta))] + o(\varepsilon).\end{aligned}\quad (22)$$

Если предполагать, что $u(t)$ – оптимальное управление, то из разложения (22) следует, что

$$H(\theta, x(\theta), v, \psi(\theta)) - H(\theta, x(\theta), u(\theta), \psi(\theta)) \geq 0,$$

т.е.

$$\max_{v \in U} H(\theta, x(\theta), v, \psi(\theta)) = H(\theta, x(\theta), u(\theta), \psi(\theta)) \geq 0. \quad (23)$$

Следовательно, имеет место следующая теорема.

Теорема (принцип максимума Понтрягина). Для оптимальности допустимого управления $u(t)$ необходимо, чтобы условие максимума (23) выполнялось для всех $v \in U$, $\theta \in [t_0, t_1]$.

Доказанная теорема является аналогом принципа максимума Понтрягина в рассматриваемой задаче.

Заключение

В работе рассмотрена задача оптимального управления для объекта, описываемого интегро-дифференциальным уравнением типа Вольтерра с дробной производной Капуто. Минимизируемый критерий качества является функционалом терминального типа. Область управления – произвольное, непустое и ограниченное множество, а управляющая функция является кусочно-непрерывной (с конечным числом точек разрыва первого порядка) вектор-функцией. Используя введенную сопряженную систему, получена формула для приращения функционала качества. Установлен соответствующий аналог принципа максимума Понтрягина.

Список источников

1. Васильев Ф.П. Методы оптимизации. М. : Изд-во МЦНМО, 2011. Кн. 2. 434 с.
2. Габасов Р., Кириллова Ф.М., Альсевич В.В. Методы оптимизации. Минск : Четыре четверти, 2011, 472 с.
3. Милютин А.А., Дмитрук А.В., Осмоловский Н.П. Принцип максимума в оптимальном управлении. М. : МГУ, 2004, 168 с.
4. Постнов С.С. Исследование задач оптимального управления динамическими системами дробного порядка методом моментов : автореферат дис. ... канд. физ.-мат. наук. М. : ИПУ РАН, 2015. 26 с.
5. Bahaa G.M. Fractional optimal control problem for differential system with delay argument // Advances in Difference Equations. 2017. № 1. P. 32–51.
6. Agrawal O.P. A General Formulation and Solution Scheme for Fractional Optimal Control Problems // Nonlinear Dynamics. 2004. V. 38. P. 323–337.
7. Кубышкин В.А., Постнов С.С. Оптимальное управление линейными динамическими системами нецелого порядка // XII Всерос. совещание по проблемам управления – ВСПУ 2014. Москва, 16–19 июня. М., 2014. С. 2562–2573.
8. Ali H.M., Pereira F.L., Gama S.M.A. A new approach to the Pontryagin maximum principle for nonlinear fractional optimal control problems // Mathematics. Optimization and Control. arXiv:1503.07720. 2015. pp. 1–17. doi: 10.1002/mma.3811
9. Мансимов К.Б., Ахмедова Ж.Б. Аналог принципа максимума Понтрягина в задаче оптимального управления системой дифференциальных уравнений с дробной производной Капуто и многоточечным критерием качества // Вестник Пермского университета. Математика. Механика. Информатика. 2022. Вып. 3 (58). С. 5–10.
10. Sabermahani S., Ordokhani Y., Rahimkhani P. Application of generalized Lucas wavelet method for solving nonlinear fractal-fractional optimal control problems // Chaos, Solitons & Fractals. 2023. V. 170. Art. 113348. doi: 10.1016/j.chaos.2023.113348
11. 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.
12. Samko S.G., Kilbas A.A., Marichev O.I. Fractional integrals and derivatives: Theory and applications. Yverdon ; Philadelphia, Pa : Gordon and Breach Science publishers. 1993. xxxvi, 976 p.

13. Usero D. Fractional Taylor series for Caputo fractional derivatives Construction of numerical schemes : preprint // Universidad Complutense. Informática. Madrid, 2008. 1–18 p. URL: http://www.fdi.ucm.es/profesor/lvazquez/calfrac/docs/paper_usero.pdf
14. Odibat Z., Shawagfeh N. Generalized Taylor's formula // Appl. Math. Comput. 2007. V. 186. P. 286–293.
15. Lin S.Y. Generalized Gronwall inequalities and their applications to fractional differential equations // Journal of Inequalities and Applications. 2013. Art. 549. doi: 10.1186/1029-242X-2013-549

References

1. Vasiliev, F.P. (2011) *Metody optimizatsii* [Optimization Methods]. Vol. 2. Moscow: MTsNMO.
2. Gabasov, R., Kirillova, F.M. & Alseovich, V.V. (2011) *Metody optimizatsii* [Optimization Methods]. Minsk: Chetyre chetverti.
3. Milyutin, A.A., Dmitruk, A.V. & Osmolovsky, N.P. (2004) *Printsip maksimuma v optimal'nom upravlenii* [The maximum principle in optimal control]. Moscow: MSU.
4. Postnov, S.S. (2015) *Issledovanie zadach optimal'nogo upravleniya dinamicheskimi sistemami drobnogo poryadka metodom momentov* [Investigation of problems of optimal of dynamic systems of fractional order by the method of moments]. Abstract of Physics and Mathematics Cand. Diss. Moscow.
5. Bahaa, G.M. (2017) Fractional optimal control problem for differential system with delay argument. *Advances in Difference Equations*. 1. pp. 32–51. DOI: 10.1186/s13662-017-1121-6
6. Agrawal, O.P. (2004) A General Formulation and Solution Scheme for Fractional Optimal Control Problems. *Nonlinear Dynamics*. 38. pp. 323–337. DOI: 10.1007/s11071-004-3764-6
7. Kubyshev, V.A. & Postnov, S.S. (2014) Optimal'noe upravlenie lineynymi dinamicheskimi sistemami netselogo poryadka [Optimal control of linear dynamical systems of non-integer order]. *XII Vseros. soveshchanie po problemam upravleniya – VSPU 2014* [The 12th All-Russian meeting on control problems. VSPU–2014]. Moscow, June 16-19. pp. 2562–2573.
8. Ali, H.M., Pereira, F.L. & Gama, S.M.A. (2015) A new approach to the Pontryagin maximum principle for nonlinear fractional optimal control problems. *Mathematics. Optimization and Control*. arXiv:1503.07720. pp. 1–17. <https://doi.org/10.1002/mma.3811>
9. Mansimov, K.B. & Akhmedova, Zh.B. (2022) An analogue of the Pontryagin maximum principle in the optimal control problem for a system of differential equations with a fractional Caputo derivative and with a multipoint quality criterion. *Vestnik Permskogo universiteta. Matematika. Mekhanika. Informatika – Perm University Bulletin. Mathematics. Mechanics. Computer science*. 3(58). pp. 5–10.
10. Sabermahani, S., Ordokhani, Y. & Rahimkhani, P. (2023) Application of generalized Lucas wavelet method for solving nonlinear fractal-fractional optimal control problems. *Chaos, Solitons & Fractals*. 170. Art. 113348. DOI: 10.1016/j.chaos.2023.113348
11. Alieva, S.T. (2021) The Pontryagin maximum principle for nonlinear fractional order difference equations. *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
12. Samko, S.G., Kilbas, A.A. & Marichev, O.I. (1993) *Fractional integrals and derivatives: Theory and applications*. Yverdon, Switzerland: Gordon and Breach Science.
13. Usero, D. (2008) *Fractional Taylor series for Caputo fractional derivatives Construction of numerical schemes*. Preprint. Madrid. Spain. pp. 1–18. [Online] Available from: http://www.fdi.ucm.es/profesor/lvazquez/calfrac/docs/paper_usero.pdf
14. Odibat, Z. & Shawagfeh, N. (2007) Generalized Taylor's formula. *Applied Mathematics and Computation*. 186. pp. 286–293. DOI: 10.1016/j.amc.2006.07.102
15. Lin, S.Y. (2013) Generalized Gronwall inequalities and their applications to fractional differential equations. *Journal of Inequalities and Applications*. Art. 549. DOI: 10.1186/1029-242X-2013-549

Информация об авторе:

Ахмедова Жалля Билал кызы – кандидат физико-математических наук, доцент кафедры «Математическая кибернетика» Бакинского государственного университета (Баку, Азербайджан); Институт систем управления Министерства науки и образования Азербайджана (Баку, Азербайджан). E-mail: akja@rambler.ru

Автор заявляет об отсутствии конфликта интересов.

Information about the author:

Ahmadova Zhalla B. (Candidate of Physical and Mathematical Sciences, Associate Professor of the Department of Mathematical Cybernetics, Baku State University, Baku, Azerbaijan; Institute of Control Systems of the Ministry of Science and Education of Azerbaijan, Baku, Azerbaijan). E-mail: akja@rambler.ru

The author declares no conflicts of interests.

Поступила в редакцию 30.03.2023; принята к публикации 04.09.2023

Received 30.03.2023; accepted for publication 04.09.2023

Научная статья

УДК 519.3. 62-50

doi: 10.17223/19988605/64/2

Необходимые условия оптимальности первого и второго порядков в одной задаче оптимального управления с нетиповым критерием качества

Камиль Байрамали оглы Мансимов¹, Илаха Фирдовис кызы Нагиева²

^{1,2} Бакинский государственный университет, Баку, Азербайджан

² Институт систем управления Министерства науки и образования Азербайджана, Баку, Азербайджан

¹ kamilbmansimov@gmail.com

² ilaha_nagiyeva@gmail.com

Аннотация. Рассматривается задача оптимального управления, описываемая системой интегро-дифференциальных уравнений типа Вольтерра с нетиповым критерием качества. При предположении открытости области управления вычислены первая и вторая вариации функционала качества. Доказан аналог уравнения Эйлера. Из условия неотрицательности второй вариации функционала качества получены различные конструктивно проверяемые необходимые условия оптимальности второго порядка. Отдельно изучен случай особых, в классическом смысле, управлений.

Ключевые слова: нетиповая задача оптимального управления; задача типа Н.Н. Моисеева; допустимое управление; интегро-дифференциальное уравнение типа Вольтерра; классическая экстремаль; аналог уравнения Эйлера; особые, в классическом смысле, управления; оптимальное управление.

Для цитирования: Мансимов К.Б., Нагиева И.Ф. Необходимые условия оптимальности первого и второго порядков в одной задаче оптимального управления с нетиповым критерием качества // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2023. № 64. С. 11–20. doi: 10.17223/19988605/64/2

Original article

doi: 10.17223/19988605/64/2

Necessary conditions for the optimality of the first and second orders in one optimal control problem with atypical quality criteria

Kamil B. Mansimov¹, Ilaha F. Nagiyeva²

^{1,2} Baku State University, Baku, Azerbaijan

² Institute of Management Systems of the Ministry of Science and Education of Azerbaijan, Baku, Azerbaijan

¹ kamilbmansimov@gmail.com

² ilaha_nagiyeva@gmail.com

Abstract. An optimal control problem described by a system of Volterra type integro-differential equations with an atypical quality criterion is considered. Assuming the openness of the control domain the first and second variations of the quality functional are calculated. An analogue of the Euler equation is proved. From the non-negativity condition of the second variation of the quality functional, various constructively verifiable necessary second-order optimality conditions are obtained. The case of singular controls in the classical sense is studied separately.

Keywords: an atypical optimal control problem; N.N. Moiseev type problem, admissible control; an integro-differential equation of the Volterra-type; classical extremal; an analogue of the Euler equation; singular in the classical sense of controls; optimal control.

For citation: Mansimov, K.B., Nagiyeva, I.F. (2023) Necessary conditions for the optimality of the first and second orders in one optimal control problem with atypical quality criteria. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 64. pp. 11–20. doi: 10.17223/19988605/64/2

Введение

К настоящему времени для различных задач оптимального управления с типовыми критериями качества (задачи управления типа Лагранжа, Майера, Больца) установлен ряд необходимых условий оптимальности типа принципа максимума Понтрягина и их следствий. Обзоры соответствующих работ представлены, например, в [1–5] и др.

В работе [6. С. 369–414] Н.Н. Моисеев при исследовании задач синтеза, сводящихся к задачам вариационного исчисления и оптимального управления, пришел к необходимости исследования задач оптимального управления объектами, описываемыми обыкновенными дифференциальными уравнениями с критерием качества в виде двойного интеграла (см.: [6. С. 393–414]).

Подобные задачи оптимального управления (задачи типа Моисеева) являются задачами оптимального управления с нетиповым критерием качества. В работе [6] Н.Н. Моисеев для такой задачи доказал необходимое условие оптимальности в форме принципа максимума Понтрягина.

Позднее в работе [7] также рассмотрена задача оптимального управления для систем, описываемых обыкновенными дифференциальными уравнениями с нетиповым критерием качества, в которой исследован случай вырождения необходимого условия оптимальности первого порядка (особый случай [8]).

В предлагаемой работе рассматривается задача, аналогичная [6, 7], в предположении, что процесс описывается системой интегро-дифференциальных уравнений типа Вольтерра.

В случае открытого множества возможных значений управлений получены необходимые условия оптимальности первого и второго порядков, носящие конструктивный характер. Отдельно изучен случай вырождения аналога условия Лежандра–Клебша и доказано необходимое условие оптимальности особых, в классическом смысле, управлений [8].

1. Постановка задачи

Предположим, что управляемый процесс на заданном отрезке времени $T = [t_0, t_1]$ описывается системой интегро-дифференциальных уравнений типа Вольтерра

$$\dot{x} = f(t, x, u) + \int_{t_0}^t K(t, \tau, x(\tau), u(\tau)) d\tau, \quad t \in T \quad (1)$$

с начальным условием

$$x(t_0) = x_0, \quad (2)$$

Здесь $f(t, x, u)$ и $K(t, \tau, x, u)$ – n -мерные заданные вектор-функции, непрерывные по совокупности переменных вместе с частными производными по (x, u) до второго порядка включительно, x_0 – заданный постоянный вектор, $u(t)$ r -мерный, кусочно-непрерывный (с конечным числом точек разрыва первого рода) вектор управляющих воздействий со значениями из заданного непустого, ограниченного и открытого множества $U \subset R^r$, т.е.

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

Каждую управляющую функцию с вышеприведенными свойствами назовем допустимым управлением.

Предполагается, что каждому допустимому управлению $u(t)$ соответствует единственное, непрерывное и кусочно-гладкое (в смысле [5]) решение $x(t)$ задачи Коши (1), (2).

Пусть $\varphi(t)$ – заданная дважды непрерывно дифференцируемая скалярная функция, $F(t, s, a, b)$ – заданная скалярная функция, непрерывная по совокупности переменных вместе с частными производными по (a, b) до второго порядка включительно.

Рассмотрим задачу о минимуме функционала

$$S(u) = \varphi(x(t_1)) + \int_{t_0}^{t_1} \int_{t_0}^t F(t, s, x(t), x(s)) ds dt, \quad (4)$$

определенного на решениях задачи Коши (1), (2), порожденных всевозможными допустимыми управлениями.

Допустимое управление $u(t)$ доставляющее минимальное значение функционалу (4), при ограничениях (1)–(3), назовем оптимальным управлением, а соответствующий процесс $(u(t), x(t))$ – оптимальным процессом.

Целью работы является получение необходимых условий оптимальности первого и второго порядков в рассматриваемой задаче.

2. Вычисление первой и второй вариаций функционала качества

Пусть $(u(t), x(t))$ и $(\bar{u}(t) = u(t) + \Delta u(t), \bar{x}(t) = x(t) + \Delta x(t))$ – два допустимых процесса. Запишем приращение функционала качества (4)

$$\Delta S(u) = S(\bar{u}) - S(u) = \varphi(\bar{x}(t_1)) - \varphi(x(t_1)) + \int_{t_0}^{t_1} (F(t, s, \bar{x}(t), \bar{x}(s)) - F(t, s, x(t), x(s))) ds dt. \quad (5)$$

Из введенных обозначений ясно, что приращение $\Delta x(t)$ траектории $x(t)$ является решением задачи

$$\Delta \dot{x}(t) = f(t, \bar{x}(t), \bar{u}(t)) - f(t, x(t), u(t)) + \int_{t_0}^t [K(t, \tau, \bar{x}(\tau), \bar{u}(\tau)) - K(t, \tau, x(\tau), u(\tau))] d\tau, \quad (6)$$

$$\Delta x(t_0) = 0. \quad (7)$$

Через $\psi(t)$ обозначим пока произвольную n -мерную вектор-функцию. Из тождества (6) получаем, что

$$\begin{aligned} \int_{t_0}^{t_1} \psi'(t) \Delta \dot{x}(t) dt &= \int_{t_0}^{t_1} \psi'(t) (f(t, \bar{x}(t), \bar{u}(t)) - f(t, x(t), u(t))) dt + \\ &+ \int_{t_0}^{t_1} \psi'(t) \left[\int_{t_0}^t [K(t, \tau, \bar{x}(\tau), \bar{u}(\tau)) - K(t, \tau, x(\tau), u(\tau))] d\tau \right] dt. \end{aligned} \quad (8)$$

Учитывая начальное условие (7) и применяя формулу Ньютона–Лейбница, получим, что

$$\int_{t_0}^{t_1} \psi'(t) \Delta \dot{x}(t) dt = \psi'(t_1) \Delta x(t_1) - \int_{t_0}^{t_1} \dot{\psi}'(t) \Delta x(t) dt. \quad (9)$$

Далее на основе формулы Фубини (см., напр.: [4, 9]) имеем

$$\begin{aligned} \int_{t_0}^{t_1} \psi'(t) \left[\int_{t_0}^t [K(t, \tau, \bar{x}(\tau), \bar{u}(\tau)) - K(t, \tau, x(\tau), u(\tau))] d\tau \right] dt &= \\ = \int_{t_0}^{t_1} \left[\int_t^{t_1} \psi'(t) [K(\tau, t, \bar{x}(t), \bar{u}(t)) - K(\tau, t, x(t), u(t))] d\tau \right] dt. \end{aligned} \quad (10)$$

Принимая во внимания тождества (8)–(10) в формуле приращения (5) функционала (4), получим, что

$$\begin{aligned} \Delta S(u) = S(\bar{u}) - S(u) &= \varphi(\bar{x}(t_1)) - \varphi(x(t_1)) + \int_{t_0}^{t_1} \int_{t_0}^{t_1} (F(t, s, \bar{x}(t), \bar{x}(s)) - F(t, s, x(t), x(s))) ds dt + \\ &+ \psi'(t_1) \Delta x(t_1) - \int_{t_0}^{t_1} \dot{\psi}'(t) \Delta x(t) dt - \int_{t_0}^{t_1} \psi'(t) (f(t, \bar{x}(t), \bar{u}(t)) - f(t, x(t), u(t))) dt - \\ &- \int_{t_0}^{t_1} \left[\int_t^{t_1} \psi'(t) [K(\tau, t, \bar{x}(t), \bar{u}(t)) - K(\tau, t, x(t), u(t))] d\tau \right] dt. \end{aligned} \quad (11)$$

Введем аналог функции Гамильтона–Понтрягина в виде:

$$H(t, x(t), u(t), \psi(t)) = \psi'(t) f(t, x(t), u(t)) + \int_t^{t_1} \psi'(\tau) K(\tau, t, x(t), u(t)) d\tau.$$

Тогда формула приращения (11) примет вид:

$$\begin{aligned} \Delta S(u) = & \varphi(\bar{x}(t_1)) - \varphi(x(t_1)) + \psi'(t_1) \Delta x(t_1) - \int_{t_0}^{t_1} \dot{\psi}'(t) \Delta x(t) dt + \\ & + \int_{t_0}^{t_1} \int_{t_0}^{t_1} (F(t, s, \bar{x}(t), \bar{x}(s)) - F(t, s, x(t), x(s))) ds dt - \\ & - \int_{t_0}^{t_1} [H(t, \bar{x}(t), \bar{u}(t), \psi(t)) - H(t, x(t), u(t), \psi(t))] dt. \end{aligned} \quad (12)$$

Применяя формулу Тейлора для отдельных слагаемых в формуле приращения (12), получаем

$$\begin{aligned} \Delta S(u) = & \frac{\partial \varphi'(x(t_1))}{\partial x} \Delta x(t_1) + \frac{1}{2} \Delta x'(t_1) \frac{\partial^2 \varphi(x(t_1))}{\partial x^2} \Delta x(t_1) + o_1(\|\Delta x(t_1)\|^2) + \psi'(t_1) \Delta x(t_1) - \int_{t_0}^{t_1} \dot{\psi}'(t) \Delta x(t) dt + \\ & + \int_{t_0}^{t_1} \int_{t_0}^{t_1} \left[\frac{\partial F'(t, s, x(t), x(s))}{\partial a} \Delta x(t) + \frac{\partial F'(t, s, x(t), x(s))}{\partial b} \Delta x(s) \right] ds dt + \\ & + \frac{1}{2} \int_{t_0}^{t_1} \int_{t_0}^{t_1} \left[\Delta x'(t) \frac{\partial^2 F(t, s, x(t), x(s))}{\partial a^2} \Delta x(t) + 2 \Delta x'(t) \frac{\partial^2 F(t, s, x(t), x(s))}{\partial a \partial b} \Delta x(s) + \right. \\ & \left. + \Delta x'(s) \frac{\partial^2 F(t, s, x(t), x(s))}{\partial b^2} \Delta x(s) \right] ds dt + \int_{t_0}^{t_1} \int_{t_0}^{t_1} o_2(\|\Delta x(t)\| + \|\Delta x(s)\|) ds dt - \\ & - \int_{t_0}^{t_1} \left[\frac{\partial H'(t, x(t), u(t), \psi(t))}{\partial x} \Delta x(t) + \frac{\partial H'(t, x(t), u(t), \psi(t))}{\partial u} \Delta u(t) \right] dt - \\ & - \frac{1}{2} \int_{t_0}^{t_1} \left[\Delta x'(t) \frac{\partial^2 H(t, x(t), u(t), \psi(t))}{\partial x^2} \Delta x(t) + 2 \Delta u'(t) \frac{\partial^2 H(t, x(t), u(t), \psi(t))}{\partial u \partial x} \Delta x(t) + \right. \\ & \left. + \Delta u'(t) \frac{\partial^2 H(t, x(t), u(t), \psi(t))}{\partial u^2} \Delta u(t) \right] dt - \int_{t_0}^{t_1} o_2(\|\Delta x(t)\| + \|\Delta x(s)\|) dt. \end{aligned} \quad (13)$$

Здесь штрих (') – операция транспонирования, $\|\alpha\|$ есть норма вектора $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_n)'$ в линейном n -мерном пространстве R^n , определяемая формулой $\|\alpha\| = \sum_{i=1}^n |\alpha_i|$, а $o(\alpha)$ – величина, имеющая более

высокий порядок малости, чем α , т.е. $\frac{o(\alpha)}{\alpha} \rightarrow 0$ при $\alpha \rightarrow 0$.

Легко видеть, что

$$\int_{t_0}^{t_1} \int_{t_0}^{t_1} \frac{\partial F'(t, s, x(t), x(s))}{\partial b} \Delta x(s) ds dt = \int_{t_0}^{t_1} \int_{t_0}^{t_1} \frac{\partial F'(s, t, x(s), x(t))}{\partial b} \Delta x(t) ds dt.$$

Поэтому, если предполагать, что $\psi(t)$ является решением уравнения

$$\dot{\psi}(t) = - \frac{\partial H(t, x(t), u(t), \psi(t))}{\partial x} - \int_{t_0}^{t_1} \left[\frac{\partial F(s, t, x(t), x(s))}{\partial a} - \frac{\partial F(s, t, x(s), x(t))}{\partial b} \right] ds, \quad (14)$$

$$\psi(t_1) = - \frac{\partial \varphi(x(t_1))}{\partial x}, \quad (15)$$

то формула приращения (13) примет вид:

$$\begin{aligned} \Delta S(u) = & - \int_{t_0}^{t_1} \frac{\partial H'(t, x(t), u(t), \psi(t))}{\partial u} \Delta u(t) dt + \frac{1}{2} \Delta x'(t_1) \frac{\partial^2 \varphi(x(t_1))}{\partial x^2} \Delta x(t_1) + \\ & + \frac{1}{2} \int_{t_0}^{t_1} \int_{t_0}^{t_1} \left[\Delta x'(t) \frac{\partial^2 F(t, s, x(t), x(s))}{\partial a^2} \Delta x(t) + 2 \Delta x'(t) \frac{\partial^2 F(t, s, x(t), x(s))}{\partial a \partial b} \Delta x(s) + \right. \\ & \quad \left. + \Delta x'(s) \frac{\partial^2 F(t, s, x(t), x(s))}{\partial b^2} \Delta x(s) \right] ds dt - \\ & - \frac{1}{2} \int_{t_0}^{t_1} \left[\Delta x'(t) \frac{\partial^2 H(t, x(t), u(t), \psi(t))}{\partial x^2} \Delta x(t) + 2 \Delta u'(t) \frac{\partial^2 H(t, x(t), u(t), \psi(t))}{\partial u \partial x} \Delta x(t) + \right. \\ & \quad \left. + \Delta u'(t) \frac{\partial^2 H(t, x(t), u(t), \psi(t))}{\partial u^2} \Delta u(t) \right] dt + o_1(\|\Delta x(t_1)\|^2) - \int_{t_0}^{t_1} o_2(\|\Delta x(t)\| + \|\Delta u(t)\|)^2 dt + \\ & \quad + \int_{t_0}^{t_1} \int_{t_0}^{t_1} o_3(\|\Delta x(t)\| + \|\Delta x(s)\|)^2 ds dt. \end{aligned} \quad (16)$$

По аналогии с работой [5] можно показать, что

$$\begin{aligned} \Delta x(t) dt \leq L_1 \left[\int_{t_0}^{t_1} (f(\tau, \bar{x}(\tau), \bar{u}(\tau)) - f(\tau, x(\tau), u(\tau))) d\tau + \right. \\ \left. + \int_{t_0}^{t_1} \int_t^{t_1} [K(s, \tau, \bar{x}(\tau), \bar{u}(\tau)) - K(s, \tau, x(\tau), u(\tau))] ds d\tau \right], \end{aligned}$$

где $L_1 = \text{const} > 0$ – некоторая постоянная.

Отсюда используя условие Липшица, после некоторых преобразований получаем

$$\Delta x(t) \leq L_2 \int_{t_0}^t \|u(\tau)\| d\tau, \quad t \in T. \quad (17)$$

В (17) $L_2 = \text{const} > 0$ – некоторая постоянная.

Пусть ε – достаточно малое по абсолютной величине число, а $\delta u(t) \in R^r, t \in [t_0, t_1]$ – произвольная кусочно-непрерывная и ограниченная r -мерная вектор-функция. Тогда специальное приращение допустимого управления $u(t)$ можно определить по формуле

$$\Delta u_\varepsilon(t) = \varepsilon \delta u(t), \quad t \in T. \quad (18)$$

Через $\Delta x_\varepsilon(t)$ обозначим специальное приращение траектории $x(t)$, отвечающее приращению (18) управления $u(t)$.

Из оценки (17) следует, что

$$\|\Delta x_\varepsilon(t)\| \leq L_3 |\varepsilon|, \quad t \in T, \quad (19)$$

где $L_3 = \text{const} > 0$ – некоторая постоянная.

Принимая во внимания формулы (18) и (19), с помощью задачи (6)–(7) доказываем, что

$$\Delta x_\varepsilon(t) = \varepsilon \delta x(t) + o(\varepsilon; t), \quad t \in T. \quad (20)$$

Здесь $\delta x(t)$ (вариация траектории $x(t)$) является решением уравнения в вариациях

$$\delta \dot{x}(t) = \frac{\partial f(t, x(t), u(t))}{\partial x} \delta x(t) + \frac{\partial f(t, x(t), u(t))}{\partial u} \delta u(t), \quad (21)$$

$$\delta x(t_0) = 0. \quad (22)$$

Принимая во внимания формулу (18) и разложение (20), из формулы приращения функционала (4) получаем справедливость разложения

$$\begin{aligned}
 S(u + \Delta u_\varepsilon) - S(u) = & -\varepsilon \int_{t_0}^{t_1} \frac{\partial H'(t, x(t), u(t), \psi(t))}{\partial u} \delta u(t) dt + \frac{\varepsilon^2}{2} \left[\delta x'(t_1) \frac{\partial^2 \varphi(x(t_1))}{\partial x^2} \delta x(t_1) + \right. \\
 & + \int_{t_0}^{t_1} \int_{t_0}^{t_1} \left[\delta x'(t) \frac{\partial^2 F(t, s, x(t), x(s))}{\partial a^2} \delta x(t) + 2\delta x'(t) \frac{\partial^2 F(t, s, x(t), x(s))}{\partial a \partial b} \delta x(s) + \right. \\
 & \left. \left. + \delta x'(s) \frac{\partial^2 F(t, s, x(t), x(s))}{\partial b^2} \delta x(s) \right] ds dt \right] - \\
 & - \frac{\varepsilon^2}{2} \int_{t_0}^{t_1} \left[\delta x'(t) \frac{\partial^2 H(t, x(t), u(t), \psi(t))}{\partial x^2} \delta x(t) + 2\delta u'(t) \frac{\partial^2 H(t, x(t), u(t), \psi(t))}{\partial u \partial x} \delta x(t) + \right. \\
 & \left. + \delta u'(t) \frac{\partial^2 H(t, x(t), u(t), \psi(t))}{\partial u^2} \delta u(t) \right] dt + o(\varepsilon^2).
 \end{aligned} \tag{23}$$

Из разложения (23) получаем, что первая и вторая вариации функционала (4) имеют соответственно следующий вид:

$$\delta^1 S(u : \delta u) = - \int_{t_0}^{t_1} \frac{\partial H'(t, x(t), u(t), \psi(t))}{\partial u} \delta u(t) dt, \tag{24}$$

$$\begin{aligned}
 \delta^2 S(u : \delta u) = & \delta x'(t_1) \frac{\partial^2 \varphi(x(t_1))}{\partial x^2} \delta x(t_1) + \int_{t_0}^{t_1} \int_{t_0}^{t_1} \left[\delta x'(t) \frac{\partial^2 F(t, s, x(t), x(s))}{\partial a^2} \delta x(t) + \right. \\
 & + 2\delta x'(t) \frac{\partial^2 F(t, s, x(t), x(s))}{\partial a \partial b} \delta x(s) + \delta x'(s) \frac{\partial^2 F(t, s, x(t), x(s))}{\partial b^2} \delta x(s) \left. \right] ds dt - \\
 & - \int_{t_0}^{t_1} \left[\delta x'(t) \frac{\partial^2 H(t, x(t), u(t), \psi(t))}{\partial x^2} \delta x(t) + 2\delta u'(t) \frac{\partial^2 H(t, x(t), u(t), \psi(t))}{\partial u \partial x} \delta x(t) + \right. \\
 & \left. + \delta u'(t) \frac{\partial^2 H(t, x(t), u(t), \psi(t))}{\partial u^2} \delta u(t) \right] dt.
 \end{aligned} \tag{25}$$

3. Необходимые условия оптимальности

В силу открытости области управления из (24) следует, что вдоль оптимального управления

$$\int_{t_0}^{t_1} \frac{\partial H'(t, x(t), u(t), \psi(t))}{\partial u} \delta u(t) dt = 0.$$

Отсюда в силу произвольности $\delta u(t) \in R^r$, $t \in [t_0, t_1]$ следует

Теорема 1. Для оптимальности допустимого управления $u(t)$ необходимо, чтобы соотношение

$$\frac{\partial H(\theta, x(\theta), u(\theta), \psi(\theta))}{\partial u} = 0 \tag{26}$$

выполнялось для всех $\theta \in [t_0, t_1]$.

Здесь $\theta \in [t_0, t_1]$ – произвольная точка непрерывности управления $u(t)$. Соотношение (26) является аналогом уравнения Эйлера для рассматриваемой задачи. Каждое допустимое управление $u(t)$ являющееся решением уравнения Эйлера, следуя, например, [8], назовем классической экстремалью.

Число классических экстремалей может быть достаточно большим (см., напр.: [8]). Поэтому надо иметь необходимые условия оптимальности второго порядка, позволяющие сузить множество классических экстремалей, подозрительных на оптимальность.

Поскольку вдоль оптимального управления вторая вариация функционала качества должна быть неотрицательной, то, учитывая формулу (25), получаем, что для оптимальности классической экстремали необходимо, чтобы для всех $\delta u(t) \in R^r$, $t \in [t_0, t_1]$, выполнялось неравенство

$$\begin{aligned} & \delta x'(t_1) \frac{\partial^2 \varphi(x(t_1))}{\partial x^2} \delta x(t_1) + \int_{t_0}^{t_1} \int_{t_0}^{t_1} \left[\delta x'(t) \frac{\partial^2 F(t, s, x(t), x(s))}{\partial a^2} \delta x(t) + \right. \\ & \left. + 2\delta x'(t) \frac{\partial^2 F(t, s, x(t), x(s))}{\partial a \partial b} \delta x(s) + \delta x'(s) \frac{\partial^2 F(t, s, x(t), x(s))}{\partial b^2} \delta x(s) \right] ds dt - \\ & - \int_{t_0}^{t_1} \left[\delta x'(t) \frac{\partial^2 H(t, x(t), u(t), \psi(t))}{\partial x^2} \delta x(t) + 2\delta u'(t) \frac{\partial^2 H(t, x(t), u(t), \psi(t))}{\partial u \partial x} \delta x(t) + \right. \\ & \left. + \delta u'(t) \frac{\partial^2 H(t, x(t), u(t), \psi(t))}{\partial u^2} \delta u(t) \right] dt \geq 0. \end{aligned} \quad (27)$$

Неравенство (27) является неявным необходимым условием оптимальности второго порядка. Но из него можно получить конструктивно проверяемые необходимые условия оптимальности второго порядка.

Запишем представление решения уравнения в вариациях (21)–(22). По формуле Коши (см., напр.: [10]) имеем

$$\delta x(t) = \int_{t_0}^t F(t, \tau) \frac{\partial f(\tau, x(\tau), u(\tau))}{\partial u} \delta u(\tau) d\tau + \int_{t_0}^t \int_{\tau}^t F(t, s) \frac{\partial K(s, \tau, x(\tau), u(\tau))}{\partial u} ds \delta u(\tau) d\tau. \quad (28)$$

Здесь $F(t, \tau)$ – $(n \times n)$ матричная функция, являющаяся решением матричного интегро-дифференциального уравнения

$$F_t(t, \tau) = -F(t, \tau) \frac{\partial f(\tau, x(\tau), u(\tau))}{\partial x} - \int_{\tau}^t F(t, s) \frac{\partial K(s, \tau, x(\tau), u(\tau))}{\partial x} ds$$

с начальным условием $F(t, t) = E$, где E – $(n \times n)$ единичная матрица.

Введя обозначение

$$Q(t, \tau) = F(t, \tau) \frac{\partial f(\tau, x(\tau), u(\tau))}{\partial u} + \int_{\tau}^t F(t, s) \frac{\partial K(s, \tau, x(\tau), u(\tau))}{\partial u} ds,$$

формулу (28) запишем в виде:

$$\delta x(t) = \int_{t_0}^t Q(t, \tau) \delta u(\tau) d\tau. \quad (29)$$

Используя представление (29), займемся преобразованием отдельных слагаемых в неравенстве (27).

По схеме из [11] имеем

$$\delta x'(t_1) \frac{\partial^2 \varphi(x(t_1))}{\partial x^2} \delta x(t_1) = \int_{t_0}^{t_1} \int_{t_0}^{t_1} \delta u'(t) Q'(t_1, \tau) \frac{\partial^2 \varphi(x(t_1))}{\partial x^2} Q(t_1, s) \delta u(t) ds d\tau, \quad (30)$$

$$\begin{aligned} & \int_{t_0}^{t_1} \int_{t_0}^{t_1} \delta x'(t) \frac{\partial^2 F(t, s, x(t), x(s))}{\partial a^2} \delta x(t) ds dt = \\ & = \int_{t_0}^{t_1} \int_{t_0}^{t_1} \int_{t_0}^{t_1} \delta u'(\alpha) \left[\int_{\max(\alpha, \beta)}^{t_1} Q'(t, \alpha) \frac{\partial^2 F(t, s, x(t), x(s))}{\partial a^2} Q(t, \beta) dt \right] \delta u(\beta) d\alpha d\beta ds, \\ & \int_{t_0}^{t_1} \int_{t_0}^{t_1} \delta x'(t) \frac{\partial^2 F(t, s, x(t), x(s))}{\partial a \partial b} \delta x(s) ds dt = \\ & = \int_{t_0}^{t_1} \int_{t_0}^{t_1} \left[\int_{\alpha}^{t_1} \int_{\beta}^{t_1} \delta u'(\alpha) Q'(t, \alpha) \frac{\partial^2 F(t, s, x(t), x(s))}{\partial a \partial b} Q(s, \beta) ds dt \right] \delta u(\beta) d\alpha d\beta, \end{aligned} \quad (31)$$

$$\begin{aligned}
 & \int_{t_0}^{t_1} \int_{t_0}^{t_1} \delta x'(s) \frac{\partial^2 F(t, s, x(t), x(s))}{\partial b \partial a} \delta x(t) ds dt = \\
 & = \int_{t_0}^{t_1} \int_{t_0}^{t_1} \delta u'(\alpha) \left[\int_{\alpha}^{t_1} \int_{\beta}^{t_1} Q'(s, \alpha) \frac{\partial^2 F(t, s, x(t), x(s))}{\partial b \partial a} Q(t, \beta) ds dt \right] \delta u(\beta) d\alpha d\beta, \\
 & \int_{t_0}^{t_1} \int_{t_0}^{t_1} \delta x'(s) \frac{\partial^2 F(t, s, x(t), x(s))}{\partial b^2} \delta x(s) ds dt = \\
 & = \int_{t_0}^{t_1} \int_{t_0}^{t_1} \delta u'(\alpha) \left[\int_{t_0}^{t_1} \int_{\max(\alpha, \beta)}^{t_1} Q'(s, \alpha) \frac{\partial^2 F(t, s, x(t), x(s))}{\partial b^2} Q(s, \beta) dt ds \right] \delta u(\beta) d\alpha d\beta, \quad (32)
 \end{aligned}$$

$$\begin{aligned}
 & \int_{t_0}^{t_1} \delta x'(t) \frac{\partial^2 H(t, x(t), u(t), \psi(t))}{\partial x^2} \delta x(t) dt = \\
 & = \int_{t_0}^{t_1} \int_{t_0}^{t_1} \delta u'(\alpha) \left[\int_{\max(\alpha, \beta)}^{t_1} Q'(t, \alpha) \frac{\partial^2 H(t, x(t), u(t), \psi(t))}{\partial x^2} Q(t, \beta) dt \right] \delta u(\beta) d\alpha d\beta, \quad (33)
 \end{aligned}$$

$$\begin{aligned}
 & \int_{t_0}^{t_1} \delta u'(t) \frac{\partial^2 H(t, x(t), u(t), \psi(t))}{\partial u \partial x} \delta x(t) dt = \\
 & = \int_{t_0}^{t_1} \left[\int_{t_0}^{t_1} \delta u'(t) \frac{\partial^2 H(t, x(t), u(t), \psi(t))}{\partial u \partial x} Q(t, \alpha) \delta u(\tau) d\tau \right] dt. \quad (34)
 \end{aligned}$$

Введем обозначение

$$\begin{aligned}
 K(\alpha, \beta) = & Q'(t_1, \alpha) \frac{\partial^2 \varphi(x(t_1))}{\partial x^2} Q(t_1, \beta) + \int_{t_0}^{t_1} \left[\int_{\max(\alpha, \beta)}^{t_1} Q'(t, \alpha) \frac{\partial^2 F(t, s, x(t), x(s))}{\partial a^2} Q(t, \beta) dt \right] ds - \\
 & - \int_{\alpha}^{t_1} \left[\int_{\beta}^{t_1} Q'(t, \alpha) \frac{\partial^2 F(t, s, x(t), x(s))}{\partial a \partial b} Q(s, \beta) ds \right] dt - \int_{\alpha}^{t_1} \left[\int_{\beta}^{t_1} Q'(s, \alpha) \frac{\partial^2 F(t, s, x(t), x(s))}{\partial b \partial a} Q(t, \beta) ds \right] dt + \\
 & + \int_{t_0}^{t_1} \left[\int_{\max(\alpha, \beta)}^{t_1} Q'(s, \alpha) \frac{\partial^2 F(t, s, x(t), x(s))}{\partial b^2} Q(s, \beta) dt \right] ds + \int_{\max(\alpha, \beta)}^{t_1} Q'(t, \alpha) \frac{\partial^2 H(t, x(t), u(t), \psi(t))}{\partial x^2} Q(x, \beta) dt.
 \end{aligned}$$

С учетом введенных обозначений и тождеств (30)–(34) неравенство (27) представляется в виде:

$$\begin{aligned}
 & \int_{t_0}^{t_1} \int_{t_0}^{t_1} \delta u'(\alpha) K(\alpha, \beta) \delta u(\beta) d\alpha d\beta + 2 \int_{t_0}^{t_1} \left[\int_{t_0}^t \delta u'(t) \frac{\partial^2 H(t, x(t), u(t), \psi(t))}{\partial u \partial x} Q(t, \tau) \delta u(\tau) d\tau \right] dt + \\
 & + \int_{t_0}^{t_1} \delta u'(t) \frac{\partial^2 H(t, x(t), u(t), \psi(t))}{\partial u^2} \delta u(t) dt \leq 0. \quad (35)
 \end{aligned}$$

Таким образом, доказана следующая теорема.

Теорема 2. Для оптимальности классической экстремали $u(t)$ необходимо, чтобы неравенство (35) выполнялось для всех $\delta u(t) \in R^r$, $t \in [t_0, t_1]$.

Как видно, в отличие от неравенства (27) неравенство (35) носит конструктивный характер. Из него, используя произвольность допустимой вариации $\delta u(t)$, можно получить более легко проверяемые необходимые условия оптимальности второго порядка. Одним из них является аналог условия Лежандра–Клебша [4, 5]

Теорема 3 (аналог условия Лежандра–Клебша). Для оптимальности классической экстремали $u(t)$ необходимо, чтобы неравенство

$$v' \frac{\partial^2 H(t, x(t), u(t), \psi(t))}{\partial u^2} v \leq 0 \quad (36)$$

выполнялось для всех $v \in R^r$ и $\theta \in [t_0, t_1]$. Здесь $\theta \in [t_0, t_1]$ – произвольная точка непрерывности управления $u(t)$.

Отметим, что в некоторых случаях аналог условия Лежандра–Клебша (36) может вырождаться, выполняясь как тождественное равенство (см., напр.: [12]).

Определение. Классическую экстремаль $u(t)$ назовем особым, в классическом смысле, управлением, если для всех $v \in R^r$ и $\theta \in [t_0, t_1]$

$$v' \frac{\partial^2 H(\theta, x(\theta), u(\theta), \psi(\theta))}{\partial u^2} v \equiv 0.$$

Из неравенства (35) следует теорема.

Теорема 4. Для оптимальности особого, в классическом смысле, управления $u(t)$ необходимо, чтобы неравенство

$$v' \left(K(\theta, \theta) + \frac{\partial^2 H(\theta, x(\theta), u(\theta), \psi(\theta))}{\partial u \partial x} \right) v \leq 0 \quad (37)$$

выполнялось для всех $v \in R^r$ и $\theta \in [t_0, t_1]$.

Доказательство. Пусть $u(t)$ особое, в классическом смысле, оптимальное управление.

Допустимую вариацию управления $u(t)$ определим следующим образом:

$$\delta u_\varepsilon(t) = \begin{cases} v, & t \in [\theta, \theta + \varepsilon), \\ 0, & t \in [t_0, t_1] \setminus [\theta, \theta + \varepsilon). \end{cases} \quad (38)$$

Здесь $v \in R^r$ – произвольный вектор, а $\varepsilon > 0$ – произвольное достаточно малое число.

Принимая во внимания формулу (38), в неравенстве (35) после некоторых преобразований получим, что

$$\varepsilon^2 \left(v' K(\theta, \theta) v + v' \frac{\partial^2 H(\theta, x(\theta), u(\theta), \psi(\theta))}{\partial u \partial x} v \right) + o(\varepsilon^2) \leq 0.$$

Разделяя обе части этого неравенства на ε^2 и переходя к пределу при $\varepsilon \rightarrow 0$, получим неравенство (37). Теорема доказана.

Заметим, что неравенство (37) является аналогом условия оптимальности Габасова–Кирилловой, полученного ими в [12] другим способом для задачи оптимального управления, описываемой обыкновенным дифференциальным уравнением.

Заключение

В работе при предположении открытости области управления вычислены первая и вторая вариации функционала качества. Доказано необходимое условие оптимальности первого порядка в форме аналога уравнения Эйлера. Из условия неотрицательности второй вариации критерия качества вдоль оптимального управления получены конструктивно проверяемые необходимые условия оптимальности второго порядка, позволяющие сузить множество классических экстремалей, подозрительных на оптимальность.

Из общего результата получено легко проверяемое необходимое условие оптимальности, являющееся аналогом условия Лежандра–Клебша.

Изучен случай вырождения аналога условия Лежандра–Клебша (классический особый случай) и доказано необходимое условие оптимальности особых, в классическом смысле, управлений.

Список источников

1. Габасов Р., Кириллова Ф.М. Принцип максимума в теории оптимального управления М. : Либроком, 2011. 272 с.
2. Ведякова А.О., Миланович Е.В., Слита О.В., Тергичный-Даури О.В. Методы оптимального управления. СПб. : Университет ИТМО, 2011. 219 с.
3. Лутманов С.В., Остапенко Е.Н. Оптимальное управление динамическими объектами. Пермь : Изд-во ПГУ, 2022. 163 с.
4. Алексеев В.М., Тихомиров В.М., Фомин С.В. Оптимальное управление. М. : Физматлит, 2005. 432 с.
5. Габасов Р., Кириллова Ф.М., Альсевич В.В. Методы оптимизации. Минск : Четыре четверти, 2011, 472 с.
6. Моисеев Н.Н. Элементы теории оптимальных систем. М. : Наука, 1975. 528 с.
7. Нагиева И.Ф., Мансимов К.Б. Необходимые условия оптимальности особых управлений типа Моисеева // Проблемы управления и информатики. 2006. № 5. С. 52–63.
8. Габасов Р.Ф., Кириллова Ф.М. Особые оптимальные управления. М. : Либроком, 2011. 256 с.
9. Колмогоров А.Н., Фомин С.В. Элементы теории функций и функционального анализа. М. : Физматлит, 2004. 752 с.
10. Ведь Ю.А., Пахыров З. Об ограниченности и устойчивости решений интегро-дифференциальных уравнений с запаздывающим аргументом // Дифференциальные уравнения. 1969. №11. С. 1050–1061.
11. Мансимов К.Б. Особые управления в системах с запаздыванием. Баку : ЭЛМ, 1999. 176 с.
12. Габасов Р., Кириллова Ф.М. К теории необходимых условий оптимальности высокого порядка // Дифференциальные уравнения. 1970. № 4. С. 964–970.

References

1. Gabasov, R. & Kirillova, F.M. (2011) *Printsip maksimuma v teorii optimal'nogo upravleniya* [The maximum principle in the theory of optimal control. Moscow: Librokom.
2. Vedyakova, A.O., Milanovich, E.V., Slita, O.V. & Tertychny-Dauri, O.V. (2011) *Metody optimal'nogo upravleniya* [Methods of optimal control]. St. Petersburg: St. Petersburg State University of Information Technologies, Mechanics and Optics.
3. Lutmanov, S.V. & Ostapenko, E.N. (2022) *Optimal'noe upravlenie dinamicheskimi ob'ektami* [Optimal control of dynamic objects]. Perm: PSU.
4. Alekseev, V.M., Tikhomirov, V.M. & Fomin, S.V. (2005) *Optimal'noe upravlenie* [Optimal control]. Moscow: Fizmatlit.
5. Gabasov, R., Kirillova, F.M. & Alseovich, V.V. (2011) *Metody optimizatsii* [Optimization Methods]. Minsk: Chetyre chetverti.
6. Moiseev, N.N. (1975) *Elementy teorii optimal'nykh sistem* [Elements of the Theory of Optimal Systems]. Moscow: Nauka.
7. Nagieva, I.F. & Mansimov K.B. (2006) Necessary conditions for the optimality of special controls of the Moiseev type. *Problemy upravleniya i informatiki – Problems of Control and Informatics*. 5. pp. 52–63.
8. Gabasov, R.F. & Kirillova, F.M. (2011) *Osobyie optimal'nye upravleniya* [Singular Optimal Controls]. Moscow: Librokom.
9. Kolmogorov, A.N. & Fomin, S.V. (2004) *Elementy teorii funktsiy i funktsional'nogo analiza* [Elements of the theory of functions and functional analysis]. Moscow: Fizmat.
10. Ved', Yu.A. & Pakhyrov, Z. (1969) Ob ogranichennosti i ustoychivosti resheniy integro-differentsial'nykh uravneniy s zapazdyvayushchim argumentom [On boundedness and stability of solutions of integro-differential equations with retarded argument]. *Differentsial'nye uravneniya*. 11. pp. 1050–1061.
11. Mansimov, K.B. (1999) *Osobyie upravleniya v sistemakh s zapazdyvaniem* [Singular controls in systems with delay]. Baku: ELM.
12. Gabasov, R. & Kirillova, F.M. (1970) K teorii neobkhodimykh usloviy optimal'nosti vysokogo poryadka [On the theory of high order necessary optimality conditions]. *Differentsial'nye uravneniya*. 4. pp. 964–970.

Информация об авторах:

Мансимов Камил Байрамали оглы – профессор, доктор физико-математических наук, заведующий лабораторией «Методы управления в сложных динамических системах» Института систем управления Министерства науки и образования Азербайджана (Баку, Азербайджан); заведующий кафедрой «Математическая кибернетика» Бакинского государственного университета (Баку, Азербайджан), E-mail: kamilbmansimov@gmail.com

Нагиева Илаха Фирдовис кызы – кандидат физико-математических наук, старший научный сотрудник Института систем управления Министерства науки и образования Азербайджана (Баку, Азербайджан); E-mail: ilaha_nagiyeva@gmail.com

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

Information about the authors:

Mansimov Kamil B. (Professor, Doctor of Physical and Mathematical Sciences, Head of the Laboratory "Control Methods in Complex Dynamic Systems" of the Institute of Control Systems of the Ministry of Science and Education of Azerbaijan (Baku, Azerbaijan); Head Department of Mathematical Cybernetics, Baku State University (Baku, Azerbaijan), E-mail: kamilbmansimov@gmail.com

Nagiyeva Ilaha F. (Candidate of Physical and Mathematical Sciences, Senior Researcher at the Institute of Control Systems of the Ministry of Science and Education of Azerbaijan, Baku, Azerbaijan); E-mail: ilaha_nagiyeva@gmail.com

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

Поступила в редакцию 28.03.2023; принята к публикации 04.09.2023

Received 28.03.2023; accepted for publication 04.09.2023

Научная статья

УДК 517.9

doi: 10.17223/19988605/64/3

Метод синтеза модального регулятора в условиях структурных возмущений в объекте управления

Андрей Николаевич Паршуков

Тюменский индустриальный университет, Тюмень, Россия, anparshukov@mail.ru

Аннотация. Разработан новый метод обоснованного понижения порядка модального регулятора. Данный метод основан на разделении мод объекта управления по отношению к цели управления на доминирующую динамику (подлежащую регулированию) и структурные возмущения (уже удовлетворяющие цели управления). При синтезе модального регулятора пониженного порядка учитываются операторы структурных возмущений. Расчет модального регулятора пониженного порядка получен в виде аналитической формулы, что позволяет реализовать данный метод на ЭВМ. Эффективность метода проиллюстрирована примером.

Ключевые слова: редуцированная модель объекта управления; доминирующая динамика; структурные возмущения; модальный регулятор; регулятор пониженного порядка.

Для цитирования: Паршуков А.Н. Метод синтеза модального регулятора в условиях структурных возмущений в объекте управления // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2023. № 64. С. 21–29. doi: 10.17223/19988605/64/3

Original article

doi: 10.17223/19988605/64/3

Method of synthesis of a modal regulator under conditions of structural disturbances in a control object

Andrej N. Parshukov

Industrial University of Tyumen, Tyumen, Russian Federation, anparshukov@mail.ru

Abstract. The article develops a new method for justifiably lowering the order of the modal regulator. This method is based on the separation of the modes of the control object in relation to the control goal into dominant dynamics (subject to regulation) and structural disturbances (already satisfying the control goal). When synthesizing a modal regulator of a reduced order, structural perturbation operators are taken into account. The calculation of the modal regulator of a reduced order is obtained in the form of an analytical formula, which makes it possible to implement this method on a computer. The effectiveness of the method is illustrated by an example.

Keywords: reduced model of the control object; dominant dynamics; structural disturbances; modal regulator; low-order regulator.

For citation: Parshukov, A.N. (2023) Method of synthesis of a modal regulator under conditions of structural disturbances in a control object. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naja tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 64. pp. 21–29. doi: 10.17223/19988605/64/3

Введение

В литературе, посвященной вопросам синтеза систем автоматического управления, в подавляющем большинстве случаев подразумевается синтез ПИ- и ПИД-законов регулирования (см., напр.: [1–3]). ПИД-регуляторы доказали свою эффективность при управлении объектами, которые хорошо описываются дифференциальными уравнениями до 2-го порядка включительно. Однако, с одной стороны, интенсивное использование средств информатики и цифровой автоматики – микропроцессорных контроллеров, SCADA-систем и т.п. – при управлении технологическими процессами в промышленности позволяет перейти от традиционных ПИ- и ПИД-законов регулирования к более сложным, например модальным. С другой – все большее число технологических процессов (объектов управления) описывается дифференциальными уравнениями высокого (начиная с 3-го) порядка [4–7]. Для объектов управления высокого порядка регулировочных возможностей ПИД-регулятора может быть недостаточно, поскольку очевидно, что, повышая порядок модели объекта, следует адекватно повышать порядок регулятора.

Метод модального управления, изложенный в работах [8–10], позволяет синтезировать регулятор для объектов управления произвольного порядка; данный метод предполагает, что объект управления описывается линейным дифференциальным уравнением n -го порядка (n – любое целое неотрицательное число) без запаздывания. Модальный регулятор ищется также в виде линейного дифференциального уравнения. Качество управления задается в виде области S на комплексной плоскости, определяющей желаемое расположение нулей характеристического полинома замкнутой системы. В книге [9, С. 11] доказано, что модальный регулятор порядка $n - 1$ и выше обеспечивает любое заданное расположение нулей характеристического полинома замкнутой системы и тем самым гарантирует устойчивость и заданные корневые показатели качества для замкнутой системы. Регулятор $(n - 1)$ -го порядка называется модальным регулятором *полного порядка*.

В настоящей статье рассматривается задача понижения порядка модального регулятора. Данная задача является актуальной: понижение порядка регулятора позволит снизить влияние возможных ошибок при реализации закона регулирования, повысить надежность замкнутой системы (за счет понижения порядка замкнутой системы) и сэкономить вычислительные ресурсы при расчете регулятора. Самым простым способом понижения порядка регулятора является редукция (упрощение) исходной модели объекта управления. Методам редукции посвящено значительное число публикаций (см., напр.: [11–14]); многие методы подразумевают разделение *быстрых* и *медленных движений* системы. В данной работе на основе подобного разделения динамик в объекте управления разработан метод обоснованного понижения порядка модального регулятора.

В статье приняты обозначения: $\doteq$ – равно по определению; T – транспонирование; $*$ – комплексное сопряжение; j – мнимая единица; R^n, C^n – пространства n -мерных векторов; s – комплексная переменная; S – область на C^1 ; ∂S – граница области S ; $\text{int} S$ – внутренняя часть области S ; t – непрерывное время; p^i – оператор i -й степени дифференцирования по времени:

$$p^i \doteq d^i / dt^i; \quad i \in \overline{0, n}, \quad p^0 \doteq 1.$$

Полиномиальным оператором степени n назовем дифференциальный оператор вида:

$$f(n, p) = \sum_{i=0}^n f_i \cdot p^i,$$

где f_i – постоянные коэффициенты ($i \in \overline{0, n}$). В изображениях по Лапласу оператору $f(n, p)$ соответствует алгебраический полином

$$f(n, s) = \sum_{i=0}^n f_i \cdot s^i,$$

определенный на C^1 ; здесь за s обозначена переменная преобразования Лапласа ($s \in C^1$).

Множество нулей полинома $f(n, s)$ будем обозначать $\Lambda(f)$:

$$\Lambda(f) \doteq \{ \lambda_i : f(n, \lambda_i) = 0, \quad i \in \overline{1, n} \}.$$

1. Метод синтеза модального регулятора в условиях структурных возмущений в объекте управления

1.1. Классический метод синтеза модального регулятора полного порядка

Пусть одномерный линейный стационарный динамический объект управления задан дифференциальным уравнением n -го порядка

$$a(n, p)y(t) = b(m, p)u(t), \quad n > m,^1 \quad (1)$$

где $y(t)$ – управляемая переменная (выходной сигнал), $u(t)$ – управляющая переменная (входной сигнал). Модальный регулятор ищется в виде дифференциального уравнения l -го порядка

$$\beta(l, p)u(t) = \alpha(l, p)y(t) + \chi(l, p)g(t), \quad \beta_l = 1, \quad (2)$$

где $g(t)$ – входной сигнал для замкнутой системы (рис. 1).

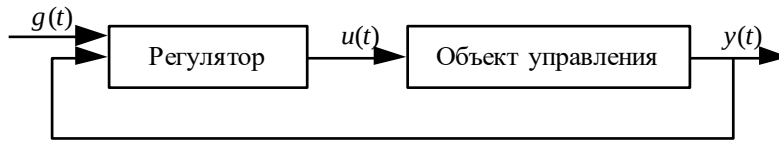


Рис. 1. Структурная схема замкнутой системы управления
Fig. 1. Block diagram of a closed control system

Уравнение замкнутой системы

$$a^{cl}(n+l, p)y(t) = b^{cl}(m+l, p)g(t), \quad (3)$$

где

$$\begin{aligned} a^{cl}(n+l, p) &= a(n, p) \cdot \beta(l, p) - b(m, p) \cdot \alpha(l, p), \\ b^{cl}(m+l, p) &= b(m, p) \cdot \chi(l, p). \end{aligned}$$

Уравнению (3) соответствует передаточная функция замкнутой системы

$$h^{cl}(s) = \frac{b(m, s) \cdot \chi(l, s)}{a(n, s) \cdot \beta(l, s) - b(m, s) \cdot \alpha(l, s)},$$

характеристический полином замкнутой системы равен

$$a^{cl}(n+l, s) = a(n, s) \cdot \beta(l, s) - b(m, s) \cdot \alpha(l, s).$$

Качество управления назначается в виде области S , определяющей допустимое расположение нулей характеристического полинома a^{cl} на C^1 . Таким образом, требования к качеству управления могут быть записаны в виде целевого условия

$$\Lambda(a^{cl}) \subset S.$$

Предполагается, что область S удовлетворяет условиям: расположена в ограниченной части C^1 слева от мнимой оси; односвязна; для любой точки $s \in S$ также выполняется $s^* \in S$.

Поскольку выбор полинома χ регулятора не влияет на расположение нулей полинома a^{cl} замкнутой системы, вопрос расчета полинома χ в данном методе не рассматривается. Таким образом, расчету подлежит вектор коэффициентов модального регулятора (2)

$$\mathbf{r} = \text{col}(\beta_0, \dots, \beta_{l-1}, \alpha_0, \dots, \alpha_l) \in R^{2l+1}.$$

В работе [9. С. 11] доказано, что при $l \geq n-1$ в (2) может быть обеспечено любое заданное расположение нулей полинома a^{cl} на комплексной плоскости. В этом смысле модальные регуляторы порядка

$$l = n-1 \quad (4)$$

называются регуляторами *полного порядка*.

¹ Условие физической реализуемости математической модели объекта управления.

При синтезе модального регулятора возможны следующие три случая.

Случай 1. Все нули полинома a объекта управления находятся внутри области S . В таком случае нет необходимости в синтезе регулятора, так как объект управления уже удовлетворяет заданному качеству управления.

Случай 2. Все нули полинома a лежат вне области S . В этом случае следует регулировать все моды¹ объекта управления, что, очевидно, можно сделать только регулятором полного порядка.

Случай 3. Часть нулей полиномов a и b объекта управления лежит внутри области S . Следовательно, регулировке подлежат не все моды объекта, а только те, которые не удовлетворяют заданному целевому условию. Очевидно, что порядок модального регулятора должен определяться количеством «неудовлетворительных» мод объекта.

Таким образом, основанием для понижения порядка модального регулятора служит наличие «удовлетворительных» нулей полиномов a и b объекта управления. Отметим, что подобные рассуждения ранее рассматривались в литературе [11–14], причем в качестве «неудовлетворительной» выступала так называемая *доминирующая динамика*², определяющая основные свойства системы и подлежащая регулированию.

Метод синтеза модального регулятора пониженного порядка изложен в следующем разделе.

1.2. Метод синтеза модального регулятора в условиях структурных возмущений в объекте управления

Для удобства дальнейших рассуждений модель объекта управления (1) представим в следующем виде:

$$\begin{aligned} v(n_2, p) \cdot a^r(n_1, p)y(t) &= w(m_2, p) \cdot b^r(m_1, p)u(t), \\ n_1 > m_1, \quad n_2 \geq m_2, \quad a_{n_1}^r &= 1, \quad v_0 = w_0 = 1^3. \end{aligned} \quad (5)$$

В записи (5) явно выделены операторы v и w , нули которых лежат внутри области S , т. е.

$$\Lambda(v) \subset \text{int } S, \quad \Lambda(w) \subset \text{int } S.$$

Следуя работе [15. С. 15], пару операторов $\langle v, w \rangle$ назовем *структурными возмущениями*.

Операторы $\langle a^r, b^r \rangle$ описывают ту часть объекта управления, которая желаемым качеством управления может не обладать, и, следовательно, подлежит регулированию. Далее редуцированную модель

$$a^r(n_1, p)y(t) = b^r(m_1, p)u(t) \quad (6)$$

назовем *доминирующей динамикой* исходного объекта управления (1). Редуцированная модель (6) будет совпадать с исходной моделью (1) объекта управления при выполнении

$$n_1 = n, \quad m_1 = m, \quad v(n_2, p) = w(m_2, p) \equiv 1.$$

В классической схеме синтеза модального управления порядок l модального регулятора определяется по формуле (4), где n – порядок модели объекта управления. Таким образом, понижая порядок модели объекта управления, мы одновременно понижаем порядок модального регулятора и, как следствие, порядок характеристического полинома замкнутой системы.

На основании изложенного модальный регулятор

$$\beta(n_1 - 1, p)u(t) = \alpha(n_1 - 1, p)y(t) + \chi(n_1 - 1, p)g(t), \quad \beta_{n_1-1} = 1, \quad (7)$$

будем называть *регулятором пониженного порядка*. Под вектором коэффициентов модального регулятора (7) понимается вектор

$$\mathbf{r} = \text{col}(\beta_0, \dots, \beta_{n_1-2}, \alpha_0, \dots, \alpha_{n_1-1}) \in R^{2n_1-1}.$$

¹ *Модами* называются слагаемые в свободной составляющей реакции системы, эти составляющие зависят от корней ее характеристического полинома [8. С. 5].

² Под *доминирующей динамикой* понимаются те моды, которые вносят наибольший вклад («главные моды») в свободную составляющую реакции системы.

³ Это условие обеспечивает одинаковый коэффициент передачи моделей (1) и (6).

В статье [15. С. 14, 15] модальный регулятор пониженного порядка предлагается рассчитывать только по модели (6), без учета операторов структурных возмущений. При этом после замыкания исходного объекта (1) синтезированным регулятором нули характеристического полинома замкнутой системы могут выйти из области S .

В настоящей работе вектор коэффициентов модального регулятора пониженного порядка предлагается рассчитывать по формуле

$$\mathbf{r}^* = \arg \min_{\mathbf{r}} J(\mathbf{r}),$$

$$J(\mathbf{r}) \doteq (\mathbf{a}^{et} - \mathbf{a}^{cl}(\mathbf{r}))^T \mathbf{Q}(\mathbf{a}^{et} - \mathbf{a}^{cl}(\mathbf{r})), \quad (8)$$

здесь

$$\mathbf{a}^{cl} \doteq \text{col}(a_0^{cl}, \dots, a_{n+n_1-2}^{cl}) \in R^{n+n_1-1}$$

есть вектор коэффициентов характеристического полинома

$$a^{cl}(n+n_1-1, s) = v(n_2, s) \cdot a^r(n_1, s) \cdot \beta(n_1-1, s) - w(m_2, s) \cdot b^r(m_1, s) \cdot \alpha(n_1-1, s)$$

системы, полученной после замыкания объекта управления (5) модальным регулятором (7);

$$\mathbf{a}^{et} \doteq \text{col}(a_0^{et}, \dots, a_{n+n_1-1}^{et}) \in R^{n+n_1-1}$$

есть вектор коэффициентов характеристического полинома эталонной системы (далее – *эталона*)

$$a^{et}(n+n_1-1, s) = s^{n+n_1-1} + \sum_{i=0}^{n+n_1-2} a_i^{et} \cdot s^i,$$

он предполагается заданным и удовлетворяющим условию

$$\Lambda(a^{et}) \subset \text{int } S; \quad (9)$$

$\mathbf{Q} \in R^{(n+n_1-1) \times (n+n_1-1)}$ – положительно определенная симметричная матрица (предполагается заданной).

Вектор $\mathbf{a}^{cl}$ может быть явно выражен через вектор $\mathbf{r}$:

$$\mathbf{a}^{cl}(\mathbf{r}) = \mathbf{C}\mathbf{r} - \mathbf{d}, \quad (10)$$

здесь

$$\mathbf{C} = \{c_{ik}, i \in \overline{1, n+n_1-1}, k \in \overline{1, 2n_1-1}\}$$

есть матрица, состоящая из коэффициентов полиномов $\langle a, b \rangle$ модели (1):

$$c_{ik} = \begin{cases} 0, & i < k, \\ a_{i-k}, & i \in \overline{k, n+k}, \\ 0, & i > n+k, \end{cases}$$

при $i \in \overline{1, n+n_1-1}$, $k \in \overline{1, 2n_1-1}$, и

$$c_{ik} = \begin{cases} 0, & i < k - n_1 + 1, \\ -b_{n_1+i-k-1}, & i \in \overline{k - n_1 + 1, m + k - n_1 + 1}, \\ 0, & i > m + k - n_1 + 1, \end{cases}$$

при $i \in \overline{1, n+n_1-1}$, $k \in \overline{1, 2n_1-1}$, а вектор $\mathbf{d} = \text{col}(d_1; \dots; d_{n+n_1-1})$ состоит из коэффициентов полиномов $\langle a, a^{et} \rangle$:

$$d_i = \begin{cases} a_{i-1}^{et}, & i \in \overline{1, n_1-1}, \\ a_{i-1}^{et} - a_{i-n_1}, & i \in \overline{n_1, n+n_1-1}. \end{cases}$$

Выражение (8) для расчета вектора $\mathbf{r}^*$ с учетом (10) принимает вид:

$$\mathbf{r}^* = \arg \min_{\mathbf{r}} J(\mathbf{r}),$$

$$J(\mathbf{r}) \doteq (\mathbf{C}\mathbf{r} - \mathbf{d} - \mathbf{a}^{et})^T \mathbf{Q}(\mathbf{C}\mathbf{r} - \mathbf{d} - \mathbf{a}^{et}). \quad (11)$$

Таким образом, расчет вектора $\mathbf{r}^*$ в (11) сводится к минимизации квадратичного функционала $J(\mathbf{r})$; решение данной задачи определяется выражением

$$\mathbf{r}^* = (\mathbf{C}^T \cdot \mathbf{Q} \cdot \mathbf{C})^{-1} \mathbf{C}^T \cdot \mathbf{Q} \cdot \mathbf{d} \quad (12)$$

при условии, что матрица $(\mathbf{C}^T \cdot \mathbf{Q} \cdot \mathbf{C})^{-1}$ существует (в противном случае решения не существует).

Предложенный метод синтеза модального регулятора пониженного порядка позволяет:

- 1) обосновать выбор порядка l модального регулятора;
- 2) из всех модальных регуляторов порядка l рассчитать такой, параметры которого будут оптимальны по критерию (8), т.е. близки к выбранному эталону a^{et} .

Однако данный метод не может гарантировать, что рассчитанный таким образом модальный регулятор будет обеспечивать желаемое качество управления.

Исходя из изложенного, предлагается следующий метод поиска модального регулятора пониженного порядка, обеспечивающего заданное качество управления.

Этап 1 (поиск регулятора заданного порядка): воспользоваться свободой в выборе эталона a^{et} (которая ограничена условием (9)), с тем чтобы попытаться разместить нули характеристического полинома замкнутой системы внутри заданной области S . В том случае, если после ряда попыток выбора эталона a^{et} не удалось обеспечить для замкнутой системы требуемого качества управления, переходим к следующему этапу.

Этап 2 (повышение порядка регулятора): в представлении (5) объекта управления следует отнести часть нулей полиномов структурных возмущений $\langle v, w \rangle$ в доминирующую динамику $\langle a^r, b^r \rangle$, при этом повысится порядок модели доминирующей динамики (6) и порядок модального регулятора. Очевидно, что данная операция применяется счетное число раз, поскольку в том случае, когда все нули полиномов v и w будут отнесены соответственно в полиномы a^r и b^r , мы получим исходный объект управления (1) и модальный регулятор полного порядка (устойчивость и требуемое качество управления гарантирует только модальный регулятор полного порядка).

2. Пример синтеза модального регулятора пониженного порядка

Объект управления задан дифференциальным уравнением [15. С. 15, 16]

$$a(3, p)y(t) = b(2, p)u(t), \quad (13)$$

где

$$\begin{aligned} a(3, p) &= (s + 21)(s + 3)(s - 2) = p^3 + 22p^2 + 15p - 126, \\ b(2, p) &= 10(s + 20)(s + 6) = 10p^2 + 260p + 1200. \end{aligned}$$

Требования к качеству управления замкнутой системой задаются в виде области (см.: [15. С. 15, 16])

$$S = \{s : \eta_2 \leq -\operatorname{Re}(s) \leq \eta_1, \quad |\operatorname{Im}(s_i)/\operatorname{Re}(s_i)| \leq \zeta_1\} \subset C^1,$$

где $\eta_1 = 25$, $\eta_2 = 2$, $\zeta_1 = 1$. Область S на комплексной плоскости расположена в левой полуплоскости, имеет трапециевидную форму и удовлетворяет всем вышеперечисленным требованиям. Параметр η_2 определяет допустимый запас устойчивости, а ζ_1 – допустимую колебательность.

Для сравнения рассчитаем модальные регуляторы полного и пониженного порядков. Отметим, что точка $(-\omega_0; 0j)$, где $\omega_0 = 5$, находится внутри области S . При расчетах характеристический полином эталонной системы будем назначать по биномиальной схеме [8. С. 6] с параметром $\omega_0 = 5$.

1. *Синтез регулятора полного порядка.* Регулятор полного порядка ищется на классе дифференциальных уравнений 2-го порядка

$$u^{(2)}(t) + \beta_1 u^{(1)}(t) + \beta_0 u(t) = \alpha_2 y^{(2)}(t) + \alpha_1 y^{(1)}(t) + \alpha_0 y(t) + \chi_0 g(t). \quad (14)$$

Характеристический полином эталонной системы

$$a^{et}(5, s) = (s + \omega_0)^5 = s^5 + 25s^4 + 250s^3 + 1250s^2 + 3125s + 3125.$$

В соответствии с изложенной в [10. С. 10–14] схемой синтеза модального регулятора нахождение коэффициентов регулятора (14) сводится к решению системы линейных алгебраических уравнений

$$\mathbf{A}\mathbf{r} = \mathbf{b}, \quad (15)$$

где $\mathbf{A}$ и $\mathbf{b}$ – соответственно матрица и вектор, составленные из коэффициентов операторов $a(3, p)$ и $b(2, p)$ объекта управления и характеристического полинома эталона $a^{et}(5, s)$:

$$\mathbf{A} = \begin{pmatrix} a_0 & 0 & -b_0 & 0 & 0 \\ a_1 & a_0 & -b_1 & -b_0 & 0 \\ a_2 & a_1 & -b_2 & -b_1 & -b_0 \\ a_3 & a_2 & 0 & -b_2 & -b_1 \\ 0 & a_3 & 0 & 0 & -b_2 \end{pmatrix} = \begin{pmatrix} -126 & 0 & -1200 & 0 & 0 \\ 15 & -126 & -260 & -1200 & 0 \\ 22 & 15 & -10 & -260 & -1200 \\ 1 & 22 & 0 & -10 & -260 \\ 0 & 1 & 0 & 0 & -10 \end{pmatrix},$$

$$\mathbf{b} = \begin{pmatrix} a_0^{et} \\ a_1^{et} \\ a_2^{et} - a_0 \\ a_3^{et} - a_1 \\ a_4^{et} - a_2 \end{pmatrix} = \begin{pmatrix} 3125 \\ 3125 \\ 1376 \\ 235 \\ 3 \end{pmatrix},$$

а $\mathbf{r} = \text{col}(\beta_0; \beta_1; \alpha_0; \alpha_1; \alpha_2)$ – вектор коэффициентов регулятора.

Решая систему уравнений (15), получаем

$$\mathbf{r}^* = \text{col}(990,2; 171,0; -106,6; 14,9; 16,8).$$

Таким образом, регулятор полного порядка

$$u^{(2)}(t) + 171,0u^{(1)}(t) + 990,2u(t) = 16,8y^{(2)}(t) + 14,9y^{(1)}(t) - 106,6y(t) + \chi_0 g(t).$$

2. Синтез регулятора пониженного порядка. Объект управления (13) представим в виде:

$$v(1, p) \cdot a^r(2, p)y(t) = w(1, p) \cdot b^r(1, p)u(t),$$

здесь

$$v(1, p) = (p + 21)/21, \quad w(1, p) = (p + 20)/20,$$

$$a^r(2, p) = p^2 + 1p - 6, \quad b^r(1, p) = 9,5p + 57,1.$$

Для редуцированной модели

$$a^r(2, p)y(t) = b^r(1, p)u(t),$$

модальный регулятор будем искать на классе дифференциальных уравнений 1-го порядка

$$u^{(1)}(t) + \beta_0 u(t) = \alpha_1 y^{(1)}(t) + \alpha_0 y(t) + \chi_0 g(t).$$

При расчете модального регулятора пониженного порядка по (12) выберем

$$\mathbf{Q} = \begin{pmatrix} 100 & 0 & 0 & 0 \\ 0 & 75 & 0 & 0 \\ 0 & 0 & 25 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

и эталон

$$a^{et}(4, s) = (s + \omega_0)^4 = s^4 + 20s^3 + 150s^2 + 500s + 625.$$

По формулам (10) находим $\mathbf{C}$ и $\mathbf{d}$:

$$\mathbf{C} = \begin{pmatrix} a_0 & -b_0 & 0 \\ a_1 & -b_1 & -b_0 \\ a_2 & -b_2 & -b_1 \\ 1 & 0 & -b_2 \end{pmatrix} = \begin{pmatrix} -126 & -1200 & 0 \\ 15 & -260 & -1200 \\ 22 & -10 & -260 \\ 1 & 0 & -10 \end{pmatrix},$$

$$\mathbf{d} = \text{col}(a_0^{et}; a_1^{et} - a_0; a_2^{et} - a_1; a_3^{et} - a_2) = \text{col}(625; 626; 135; -2).$$

Матрица $(\mathbf{C}^T \cdot \mathbf{Q} \cdot \mathbf{C})^{-1}$ существует; вектор $\mathbf{r}^*$, вычисленный по формуле (12), равен

$$\mathbf{r}^* = \text{col}(1,69; -0,69; -0,35).$$

Таким образом, регулятор пониженного порядка

$$u^{(1)}(t) + 1,69u(t) = -0,35y^{(1)}(t) - 0,69y(t) + \chi_0 g(t). \quad (16)$$

После замыкания объекта управления (13) регулятором (16) получаем следующий характеристический полином замкнутой системы:

$$a^{c.l.}(4, s) = s^4 + 27,18s^3 + 150,0s^2 + 500,0s + 625,0.$$

Нули характеристического полинома замкнутой системы

$$s_1 = -2,12, \quad s_{2,3} = -3,18 \pm j 1,96, \quad s_4 = -21,1$$

лежат внутри заданной области S . Корневые показатели качества замкнутой системы $\eta = 2,12$, $\zeta = 0,62$ несколько лучше заданных допустимых значений $\eta_2 = 2$, $\zeta_1 = 1$.

Заключение

В статье разработан новый метод синтеза модального регулятора пониженного порядка. Данный метод основан на разделении мод объекта управления по отношению к цели управления на доминирующую динамику (подлежащую регулированию) и структурные возмущения (уже удовлетворяющие цели управления). Расчет модального регулятора пониженного порядка получен в виде аналитической формулы, что позволяет реализовать данный метод на ЭВМ. Эффективность метода проиллюстрирована примером.

Список источников

1. Wang Q.-G., Ye Z., Cai W.-J. PID control for multivariable processes. Berlin : Springer, 2008. 266 p.
2. Бураков М.В., Коновалов А.С. Нечеткий супервизор ПИД-регулятора // Информационно-управляющие системы. 2018. № 5 (96). С. 13–21.
3. Поляк Б.Т., Хлебников М.В. Новые критерии настройки ПИД-регуляторов // Автоматика и телемеханика. 2022. № 11. С. 62–82.
4. Абрамкин С.Е., Душин С.Е. Математическое моделирование управляемых технологических процессов осушки природного газа // Информационно-управляющие системы. 2015. № 4 (77). С. 41–49.
5. Абрамкин С.Е., Душин С.Е., Сирота Д.Д. Разработка математической модели системы «пласт – газовая скважина» // Известия СПбГЭТУ «ЛЭТИ». 2019. № 3 (77). С. 10–15.
6. Брикова О.И., Душин С.Е. Анализ влияния внешних факторов на процессы биологической очистки в моделях нитрификации и денитрификации // Известия Южного федерального университета. Технические науки. 2018. № 5 (199). С. 79–88.
7. Брикова О.И., Душин С.Е. Исследование влияния температуры среды на биологические процессы в моделях типа ASM1 // Известия СПбГЭТУ «ЛЭТИ». 2019. № 5. С. 144–148.
8. Кузовков Н.Т. Модальное управление и наблюдающие устройства. М. : Машиностроение, 1976. 184 с.
9. Соловьев И.Г. Методы мажоризации в анализе и синтезе адаптивных систем. Новосибирск : Наука, 1992. 191 с.
10. Паршуков А.Н. Методы синтеза модальных регуляторов. Тюмень : ТюмГНГУ, 2009. 84 с.
11. Воропаева Н.В., Соболев В.А. Геометрическая декомпозиция сингулярно возмущенных систем. М. : Физматлит, 2009. 255 с.
12. Воропаева Н.В., Соболев В.А. Декомпозиция линейно-квадратичной задачи оптимального управления с быстрыми и медленными переменными // Автоматика и телемеханика. 2006. № 8. С. 3–11.
13. Дмитриев М.Г., Курина Г.А. Сингулярные возмущения в задачах управления // Автоматика и телемеханика. 2006. № 1. С. 3–51.
14. Kokotovic P.V., Khalil H.K., O'Reilly J. Singular Perturbations Methods in Control: Analysis and Design. New York : Academic Press, 1986. 371 p.
15. Паршуков А.Н. Метод синтеза модального регулятора пониженного порядка // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2021. № 3(56). С. 12–19.

References

1. Wang, Q.-G., Ye, Z. & Cai, W.-J. (2008) *PID control for multivariable processes*. Berlin: Springer.
2. Burakov, M.V. & Kononov, A.S. (2018) Nechetkiy supervisor PID-regulyatora [Fuzzy supervisor for PID controller]. *Informatsionno-upravlyayushchie sistemy – Information and Control Systems*. 5(96). pp. 13–21.
3. Polyak, B.T. & Khlebnikov, M.V. (2022) New criteria for tuning PID controllers. *Automation and Remote Control*. 11(83). pp. 62–82. DOI: 10.31857/S0005231022110022

4. Abramkin, S.E. & Dushin, S.E. (2015) Mathematical modeling of controlled technological processes of natural gas drying. *Informatsionno-upravlyayushchie sistemy – Information and Control Systems*. 4(77). pp. 41–49.
5. Abramkin, S.E., Dushin, S.E. & Sirota, D.D. (2019) Development of a mathematical model of the plast – gas well system. *Izvestiya SPBGETU LETI*. 3(77). pp. 10–15.
6. Brikova, O.I. & Dushin, S.E. (2018) Analysis of the influence of external factors on biological purification processes in nitrification and denitrification models. *Izvestiya Yuzhnogo federal'nogo universiteta. Tekhnicheskie nauki – Izvestiya Sfedu. Engineering Sciences*. 5(199). pp. 79–88.
7. Brikova, O.I. & Dushin, S.E. (2019) Investigation of the influence of ambient temperature on biological processes in models of the type ASM1. *Izvestiya SPBGETU LETI*. 5. pp. 144–148.
8. Kuzovkov, N.T. (1976) *Modal'noe upravlenie i nablyudayushchie ustroystva* [Modal Control and Monitoring Devices]. Moscow: Mashinostroenie.
9. Soloviev, I.G. (1992) *Metody mazhorizatsii v analize i sinteze adaptivnykh sistem* [Methods of majorization in the analysis and synthesis of adaptive systems]. Novosibirsk: Nauka.
10. Parshukov, A.N. (2009) *Metody sinteza modal'nykh regulyatorov* [Methods of Synthesis of Modal Regulators]. Tyumen: TyumGNGU.
11. Voropaeva, N.V. & Sobolev, V.A. (2009) *Geometricheskaya dekompozitsiya singulyarno vozmushchennykh sistem* [Geometric decomposition of singularly perturbed systems]. Moscow: Fizmatlit.
12. Voropaeva, N.V. & Sobolev, V.A. (2006) Decomposition of a linear-quadratic optimal control problem with fast and slow variables. *Automation and Remote Control*. 67(8). pp. 1185–1193.
13. Dmitriev, M.G. & Kurina, G.A. (2006) Singular perturbations in control problems. *Automation and Remote Control*. 67(1). pp. 1–43.
14. Kokotovic, P.V., Khalil, H.K. & O'Reilly, J. (1986) *Singular Perturbations Methods in Control: Analysis and Design*. New York: Academic Press.
15. Parshukov A.N. (2021) Method of synthesis of a reduced-order modal regulator. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 3(56). pp. 12–19. DOI: 10.17223/19988605/56/2

Информация об авторе:

Паршуков Андрей Николаевич – доцент, кандидат технических наук, доцент кафедры «Электроэнергетика» Тюменского индустриального университета (Тюмень, Россия). E-mail: anparshukov@mail.ru

Автор заявляет об отсутствии конфликта интересов.

Information about the author:

Parshukov Andrej N. (Candidate of Technical Sciences, Associate Professor, Industrial University of Tyumen, Tyumen, Russian Federation). E-mail: anparshukov@mail.ru

The author declares no conflicts of interests.

Поступила в редакцию 03.03.2023; принята к публикации 04.09.2023

Received 03.03.2023; accepted for publication 04.09.2023

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ

MATHEMATICAL MODELING

Научная статья

УДК 004.94

doi: 10.17223/19988605/64/4

Аспектно-предикатное представление компонентных цепей
для адаптивного анализаВячеслав Михайлович Дмитриев¹, Максим Игоревич Кочергин², Тарас Викторович Ганджа³^{1, 2, 3} Томский государственный университет систем управления и радиоэлектроники, Томск, Россия¹ dmitriewvm@gmail.com² maksim.i.kochergin@tusur.ru³ gtv@main.tusur.ru

Аннотация. Предлагается подход к построению формального описания портрета исследуемой системы для последующего адаптивного анализа ее математической модели. Подход заключается в описании классов задач группами предикатов, характеризующими их поведение и особенности, с дальнейшим сведением их моделей к композиции ограниченного набора канонических форм в виде системы линейных алгебраических уравнений, эквивалентной исходной математической модели системы и задающих модель цепи компонентов и узловых законов сохранения. Приводятся аспектно-предикатная характеристика моделей, предикатная классификация решаемых задач, а также примеры задач с их предикатной характеристикой.

Ключевые слова: математическое моделирование; компонентные цепи; аспектно-предикатное описание; адаптивный анализ.

Для цитирования: Дмитриев В.М., Кочергин М.И., Ганджа Т.В. Аспектно-предикатное представление компонентных цепей для адаптивного анализа // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2023. № 64. С. 30–40. doi: 10.17223/19988605/64/4

Original article

doi: 10.17223/19988605/64/4

Aspect-predicate representation of component circuits for adaptive analysis

Vyacheslav M. Dmitriev¹, Maxim I. Kochergin², Taras V. Gandzha³^{1, 2, 3} Tomsk State University of Control Systems and Radioelectronics, Tomsk, Russian Federation¹ dmitriewvm@gmail.com² maksim.i.kochergin@tusur.ru³ gtv@main.tusur.ru

Abstract. The paper proposes an approach to constructing a formal description of the portrait of the system for adaptive analysis of its mathematical model. The approach consists in describing classes of systems by groups of predicates that characterize their behavior and features. Then the model of the system is reduced to a composition of a limited set of canonical forms (equivalent to the original), which consist of a system of linear algebraic equations that define the model of the circuits, components, and nodal conservation laws. The paper provides an aspect-predicate characteristic of models, a predicate classification of the systems being modeled and examples of models with their predicate characteristic.

Keywords: mathematical modeling; component circuits; aspect-predicate description; adaptive analysis.

For citation: Dmitriev, V.M., Kochergin, M.I., Gandzha, T.V. (2022) Aspect-predicate representation of component circuits for adaptive analysis. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naja tehnika i informatika – Tomsk State University Journal of Control and Computer Science*. 64. pp. 30–40. doi: 10.17223/19988605/64/4

Введение

При математическом моделировании систем со сложной структурой и подсистемами различной физической природы актуальной задачей является подбор режимов анализа модели или замены (упрощения) математических моделей некоторых подсистем, которые обеспечат требуемую точность анализа при наименьших временных затратах. В работах [1, 2] заложены основы адаптивного анализа физически неоднородных цепей (математических моделей систем, сочетающих в себе связи различной физической природы) в рамках метода компонентных цепей [3, 4]. Адаптивный анализ в общем случае предполагает следующие варианты реализации, которые могут применяться отдельно или в комбинациях:

- 1) выбор оптимального режима анализа компьютерной модели, обеспечивающего требуемую точность при простоте программно-инструментальных средств или минимальной вычислительной сложности [5–7];
- 2) выбор момента переключения режима анализа цепи с одного анализа на другой;
- 3) выбор момента переключения во время анализа цепи с одной модели компонента на другую (более простую или более сложную) и наоборот [8–10];
- 4) подбор шага интегрирования в соответствии с выбранной схемой анализа цепи [11–13].

В данной работе предлагается создание формализованного описания исследуемой цепи для выбора оптимального режима ее анализа (согласно п. 1 из списка выше) с применением предикатного описания свойств систем. Исследователи также предлагают комбинированные методы формализованного описания систем [14, 15], которые можно рассматривать в качестве инструмента.

1. Основные типы М-моделей

Рассмотрим основные типы М-моделей [16] непрерывных цепей со сосредоточенными параметрами. Сформулируем принципы классификации этого класса цепей и установим связь между классами. Полагаем, что при исследовании цепей определенной физической природы существует (имеется или строится) библиотека М-моделей. Для описания реального объекта необходимо выбрать определенные модели из библиотеки моделей, для чего следует построить систему качественных и количественных оценок.

Будем считать, что компонент цепи описывается признаками (предикатами) трех следующих множеств:

- 1) топологические предикаты: $ПТ = \{ПТ_j\}, j \in J_T$;
- 2) физические предикаты $ПФ = \{ПФ_j\}, j \in J_\Phi$;
- 3) математические предикаты: $ПМ = \{ПМ_j\}, j \in J_M$.

В принципе реальный объект (компонент цепи) должен описываться тем же набором предикатов, что и его модель. Построение набора предикатов для компонентов определяет соответствующую предикатную группу, что позволяет более точно определить класс этой модели.

Предикатное представление позволяет:

1. Более четко классифицировать модели компонентов и объектов согласно образованным в соответствии с этим представлением предикатным группам.
2. Применить к построению программно-алгоритмического инструментария системы моделирования объектно-ориентированные подходы.

3. Адаптировать оптимальным образом вычислительное ядро системы моделирования на конкретный класс задач.

4. Накапливать классы решенных задач в базах знаний с применением методов искусственного интеллекта для их эффективного решения.

Рассмотрим подробнее элементы каждого предикатного множества. Подчеркнем, что список этих элементов может уточняться и расширяться для конкретного класса цепей.

2. Топологические предикаты

Множество топологических предикатов ПТ задает характеристики компонента, связанные со способом его включения в компонентную цепь (КЦ). КЦ объекта будем рассматривать в общем виде как тройку объектов

$$C = (K, B, N). \quad (1)$$

Здесь $K = (K_E, K_M, K_P, K_S)$ – множество различных компонентов: K_E – элементарные компоненты, K_P – множество подцепей (с открытой внутренней структурой), K_M – множество макрокомпонентов (с закрытой (автономной) структурой), K_S – множество структур (из компонентов, подцепей и макрокомпонентов):

$$K_S = (K_E \cup K_P \cup K_M),$$

$B = (B_n, B_o) \in B_v$ – множество неориентированных (двусторонних – B_n) и ориентированных (B_o) связей компонентов, принадлежащих множеству векторных связей B_v ($v = 1, 2, \dots, m$), оно содержит в себе информационные $B(1)$, энергетические $B(2)$ и многопоточные (векторные) $B(m)$ связи формата $B(i)$, где i – размер вектора переменных [17]. В (1) множеством N обозначим узлы объединения связей.

3. Физические предикаты

Физические предикаты выполняют функцию качественной оценки режима работы компонента цепи, на основании которой осуществляется выбор математических предикатов. Физические предикаты модели компонента могут быть связаны с внутренними параметрами компонента (ПФ_К-предикаты) и с характером переменных, действующих на связях компонента (ПФ_С-предикаты):

ПФ_К = {состав внутренних параметров, инерционность, безынерционность, сосредоточенные и распределенные параметры, непрерывность, дискретность};

ПФ_С = {состав и типы переменных на связях, стационарность и нестационарность переменных, детерминированность и стохастичность, большая и малая амплитуды, высокая и низкая частоты}.

Между элементами ПФ_К и ПФ_С обычно существует определенная корреляция. Так, например, большой амплитуде переменных связей соответствует обычно нелинейный характер компонента.

Опишем свойства, задаваемые физическими предикатами:

а) *законы*, которые необходимы для накопления в библиотеку законов, так как на них строятся модели элементарных компонентов: электротехники, электроники, механики и гидравлики, а также модели датчиков и исполнителей;

б) *характер структуры объектов*, которая может носить как однородный, так и неоднородный характер, что необходимо учитывать как при выборе моделей законов, так и при построении элементов сопряжения различных типов неоднородности;

в) *типы сигналов*: аналоговые, дискретные, – данный признак также указывает на выбираемые для описания математические предикаты в форме непрерывного или дискретного анализа процессов;

г) *диапазоны и режимы*, которые выступают в роли ограничений на амплитудно-временные или амплитудно-частотные характеристики сигнал-переменных;

в) *переменные связей*: информационные – с одной сигнал-переменной, привязанной к определенному контакту V_{Ni} ; энергетические – с двумя переменными: потокового V_{Bi} и потенциального V_{Ni} типов; векторные (шины) – с векторным потоком переменных в связи V_v .

4. Математические предикаты

Математические предикаты (ПМ) во многом определяются совокупностью физических и топологических предикатов. Они характеризуют свойства собственно М-моделей. Будем полагать, что множество ПМ образовано следующим образом:

ПМ = {динамичность, статичность, линейность, нелинейность, временная форма, частотная форма, сосредоточенные параметры, распределенные параметры}.

Данные предикаты после детерминирования помогают выявить определенные классы уравнений, описывающих объект, а также позволяют сформулировать критерии адаптивного переключения классов уравнений модели для повышения эффективности функционирования вычислительного ядра [2].

Статические модели не содержат времени t и производных по времени. Динамические модели либо содержат время t , либо обладают инерционными свойствами, т.е. содержат производные по времени. Статические модели подразделяются на линейные и нелинейные. Первые описываются линейными относительно переменных алгебраическими уравнениями. Нелинейные модели могут включать в себя нелинейные алгебраические, тригонометрические, показательные и прочие типы нелинейных относительно внешних переменных уравнений.

Динамические М-модели также делятся на линейные и нелинейные. Линейные безынерционные модели задаются относительно переменных алгебраическими уравнениями с зависящими от времени коэффициентами. Линейные инерционные динамические модели задаются во временной форме обыкновенными дифференциальными уравнениями, линейными относительно производных, в нормальном виде, с линейной относительно переменных правой частью. Эти модели допускают представление в частотной форме (на основе преобразования Лапласа) в виде линейных относительно действительных и мнимых составляющих внешних переменных.

Методом КЦ допускается три типа уравнений относительно переменных связей компонента $V^k = V_n^k \cup V_b^k$:

– линейные

$$\Phi(V^k) = c; \quad (3)$$

– нелинейные

$$f(V^k) = 0; \quad (4)$$

– дифференциальные

$$\frac{d\Psi(V^k)}{dt} = f(V^k, t). \quad (5)$$

В (3)–(5) $\Phi(V^k) = \sum_{i=1}^{n+m} a_i V_i^k = a_1 V_{n1}^k + a_2 V_{n2}^k + \dots + a_{n+1} V_{b1}^k + \dots + a_{n+m} V_{bm}^k$ – линейная форма относительно переменных связей V^k , $f(V^k, t)$ – нелинейная функция, функция

$$\frac{d\Psi(V^k)}{dt} = \sum_{i=1}^{n+m} \psi_i \frac{dV_i^k}{dt} = \psi_1 \frac{dV_{n1}^k}{dt} + \psi_2 \frac{dV_{n2}^k}{dt} + \dots + \psi_{n+m} \frac{dV_{bm}^k}{dt}$$

является линейной формой относительно производных от переменных связей по времени, $\psi_1, \psi_2, \dots, \psi_{n+m}$ – коэффициенты при производных, n, m – количество узлов и ветвей компонента соответственно.

В общем случае правая часть уравнения с и коэффициенты при производных $\psi_i, i = 1, 2, \dots, n + m$, являются функциями времени.

Уравнения вида (3)–(5) представляют собой необходимый и достаточный набор уравнений, которыми может быть описана математическая модель любого элемента непрерывных цепей с сосредоточенными параметрами, в том числе если она содержит производные высших порядков. Приведенные уравнения (3)–(5) группируют канонические формы А, Б, В, Г, Д [1] к трем типам уравнений: {Г, Д} – тип линейных уравнений вида (3), {В} – тип линейных и нелинейных статических уравнений вида (4), {А, Б} – тип нелинейных статических и дифференциальных различных порядков вида (5). Математические модели компонентов, не относящиеся к указанным классам, приводятся к канониче-

ским формам путем введения дополнительных переменных и уравнений моделей на этапе разработки вычислительной модели компонента.

5. Аспектно-предикатное представление математических моделей

Между различными М-моделями существует взаимосвязь, обусловленная иерархическими зависимостями. Рассмотрим возможные переходы моделей из одного класса в другой. Введем следующие обозначения одноместных предикатов, принимающих значение истины, если система X обладает соответствующим свойством: $L(X)$ – статичность модели; $T(X)$ – динамичность модели; $\Theta(X)$ – линейность относительно внешних параметров; N – инерционность модели; $Q(X)$ – линейность относительно производных; $D(X)$ – признак наличия производных высших порядков; $D_1(X)$ – дифференциальные уравнения первого порядка; $B(X)$ – нормальный вид дифференциального уравнения; $\bar{C}(X)$ – частотная форма; $P(X)$ – распределенные параметры.

Основная иерархическая зависимость М-моделей приведена на рис. 1. Класс М-модели определяется совокупностью предикатов – его название будем формировать из совокупности обозначений (букв) соответствующих предикатов. Так, например, $ML\Theta$ – линейная статическая модель, а $ML\bar{\Theta}$ – нелинейная статическая модель.

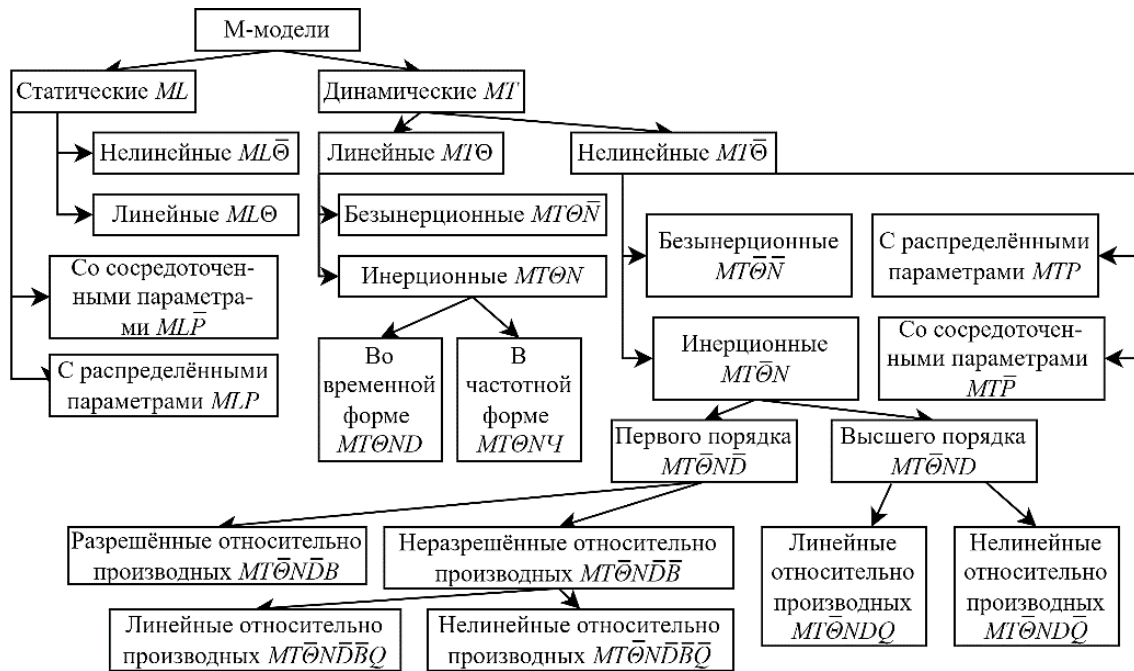


Рис. 1. Иерархическая зависимость М-моделей в предикатных обозначениях

Fig. 1. Hierarchical dependence of M-models in predicate notation

Класс модели цепи определяется как

$$KM = \bigcup_{i=1}^m KL_i, \quad (2)$$

где KM – совокупность предикатов, описывающих цепь (модель) в целом (множество предикатов, истинных для модели), KL_i – множество предикатов (без повторения), описывающих i -й компонент, входящий в рассматриваемую цепь.

Математическая модель КЦ образуется объединением моделей компонентов вида (3)–(5) и уравнений топологических законов сохранения для потоковых переменных всех узлов КЦ за исключением базового. Узловые топологические законы имеют следующую формулировку:

1) равенство одноименных потенциальных переменных, действующих на ветвях, инцидентных каждому из узлов;

2) равенство нулю суммы одноименных потоковых переменных (знак каждой из переменных задается ориентацией ветви).

6. Адаптивный анализ математических моделей

В данном разделе рассматриваются следующие варианты адаптивного анализа математических моделей из числа представленных во введении:

- 1) выбор момента переключения одной модели компонента на другую во время анализа цепи;
- 2) выбор оптимального режима анализа (расчета) компонентной цепи.

Для реализации адаптивного анализа первым подходом предлагается сформировать адаптивную модель компонента [2] как совокупность его физических, топологических и математических предикатов, которые определяют режим функционирования компонента:

$$AMK = \{(\bigcup_{i \in I_{MF}} PF_i) \cup (\bigcup_{i \in I_{MT}} PT_i) \cup (\bigcup_{i \in I_{MM}} PM_i)\}.$$

Отмечается наличие связей между предикатной структурой модели и переменными связями (величина переменной, ее форма, физический тип, скорость изменения), что позволяет сформулировать критерии переключения модели компонента. Другими словами, в пространстве состояний компонента (образованного его предикатным описанием, значением его переменных, скоростью их изменения и текущим моментом времени моделирования) можно выделить области предпочтительности, которые и задают критерий переключения модели между режимами.

Для реализации адаптивного анализа с применением второго подхода необходимо на основе аспектно-предикатного описания математической модели компонентной цепи привести ее к канонической форме с заданным режимом анализа (решения). Опишем далее общий порядок определения класса модели цепи на основе классов входящих в нее компонентов и последующий порядок анализа ее математической модели.

Каждому компоненту библиотеки моделей компонентов ставится в соответствие предикатное описание его класса. Класс модели цепи задается совокупностью таких предикатных моделей, составляющей цепь компонентов. Если в класс модели войдут взаимно противоречивые предикаты (например, линейность одного компонента и нелинейность другого), то выбирается тот предикат, который задает более общий случай. Таким образом, классификации М-моделей компонентов и цепей совпадают. Затем математическая модель цепи сводится к определенной совокупности канонических форм согласно определенным правилам, фрагмент которых представлен в таблице. Таким образом, анализ цепи любого класса осуществляется решением системы линейных алгебраических уравнений, включающих в себя блоки А, Б, В, Г, Д.

Соответствие класса цепи методам анализа

Класс цепи	Метод анализа
$ML\bar{\Theta}, MT\bar{\Theta}C$	Имеем линейное уравнение $\Phi(V^k) = c$, где Φ – некоторая линейная функция относительно внешних параметров, а c – постоянный коэффициент. Такие модели изначально соответствуют канонической форме Г
$ML\bar{\Theta}, MT\bar{\Theta}N$	Имеем нелинейное уравнение $f(V^k) = 0$, которое приводимо к канонической форме А или Б. Основным методом решения является метод Ньютона – модель заменяется в заданной точке V^0 уравнением касательной к плоскости: $\sum_{\alpha} \frac{dF(V)}{dV^{\alpha}} \Big _{V=V^0} V^{\alpha} = F_1$
$MT\bar{\Theta}, MT\bar{\Theta}D$	Имеем линейные дифференциальные уравнения в явной форме с линейной правой частью: $\frac{d\Psi(V^k)}{dt} = f(V^k, t)$. Переход в каноническую форму осуществляется на основе выбранного метода интегрирования (явный метод Эйлера, с применением разложения в ряд Тейлора, по четырехточечной схеме Рунге–Кутты и пр.).
$ML\bar{\Theta}, MT\bar{\Theta}N, MT\bar{\Theta}ND, MT\bar{\Theta}ND_1, MT\bar{\Theta}ND\bar{B}$	Имеем линейные дифференциальные в неявной форме с нелинейной правой частью: $\frac{d\Psi(V^k)}{dt} = f(V^k, t)$, где $\Psi(V)$ – линейная форма с постоянными коэффициентами, f – произвольная функция. Приведение к канонической форме осуществляется аналогичным образом

Реализация такого механизма классификации компонентной цепи и ее приведения к канонической форме с минимальным участием человека-исследователя (в автоматизированном режиме) представляется актуальной задачей, решаемой с применением технологии баз знаний и экспертных систем и является перспективой дальнейших исследований, выходящих за рамки данной статьи.

На основании данных таблицы можно сделать вывод о возможности представления обширного класса моделей со сосредоточенными параметрами ограниченным количеством канонических форм А, Б, В, Г, Д.

Рассмотрим ряд иллюстративных примеров компонентов различной физической природы, классифицируемых различными предикатными группами.

Пример 1. Электромеханический преобразователь

На рис. 2 приведены принципиальная схема (а) и графическое обозначение компонента (b) электромеханического преобразователя.

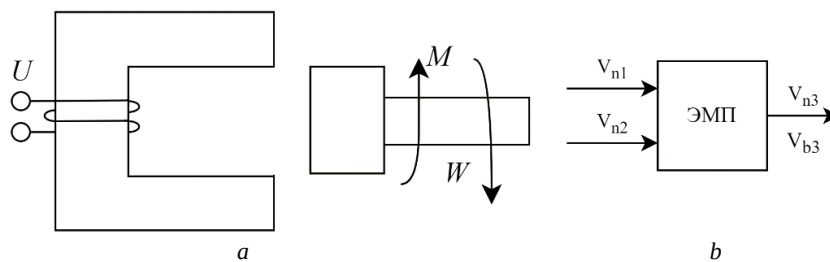


Рис. 2. Электромеханический преобразователь: а – схема, b – компонент

Fig. 2. Electromechanical transducer: a) scheme, b) component

Данный компонент описывается следующей системой уравнений:

$$\begin{cases} V_{n3} = \frac{(V_{n2} - V_{n1})}{H}, \\ V_{b3} = H \cdot V_{b2}, \\ V_{b2} = V_{b1}, \end{cases}$$

где V_{n1} и V_{n2} – напряжения на зажимах обмотки намагничивания; $V_{n3} = w$ – угловая скорость вращения якоря; V_{b1} и V_{b2} – токи в обмотке подмагничивания; $V_{b3} = M$ – момент на якоре; H – постоянная электромагнитной связи. Уравнения F_1 , F_2 электромагнитного привода соответствуют модели класса $ML\Theta$.

Пример 2. Инерционный элемент для измерения параметров вибрации

Схема инерционного элемента вибратора и его графическое обозначение приведены на рис. 3.

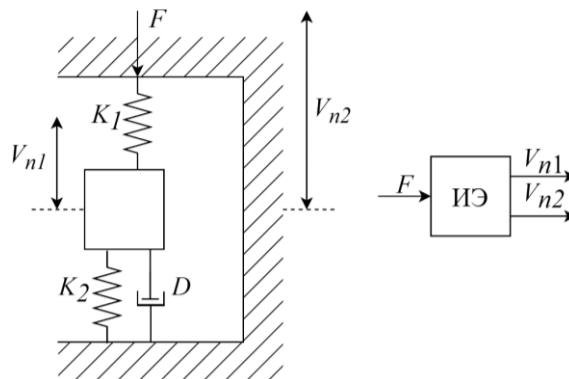


Рис. 3. Схема и компонент инерционного вибратора

Fig. 3. Schematic and component representation of an inertial vibrator

Данный компонент описывается следующей системой уравнений:

$$\begin{cases} m \frac{d^2(V_{n2} - V_{n1})}{dt^2} + D \frac{d(V_{n2} - V_{n1})}{dt} + K(V_{n2} - V_{n1}) = 0, \\ m \frac{d^2(V_{n2} - V_{n1})}{dt^2} = F, \end{cases}$$

где $K = K_1 + K_2$ – коэффициенты жесткости пружин; m – масса инерционного тела; D – коэффициент демпфирования, V_{n2} – перемещение тела по отношению к корпусу прибора, V_{n1} – усилие, передаваемое на вибратор.

Модель рассматриваемого компонента относится к классу $MTNQD$.

Пример 3. Зарядовая модель p – n -перехода

На рис. 4 представлена схема модели p – n -перехода и эк компонент.

Уравнения модели p – n -перехода, соответствующие этой эквивалентной схеме:

$$\frac{d(V_{n1} - V_{n2})}{dt} = \frac{I - I_d - (V_{n1} - V_{n2}) / R_y}{C_y},$$

где I_d – ток диода; $(V_{n1} - V_{n2})$ – напряжение на переходе; R_y – сопротивление утечки перехода; C_y – суммарная барьерная и диффузионная емкости перехода.

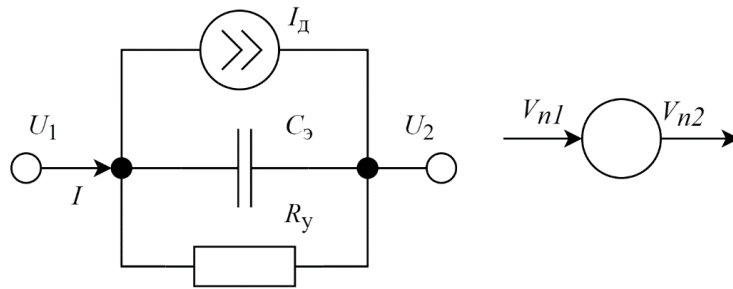


Рис. 4. Схема и компонент зарядовой модели p – n -перехода
Рис. 4. Schematic and component representation of an p - n junction model

Модель перехода, соответствующая рис. 4, относится к классу $MT\bar{\Theta}ND$.

Пример 4. Выпрямитель с фильтром

Далее рассмотрим пример определения класса компонентной цепи на примере схемы выпрямителя с фильтром. На рис. 5 представлена его схема.

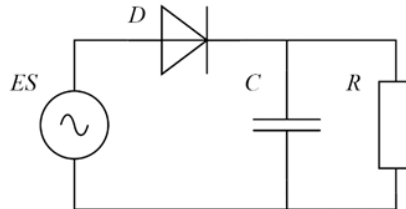


Рис. 5. Схема выпрямителя с фильтром
Fig. 5. Rectifier with filter

Модели элементы схемы соответствуют следующим классам: ES – безынерционная, линейная относительно внешних параметров (класс $MTN\bar{\Theta}$); D – динамическая, нелинейная относительно внешних параметров, инерционная, с дифференциальным уравнением первого порядка (класс $MT\bar{\Theta}ND_1$); C – динамическая, инерционная, линейная относительно производных, с дифференциальным уравнением первого порядка (класс $MTNQD_1$); R – безынерционная линейная модель (класс $MN\bar{\Theta}$).

Класс модели цепи определяется опросом этих предикатных групп и их обработкой по формуле (2). Класс данной цепи образован следующей совокупностью предикатов $MT\bar{O}NQD_1$.

Заключение

Предложенное в работе аспектно-предикатное представление математических моделей систем позволяет описать широкий класс задач в области физики и техники. На основании предикатного описания систем становится возможной реализация адаптивного анализа ее математической модели. В общем случае адаптивный анализ состоит в подборе оптимального режима анализа модели (компонентной цепи) или ее субмоделей (моделей отдельных компонентов цепи) в зависимости от структуры анализируемой математической модели и критерия оптимальности, который заключается, как правило, в уменьшении вычислительной сложности модели. Описанный в работе подход позволяет привести математическую модель исследуемой системы к определенной совокупности канонических форм, что дает возможность производить анализ цепи любого класса решением системы линейных алгебраических уравнений. Таким образом, рассматриваемые в работе канонические формы моделей позволяют ограничиться при анализе математической модели минимальным набором основных операций. Различные методы анализа в таком случае сводятся к различным комбинациям основных операций. Наличие канонических форм дает возможность организовать эффективное модульное представление алгоритмов анализа. Это позволяет реализовывать новые алгоритмы с минимальными затратами, допуская широкое численное экспериментирование для выбора оптимальных методов исследования конкретных классов цепей. На основании предложенной структуры моделей компонентов и модели цепи можно осуществлять частичную перестройку модели цепи по блокам канонических форм, реализуя тот или иной шаг алгоритма моделирования. Приведенные в работе примеры алгоритмов анализа цепей отдельных классов подтверждают модульность и адаптивность предложенного подхода для представленных классов моделей.

Список источников

1. Арайс Е.А., Дмитриев В.М. Моделирование неоднородных цепей и систем на ЭВМ. М. : Радио и связь, 1982. 160 с.
2. Арайс Е.А. Адаптивный анализ цепей // Вопросы программирования и автоматизации проектирования. Томск : Изд-во ТГУ, 1979. Вып. 4. С. 3–24.
3. Дмитриев В.М., Шутенков А.В., Зайченко Т.Н., Ганджа Т.В. MAPC – среда моделирования технических устройств и систем. Томск : В-спектр, 2011. 277 с.
4. Зайченко Т.Н. Объектная ориентация универсального метода компонентных цепей для решения задач моделирования электротехнических устройств и систем // Информационные технологии. 2006. № 1. С. 18–26.
5. Долгонос А.М., Прудковский А.Г., Колотилина Н.К. Моделирование новой хроматографической системы с помощью программы IONCHROM и выбор оптимального режима хроматографического анализа // Журнал аналитической химии. 2016. Т. 71, № 7. С. 731–738.
6. Стремилова О.С., Кац И.М., Дикович В.В., Завьялов П.Б. Определение оптимального способа анализа электрических режимов и выбор средств снижения несимметрии напряжений в районах с высокой долей потребления электроэнергии тяговой нагрузкой // Электроэнергетика глазами молодежи – 2017 : материалы VIII Междунар. науч.-техн. конф. Самара, 2017. С. 14–17.
7. Банков С.Е., Курушин А.А. Проектирование СВЧ устройств и антенн с Ansoft HFSS. М., 2009. 736 с.
8. Петунин В.И. Математические модели многосвязных систем автоматического управления с селекторами каналов // Вестник Уфимского государственного авиационного технического университета. 2011. Т. 15, № 2 (42). С. 52–58.
9. Александрова М.И., Антонов В.И. Оценивание фазы включения шунтирующего реактора методами адаптивного структурного анализа // Электрические станции. 2022. № 4 (1089). С. 40–47.
10. Воробьев Е.С., Антонов В.И., Иванов Н.Г., Наумов В.А., Солдатов А.В. Многоканальный адаптивный структурный анализ // Современные тенденции развития цифровых систем релейной защиты и автоматики : материалы науч.-техн. конф. молодых специалистов в рамках форума «РЕЛАВЭКСПО-2021». Чебоксары, 2021. С. 114–120.
11. Krishnamoorthy C.S., Meher Prasad A., Vinu Unnithan U. Adaptive Finite Element Analysis of Kinetically Non-Linear Elastoplastic Problems in Two Dimensions // Computational Mechanics – New Frontiers for the New Millennium / S. Valliappan, N. Khalili (eds.). Elsevier Science, 2001. P. 79–84. doi: 10.1016/B978-0-08-043981-5.50017-1

12. Chang-Koon C., Tae-Yeol L. High Performance Variable-Node Element Libraries for Structural Engineering Applications // Computational Mechanics – New Frontiers for the New Millennium / S. Valliappan, N. Khalili (eds.). Elsevier Science, 2001. P. 187–194. doi: 10.1016/B978-0-08-043981-5.50031-6
13. Boroomand B., Zienkiewicz O.C. Recovery procedures in error estimation and adaptivity: Adaptivity in non-linear problems of elasto-plasticity behaviour // Studies in Applied Mechanics. 1998. V. 47. P. 383–410. doi: 10.1016/S0922-5382(98)80022-6
14. Черненко В.М., Гапанюк Ю.Е., Ревунков Г.И., Терехов В.И., Каганов Ю.Т. Метаграфовый подход для описания гибридных интеллектуальных информационных систем // Прикладная информатика. 2017. Т. 12, № 3 (69). С. 57–79.
15. Мелехин В.Б., Хачумов В.М., Хачумов М.В. Логика условно-зависимых предикатов в решателях задач автономных интеллектуальных систем // Сборник трудов XIII Всероссийского совещания по проблемам управления ВСПУ-2019. М. : Ин-т проблем управления им. В.А. Трапезникова РАН, 2019. С. 1936–1940.
16. Зайченко Т.Н. Разработка новой методологии анализа функционирования систем управления // Доклады Томского государственного университета систем управления и радиоэлектроники. 2012. № 2-1 (26). С. 151–156.
17. Ганджа Т.В. Формализованное представление обобщенного технически сложного объекта с компьютерной моделью в контуре управления // Приборы и системы. Управление, контроль, диагностика. 2012. № 2. С. 29–35.

References

1. Arays, E.A. & Dmitriev, V.M. (1982) *Modelirovanie neodnorodnykh tsepey i sistem na EVM* [Modeling of heterogeneous circuits and systems on a computer]. Moscow: Radio i svyaz'.
2. Arays, E.A. (1979) Adaptivnyy analiz tsepey [Adaptive circuit analysis]. In: Arays, E.A. (ed.) *Voprosy programmirovaniya i avtomatizatsii proektirovaniya* [Issues of programming and design automation]. Vol. 4. Tomsk: TSU. pp. 3–24.
3. Dmitriev, V.M., Shutenkov, A.V., Zaychenko, T.N. & Gandzha, T.V. (2011) *MARS – sreda modelirovaniya tekhnicheskikh ustroystv i sistem* [MARS – an environment for modeling technical devices and systems]. Tomsk: V-spektr.
4. Zaychenko, T.N. (2006) Ob"ektnaya orientatsiya universal'nogo metoda komponentnykh tsepey dlya resheniya zadach modelirovaniya elektrotekhnicheskikh ustroystv i sistem [Object orientation of the universal method of component circuits for solving problems of modeling electrical devices and systems]. *Informatsionnye tekhnologii – Information Technologies*. 1. pp. 18–26.
5. Dolgonosov, A.M., Prudkovskiy, A.G. & Kolotilina, N.K. (2016) Modelirovanie novoy khromatograficheskoy sistemy s pomoshch'yu programmy IONCHROM i vybor optimal'nogo rezhima khromatograficheskogo analiza [Modeling a new chromatographic system using the IONCHROM program and choosing the optimal chromatographic analysis mode]. *Zhurnal analiticheskoy khimii – Journal of Analytical Chemistry*. 71(7). pp. 731–738.
6. Stremilova, O.S., Kats, I.M., Dikovich, V.V. & Zavyalov, P.B. (2017) Opredelenie optimal'nogo sposoba analiza elektricheskikh rezhimov i vybor sredstv snizheniya nesimmetrii napryazheniy v rayonakh s vysokoy doley potrebleniya elektroener-gii tyagovoy nagruzkoy [Determining the optimal method for analyzing electrical modes and choosing means to reduce voltage unbalance in areas with a high share of electricity consumption by traction load]. *Elektroenergetika glazami molodezhi – 2017* [Electric power Industry Through the Eyes of Youth – 2017]. Proc. of the 8th International Conference. Samara. pp. 14–17.
7. Bankov, S.E. & Kurushin, A.A. (2009) *Proektirovanie SVCh ustroystv i anten s Ansoft HFSS* [Designing microwave devices and antennas with Ansoft HFSS]. Moscow: [s.n.].
8. Petunin, V.I. (2011) Mathematical Models of Multiconnected Automatic Control Systems with Channel Selectors. *Vestnik Ufimskogo gosudarstvennogo aviatsionnogo tekhnicheskogo universiteta*. 2(42). pp. 52–58.
9. Aleksandrova, M.I. & Antonov, V.I. (2022) Shunt reactor switch-on phase estimation by adaptive structural analysis methods. *Elektricheskie stantsii – Electrical Stations*. 4(1089). pp. 40–47.
10. Vorobiev, E.S., Antonov, V.I., Ivanov, N.G., Naumov, V.A. & Soldatov, A.V. (2021) Mnogokanal'nyy adaptivnyy strukturnyy analiz [Multichannel adaptive structural analysis]. *Sovremennye tendentsii razvitiya tsifrovyykh sistem releynoy zashchity i avtomatiki* [Modern trends in the development of digital systems of relay protection and automation]. Proc. of the Conference of Young Professionals within RELAVEXPO-2021. Cheboksary. pp. 114–120.
11. Krishnamoorthy, C.S., Meher Prasad, A. & Vinu Unnithan, U. (2001) Adaptive Finite Element Analysis of Kinematically Non-Linear Elasto-Plastic Problems in Two Dimensions. In: Valliappan, S. & Khalili, N. (eds) *Computational Mechanics–New Frontiers for the New Millennium*. Elsevier Science. pp. 79–84. DOI: 10.1016/B978-0-08-043981-5.50017-1
12. Chang-Koon, C. & Tae-Yeol, L. (2001) High Performance Variable-Node Element Libraries for Structural Engineering Applications. In: Valliappan, S. & Khalili, N. (eds) *Computational Mechanics–New Frontiers for the New Millennium*. Elsevier Science. pp. 187–194. DOI: 10.1016/B978-0-08-043981-5.50031-6
13. Boroomand, B. & Zienkiewicz, O.C. (1998) Recovery procedures in error estimation and adaptivity: Adaptivity in non-linear problems of elasto-plasticity behaviour. *Studies in Applied Mechanics*. 47. pp. 383–410. DOI: 10.1016/S0922-5382(98)80022-6
14. Chernenkiy, V.M., Gapanyuk, Yu.E., Revunkov, G.I., Terekhov, V.I. & Kaganov, Yu.T. (2017) Metaografyvy podkhod dlya opisaniya gibridnykh intellektual'nykh informatsionnykh sistem [Metagraph Approach for Describing Hybrid Intelligent Information Systems]. *Prikladnaya informatika – Applied Informatics*. 3(69). pp. 57–79.
15. Melekhin, V.B., Khachumov, V.M. & Khachumov, M.V. (2019) Logika uslovno-zavisimyykh predikatov v reshatel'nykh zadach avtonomnykh intellektual'nykh sistem [Logic of Conditionally Dependent Predicates in Problem Solvers of Autonomous Intelligent Systems]. In: *Sbornik trudov XIII Vserossiyskogo soveshchaniya po problemam upravleniya VSPU-2019* [Proceedings of the 13th All-Russian Conference on Management Problems of the VSPU-2019]. Moscow: Institute of Management Problems. pp. 1936–1940.

16. Zaychenko, T.N. (2012) Development of new methodology of control systems operation analysis. *Doklady Tomskogo gosudarstvennogo universiteta sistem upravleniya i radioelektroniki – Proceedings of TUSUR University*. 2-1(26). pp. 151–156.
17. Gandzha, T.V. (2012) Formalized representation of a generalized technically complex object with a computer model in the control loop. *Pribory i sistemy. Upravlenie, kontrol', diagnostika – Instruments and Systems: Monitoring, Control, and Diagnostics*. 2. pp. 29–35.

Информация об авторах:

Дмитриев Вячеслав Михайлович – профессор, доктор технических наук, профессор кафедры компьютерных систем в управлении и проектировании Томского государственного университета систем управления и радиоэлектроники (Томск, Россия). E-mail: dmitriewvm@gmail.com

Кочергин Максим Игоревич – кандидат технических наук, доцент кафедры компьютерных систем в управлении и проектировании Томского государственного университета систем управления и радиоэлектроники (Томск, Россия). E-mail: maksim.i.kochergin@tusur.ru

Ганджа Тарас Викторович – доцент, доктор технических наук, профессор кафедры компьютерных систем в управлении и проектировании Томского государственного университета систем управления и радиоэлектроники (Томск, Россия). E-mail: gtv@main.tusur.ru

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

Information about the authors:

Dmitriev Vyacheslav M. (Professor, Doctor of Technical Sciences, Professor of the Department of Computer Systems in Control and Design of Tomsk State University of Control Systems and Radioelectronics, Tomsk, Russian Federation). E-mail: dmitriewvm@gmail.com

Kochergin Maxim I. (Candidate of Technical Sciences, Associate Professor, Department of Computer Systems in Control and Design, Tomsk State University of Control Systems and Radioelectronics, Tomsk, Russian Federation). E-mail: maksim.i.kochergin@tusur.ru

Gandzha Taras V. (Associate Professor, Doctor of Technical Sciences, Professor of the Department of Computer Systems in Control and Design, Tomsk State University of Control Systems and Radioelectronics, Tomsk, Russian Federation). E-mail: gtv@main.tusur.ru

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

Поступила в редакцию 14.03.2023; принята к публикации 04.09.2023

Received 14.03.2023; accepted for publication 04.09.2023

Научная статья

УДК 53.083.2; 53.082.73

doi: 10.17223/19988605/64/5

Математическая модель и модальный анализ термоиндикаторного полимерного покрытия

Андрей Анатольевич Паньков¹, Павел Викторович Писарев², Станислав Равилевич Баяндин³

^{1, 2, 3} Пермский национальный исследовательский политехнический университет, Пермь, Россия

¹ a_a_pankov@mail.ru

² pisarev@pstu.ru

³ krd2874@mail.ru

Аннотация. Разработана математическая модель представительного фрагмента – ячейки периодичности «полимерный слой / датчик» термоиндикаторного покрытия – в рамках теории электротермоупругости механики композитов с учетом структурной неоднородности области ячейки, анизотропии элементов структуры и связанности электрических и деформационных полей. Осуществлен модальный анализ – определение собственных частот и форм (мод) собственных колебаний представительного фрагмента термоиндикаторного покрытия при различных приращениях температуры. Найдено численное значение коэффициента пропорциональности для функциональной зависимости приращения первой резонансной частоты покрытия от приращения температуры. Эта характеристика покрытия используется для диагностирования температурного спектра покрытия по результатам измерения спектра амплитуды интенсивности результирующего светового потока на выходе из оптоволокна встроенного PEL-датчика.

Ключевые слова: численное моделирование; модальный анализ; механика композитов.

Благодарности: Результаты получены при выполнении государственного задания Министерства науки и высшего образования Российской Федерации на выполнение фундаментальных научных исследований (проект № FSNM-2023-0006).

Для цитирования: Паньков А.А., Писарев П.В., Баяндин С.Р. Математическая модель и модальный анализ термоиндикаторного полимерного покрытия // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2023. № 64. С. 41–49. doi: 10.17223/19988605/64/5

Original article

doi: 10.17223/19988605/64/5

Mathematical model and modal analysis of thermoindicator polymer coating

Andrey A. Pan'kov¹, Pavel V. Pisarev², Stanislav R. Bayandina³

^{1, 2, 3} Perm National Research Polytechnic University, Russian Federation

¹ a_a_pankov@mail.ru

² pisarev@pstu.ru

³ krd2874@mail.ru

Abstract. A numerical model of a representative fragment “polymer layer/sensor” (a periodicity cell) of a thermal indicator coating has been developed within the framework of the theory of electrothermoelasticity of mechanics of composites, taking into account structural heterogeneity of the cell, anisotropy of the structure elements and coupling of electrical and deformation fields. Modal analysis was carried out - determination of natural frequencies and forms (modes) of natural oscillations of a representative fragment of a thermal indicator coating at different temperature increments. A numerical value of the proportionality coefficient was found for the functional dependence of the

increment of the first resonant frequency of the coating on the temperature increment. This coating characteristic is used to diagnose the temperature spectrum of the coating by measuring the amplitude spectrum of the resulting light flux at the output of the PEL sensor optical fiber.

Keywords: numerical modeling; modal analysis; mechanics of composites.

Acknowledgments: The results were obtained within the framework of the State task of the Ministry of Science and Higher Education of the Russian Federation (project No. FSNM-2023-0006).

For citation: Pan'kov, A.A., Pisarev, P.V., Bayandin, S.R. (2023) Mathematical model and modal analysis of thermoindicator polymer coating. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naja tehnika i informatika – Tomsk State University Journal of Control and Computer Science*. 64. pp. 41–49. doi: 10.17223/19988605/64/5

Введение

Датчики широко используются в системах автоматизированного управления для получения необходимой информации о характеристиках (температура, напряженное состояние, повреждения и др.) и поведении объекта, при этом представление измерений в виде информативных оптических сигналов допускает оптоволоконную передачу информации на большие расстояния, что необходимо для удаленного мониторинга и управления объектом. Актуальность совершенствования датчиков температуры обусловливается тем, что температура существенным образом влияет на многие физико-механические и эксплуатационные свойства современных полимерных композитных конструкций, в частности аэрокосмической техники, для которых необходим мониторинг неоднородных протяженных температурных полей с высоким пространственным разрешением. Современные датчики температуры основаны на различных физических принципах, и измерение температуры осуществляется опосредованно с использованием известных зависимостей тех или иных физико-механических свойств чувствительного элемента датчика от температуры, в частности спектра отражения оптоволоконной брэгговской решетки [1–4], резонансной частоты пьезоэлемента [5–11], что обуславливает высокую точность измерений в широком диапазоне температур. Распространенными волоконно-оптическими датчиками температуры являются датчики амплитудного типа, в которых осуществляется модуляция проходящего по оптоволокну светового потока в результате зависимости свойств поглощения и отражения чувствительного элемента от температуры, например для случая встроенного внутрь или расположенного у торца оптоволокну чувствительного элемента [12]. Известны чувствительные элементы с температурной зависимостью светоотдачи (генерации) информативных световых сигналов на основе эффекта электролюминесценции или флюоресценции [13–16]. Пьезоэлектрические датчики температуры основаны на температурной зависимости резонансной частоты вибраций пьезоэлектрического чувствительного элемента, что обуславливает сравнительно большую мощность и простоту обработки их информативных электрических сигналов. В пьезоэлектрических датчиках используются специально разработанные резонаторы с сильной и линейной зависимостью резонансной частоты от температуры в широком диапазоне частот [17]. Для мониторинга температуры протяженных конструкций используют распределенные и квазираспределенные датчики и сенсорные системы, при этом актуальными остаются задачи нахождения не только интегральных характеристик температурного поля, в частности усредненного по длине протяженного датчика значения температуры, температурного спектра - функции плотности распределения значений температуры по длине датчика, но и реального неоднородного распределения температуры с высоким пространственным разрешением, например по длине оптоволоконного пьезоэлектролюминесцентного датчика [18, 19], внедренного внутрь полимерной композитной конструкции или внутрь термоиндикаторного полимерного покрытия, закрепленного на поверхности конструкции. Отметим, что в [20, 21] разработана математическая модель функционирования индикаторного полимерного покрытия со встроенным оптоволоконным PEL-датчиком для мониторинга температуры и индикации, локации и резонансной самоочистки появляющихся участков обледенения протяженных аэродинамических поверхностей.

Цель настоящей статьи – разработка математической модели, определение собственных частот и форм (мод) колебаний и, как результат, нахождение численных значений характеристик термоиндикаторного полимерного покрытия со встроенным оптоволоконным пьезоэлектrolюминесцентным датчиком [18] для резонансного режима функционирования покрытия с учетом нелинейных температурных зависимостей электроупругих свойств его структурных элементов.

1. Термоиндикаторное пьезоэлектrolюминесцентное покрытие

Рассматриваемое термоиндикаторное полимерное покрытие устанавливается на поверхности конструкции и представляет собой полимерный слой со встроенным оптоволоконным пьезоэлектrolюминесцентным (PEL) датчиком [18] (или системой из сонаправленных с периодическим взаимным расположением в плоскости покрытия таких датчиков) (рис. 1), в котором оптоволоконно 1 и электrolюминесцентный слой 2 разделены внутренним светопрозрачным (или перфорированным) управляющим электродом 4, на поверхности пьезоэлектрического слоя 3 с радиальной поляризацией расположен внешний управляющий электрод 5. Управляющие электроды 4, 5 образуют «двухпроводную» электрическую линию, протяженную вдоль оси PEL-датчика, на входе которой прикладывается электрическое напряжение $U_{con}(v, t)$ с частотой v . Информативная величина I – интенсивность результирующего светового потока на выходе из оптоволоконна PEL-датчика.

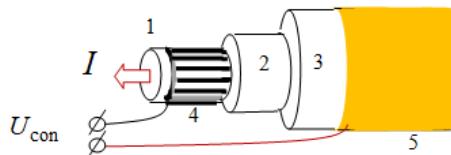


Рис. 1. Оптоволоконный PEL-датчик: 1 – оптоволоконно, 2 – люминофор, 3 – пьезоэлектрик, 4, 5 – управляющие электроды
Fig. 1. Fiber optic PEL-sensor, 1 – fiber optic, 2 – phosphor, 3 – piezoelectric, 4, 5 – control electrodes

Механолюминесцентный эффект возникает в результате взаимодействия между собой электrolюминесцентного 2 и пьезоэлектрического 3 слоев при вынужденных электромеханических колебаниях (вибрациях) PEL-датчика и в целом термоиндикаторного полимерного покрытия. Информативные световые сигналы возникают в электrolюминесцентном слое 2 и далее проникают через светопрозрачный электрод 4, расположенный на межфазной поверхности «оптоволоконно / электrolюминесцентный слой», внутрь оптического волокна и распространяются по нему к приемнику-анализатору интенсивности I световых сигналов на выходе из оптического волокна.

2. Алгоритм диагностирования температурного спектра

Пусть искомый «температурный спектр» $f_T(\theta)$ – функция плотности распределения значений отклонений ΔT диагностируемой температуры T по длине l оптоволоконна PEL-датчика, встроенного в полимерное покрытие, от его номинального (известного) значения T_0 . Для нахождения температурного спектра $f_T(\theta)$ используем резонансный метод диагностирования [20], который основан на рассмотрении вынужденных осесимметричных электромеханических колебаний датчика под действием приложенного к его электродам гармонического управляющего электрического напряжения $U_{con}(v, t)$ с частотой v . Для случая начального ($\Delta T = 0$) состояния датчика имеем однородные по всей длине l колебания датчика и, как следствие, равномерную погонную (с единицы длины) светоотдачу $i_{lum}^{(0)}(v, t)$ электrolюминесцентного слоя датчика со считающим известным спектром амплитуды $\tilde{i}_0(v) \equiv \tilde{I}_0(v)/l$ погонной интенсивности светоотдачи $i_{lum}^{(0)}(v, t)$ с резонансной частотой v_0 , где $\tilde{I}_0(v)$ – спектр амплитуды интенсивности результирующего светового потока на выходе из оптоволоконна датчика. Полага-

ем, что при однородном нагреве на температуру $\theta \neq 0$ произвольного локального участка датчика на нем реализуется спектр амплитуды погонной интенсивности светоотдачи

$$\tilde{i}_\theta(v) \approx \tilde{i}_0(v - \Delta_\theta) \quad (1)$$

как результат смещения графика начального спектра $\tilde{i}_0(v)$ по оси абсцисс (частоты v) на величину изменения резонансной частоты $\Delta_\theta \equiv v_\theta - v_0$ (1) на этом локальном участке датчика. Считаем, что имеем пропорциональную зависимость для отклонений резонансных частот

$$\Delta_\theta \approx k_v \theta \quad (2)$$

от значений нагрева θ , где коэффициент k_v - характеристика термоиндикаторного покрытия для рассматриваемого диапазона значений нагрева $\Delta T \in (\theta_{\min}; \theta_{\max})$ с учетом $v_0 = v_\theta(0)$.

Амплитуду интенсивности $\tilde{I}(v)$ результирующего светового потока на выходе из оптоволоконна датчика запишем в виде:

$$\tilde{I}(v) = \int_{\theta} d\tilde{l}_\theta \approx \int_{\theta} \tilde{i}_0(v - k_v \theta) d\tilde{l}_\theta, \quad (3)$$

как результат интегрирования по множеству всех значений нагрева θ , реализующихся по длине l датчика, с учетом вида Δ_θ (2). Здесь $d\tilde{l}_\theta$ - это некоторая малая длина участка (или суммарная длина участков) датчика со значениями нагрева $\Delta T \in (\theta, \theta + d\theta)$, при этом имеем равенство дифференциалов

$$d\tilde{l}_\theta = d\tilde{l}_\theta, \quad (4)$$

где $\tilde{l}_\theta = \tilde{l}_\theta(\theta)$ - суммарная длина участков датчика со значениями нагрева $\Delta T < \theta$, приращение $d\tilde{l}_\theta$ обусловлено приращением $d\theta$, область значений $0 < \tilde{l}_\theta \leq l$, длина оптоволоконна l . Приращение $d\tilde{l}_\theta$ (4) представим в виде:

$$d\tilde{l}_\theta = l f_T(\theta) d\theta \quad (5)$$

на основе двойного равенства

$$dP = d\tilde{l}_\theta / l = f_T(\theta) d\theta \quad (6)$$

для геометрической вероятности dP , где искомая функция плотности распределения $f_T(\theta)$ значений неоднородного поля нагрева ΔT по всей длине l датчика; вид функции

$$\tilde{l}_\theta(\theta) = l \int_{-\infty}^{\theta} f_T(\theta') d\theta' \quad (7)$$

следует из (5), (6). Для случая, когда диагностируемые величины $\Delta T \in (\theta_{\min}; \theta_{\max})$, нижний предел интегрирования в (7) может быть заменен на $\theta_{\min}$, в предельных случаях имеем $\tilde{l}_\theta(\theta_{\min}) = 0$, $\tilde{l}_\theta(\theta_{\max}) = l$.

В результате подстановки (5) в (3) получим уравнение Фредгольма 1-го рода:

$$\tilde{I}(v) = l \int_{-\infty}^{\infty} \tilde{i}_0(v - k_v \theta) f_T(\theta) d\theta,$$

или в виде:

$$\tilde{i}(v) = \int_{-\infty}^{\infty} \tilde{i}_0(v - k_v \theta) f_T(\theta) d\theta, \quad (8)$$

где $\tilde{i}(v) \equiv \tilde{I}(v)/l$ - амплитуда осредненной по длине l датчика погонной интенсивности светоотдачи электролюминесцентного слоя, нагруженного неоднородным по длине датчика нагревом ΔT , $\tilde{i}_0(v)$ - амплитуда погонной интенсивности светоотдачи электролюминесцентного слоя (с единицы длины) при номинальной температуре T_0 ($\Delta T = 0$) всего датчика. Для случая $\Delta T = 0$ имеем равенство $\tilde{i}(v) = \tilde{i}_0(v)$, и искомая функция температурного спектра $f_T(\theta) = \delta(\theta)$ - дельта-функция Дирака.

3. Результаты численного моделирования

Расчетная область представительной ячейки термоиндикаторного покрытия изображена на рис. 2, ее компоненты – кварцевое оптоволокно, полимерный электролюминофор, пьезоэлектрик PVDF с радиальной поляризацией и силиконовый корпус с учетом заданных зависимостей их характеристик от температуры.

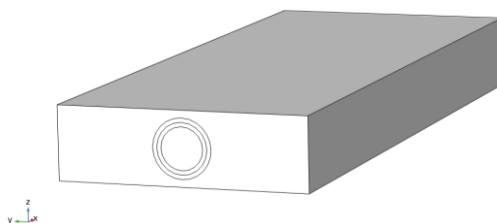


Рис. 2. Расчетная область представительной ячейки термоиндикаторного покрытия

Fig. 2. Design area of the representative cell of the thermal indicator coating

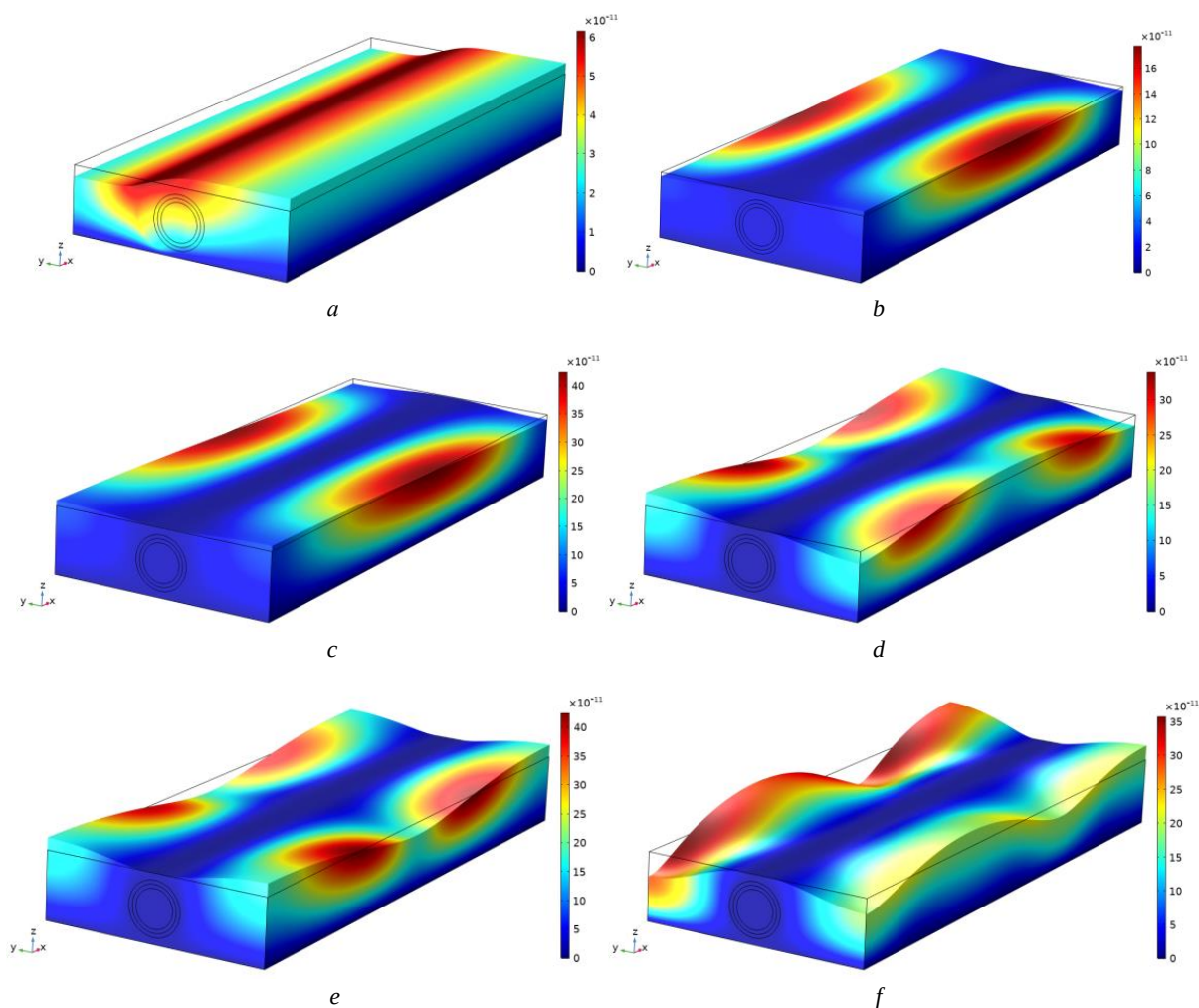


Рис. 3. Собственные формы колебаний представительной ячейки термоиндикаторного покрытия

Fig. 3. Natural forms of oscillations of the representative cell of the thermal indicator coating

Численное моделирование резонансных характеристик – первых шести частот и собственных форм колебаний термоиндикаторного покрытия – реализовано на основе численного решения связанной стационарной краевой задачи электроупругости для представительной ячейки (см. рис. 2). Использован многопроцессорный вычислительный комплекс Центра высокопроизводительных вы-

числительных систем Пермского национального исследовательского политехнического университета в программной системе конечно-элементного анализа ANSYS. Для решения систем линейных алгебраических уравнений использован итерационный решатель метода сопряженных градиентов Jacobi Conjugate Gradient (JCG) solver, который позволил существенно сократить время счета в сравнении с решателем Sparse direct equation solver (SPARSE); время счета для базового варианта задачи составило около 18 ч для SPARSE и 7,5 ч для JCG.

Температурные зависимости собственных частот ν_n термоиндикаторного покрытия

$\Delta T, ^\circ\text{K}$	Собственные частоты ν_n , Гц					
	$n = 1$	$n = 2$	$n = 3$	$n = 4$	$n = 5$	$n = 6$
0	3 359	3 595	3 608	3 842	3 843	4 118
20	3 360	3 595	3 602	3 843	3 843	4 119
40	3 412	3 592	3 603	3 879	3 881	4 206
60	3 451	3 594	3 608	3 907	3 910	4 277
80	3 486	3 601	3 617	3 935	3 939	4 343
100	3 531	3 615	3 634	3 971	3 976	4 422
120	3 603	3 638	3 660	4 021	4 027	4 532
140	3 670	3 697	3 717	4 089	4 098	4 670
160	3 713	3 747	3 894	4 178	4 189	4 799
180	3 738	3 777	4 010	4 228	4 240	4 858
200	3 738	3 777	4 010	4 228	4 240	4 858

Результаты поиска первых шести форм колебаний представительной ячейки (см. рис. 2) даны на рис. 3, в таблице – численные значения соответствующих шести значений резонансных (собственных) частот в зависимости от приращения температуры $\Delta T = T - T_0$ термоиндикаторного покрытия, где начальная температура $T_0 = 273,15$ К (или 15°C). На рис. 3, а–f приведены результаты моделирования для этих шести случаев. На рис. 4 показаны зависимости собственных частот ν_n от их порядкового номера $n = 1, \dots, 6$ при различных приращениях температуры ΔT термоиндикаторного покрытия. На рис. 5 даны графики зависимостей каждой из шести собственных частот от приращения температуры ΔT . Выявлено, что с ростом температуры значения собственных частот увеличиваются по нелинейным законам. На основе результатов численного моделирования (см. таблицу, рис. 5) может быть определено численное значение характеристики термоиндикаторного покрытия - коэффициент пропорциональности k_v в зависимости $\Delta\theta \approx k_v\theta$ (2) для различных резонансных форм и частот ν_n , где $n = 1, \dots, 6$.

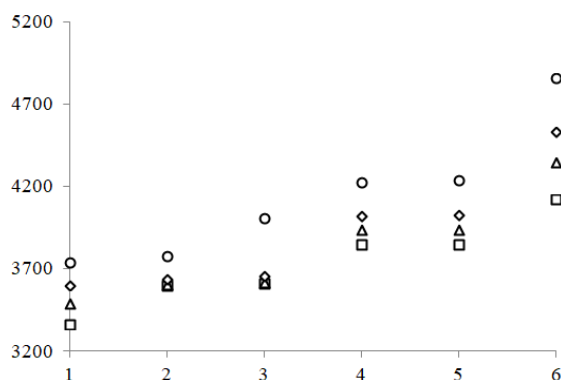


Рис. 4. Зависимости собственных частот от порядкового номера $n = 1, \dots, 6$ форм при различных приращениях температуры $\Delta T = 0$ ($\square$), 60 (Δ), 100 ($\diamond$), 200 ($\circ$) термоиндикаторного покрытия
Fig. 4. Dependence of natural frequencies on sequence number $n = 1, \dots, 6$ forms at different temperature $\Delta T = 0$ ($\square$), 60 (Δ), 100 ($\diamond$), 200 ($\circ$) increments ΔT thermal indicator coating

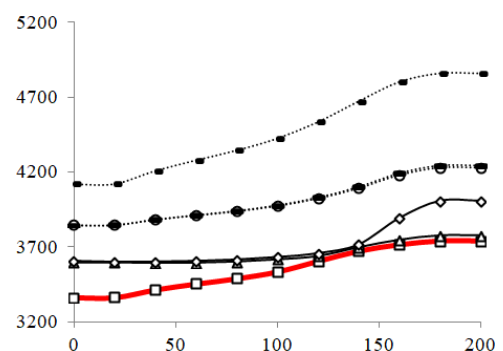


Рис. 5. Зависимости собственных частот ν_n от приращения температуры ΔT термоиндикаторного покрытия, $n = 1$ ($\square$), 2 (Δ), 3 ($\diamond$), 4 ($\circ$), 5 ($-$), 6 ($-$)
Fig. 5. Dependence of natural ν_n frequencies on temperature increment ΔT thermal indicator coating, $n = 1$ ($\square$), 2 (Δ), 3 ($\diamond$), 4 ($\circ$), 5 ($-$), 6 ($-$)

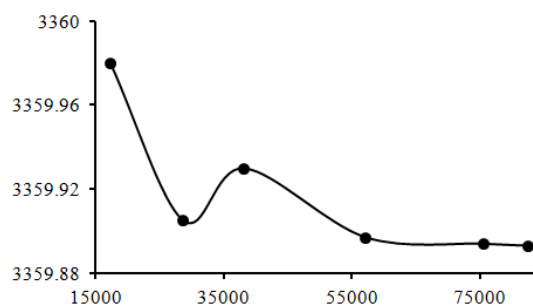


Рис. 6. Зависимость первой собственной частоты ν_1 от числа N элементов дискретизации расчетной области
Fig. 6. Dependence of the first natural frequency ν_1 on the number N of sampling elements of the representative cell

Например, для первой резонансной частоты ν_1 численное значение коэффициента пропорциональности $k_\nu \approx 1.724$ Гц/К. Осуществлена проверка сходимости результатов расчета собственных частот ν_n к искомым значениям (рис. 5, 6, таблица), и в качестве примера на рис. 6 приведен график зависимости первой собственной частоты ν_1 от числа N элементов дискретизации расчетной области (см. рис. 2) термоиндикаторного покрытия.

Заключение

Разработана математическая модель представительного фрагмента – ячейки периодичности «полимерный слой / датчик» термоиндикаторного покрытия – в рамках теории электротермоупругости механики композитов с учетом наличия электродов и особенностей взаимного расположения, анизотропии и температурных зависимостей свойств элементов структуры: кварцевого оптоволокна, полимерного электролюминофора и пьезоэлектрика PVDF с радиальной поляризацией, обуславливающего связанность электрических и деформационных полей, и силиконового корпуса в виде слоя (покрытия) на жестком основании. Осуществлен модальный анализ ячейки термоиндикаторного покрытия по определению собственных частот и форм (мод) собственных колебаний покрытия при различных приращениях температуры, исследованы закономерности влияния приращения температуры ΔT на резонансные формы и частоты колебаний покрытия в результате действия управляющего гармонического электрического напряжения U_{con} на электроды датчика.

Список источников

1. Каблов Е.Н., Сиваков Д.В., Гуляев И.Н., Сорокин К.В., Федотов М.Ю., Дианов Е.М., Васильев С.А., Медведков О.И. Применение оптического волокна в качестве датчиков деформации в полимерных композиционных материалах // Все материалы : энциклопедический справочник. 2010. № 3. С. 10–15.
2. Зуев М.А., Махсидов В.В., Федотов М.Ю., Шиенок А.М. К вопросу об интеграции оптоволокна в ПКМ и измерении деформации материала с помощью волоконных брэгговских решеток // Механика композиционных материалов и конструкций. 2014. Т. 20, № 4. С. 568–574.
3. Аношкин А.Н., Воронков А.А., Кошелева Н.А., Матвеев В.П., Сероваев Г.С., Спаскова Е.М., Шардаков И.Н., Шипунов Г.С. Измерение неоднородных полей деформаций встроенными в полимерный композиционный материал волоконно-оптическими датчиками // Известия РАН. Механика твердого тела. 2016. № 5. С. 42–51.
4. Якушин С.С., Достовалов А.В., Вольф А.А., Парыгин А.В., Бабин С.А. Измерение величины и положения точечных температурных воздействий на длинные ВБР // I Всерос. науч.-практ. конф. «Оптическая рефлектометрия 2016» 26–27 мая 2016 г. : сб. тез. докл. Пермь, 2016. С. 39–40.
5. Fraden J. Handbook of Modern Sensors. New York : Springer-Verlag, 2004. 589 p.
6. Next Generation Sensors and Systems / S.Ch. Mukhopadhyay (ed.). Springer International Publishing, 2016. 330 p.
7. Fiber Optic Sensors / I.R. Matias, S. Ikezawa, J. Corres (eds.) Springer International Publishing, 2017. 381 p.
8. Sensors and Microsystems : Proc. of the 19th AISEM 2017 National Conference. Springer International Publishing / A. Leone et al. (eds.). Springer International Publishing, 2018. 330 p.
9. Smart Sensors and MEMS / S. Nihitani, A. Luque (eds.). Woodhead Publishing, 2018. 604 p.
10. Козлов В.Л. Оптоэлектронные датчики. Минск : Белгосуниверситет, 2005. 116 с.
11. Волоконно-оптические датчики : вводный курс для инженеров и научных работников / под ред. Э. Удда. М. : Техносфера, 2008. 520 с.

12. Гармаш В.Б., Егоров Ф.А., Коломиец Л.Н., Неугодников А.П., Поспелов В.И. Возможности, задачи и перспективы волоконно-оптических измерительных систем в современном приборостроении // Фотон-Экспресс. Наука. 2005. № 6, спецвып. С. 128–140.
13. Карякин А.В., Боровиков А.С. Люминесцентная и цветная дефектоскопия. М. : Машиностроение, 1972. 240 с.
14. Левшин Л.В., Салецкий А.М. Люминесценция и ее измерения. Молекулярная люминесценция. М. : Изд-во МГУ, 1989. 272 с.
15. Гришаева Т.И. Методы люминесцентного анализа. СПб. : Профessional, 2003. 226 с.
16. Карицкая С.Г. Диагностика полей температур и скоростей люминесцентными методами : дис. ... канд. физ.-мат. наук. М., 1997. 103 с.
17. Андросова В.Г., Банков В.Н., Дикиджи А.Н. и др. Справочник по кварцевым резонаторам / под ред. П.Г. Позднякова. М. : Связь, 1978. 288 с.
18. Патент RU № 2630537. Волоконно-оптический датчик давления / Паньков А.А., опубл. 11.09.2017; Бюл. № 26.
19. Pan'kov A.A. Piezoelectroluminescent fiber-optic sensors for temperature and deformation fields // Sensors and Actuators A: Physical. 2019. V. 288. P. 171–176.
20. Pan'kov A.A. Resonant piezoelectroluminescent fiber-optical sensor of a temperature field in composite structures // Mechanics of Composite Materials. 2019. V. 55 (4). P. 547–556.
21. Паньков А.А., Писарев П.В. Антиобледенительное индикаторное полимерное покрытие со встроенным оптоволоконным PEL-датчиком для индикации, локации и очистки от обледенения аэродинамических поверхностей // Вестник Пермского национального исследовательского политехнического университета. Механика. 2021. № 4. С. 111–121.

References

1. Kablov, E.N., Sivakov, D.V., Gulyaev, I.N., Sorokin, K.V., Fedotov, M.Y., Dianov, E.M., Vasiliev, S.A. & Medvedkov, O.I. (2010) *Primenenie opticheskogo volokna v kachestve datchikov deformatsii v polimernykh kompozitsionnykh materialakh* [The use of optical fiber as deformation sensors in polymer composite materials]. *Vse materialy: entsiklopedicheskiy spravochnik – All Materials. Encyclopedic Handbook*. 3. pp. 10–15.
2. Zuev, M.A., Makhidov, V.V., Fedotov, M.Y. & Shienok, A.M. (2014) *K voprosu ob integratsii optovolokna v PKM i izmerenii deformatsii materiala s pomoshch'yu volokonnykh breggovskikh reshetok* [On the issue of fiber integration in PCM and material deformation measurement using fiber Bragg grids]. *Mekhanika kompozitsionnykh materialov i konstruktivnykh – Mechanics of Composite Materials and Structures*. 20(4). pp. 568–574.
3. Anoshkin, A.N., Voronkov, A.A., Kosheleva, N.A., Matveenko, V.P., Serovaev, G.S., Spaskova, E.M., Shardakov, I.N. & Shipunov, G.S. (2016) *Izmerenie neodnorodnykh poley deformatsiy vstroennymi v polimernyy kompozitsionnyy material volokonno-opticheskimi datchikami* [Measurement of heterogeneous deformation fields by fiber-optic sensors built into the polymer composite material]. *Izvestiya RAN. Mekhanika tverdogo tela*. 5. pp. 42–51.
4. Yakushin, S.S., Dostovalov, A.V., Volf, A.A., Parygin, A.V. & Babin, S.A. (2016) *Measuring the magnitude and position of point temperature effects on long VBR. Opticheskaya reflektometriya 2016* [Optical Reflectometry 2016]. Proc. of the First All-Russian Conference. Perm, May 26-27, 2016. Perm. pp. 39–40.
5. Fraden, J. (2004) *Handbook of Modern Sensors*. Springer-Verlag New York.
6. Mukhopadhyay, S.Ch. (ed.) (2016) *Next Generation Sensors and Systems*. Springer.
7. Matias, I.R., Ikezawa, S. & Corres, J. (eds) (2017) *Fiber Optic Sensors*. Springer International Publishing.
8. Leone A., Forleo, A., Francioso, L., Capone, S., Siciliano, P. & Di Natale, C. (eds) (2018) *Sensors and Microsystems*. Proceedings of the 19th AISEM 2017 National Conference. Springer International Publishing.
9. Nihtianov, S. & Luque, A. (eds) (2018) *Smart Sensors and MEMS*. Woodhead Publishing.
10. Kozlov, V.L. (2005) *Optoelektronnye datchiki* [Optoelectronic sensors]. Minsk: Belarusian State University.
11. Udd, E. (ed.) (2008) *Volokonno-opticheskie datchiki : vvodnyy kurs dlya inzhenerov i nauchnykh rabotnikov* [Fiber optic sensors. Introductory Course for Engineers and Scientists]. Translated from English by I.Yu. Shkadina. Moscow: Tekhnosfera.
12. Garmash, V.B., Egorov, F.A., Kolomiets, L.N., Neugodnikov, A.P. & Pospelov, V.I. (2005) *Vozmozhnosti, zadachi i perspektivy volokonno-opticheskikh izmeritel'nykh sistem v sovremennom priborostroenii* [Opportunities, problems and prospects of fiber-optic measuring systems in modern instrumentation]. *Foton-Ekspress. Nauka*. 6. pp. 128–140.
13. Karyakin, A.V. & Borovikov, A.S. (1972) *Lyuminescentnaya i tsvetnaya defektoskopiya* [Luminescent and Color Flaw Detection]. Moscow: Mashinostroenie.
14. Levshin, L.V. & Saletskiy, A.M. (1989) *Lyuminesentsiya i ee izmereniya. Molekulyarnaya lyuminesentsiya* [Luminescence and its measurements. Molecular luminescence]. Moscow: Moscow State University.
15. Grishaeva, T.I. (2003) *Metody lyuminescentnogo analiza* [Methods of Luminescent Analysis]. St. Petersburg: Professional.
16. Karitskaya, S.G. (1997) *Diagnostika poley temperatur i skorostey lyuminescentnymi metodami* [Diagnostics of temperature and velocity fields by luminescent methods]. Physics and Mathematics Cand. Diss. Moscow: MSU.
17. Androsova, V.G., Bankov, V.N., Dikidzhi, A.N. et al. (1978) *Spravochnik po kvartsevykh rezonatoram* [Quartz Resonator Handbook]. Moscow: Svyaz'.
18. Pankov, A.A. (2017) *Patent RU no. 2630537. Fiber optic pressure sensor*. 11th September. Bul. no. 26.
19. Pankov, A.A. (2019) *Piezoelectroluminescent fiber-optic sensors for temperature and deformation fields. Sensors and Actuators A: Physical*. 288. pp. 171–176.

20. Pankov, A.A. (2019) Resonant piezoelectroluminescent fiber-optical sensor of a temperature field in composite structures. *Mechanics of Composite Materials*. 55(4). pp. 547–556.
21. Pankov, A.A. & Pisarev, P.V. (2021) Antiobledenitel'noe indikatornoe polimernoe pokrytie so vstroennym optovolokonnym PEL-datchikom dlya indikatsii, lokatsii i ochistki ot obledeneniya aerodinamicheskikh poverkhnostey [Anti-icing indicator polymer coating with built-in fiber optic PEL sensor for indication, location and deicing of aerodynamic surfaces]. *Vestnik Permskogo natsional'nogo issledovatel'skogo politekhnicheskogo universiteta. Mekhanika*. 4. pp. 111–121.

Информация об авторах:

Паньков Андрей Анатольевич – доктор физико-математических наук, профессор кафедры механики композиционных материалов и конструкций Пермского национального исследовательского политехнического университета (Пермь, Россия). E-mail: a_a_pankov@mail.ru

Писарев Павел Викторович – кандидат технических наук, доцент кафедры механики композиционных материалов и конструкций Пермского национального исследовательского политехнического университета (Пермь, Россия). E-mail: pisarev@pstu.ru

Баяндин Станислав Равилевич – аспирант кафедры механики композиционных материалов и конструкций Пермского национального исследовательского политехнического университета (Пермь, Россия). E-mail: krd2874@mail.ru

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

Information about the authors:

Pan'kov Andrey A. (Doctor of Physical and Mathematical Sciences, Professor, Associate Professor, Department of the mechanics of composite materials and structures, Perm National Research Polytechnic University, Perm, Russian Federation). E-mail: a_a_pankov@mail.ru

Pisarev Pavel V. (Candidate of Technical Sciences, Associate Professor, Department of the mechanics of composite materials and structures, Perm National Research Polytechnic University, Perm, Russian Federation). E-mail: pisarev@pstu.ru

Bayandin Stanislav R. (Post-graduate Student, Department of the mechanics of composite materials and structures, Perm National Research Polytechnic University, Perm, Russian Federation). E-mail: krd2874@mail.ru

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

Поступила в редакцию 11.02.2023; принята к публикации 04.09.2023

Received 11.02.2023; accepted for publication 04.09.2023

ОБРАБОТКА ИНФОРМАЦИИ

DATA PROCESSING

Научная статья

УДК 004.032.26

doi: 10.17223/19988605/64/6

Нейросетевой анализ данных ультразвуковой дефектоскопии магистральных газопроводов с использованием программно-аппаратной системы ДВУ**Игорь Георгиевич Боровской¹, Елена Александровна Шельмина²,
Андрей Анатольевич Матолыгин³, Евгений Петрович Ильин⁴**^{1, 2, 3, 4} Томский государственный университет систем управления и радиоэлектроники, Томск, Россия¹ igor.g.borovskoi@tusur.ru² elena.a.shelmina@tusur.ru³ andrei.a.matolygin@tusur.ru⁴ eugenecalls@gmail.com

Аннотация. Рассматривается задача определения зон, подвергшихся коррозионному воздействию, полученных при неразрушающем контроле с использованием методов ультразвуковой дефектоскопии магистральных газопроводов и программно-аппаратной системы ДВУ (дефектоскоп внутритрубный ультразвуковой).

Предлагается обработку большого объема данных, полученных в результате контроля с помощью системы ДВУ, осуществлять с использованием нейросетевой модели анализа данных. В качестве основной нейронной сети выбрано семейство сверточных нейросетей. Обработываемые данные были разбиты на обучающую, валидационную и тестовую выборки. Данные внутри выборок перемешивались, что позволило процесс обучения сделать более качественным. Как показал анализ тестирования, предложенный подход позволяет с достаточной точностью определять зоны дефектов газопроводов.

Ключевые слова: нейросетевой анализ данных; сверточные нейронные сети; ультразвуковая дефектоскопия.

Благодарности: Исследование выполнено при финансовой поддержке гранта на реализацию программы стратегического академического лидерства «Приоритет-2030» в рамках научного проекта Пр2030-МОН наука БЧ/СП2/Б/196.

Для цитирования: Боровской И.Г., Шельмина Е.А., Матолыгин А.А., Ильин Е.П. Нейросетевой анализ данных ультразвуковой дефектоскопии магистральных газопроводов с использованием программно-аппаратной системы ДВУ // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2023. № 64. С. 50–60. doi: 10.17223/19988605/64/6

Original article

doi: 10.17223/19988605/64/6

Neural network analysis of ultrasonic flaw detection data of main gas pipelines using a hardware and software DVU system**Igor G. Borovskoy¹, Elena A. Shelmina², Andrey A. Matolygin³, Evgenyi P. Ilin⁴**^{1, 2, 3, 4} Tomsk State University of Control Systems and Radioelectronics, Tomsk, Russian Federation¹ igor.g.borovskoi@tusur.ru² elena.a.shelmina@tusur.ru

Abstract. The paper considers the problem of determining the zones exposed to corrosion, obtained during non-destructive testing using ultrasonic flaw detection methods of main gas pipelines and the hardware and software system of the DVU (in-tube ultrasonic flaw detector).

It is proposed that the processing of a large amount of data obtained as a result of control using the DVU system is carried out using a neural network model of data analysis. A family of convolutional neural networks was chosen as the main neural network. The processed data were divided into training, validation and test sets. The data within the samples were mixed, which allowed the learning process to be made more qualitative. As the testing analysis showed, the proposed approach makes it possible to determine the defect zones of gas pipelines with sufficient accuracy.

Keywords: neural network data analysis; convolutional neural networks; ultrasonic flaw detection.

Acknowledgments: The research was carried out with the financial support of a grant for the implementation of the strategic academic leadership program “Priority 2030” within the framework of the scientific project Пр2030-МОН Science БЧ/СП2/Б/196.

For citation: Borovskoy, I.G., Shelmina, E.A., Matolygin, A.A., Ilin, E.P. (2023) Neural network analysis of ultrasonic flaw detection data of main gas pipelines using a hardware and software DVU system. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 64. pp. 50–60. doi: 10.17223/19988605/64/6

Введение

На текущий период времени остро стоит вопрос энергетической безопасности как стран в целом, так и отдельных регионов. Одним из эффективных средств решения данной проблемы является использование природного газа. Сегодня основным безопасным, независимым от погодных условий и ритмичным средством доставки является трубопроводный транспорт. Кроме того, магистральные газопроводы транспортируют газ на большие расстояния. Для обеспечения передачи заданных объемов газа используется рабочее давление, превышающее 100 атмосфер. Соответственно, предъявляются высокие требования к механической прочности трубопроводов, и это обеспечивается за счет использования стальных труб. Известно [1], что транспортируемый природный газ состоит в основном из метана, до 98%. Помимо этого, он также может содержать всевозможные примеси как в твердом состоянии, механические примеси в виде песка и возможной окалины, так и в жидком состоянии, в виде водяных паров и конденсатов тяжелых фракций углеводородов. Жидкие примеси, присутствующие в природном газе, оказывают корродирующее действие на внутреннюю поверхность газопровода, что приводит к утончению его несущей стенки и, соответственно, ее механической прочности. Данная проблема является чрезвычайно актуальной, и ей посвящен ряд современных публикаций по выявлению причин данного явления и борьбы с ним [2, 3]. Критические значения утончения стенки в конечном результате приведут к невозможности эксплуатации трубопровода и, соответственно, уменьшению времени его эксплуатации.

В связи с этим актуальной задачей становится определение текущего состояния стенок газопровода. Наиболее актуальными являются методики неразрушающего контроля, одна из которых – ультразвуковая дефектоскопия. При таких исследованиях определяют в обобщенном виде времена прихода отраженных сигналов от внутренней и внешней стенок газопровода [4]. Методы и техническая реализация на сегодняшний день чрезвычайно развиты, но применение на объектах гигантской протяженности чрезвычайно затруднительно. Решением данной проблемы является создание автоматизированных внутритрубных инспекционных приборов (ВИП) [5]. Одним из вариантов реализации таких приборов является дефектоскоп внутритрубный ультразвуковой (ДВУ), примерами реализации могут быть внутритрубные дефектоскопы 10-УСК.04-00.000 [6] и 40-ДФР.00-00.000 [7].

Цель настоящей работы – проведение нейросетевого анализа данных, направленного на автоматизацию обнаружения дефектов стенок газопроводов по результатам их ультразвукового исследования, при том что обрабатываемые данные представлены показаниями нескольких десятков тысяч

точек измерений. Исходя из этого, к задачам исследования можно отнести выбор нейросетевой модели, подготовку исходных данных для выбранной нейронной сети (НС), обучение НС и проведение вычислительных экспериментов для достоверного определения зон дефектов.

1. Программно-аппаратная система ДВУ

Программно-аппаратная система ДВУ [7] позволяет генерировать ультразвуковые сигналы и регистрировать данные их прохождения с 400 датчиков, расположенных по окружности трубы, при этом устройство с датчиками перемещается вдоль оси трубы с заданной скоростью. Соответственно, генерируется гигантский объем данных, который необходимо проанализировать. Совместно с зарегистрированными данными для исследования были получены данные о дефектах, их расположение в координатах «датчики–продольная ось трубы». Следует заметить, что доля данных, отвечающих «дефектам», в общей выборке очень мала и, как правило, не превышает единиц процентов.

В программно-аппаратной системе ДВУ для каждого датчика должно быть зарегистрировано по 31 значению времени и амплитуды отраженного сигнала. Для анализа исходных данных и выявления признаков на данных, соответствующих дефектам, проведена визуализация измерений в виде распределения амплитуд возвращенных волн во времени. Как правило, получается следующая картина (рис. 1). Можно видеть, что на графике присутствуют характерные группировки амплитуд, соответствующие интенсивностям возвращенного сигнала, и некоторые данные, которые явно можно отнести к шумам, вызванным различными физическими процессами, связанными как с волновыми процессами, так и процессами в системе измерения. Амплитуда отраженного сигнала при многократных отражениях уменьшается. На графике точками помечены центры отраженных ультразвуковых сигналов, которые могут быть теоретически использованы для пересчета толщины стенки, например за счет статистической обработки данных измерений.

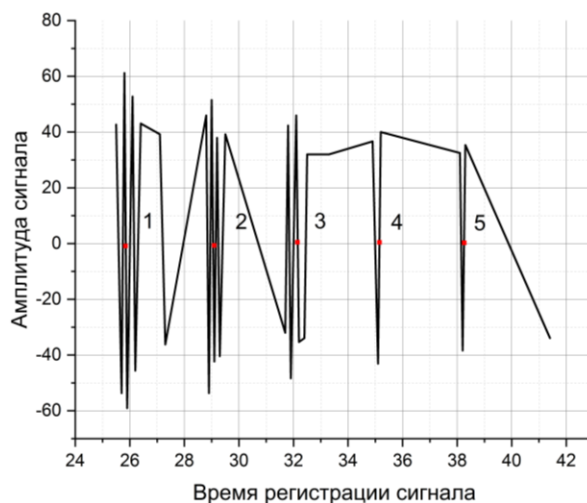


Рис. 1. Визуализация зарегистрированных данных с одного датчика внутритрубных инспекционных приборов

Fig. 1. Visualization of recorded data from one sensor of in-line inspection devices

Если рассматривать зарегистрированные данные для нескольких датчиков сразу, то получим следующую картину распределения данных о прохождении волн (рис. 2). Выделенный жирной линией замер с номером 87 отвечает сигналам из зоны с дефектом. На рисунке видны для номеров замеров положений до 88 и после 93 характерные области с данными, отвечающие отраженным сигналам, расположенным согласно правилам распространения волн. На этом же рисунке присутствуют зарегистрированные данные со смещением центров наборов данных по оси времени для областей с одной из сторон дефекта.

Однако в некоторых случаях картина регистрации данных будет совсем удручающей для методов математической статистики (рис. 3). На рисунках представлены данные для одного сечения тру-

бы: на рис. 3, *a* – данные для начальных детекторов в сечении, а на рис. 3, *b* – конечные детекторы. Соответственно, точки замера находятся рядом. Зоне дефекта соответствует рис. 3, *a*, жирные линии, а зоне без дефектов – рис. 3, *b*. Анализ визуализированных данных позволяет сделать вывод о том, что картина не отвечает «общепринятой» картине регистрируемых данных, что может быть обусловлено различными причинами – как физического характера, так и проблемами с регистрацией данных электронными приборами.

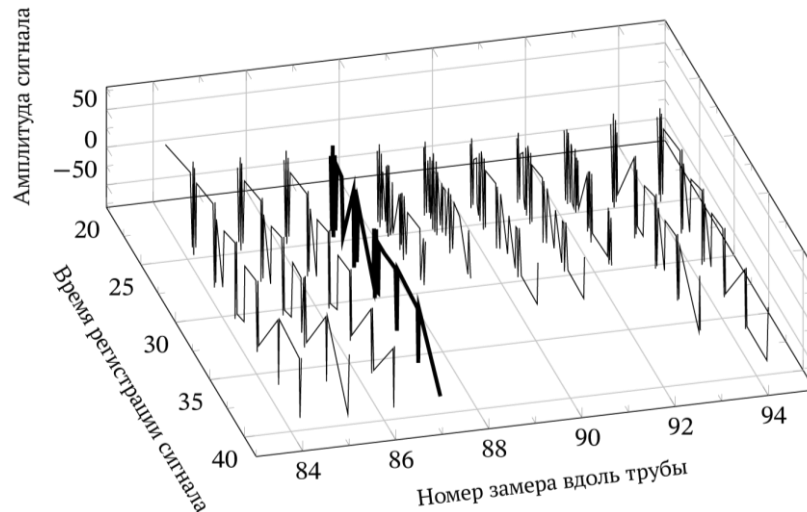
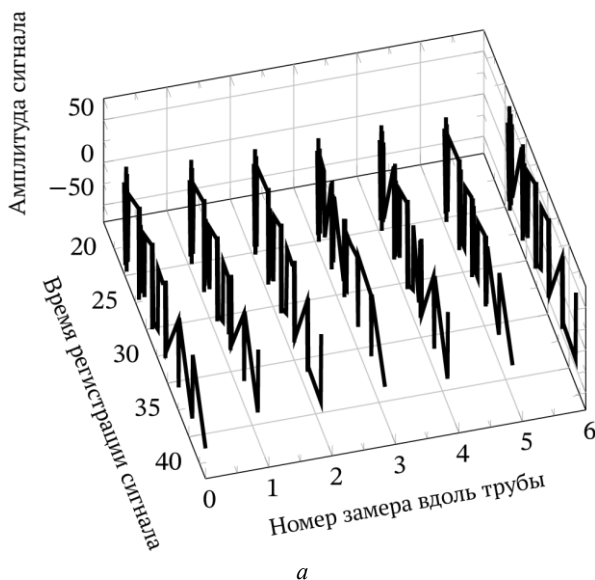
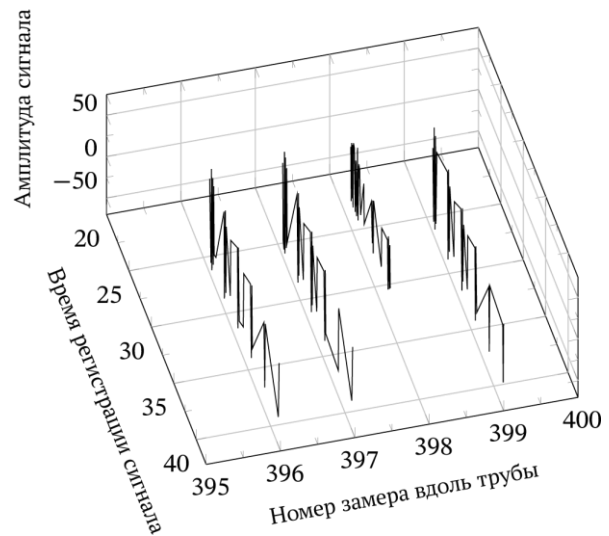


Рис. 2. Визуализация данных для группы датчиков ДВУ
Fig. 2. Visualization of data for a group of sensors of in-line inspection devices



a



b

Рис. 3. Визуализация данных для группы датчиков ДВУ: *a* – область с 0-го датчика по 6-й (с 0-го по 4-й – зона с дефектом);
b – примыкающая область трубы с 396 датчика до 0

Fig. 3. Visualization of data for a group of sensors of in-line inspection devices: *a*) the area from 0 of the sensor to 6 (from 0 to 4 the zone with a defect); *b*) the adjacent area of the pipe from 396 of the sensor to 0

Большое количество данных, неоднородность их получения и стремление к автоматизации прогнозного определения зон дефектов приводят к идее воспользоваться методами обработки данных, основанными на машинном обучении. Условно «классические» методы решения такого класса задач авторам неизвестны.

Популярной практикой при ультразвуковой дефектоскопии больших поверхностей является визуализация, т.е. создание проекций отраженных сигналов. Формируется двумерная картина, где цве-

том отражаются особенности прохождения сигнала, схема которого представлена на рис. 4. Если весь набор данных рассматривать как итог формирования некоторой картины, то для задачи определения мест дефекта можно воспользоваться сверточными нейросетями [8–13], хорошо себя зарекомендовавшими при работе с изображениями.

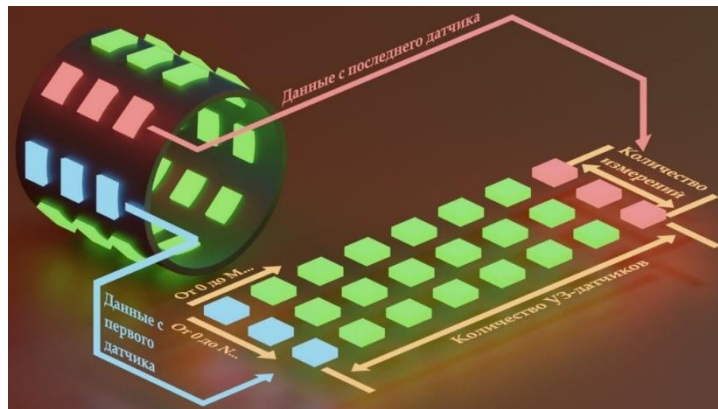


Рис. 4. Схема развертывания регистрируемых данных в плоскую картинку

Fig. 4. Scheme of deploying the recorded data into a flat picture

2. Сверточные нейронные сети

Сверточные нейронные сети, как и большинство нейронных сетей, состоят из нескольких слоев [14]: сверточные слои (convolutional layer), которые могут чередоваться со слоями подвыборки, или избирательными слоями, или слоями пулинга (pooling layers), а также завершающие полносвязные слои. Чем больше слоев, тем мощнее архитектура и лучше обучение нейронной сети. Количество слоев может достигать нескольких десятков [15]. На этапе свертки нейронная сеть агрегирует данные, что в дальнейшем поможет анализу изображения. В общем случае свертку можно создавать для каждого признака. Нейронная сеть будет подбирать их во время распознавания и классификации на каждом сверточном слое [14]. Операция пулинга (pooling) представляет собой алгоритм уменьшения входного тензора по пространственным осям и выделения наиболее важных признаков. Существует множество видов пулинга [16], среди них можно выделить: max pooling – операция выбирает наибольший элемент из части изображения, покрытой ядром пулинга; mean pooling – операция выбирает средний элемент; min pooling – соответственно, минимальный. Алгоритм предполагает выбор размера фильтра пулинга, шага его движения и типа пулинга, после чего ядро пулинга последовательно накладывается на части входного изображения с выбранным шагом, а к части выходного изображения, покрытой ядром пулинга, применяется его операция. Завершением НС служат один или несколько полносвязных слоев, предназначенных, как правило, для классификации [14].

При работе с зарегистрированными ДВУ данными предполагается подавать на вход некоторой НС тензор с размерностью, определяемой как произведение количества строк на количество столбцов и количество каналов, а на выходе ожидать информацию о наличии или отсутствии дефектов. В качестве значений для каналов предполагается использовать значения времен и амплитуд для каждого измерения, представленных одномерным массивом. Таким образом, сформированный тензор можно считать изображением с большим количеством каналов.

3. Результаты тестирования

Для визуализации принципа работы сверточной НС удобнее всего использовать развернутую карту дефектов, поскольку ее значительно проще визуализировать, чем данные, считанные ДВУ в полном объеме, так как последние представляют собой трехмерный тензор. В рамках данного подхода оба набора измерений воспринимаются как изображения, содержащие разное количество кана-

лов. В качестве входных данных НС предполагается использовать изображение, меньшее по высоте и ширине, чем исходное. А так как размер изображения задается по осям высоты и ширины, а данные, считанные ДВУ, и развернутой карты дефектов равны, то изображение, представляющее собой обрезаемую зону развернутой карты дефектов, будет соответствовать такой же зоне в данных ДВУ с той лишь разницей, что они будут иметь разное количество каналов.

Тогда, имея карту признаков, можно обрезать ее часть некоторого размера и использовать ее как выходные данные НС, а как входные данные использовать соответствующий тензор. Количество датчиков и измерений представляется как ширина и высота изображения, а числа в каждой ячейке – как каналы с той разницей, что в RGB-изображениях 3 канала, а для анализируемых данных – вплоть до 64. Развернув данные в двумерную картину, получим следующее изображение (рис. 5), где по горизонтальной оси – номера датчиков, а по вертикальной – номера измерений вдоль оси трубы. На рисунке фиолетовый пиксель соответствует состоянию «отсутствие дефекта», желтый – «дефект».

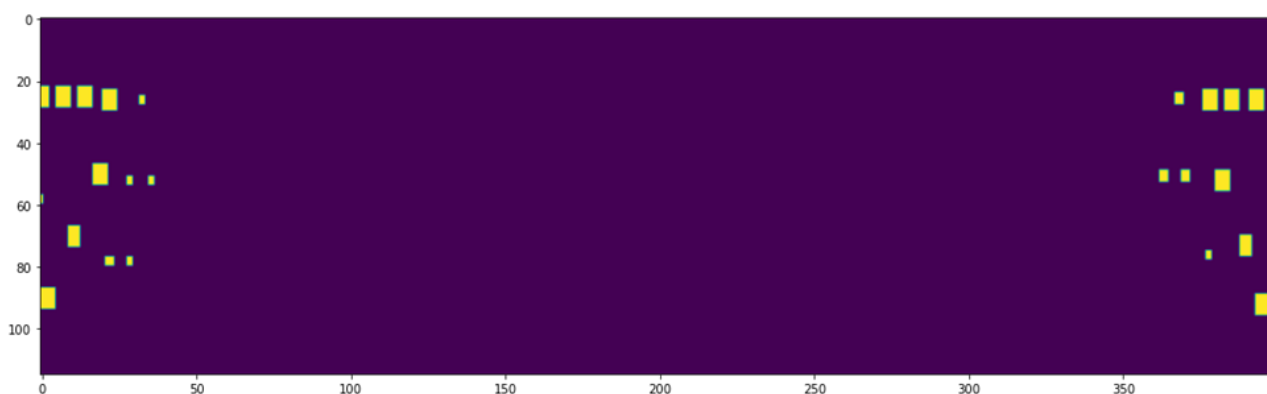


Рис. 5. Развернутая картина бинарной карты дефектов
Fig. 5. An expanded picture of a binary defect map

При подготовке данных получено одно «изображение» размера не менее ста строк, 400 столбцов и 64 канала (время и амплитуда отраженного сигнала). К этим данным формируется маска с ранее заданным количеством строк и столбцов и одним каналом, определяющим наличие и отсутствие дефекта. Для обучения сети такой вид данных не подойдет по причине малого количества наблюдений и большого размера самого наблюдения. Вывод о том, что наблюдение слишком большое, сделан на основе реализации решения задачи при работе с изображениями на нейронной сети VGG16 [17]. При обработке в нейронной сети VGG16 на вход подается, как правило, порядка сотен тысяч значений и сеть имеет глубину более 15 слоев, по этой причине использование данной сети не подходит для нашего обрабатываемого набора значений, поскольку потребует обработки более двух миллионов значений.

В качестве решения данной проблемы предлагается разбить имеющееся изображение на кропы. Кроп – часть изображения меньшего размера, но с таким же количеством каналов. Первой задачей для последующего обучения явилась задача выбора размеров кропа. В исследованиях использовались следующие значения стороны квадратного кропа: 8, 16, 24, 48. Ни одно из значений явных преимуществ перед другими не дало.

Поскольку в выборке мало дефектов, то для обучения предлагается провести аугментацию данных [18], т.е. искусственное увеличение выборки. Аугментация проводилась только для кропов с дефектами. Кроме того, стоит отметить, что проводилась только геометрическая аугментация со следующими операциями: поворот изображения кропа на 90° (три варианта), отражение изображения по горизонтали, отражение по вертикали. Для увеличения количества кропов с дефектами предложена аугментация со сдвигом изображения как по горизонтали, так и по вертикали на половину величины кропа с последующей склейкой недостающей части.

Итоговая нейросетевая модель состояла из двух блоков выделения признаков для значений времен и амплитуд, имеющих одинаковую архитектуру, но разные функции активации с учетом того, в каких пределах располагаются значения каждой выборки. Общая схема такого блока представлена на рис. 6.

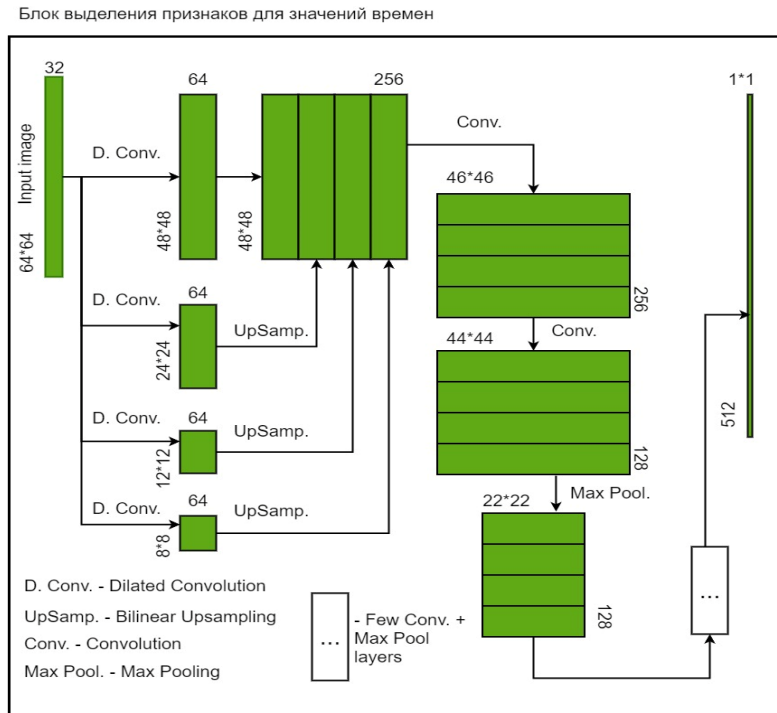


Рис. 6. Общая схема блока выделения признаков из входных данных значений времен
Fig. 6. The general scheme of the block for extracting features from the input data of time values

На вход подается исходный кроп как часть из общего набора данных, размером 64×64 «пикселя». Используя слои расширенной свертки с билинейной повышающей дискретизацией и слои пуллинга с использованием операции max pooling получаем 512 «рисунков» 1×1 . Данная процедура продлевается как для блока значений амплитуд интенсивностей отраженных волн, так и времен.

Далее выходы двух таких блоков объединялись. Общая схема архитектуры нейронной сети представлена на рис. 7.

Общая схема модели

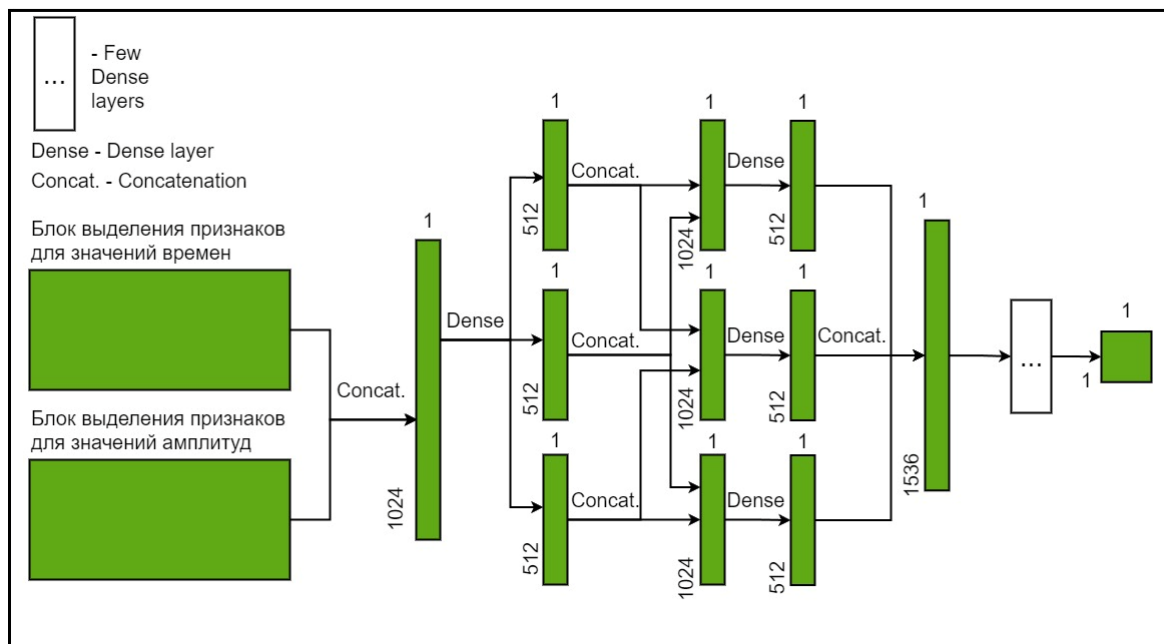


Рис. 7. Общая схема архитектуры нейронной сети
Fig. 7. General diagram of the neural network architecture

За счет операции конкатенации и использования набора полносвязанных слоев НС на выходе получим бинарный ответ: содержит ли данный кроп в своем составе точки дефектов.

Исходные данные вручную разбиты на три выборки: обучающую, валидационную и тестовую. Это позволяет говорить о том, что вероятность переобучения нейронной сети и проблемы с внешним ковариационным сдвигом предусмотрены и данные подготовлены правильно. Также данные внутри всех трех выборок были перемешаны, что позволило процесс обучения сделать более стабильным и качественным.

Стоит заметить, что выходными данными для такой НС является одномерный массив логических значений. Для обучения нейронной сети такой массив создан на основании уже имеющихся выходных данных. Исходные данные имели точки дефектов, приближенные к краям исследуемой области (см. рис. 5). Для исключения краевых эффектов при обучении и последующего тестирования массив был преобразован сдвигом на заданную величину, что позволило получить область дефектов во внутренних областях данных (рис. 8). Операция, на взгляд авторов, допустима, поскольку данные снимаются с поверхности трубы и определение начала и конца окружности чрезвычайно условно.

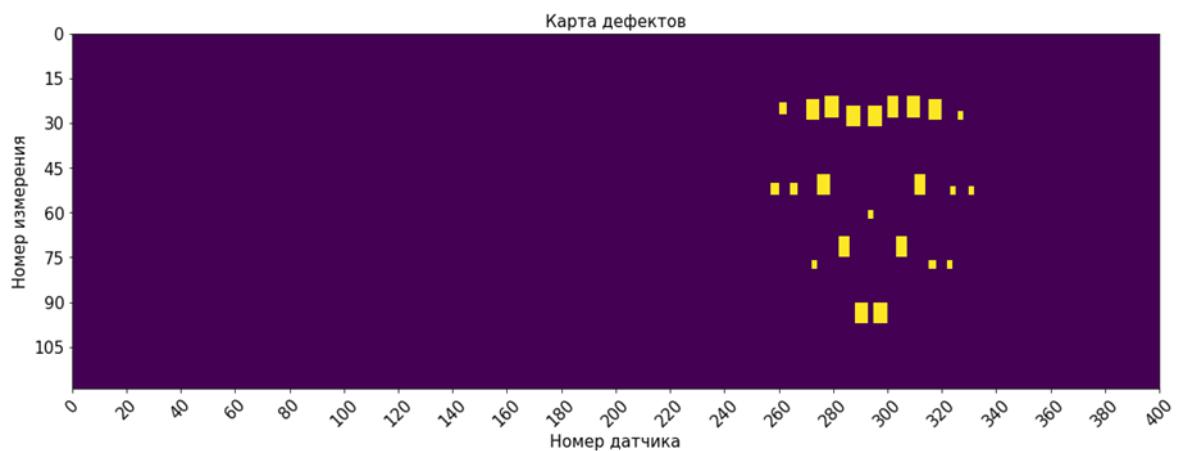


Рис. 8. Развернутая карта дефектов
Fig. 8. Expanded defect map

На подготовленных данных производилось обучение построенной нейронной сети с использованием функции потерь в виде среднего квадрата ошибки (mean squared error). График функции ошибки в процессе обучения на тренировочной и валидационной выборках представлен на рис. 9.

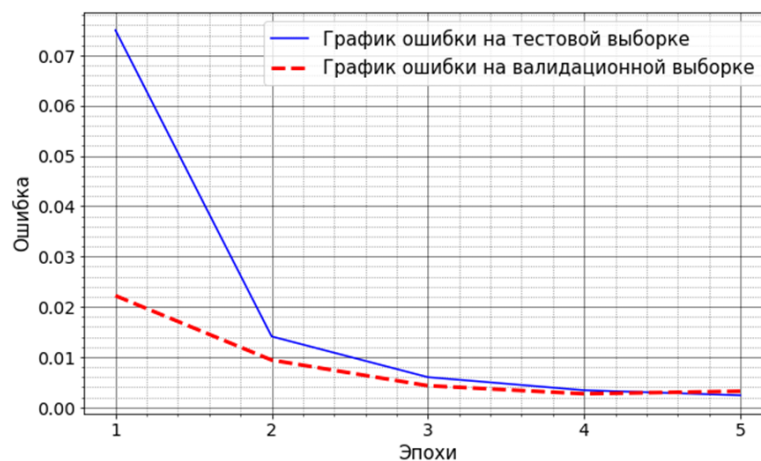


Рис. 9. График функции ошибки в процессе обучения на тренировочной и валидационной выборках
Fig. 9. Graph of the error function in the learning process on the training and validation sets

Из анализа данных можно сделать вывод об отсутствии переобучения ввиду отсутствия расхождения графиков ошибки на валидационной и тренировочной выборках. Наименьшее значение

функции ошибки на тренировочной выборке составило 0,0024, на валидационной – 0,0032. НС протестирована, в результате значение функции ошибки составило 0,0408.

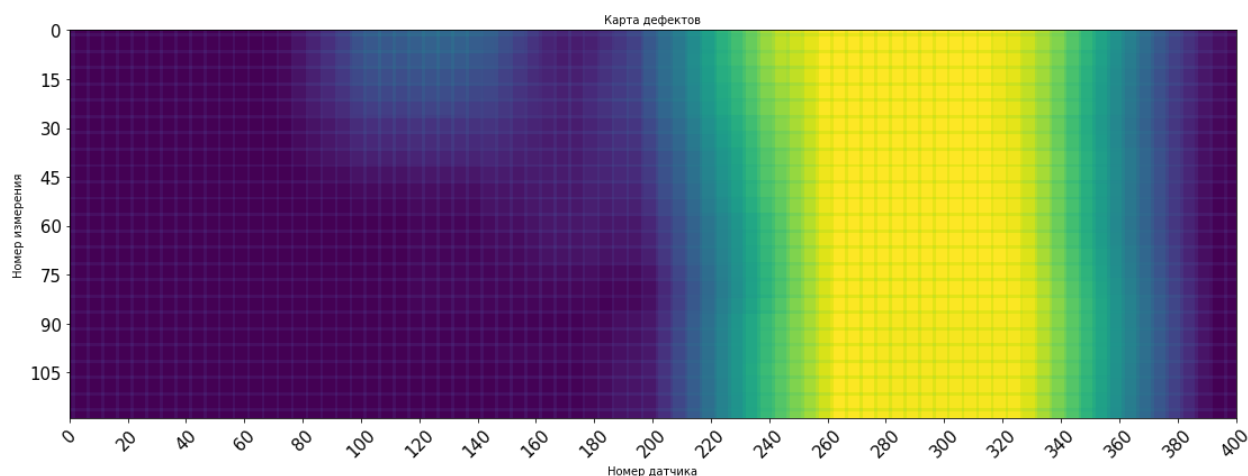


Рис. 10. Итоговая предсказанная развернутая карта дефектов
Fig. 10. Final predicted rolled defect map

На тестовом наборе данных проведено исследование расположения области дефектов. Результаты работы модели нейронной сети представлены на рис. 10. Наиболее светлые точки рисунка отвечают зонам с наибольшей вероятностью дефектов. В целом предсказанная область накрывает область дефектов тестовой выборки (рис. 8).

Заключение

В работе представлен нейросетевой анализ данных с использованием сверточных нейронных сетей. Имеющиеся входные данные были разбиты на обучающую, валидационную и тестовую выборки. Кроме того, данные внутри выборок были перемешаны, что позволило процесс обучения сделать более качественным.

Для обучения модели была использована нейросеть, построенная с использованием сверточных слоев. Функция потерь в виде среднего квадрата ошибки была применена для оценки точности модели. Из проведенного анализа данных можно сделать вывод, что у модели отсутствует переобучение, так как графики ошибки на валидационной и тренировочной выборках не расходятся. После обучения НС была протестирована на тестовом наборе данных, и значение функции ошибки составило 0,0408. Это позволяет оценить точность модели и ее способность предсказывать зоны дефектов.

В целом нейросетевая модель показала свою практическую применимость для предсказания зон дефектов магистральных газопроводов.

Список источников

1. Кальченко О.П., Сауткина Т.Н. Причины коррозии газопроводов // Техническое регулирование в транспортном строительстве. 2020. № 4 (43). С. 299–302.
2. Ma H., Zhang W., Wang Y., Ai Y., Zheng W. Advances in corrosion growth modeling for oil and gas pipelines: A review // Process Safety and Environmental Protection. 2023, V. 171. P. 71–86.
3. Mazumder R.K., Salman A.M., Li Y. Failure risk analysis of pipelines using data-driven machine learning algorithms // Structural Safety. 2021. V. 89. Art. 102047.
4. Алешин Н.П., Крысько Н.В., Щипаков Н.А., Кусый А.Г. Ультразвуковой контроль и комплексное применение методов дефектоскопии в процессе диагностирования магистральных трубопроводов // Наука и технологии трубопроводного транспорта нефти и нефтепродуктов. 2023. Т. 13, № 1. С. 8–17.
5. Жуков А.Д., Григорьев М.В., Данилов В.Н. Исследование вопроса идентификации трещиноподобного коррозионно-механического дефекта акустическими внутритручными инспекционными приборами // Контроль. Диагностика. 2020. № 2. С. 56–63.

6. Дефектоскоп внутритрубный ультразвуковой 10-USK.04-00.000. Номер в госреестре 84767-22 // Федеральный информационный фонд по обеспечению единства измерений. 2022. URL: <https://fgis.gost.ru/fundmetrology/registry/4/items/1397752> (дата обращения: 10.05.2023).
7. Дефектоскоп внутритрубный ультразвуковой 40-ДФР.00-00.000. Номер в госреестре 84766-22 // Федеральный информационный фонд по обеспечению единства измерений. 2022. URL: <https://fgis.gost.ru/fundmetrology/registry/4/items/1397751> (дата обращения: 10.05.2023).
8. Lei L., Sun S., Zhang Y., Liu H., Xu W. PSIC-Net: Pixel-Wise Segmentation and Image-Wise Classification Network for Surface Defects // *Machines*. 2021. V. 9. Art. 9100221.
9. Hu W., Li X., Li C., Li R., Jiang T., Sun H., Huang X., Grzegorzec M., Li X. A state-of-the-art survey of artificial neural networks for Whole-slide Image analysis: From popular Convolutional Neural Networks to potential visual transformers // *Computers in Biology and Medicine*. 2023. V. 161. Art. 107034.
10. Wang J., Zhang J., Cohn A.G., Wang Z., Liu H., Kang W., Jiang P., Zhang F., Chen K., Guo W., Yu Y. Arbitrarily-oriented tunnel lining defects detection from Ground Penetrating Radar images using deep Convolutional Neural networks // *Automation in Construction*. 2022. V. 133. Art. 104044.
11. Zhou S., Canchila C., Song W. Deep learning-based crack segmentation for civil infrastructure: data types, architectures, and benchmarked performance // *Automation in Construction*. 2023. V. 146. Art. 104678.
12. Kaur G., Sinha R., Tiwari P.K., Yadav S.K., Pandey P., Raj R., Vashisth A., Rakhra M. Face mask recognition system using CNN model // *Neuroscience Informatics*. 2022. V. 2, is. 3. Art. 100035.
13. Simonyan K., Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. 2014. URL: <https://arxiv.org/abs/1409.1556> (accessed: 10.05.2023).
14. Багаев И.И. Анализ понятий нейронная сеть и сверточная нейронная сеть, обучение сверточной нейросети при помощи модуля TensorFlow // *Математическое и программное обеспечение систем в промышленной и социальной сферах*. 2020. Т. 8, № 1. С. 15–22.
15. He K., Zhang X., Ren S., Sun J. Deep Residual Learning for Image Recognition // *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016. P. 770–778.
16. Фаворская М.Н. Структурные особенности сверточных нейронных сетей для задач распознавания изображений // *Цифровая обработка сигналов и ее применение DSPA-2019: докл. 21-й Междунар. конф.* 2019. Т. XXI, кн. 2 (2). С. 542–546.
17. Sharma S., Guleria K., Tiwari S., Kumar S. A deep learning based convolutional neural network model with VGG16 feature extractor for the detection of Alzheimer Disease using MRI scans // *Measurement: Sensors*. 2022. V. 24. Art. 100506.
18. Емельянов С.О., Иванова А.А., Швец Е.А., Николаев Д.П. Методы аугментации обучающих выборок в задачах классификации изображений // *Сенсорные системы*. 2018. Т. 32, № 3. С. 236–245.

References

1. Kalchenko, O.P. & Sautkina, T.N. (2020) Prichiny korrozii gazoprovodov [Causes of corrosion of gas pipelines]. *Tekhnicheskoe regulirovanie v transportnom stroitel'stve – Technical regulation in transport construction*. 55(4). pp. 299–302.
2. Ma, H., Zhang, W., Wang, Y., Ai, Y. & Zheng, W. (2023) Advances in corrosion growth modeling for oil and gas pipelines: A review. *Process Safety and Environmental Protection*. 171. pp. 71–86. DOI: 10.1016/j.psep.2022.12.054
3. Mazumder, R.K., Salman, A.M. & Li, Y. (2021) Failure risk analysis of pipelines using data-driven machine learning algorithms. *Structural Safety*. 89. Art. 102047. DOI: 10.1016/j.strusafe.2020.102047
4. Aleshin, N.P., Krysko, N.V., Shchipakov, N.A. & Kusyy, A.G. (2023) Ultrasonic testing and integrated application of non-destructive techniques in main pipelines diagnostics. *Nauka i tekhnologii truboprovodnogo transporta nefti i nefteproduktov – Science & Technologies: Oil and Oil Products Pipeline Transportation*. 13(1). pp. 8–17. DOI: 10.28999/2541-9595-2023-13-1-8-17
5. Zhukov, A.D., Grigoriev, M.V. & Danilov, V.N. (2020) The research for an identification of mechanical crack-like corrosion defect by acoustic in-line inspection tools. *Kontrol'. Diagnostika – Testing. Diagnostics*. 2. pp. 56–63. DOI: 10.14489/td.2020.02.pp.056-063
6. The Federal Information Fund for Ensuring the Uniformity of Measurements. (2022a) *Defektoskop vnutritrubnyy ul'trazvukovoy 10-USK.04-00.000. Nomer v gosreestre 84767-22* [Ultrasonic in-line flaw detector 10-USK.04-00.000. Number in the state register 84767-22]. [Online] Available from: <https://fgis.gost.ru/fundmetrology/registry/4/items/1397752> (Accessed: 10th May 2023).
7. The Federal Information Fund for Ensuring the Uniformity of Measurements. (2022b) *Defektoskop vnutritrubnyy ul'trazvukovoy 40-DFR.00-00.000. Nomer v gosreestre 84766-22* [Ultrasonic in-tube flaw detector 40-DFR.00-00.000. (2022) Number in the state register 84766-22]. [Online] Available from: <https://fgis.gost.ru/fundmetrology/registry/4/items/1397751> (Accessed: 10th May 2023).
8. Lei, L., Sun, S., Zhang, Y., Liu, H. & Xu, W. (2021) PSIC-Net: Pixel-Wise Segmentation and Image-Wise Classification Network for Surface Defects. *Machines*. 9. Art. 9100221. DOI: 10.3390/machines9100221
9. Hu, W., Li, X., Li, C., Li, R., Jiang, T., Sun, H., Huang, X., Grzegorzec, M. & Li, X. (2023) A state-of-the-art survey of artificial neural networks for Whole-slide Image analysis: From popular Convolutional Neural Networks to potential visual transformers. *Computers in Biology and Medicine*. 161. Art. 107034.
10. Wang, J., Zhang, J., Cohn, A.G., Wang, Z., Liu, H., Kang, W., Jiang, P., Zhang, F., Chen, K., Guo, W. & Yu, Y. (2022) Arbitrarily-oriented tunnel lining defects detection from Ground Penetrating Radar images using deep Convolutional Neural networks. *Automation in Construction*. 133. Art. 104044. DOI: 10.1016/j.autcon.2021.104044

11. Zhou, S., Canchila, C. & Song, W. (2023) Deep learning-based crack segmentation for civil infrastructure: data types, architectures, and benchmarked performance. *Automation in Construction*. 146. Art. 104678. DOI: 10.1016/j.autcon.2022.104678
12. Kaur, G., Sinha, R., Tiwari, P.K., Yadav, S.K., Pandey, P., Raj, R., Vashisth, A. & Rakhra, M. (2022) Face mask recognition system using CNN model. *Neuroscience Informatics*. 2(3). Art. 100035. DOI: 10.1016/j.neuri.2021.100035
13. Simonyan, K. & Zisserman, A. (2014) *Very Deep Convolutional Networks for Large-Scale Image Recognition*. [Online] Available from: <https://arxiv.org/abs/1409.1556> (Accessed: 10th May 2023).
14. Bagaev, I.I. (2020) Analiz ponyatiy neyronnaya set' i svertochnaya neyronnaya set', obuchenie svertochnoy neyroseti pri pomoshchi modulya TensorFlow [Analysis of the concepts of a neural network and a convolutional neural network, training a convolutional neural network using the TensorFlow module]. *Matematicheskoe i programnoye obespechenie sistem v promyshlennoy i sotsial'noy sferakh – Software of Systems in the Industrial and Social Field*. 8(1). pp. 15–22. DOI: 10.18503/2306-2053-2020-8-1-15-22
15. He, K., Zhang, X., Ren, S. & Sun, J. (2016) Deep Residual Learning for Image Recognition. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. pp. 770–778.
16. Favorskaya, M.N. (2019) Strukturnye osobennosti svertochnykh neyronnykh setey dlya zadach raspoznavaniya izobrazheniy [Structural features of convolutional neural networks for image recognition problems]. *Tsifrovaya obrabotka signalov i ee primeneniye DSPA-2019* [Digital signal processing and its application – DSPA-2019]. Reports of the 21st International Conference. Iss. XXI. Vol. 2(2). pp. 542–546.
17. Sharma, S., Guleria, K., Tiwari, S. & Kumar, S. (2022) A deep learning based convolutional neural network model with VGG16 feature extractor for the detection of Alzheimer Disease using MRI scans. *Measurement: Sensors*. 24. Art. 100506. DOI: 10.1016/j.measen. 2022.100506
18. Emelyanov, S.O., Ivanova, A.A., Shvets, E.A. & Nikolaev, D.P. (2018) Methods of training data augmentation in the task of image classification. *Sensornye systemy – Sensory Systems Journal*. 32(3). pp. 236–245. DOI: 10.1134/S0235009218030058

Информация об авторах:

Боровской Игорь Георгиевич – профессор, доктор физико-математических наук, заведующий кафедрой экономической математики, информатики и статистики Томского государственного университета систем управления и радиоэлектроники (Томск, Россия). E-mail: igor.g.borovskoi@tusur.ru

Шельмина Елена Александровна – кандидат физико-математических наук, доцент кафедры экономической математики, информатики и статистики Томского государственного университета систем управления и радиоэлектроники (Томск, Россия). E-mail: elena.a.shelmina@tusur.ru

Матолыгин Андрей Анатольевич – старший преподаватель кафедры экономической математики, информатики и статистики Томского государственного университета систем управления и радиоэлектроники (Томск, Россия). E-mail: andrei.a.matolygin@tusur.ru

Ильин Евгений Петрович – студент кафедры экономической математики, информатики и статистики Томского государственного университета систем управления и радиоэлектроники (Томск, Россия). E-mail: eugenecalls@gmail.com

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

Information about the authors:

Borovskoy Igor G. (Doctor of Physical and Mathematical Sciences, Professor, Tomsk State University of Control Systems and Radioelectronics, Tomsk, Russian Federation). E-mail: igor.g.borovskoi@tusur.ru

Shelmina Elena A. (Candidate of Physical and Mathematical Sciences, Associate Professor, Tomsk State University of Control Systems and Radioelectronics, Tomsk, Russian Federation). E-mail: elena.a.shelmina@tusur.ru

Matolygin Andrey A. (Senior Lecturer, Tomsk State University of Control Systems and Radioelectronics, Tomsk, Russian Federation). E-mail: andrei.a.matolygin@tusur.ru

Ilin Evgenyi P. (Student, Tomsk State University of Control Systems and Radioelectronics, Tomsk, Russian Federation). E-mail: eugenecalls@gmail.com

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

Поступила в редакцию 10.05.2023; принята к публикации 04.09.2023

Received 10.05.2023; accepted for publication 04.09.2023

Научная статья

УДК 004.95

doi: 10.17223/19988605/64/7

Геоинформационная система динамической пространственной кластеризации распределенных источников данных

Андрей Владимирович Воробьев¹, Гульнара Равилевна Воробьева²

¹ Геофизический центр РАН, Москва, Россия

^{1,2} Уфимский университет науки и технологий, Уфа, Россия

¹ geomagnet@list.ru

² gulnara.vorobeva@gmail.com

Аннотация. Представлены результаты разработки подхода к динамической кластеризации геопро пространственных объектов, распределение которых характеризуется выраженной пространственной анизотропией. Предложен подход к динамической пространственной кластеризации на основе анализа распределения значений индекса Моришита, дополненный кластеризацией на основе атрибутивных значений соответствующих точек. Предложен новый вариант архитектуры веб-приложения, реализующего динамическую пространственную кластеризацию, представляющий собой комбинирование «чистой» архитектуры и паттерна «модель–представление–контроллер», обеспечивающий повышение модульности приложения, а также его независимость от программных фреймворков. Разработан алгоритм динамической пространственной кластеризации, в асинхронном режиме обеспечивающий расчет индекса Моришита и информационной корреляции для формирования кластерной структуры пространственных объектов в виде доступного для рендеринга слоя.

Ключевые слова: геоинформатика; пространственные данные; кластеризация; индекс Моришита; геоинформационные системы.

Благодарности: Исследование выполнено при финансовой поддержке Российского научного фонда в рамках научного проекта № 21-77-30010.

Для цитирования: Воробьев А.В., Воробьева Г.Р. Геоинформационная система динамической пространственной кластеризации распределенных источников данных // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2023. № 64. С. 61–73. doi: 10.17223/19988605/64/7

Original article

doi: 10.17223/19988605/64/7

Geoinformation system for dynamic spatial clustering of distributed data sources

Andrei V. Vorobev¹, Gulnara R. Vorobeva²

¹ Geophysical Center RAS, Moscow, Russian Federation

^{1,2} Ufa University of Science and Technology, Ufa, Russian Federation

¹ geomagnet@list.ru

² gulnara.vorobeva@gmail.com

Abstract. The results of developing an approach to dynamic clustering of geospatial objects, the distribution of which is characterized by pronounced spatial anisotropy, are presented. An approach to dynamic spatial clustering is proposed based on the analysis of the distribution of Morishita index values, supplemented by clustering based on the attribute values of the corresponding points. A new version of the web application architecture that implements dynamic spatial clustering is proposed, which is a combination of a “pure” architecture and a “model-view-controller” pattern that provides an increase in the modularity of the application, as well as its independence from software frameworks. An algorithm for dynamic spatial clustering has been developed, which in an asynchronous

mode provides the calculation of the Morishita index and informational correlation for the formation of a cluster structure of spatial objects in the form of a layer available for rendering.

Keywords: geoinformatics; spatial data; clustering; Morishita index; geoinformation systems.

Acknowledgments: The research was carried out with the financial support of the Russian Science Foundation for No. 21-77-30010.

For citation: Vorobev, A.V., Vorobeva, G.R. (2023) Geoinformation system for dynamic spatial clustering of distributed data sources. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naja tehnika i informatika – Tomsk State University Journal of Control and Computer Science*. 64. pp. 61–73. doi: 10.17223/19988605/64/7

Введение

Прочно укрепившиеся тенденции к цифровизации современного общества находят свое отражение в том числе и во все возрастающих потребностях разнообразных областей деятельности к информации, имеющей геопространственную привязку, – пространственным данным. Современные реалии таковы, что практически ни одна сфера науки и производства не обходится без данных, описывающих местоположение объектов на поверхности Земли или характеризующих пространственное распределение параметров явлений и процессов различной природы. Считается, что около 80% данных, образующих национальные (в том числе государственные) информационные ресурсы, снабжено или без существенных затрат может быть снабжено координатным описанием, фактически или потенциально представляя из себя пространственные данные [1].

Задача обработки, анализа и интерпретации пространственной информации является очень важной и сложной, а ее решение невозможно без комплексного системного подхода и различных методов, характеризующих те или иные объекты, явления и процессы. Требуется разработка (преимущественно на программном уровне) комплексных интеграционных межотраслевых решений на стыке геодезии, геоинформатики и отраслевых технологий.

Одним из начальных этапов работы с пространственными данными является оценка их распределения с точки зрения регулярности распределения по соответствующей пространственной сетке. Если пространственные данные (и, соответственно, их источники) могут быть описаны нерегулярной кластерной сетью мониторинга, то необходима их декластеризация для расчета характеристик глобальной статистики. В случае высокой плотности пространственного распределения анализируемых данных имеет место кластерная сеть мониторинга; при этом выделяются пространственные области с более плотными по сравнению с остальными значениями. Тогда возникает необходимость декластеризации во избежание искажения оценок глобальной статистики [2].

Идентификация пространственных кластеров является основным методом, используемым при анализе сети мониторинга, цель которого – оценка равномерности распределения точек измерения в пространстве. Поскольку значительная часть геостатистических аналитических операций предполагает оперирование равномерной (регулярной) сеткой измерений, то процедура пространственной кластеризации – необходимое условие для последующей декластеризации, пространственной интерполяции и, как следствие, формирования регулярной сети мониторинга. На программном уровне операция кластеризации измерений должна быть реализована как в виде доступного сторонним приложениям набора данных с выделенными группами, так и в формате пространственного слоя, доступного для визуальной интерпретации пользователями.

Известно, что пространственная кластеризация в геостатистике представляет собой метод группировки набора пространственных объектов в пространственные регионы (группы), называемые кластерами. При этом пространственные объекты внутри заданного пространственного кластера демонстрируют высокую степень сходства, тогда как кластеры максимально различны. Как правило, пространственная кластеризация статична: размеры ячейки априори известны, и распределение пространственных объектов по кластерам осуществляется единожды. Вместе с тем работа с простран-

ственными объектами в интерактивном режиме предполагает, что размеры ячейки при кластеризации должны меняться динамически при манипуляции с плоскими или трехмерными картами. Иными словами, необходима так называемая динамическая кластеризация, при которой размер ячейки априори неизвестен и рассчитывается на основании параметров геопространственного изображения каждый раз при его изменении.

На сегодняшний день задача пространственной кластеризации решается во многих прикладных областях, где предполагается обработка данных с геопространственной привязкой. Так, к примеру, в работе [3] предложен метод пространственной регрессии на основе методов и моделей кластеризации, который позволяет спрогнозировать влияние внешних факторов на значение исследуемой переменной в заданном пространственном кластере. Технология представляет интерес для выявления пространственных закономерностей для исследования, к примеру, социально-экономических процессов, данные которых имеют выраженную пространственную привязку. Известны исследования, направленные на изучение пространственных закономерностей развития природных процессов на основе пространственно-кластеризованной модели переменных коэффициентов, в которой коэффициенты регрессии могут плавно изменяться в пределах каждого кластера, но резко изменяться на границах соседних кластеров [4].

В литературе описывается множество различных методов пространственной кластеризации, основанных на гипотезе компактности, согласно которой атрибутивные данные, принадлежащие одному и тому же кластеру, схожи, а измерения, относящиеся к разным кластерам, достаточно четко отделимы друг от друга. При этом в качестве меры сходства чаще всего используется функция расстояний, что, по сути, является наиболее очевидным способом введения меры сходства для множеств объектов, интерпретируемых как точки в евклидовом пространстве [5].

Несмотря на серьезные достижения исследователей в области кластерного анализа, разработанные алгоритмы оказываются слишком статичными для динамического формирования пространственных кластеров в зависимости от размеров соответствующих пространственных ячеек [6]. Были разработаны различные методы пространственной кластеризации, включая статистические подходы, триангуляцию Делоне [2], деление на основе сетки [2], методы случайного блуждания [2] и даже полный перебор [7]. Однако для эффективного исследования больших и многомерных геопространственных наборов данных необходимо разработать высоко интерактивную среду анализа, которая объединила бы лучшие академические достижения и программные технологии.

В настоящее время также существует ряд программных решений по кластеризации источников данных, основанных на комплексе геоинформационных технологий. К примеру, одним из таких подходов является программная библиотека ArcGIS API for JavaScript, которая по входным пространственным данным (к примеру, GeoJSON или CSV) объединяет измерения в визуальные кластеры [8]. При этом невозможно иным, отличным от визуального, способом оценить, равномерно ли распределены точки в пространственном слое. Для серверной (или десктопной) программной реализации известны применения языка программирования Python и библиотеки sklearn с классом cluster, однако при таком подходе не учитываются специфика рассматриваемых данных и статистическое (а не детерминистическое) описание их пространственного распределения (здесь имеют место пространственные регионализированные переменные), что может привести к ложным аналитическим выводам [9].

Вместе с тем по-прежнему актуальной и нерешенной остается задача динамической кластеризации пространственных данных (или динамической пространственной кластеризации). К примеру, существующие программные подходы по визуализации кластеров не адаптированы под изменение масштабов соответствующего пространственного изображения и остаются неизменными при работе с картографическими примитивами, что не позволяет пользователям оценить реальную картину пространственного распределения. В этой связи актуальна задача разработки программного решения для динамической пространственной кластеризации региональных пространственных переменных (в обобщенной форме – источников данных), основанного на ГИС-технологиях, наследующего достоинства известных решений и нивелирующего их недостатки.

1. Формализация задачи пространственной кластеризации

Особенностью пространственных данных является невозможность формульного описания законов их распределения на основе физической природы процессов, обуславливающих соответствующее явление. Поэтому для их анализа целесообразно применять геостатистические оценки, которые являются адаптивными и опираются на информацию о внутренней структуре самих данных [2]. В этом случае в качестве объекта анализа выступают так называемые пространственные переменные как переменные с координатной привязкой [2].

Здесь и далее предполагается, что расстояние между точками внутри пространственного кластера, а также между кластерами определяется евклидовой нормой.

Одним из статистических методов, позволяющих оценить равномерность распределения пространственных точек в сети мониторинга, является индекс Моришита I_δ [2, 10]. Индекс характеризует вероятность того, что две случайные пространственные точки могут попасть в один пространственный кластер, и может быть определен из следующего соотношения:

$$I_\delta = Q \frac{\sum_{i=1}^Q n_i(n_i-1)}{N(N-1)}, i = 1, \dots, Q, \quad (1)$$

где N – общее число пространственных точек, Q – количество ячеек разбиения, n_i – количество пространственных точек, которые попали в i -ю ячейку.

Если значение индекса Моришита колеблется около 1, то распределение пространственных точек считается равномерным. В случае, если значение I_δ превышает 1, сеть мониторинга имеет выраженную кластерную структуру, что затрудняет при анализе получение репрезентативной гистограммы распределения и требует проведения процедуры декластеризации. Каждой пространственной точке в зависимости от ее принадлежности пространственному кластеру присваивается весовое значение, которое учитывается, к примеру, при расчете усредненных и агрегированных значений в соответствующем пространственном регионе. В общем случае, когда точка располагается в регионе с высокой концентрацией измерений, ее весовое значение меньше, чем в регионе с равномерной сетью мониторинга. Кроме того, для решения задачи декластеризации также может быть применен распространенный в настоящее время метод ячейковой декластеризации (cell-declustering) [10].

Расчет индекса Моришита ориентирован преимущественно на кластеризацию на основе плотности, принимая во внимание только геодезическое расстояние между соответствующими точками. В общем виде частота попадания пространственных точек в заданный пространственный регион (ячейку) заданного размера определяется следующим образом:

$$p = \frac{Z}{S}, Z = |\{z_1, z_2, z_3, \dots\}|, \quad (2)$$

где S – площадь рассматриваемого пространственного региона (ячейки), Z – мощность множества пространственных точек, размещенных в указанном пространственном регионе.

Здесь важно заметить необходимость динамической адаптации пространственных кластеров, полученных на основе расчета индекса Моришита: при изменении размеров и количества ячеек, на которые разбивается сеть мониторинга, соответствующее значение должно быть перерассчитано. Наиболее очевидна необходимость такой адаптации непосредственно при пространственной визуализации, когда изменение масштаба виртуальной плоской карты или глобуса требует повторной идентификации пространственных кластеров в новых условиях. Таким образом, динамический характер пространственной кластеризации определяется непосредственно тем, что состав пространственного кластера адаптируется к размеру соответствующей ячейки, который, в свою очередь, может меняться при манипулировании визуализацией пространственных данных и иных выполняемых конечным пользователем операций.

Представляется целесообразным дополнить известный подход кластеризацией на основе атрибутивных значений соответствующих точек. Целью такого подхода является не выявление неравномерности сети мониторинга как совокупности расположений пространственных объектов, а определение сходных по значениям измерений в рамках выявленного анализом плотности кластера и их последующая группировка для получения средних и агрегированных статистических характеристик. Для

идентификации такого подхода предлагается ввести термин «атрибутивная кластеризация», которая, в отличие от пространственной кластеризации по плотности измерений, ориентируется на конкретные значения параметров измерений. При этом атрибутивная кластеризация может быть одномерной (при этом анализ построен на оценке значений одного атрибута пространственных точек) и многомерной (когда последовательно рассматриваются два и более атрибутов пространственных точек).

В качестве критерия оценки схожести атрибутивных значений представляется целесообразным выбрать значение коэффициента информационной корреляции, под которым понимается мера зависимости между атрибутивными значениями двух пространственных точек X и Y , определяемая как функция от количества информации в одной случайной величине относительно другой [11]:

$$R(X, Y) = \sqrt{1 - e^{-2I(X, Y)}}, \quad (3)$$

где $I(X, Y)$ – количество информации, являющееся, в свою очередь, мерой взаимной информации для соответствующих выборок атрибутивных значений в пространственных точках X и Y .

При этом взаимная информация квантифицирует количество информации (по мере Хартли), полученное от одной величины при наблюдении другой, и рассчитывается на основании известной формулы, базирующейся на постулатах теоремы Шеннона [12]:

$$I(X, Y) = \sum_{i=1}^n \sum_{j=1}^n p(x_i, y_j) \ln \frac{p(x_i, y_j)}{p(x_i)p(y_j)}, \quad (4)$$

где $p(x_i)$ – вероятность появления значения x_i в точке X , $p(y_j)$ – вероятность появления значения y_j в точке Y , $p(x_i, y_j)$ – вероятность совместного появления значения x_i в точке X и значения y_j в точке Y .

На основе анализа ретроспективных данных представляется возможным определить группы источников данных, для которых в соответствующих пространственных точках наблюдается наилучшая информационная корреляция. В зависимости от конкретной прикладной предметной области, относительно которой рассматриваются анализируемые пространственные данные, ретроспективные значения могут быть различной длительности. К примеру, для геомагнитных данных, регистрируемых наземными магнитными обсерваториями и вариационными станциями, целесообразно рассматривать годовые архивы наблюдений ввиду специфики периодической и сезонной составляющих обрабатываемых временных рядов.

Соответственно, пространственная кластеризация должна быть выполнена в два этапа. На первом из них предполагается идентификация степени кластерности сети мониторинга по итогам расчета индекса Моришита. Далее для выделенных пространственных групп определяются пространственные подгруппы, составленные на основе результатов расчета коэффициента информационной корреляции.

Вообще говоря, проведенные исследования показали, что наибольшей информационной корреляцией характеризуются источники данных, которые территориально близко располагаются друг к другу. При этом было установлено [13], что для таких источников данных целесообразно применение так называемого информационного резервирования, позволяющего восстанавливать временные ряды одного источника данных по синхронно зарегистрированным данным соседнего с ним источника. Соответственно, возможно определить пространственные кластеры с источниками данных, которые по своей геопространственной привязке могут рассматриваться как потенциальные «резервные» источники, а для повышения точности применения указанного решения внутри выделенного пространственного кластера выделить подгруппу источников с наилучшими показателями информационной корреляции.

Проблема (преимущественно техническая) при этом заключается в том, что с точки зрения вычислительной скорости процедура определения источников данных с наибольшей информационной корреляцией является весьма ресурсозатратной, что было подтверждено серией вычислительных экспериментов на серверах используемого хостинга (технические характеристики: процессоры: 88 * Intel(R) Xeon(R) CPU E5-2699 v4 @ 2.20GHz, операционная система Ubuntu Server Edition). При этом сначала последовательно рассчитывается индекс Моришита для выявления пространственных кластеров.

Согласно отраслевым рекомендациям в веб-разработке, необходимо поддерживать время отклика страницы максимум в 500 мс в среднем, чтобы иметь веб-приложение, которое пользователю комфортно использовать даже при сетевом времени 200–300 мс. При этом важно отметить, что повышение быстродействия благотворно сказывается на устойчивости сайта к высоким нагрузкам – чем

быстрее приложению удастся обработать каждый отдельный запрос, тем больше запросов за единицу времени может быть обработано и тем больше пользователей одновременно может работать с приложением.

Исследования показали, что если количество анализируемых источников данных превышает 50, то расчет в среднем занимает порядка 2–3 мин, что негативно сказывается на производительности веб-приложения в целом (в то время как с точки зрения эргономики в веб-приложении время отклика от сервера на клиентскую сторону не должно превышать 4–5 с, а с точки зрения производительности в целом – не превышать 500 мс). Если добавить к этому еще вычисление коэффициента информационной корреляции в цикле для каждой пары источников данных, то совокупное время выполнения вычислительных операций занимает более 8–9 мин в зависимости от количества рассматриваемых источников данных.

Снизить временные затраты на выполнение перечисленных операций возможно посредством сокращения рассматриваемых источников данных. Для этого расчет коэффициента информационной корреляции осуществляется не на всем множестве доступных источников данных, а локально, внутри выделенного на предшествующем этапе региона – пространственного кластера. В результате проведенных вычислительных экспериментов для источников геомагнитных данных было установлено сокращение времени расчета коэффициента информационной корреляции в среднем в 3,38 раза, что для рассматриваемого примера составляет около 1,7 минут. Если предварительно выполнить расчет пространственных кластеров по известным геопространственным данным (без привязки к атрибутивным значениям), то указанное расчетное время еще более сократится, что обеспечит соответствие скорости выполнения вычислительных операций и формирования отклика от сервера к клиенту заданным эргономическим требованиям к веб-приложениям.

2. Архитектура программного инструмента для динамической пространственной кластеризации

В общем виде программное средство для динамической пространственной кластеризации предполагает на основе входных пространственных данных формирование слоя пространственных кластеров для анализа и последующей визуальной интерпретации, в том числе сторонними программными библиотеками. При этом доступность программного решения широкому кругу конечных пользователей без привязки к конкретным аппаратно-программным платформам может быть реализована в веб-ориентированной среде на основе современных геоинформационных технологий.

Идентификация кластерной структуры пространственных данных на программном уровне реализована на базе паттерна «наблюдатель» (observer) [14]. Объект, называемый «субъектом» (subject), в данном случае является экземпляром серверного класса, через который выполняется интеграция исходных данных, полученных из распределенных гетерогенных источников (в частности, для геомагнитных данных это соответствующие программные сервисы магнитных обсерваторий и вариационных станций либо, к примеру, экземпляр класса Python API агрегатора геомагнитной информации SuperMag [15]). При этом функция расчета индекса Моришита выступает в качестве объекта, который называется «наблюдателем» (observer) и исполняется в качестве функции обратного вызова (callback) при изменении состояния объекта «субъекта» (т.е. при завершении получения информационных потоков от сторонних источников и формировании готового для дальнейшей обработки набора пространственных данных).

В общем виде процесс взаимодействия между указанными субъектом и наблюдателем можно представить в виде схемы. Здесь задан класс “Data Aggr”, предназначенный для формирования единого набора пространственных данных по запросам к удаленным распределенным источникам различного типа. На класс “Data Aggr” (на его экземпляр) подписывается (subscribe) функция “Cluster”, обеспечивающая идентификацию кластерности сети мониторинга по пространственным данным. При этом функция обратного вызова подписана на событие, определяемое завершением загрузки данных от распределенных источников и формированием готового для дальнейшей обработки набора про-

пространственных данных. При наступлении обозначенного события подписчик экземпляра класса направляет запрос на получение данных, в ответ на который формируется отклик с искомым набором данных.

Предложенное программное решение базируется на комбинировании двух видов архитектуры веб-приложений. Основой является архитектура вида «Модель–Представление–Контроллер» (Model-View-Controller, MVC), что обусловлено отчасти необходимостью реализации серверной части на языке программирования Python для упрощения последующей ее интеграции с другими проектами авторов. В дополнение к обозначенной использована так называемая «чистая» архитектура (the clean architecture), широко используемая в современных веб-приложениях для обеспечения независимости от программных фреймворков [16]. Преимуществом обоих подходов является многоуровневая реализация веб-приложения, при которой его компоненты разделяются на слои в зависимости от своего назначения или обязанностей (принцип, известный как separation of concerns [16]), что, как результат, обеспечивает независимость конечного программного продукта от фреймворка, пользовательского интерфейса, базы данных и какого-либо внешнего приложения / сервиса.

Комбинирование обозначенных подходов нашло отражение в следующей совокупности компонентов (слоев) веб-приложения, реализующего динамическую пространственную кластеризацию:

- View (представление) – обеспечивает обработку пространственных данных в контексте идентификации соответствующих пространственных кластеров и анализа информационной корреляции соответствующих источников данных;
- Controller (контроллер) – реализует взаимодействие между клиентом и сервером, фактически между пользователем и серверной частью приложения (его бизнес-логикой);
- Repository (репозиторий) – отвечает за получение пространственных данных из внешних источников различного типа, а также формирование набора данных для последующей обработки и анализа с учетом заданной в приложении бизнес-логики;
- Model (модель) – обеспечивает описание предметной области в терминах информационной модели или внутреннего домена веб-приложения;
- Service (сервис) – реализует непосредственно бизнес-логику веб-приложения и обеспечивает в том числе ее переиспользование между компонентами приложения;
- Template (шаблон) – описывает представление отклика от сервера конечному пользователю в виде совокупности визуальных элементов, доступных для рендеринга браузером.

В предложенной архитектуре внешними при этом являются два слоя: Repository, поскольку он обращается к сторонним источникам данных, и Template отвечающий непосредственно за взаимодействие с конечным пользователем через визуальные элементы управления в окне браузера. При этом применяемое в «чистой» архитектуре правило зависимостей [16] должно быть несколько адаптировано под специфику предложенной архитектуры в контексте независимости только внешних слоев, которые при незначительных корректировках могут быть использованы несколькими приложениями под разные задачи. По сути, здесь прослеживается концепция отделения бизнес-логики приложения от ее представления, также положенная в основу архитектуры типа MVC.

Известно, что «чистая» архитектура не привязана к определенному фреймворку или языку программирования, что, однако, нехарактерно для архитектуры «Модель–Представление–Контроллер»: здесь набор доступных программных технологий ограничен. В рамках проекта, рассматриваемого в статье, был применен фреймворк Django, обеспечивающий реализацию MVC-архитектуры веб-приложения и предполагающий создание серверных сценариев на языке программирования Python. При этом, согласно предлагаемой концепции, компоненты View, Model и Controller уже представлены во фреймворке Django: компонент Model с описанием используемых объектов определяет работу компонента View, представленного множеством Python-функций, вызов которых прописан в компоненте Controller и связанном с ним модуле URL Dispatcher.

В общем виде предложенный перечень компонент может быть агрегирован в виде совокупности слоев. Слой представления (presenter layer) объединяет компоненты, ответственные за визуализацию полученного с серверной стороны отклика, а также модули, реализующие используемую в приложении

модель. Слой управления (control layer) включает в себя программные компоненты, предназначенные для реализации бизнес-логики приложения, размещенные на стороне веб-сервера и непосредственно формирующие отклик для клиентской стороны. В свою очередь, слой данных (data layer) объединяет компоненты, отвечающие за информационное обеспечение приложения (в том числе и сторонние модули получения данных).

Для взаимодействия между отдельными слоями используются дополнительные программные модули, каждый из которых предполагает применение двух программных интерфейсов: один для запроса, другой – для ответа. Такие модули-переходы обеспечивают независимость модулей внутреннего слоя от компонент внешнего слоя, сохраняя при этом возможность передачи данных. При этом важно отметить, что имеет место неявная зависимость между слоями, обратными потоку данных. С точки зрения «чистой» архитектуры это соответствует принципу инверсии зависимостей [16].

Таким образом, предложенная архитектура приложения, сочетая особенности MVC и «чистой» архитектуры, с одной стороны, затрагивает все веб-приложение, усиливая его модульность, переносимость и обратную совместимость. С другой стороны, применение презентационного паттерна MVC (MVP) обеспечивает отделение бизнес-логики от визуальной компоненты приложения, что существенно упрощает работу фронтэнд-разработчика.

3. Алгоритм динамической пространственной кластеризации

В общем виде процесс динамической пространственной кластеризации может быть представлен последовательностью этапов получения и обработки соответствующих данных (рис. 1).

На начальной стадии выполнения алгоритм предполагает обращение к внешним и внутренним источникам данных посредством соответствующих протоколов передачи данных (как правило, HTTP(s) или FTP). В соответствии с прикладной спецификой обрабатываемых данных определяется период, за который должен быть проведен анализ. Так, к примеру, в случае геомагнитных данных представляется целесообразным рассматривать годовые архивы минутных данных, зарегистрированных магнитными обсерваториями и вариационными станциями. Для данных, к примеру, по вероятности наблюдения полярных сияний можно рассмотреть как годовые, так и суточные архивы 5-минутных или получасовых прогнозов.

На следующем шаге в соответствии с предлагаемым алгоритмом выделенные источники данных с учетом их пространственного распределения анализируются на предмет оценки соответствующей сети мониторинга. Для этого по имеющимся данным выполняется расчет значения индекса Моришита, который и характеризует пространственную анизотропию пространственных объектов.

В случае если установлено, что сеть мониторинга регулярная, решение задачи кластеризации источников данных является нецелесообразным и алгоритм завершает свою работу. В противном случае осуществляется расчет расстояния между точками и формируется соответствующая гистограмма. По результатам проведенных расчетов и на основании принятого по умолчанию масштаба картографической подложки определяется размер ячейки, для которой требуется вычислить пространственные кластеры по имеющимся пространственным данным. Здесь же выполняется расчет информационной корреляции источников по накопленным данным, уже в составе заданной пространственной ячейки.

На последующем шаге выполняется непосредственно формирование геопространственного слоя с выделенными на основании результатов предыдущих этапов пространственными кластерами. При этом целесообразно представить сформированный слой в одном из традиционных форматов пространственного рендеринга, например, GeoJSON, CSV, KML / KMZ и пр.

Далее выполняется непосредственно визуализация пространственного слоя на картографической подложке. При этом не имеет значения формат представления результата визуализации: это может быть как двумерная плоская карта, так и трехмерный виртуальный глобус. Единственными требованиями здесь являются поддержка рендеринга пространственного слоя поверх картографической подложки и возможность манипулирования им как отдельным графическим объектом.

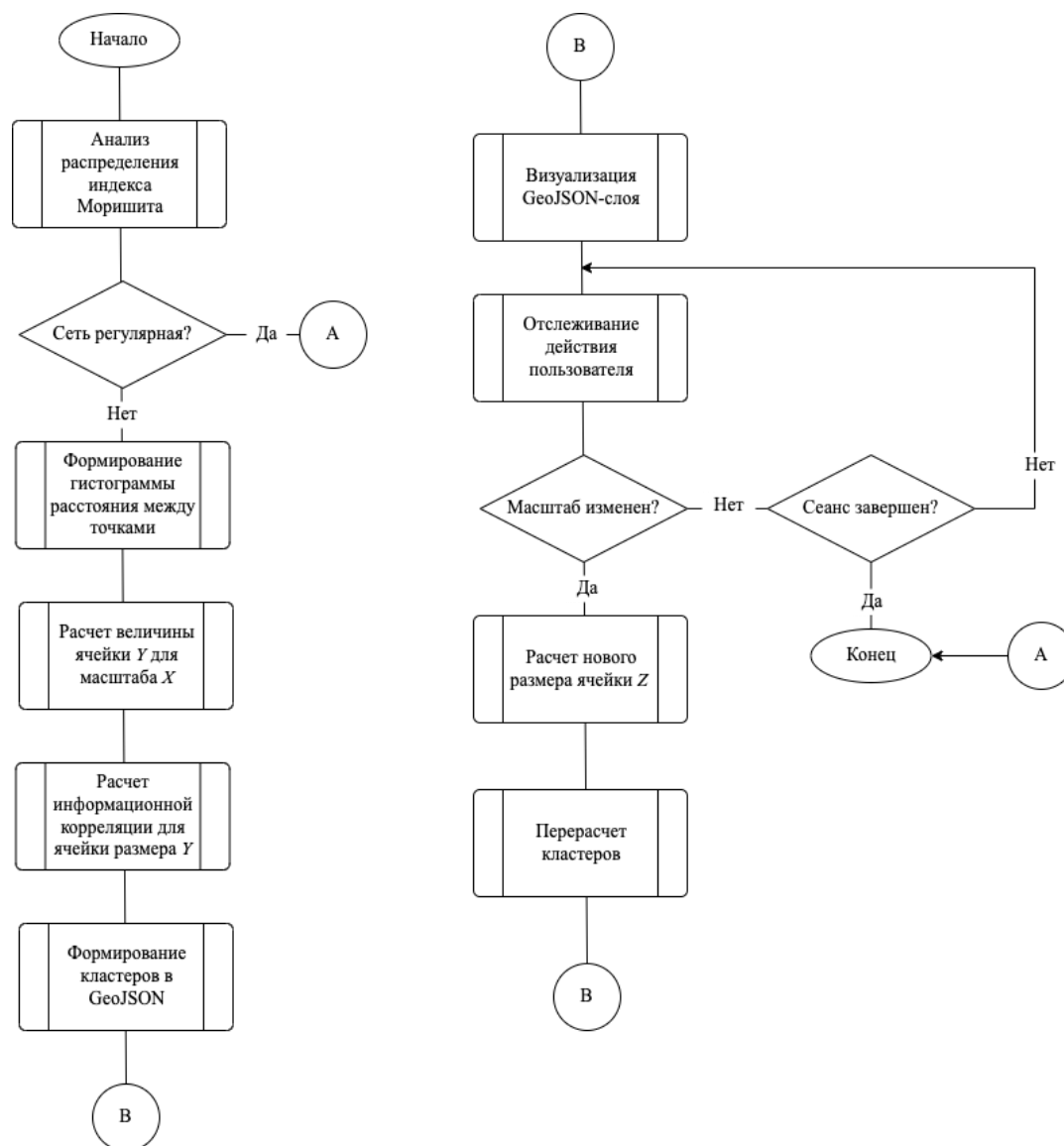


Рис. 1. Схема алгоритма динамической пространственной кластеризации
 Fig. 1. Scheme of the dynamic spatial clustering algorithm

По завершении процедуры первичного рендеринга, имеющего место на начальном этапе пользовательского сеанса взаимодействия с веб-приложением, запускается на выполнение процедуры отслеживания действий пользователя со сформированным геопространственным изображением: отслеживается изменение пользователем масштаба пространственного изображения, что характеризует непосредственно динамическую процедуру кластеризации. В случае наступления указанного события выполняется перерасчет размера ячейки и, соответственно, заново осуществляется процедура кластеризации с учетом новых значений масштаба изображения и размера ячейки.

Отслеживание изменения пользователем масштаба пространственного изображения выполняется на протяжении всего цикла работы пользователя с веб-приложением (пользовательского сеанса). По завершении сеанса работы работа алгоритма динамической пространственной кластеризации завершается.

4. Примеры реализации веб-приложений для динамической кластеризации

В качестве примера программной реализации предложенных решений рассматривается два веб-приложения, разработанных авторами статьи и обеспечивающих обработку пространственных данных.

Первый проект ориентирован на работу с геомагнитными данными, зарегистрированными магнитными обсерваториями и вариационными станциями. В качестве источника данных выступает ресурс SuperMAG (<https://supermag.jhuapl.edu/>) [15, 17]. Данные уже были использованы ранее в одном из проектов авторов статьи, который обеспечивает визуализацию геомагнитных возмущений по данным наземных магнитных станций [18].

Исходные данные представлены в CSV-подобном формате и представляют собой значения компонента вектора геомагнитного поля в пространственных точках земной поверхности, соответствующих магнитным обсерваториям и вариационным станциям. Источники данных характеризуются выраженной пространственной анизотропией: в отдельных пространственных регионах наблюдается их плотное скопление, в то время как присутствуют области, где размещение соответствующих точек представлено единичными значениями. Таким образом, очевидно, что имеет место нерегулярная сетка измерений.

Посредством разработанного программного средства был определен индекс Моришита для анализируемых источников данных. Установлено, что сеть мониторинга может быть классифицирована как произвольная слабо связанная кластеризованная сеть [2].

На следующем шаге серверным сценарием на основе предложенного авторами алгоритма был определен размер ячейки, в соответствии с которой должна быть выполнена пространственная кластеризация для заданного по умолчанию масштаба (1:20 000 000). Соответствующий результат получен на основе расчета расстояния между пространственными точками, что в общем виде может быть представлено в виде гистограммы. В соответствии с вычисленным значением и на основании заданного масштаба заданные пространственные точки были объединены в группы-кластеры. Непосредственно результат кластеризации представлен в виде GeoJSON-слоя, передаваемого в HTTP-сообщении отклика от сервера клиенту.

В клиентских сценариях приложения предусмотрена функция отслеживания пользовательского поведения при изменении масштабов пространственного изображения. При наступлении указанного события выполняется асинхронный (по технологии AJAX) запрос к соответствующему серверному сценарию. При этом в теле запроса передается новое значение параметра масштабирования пространственного изображения, для которого необходимо адаптировать кластеризацию источников данных.

Серверный сценарий выполняет соответствующую функцию перерасчета ячейки для кластеризации, что в дальнейшем используется для формирования нового пространственного слоя с кластеризованными пространственными объектами. GeoJSON-слой, в свою очередь, направляется откликом на клиентскую сторону и в процессе рендеринга заменяет собой уже отрисованный в окне браузера пространственный слой на заданной картографической подложке.

В общем виде результат такой кластеризации имеет вид, представленный на рис. 2. Здесь используется заданный по умолчанию масштаб пространственного изображения (см. рис. 2, *a*). Для явного представления динамической кластеризации выполняется изменение масштаба пространственного изображения (см. рис. 2, *b*).

Стек технологий представлен фреймворком Django, дополненным модулями для реализации предложенной интеграции программных архитектур. На серверной стороне используется язык программирования Python, клиентская сторона представлена связкой HTML5 / CSS3 / JavaScript. При этом формирование пространственных изображений реализовано посредством библиотеки ArcGIS API for JavaScript, сценарии на основе которой также выполняются на клиентской стороне.

Другое разработанное авторами приложение предназначено для визуализации пространственных точек, каждой из которых соответствует некоторое прогнозируемое значение наблюдения вероятности полярных сияний. В основе приложения лежит модель Ovation-Prime [19], программно реализованная в виде соответствующих сценариев, которые выполняются на стороне сервера. Таким образом, источником данных выступает модель, или результат работы соответствующего программного серверного сценария, возвращающего результат-отклик на клиентскую сторону по протоколу HTTP(s). Такие данные также уже были использованы авторами ранее для визуализации геофизических параметров в области аврального овала [20].

Для сравнения в качестве вычислительного эксперимента была сформирована регулярная сетка пространственных точек, соответствующих указанным атрибутивным значениям

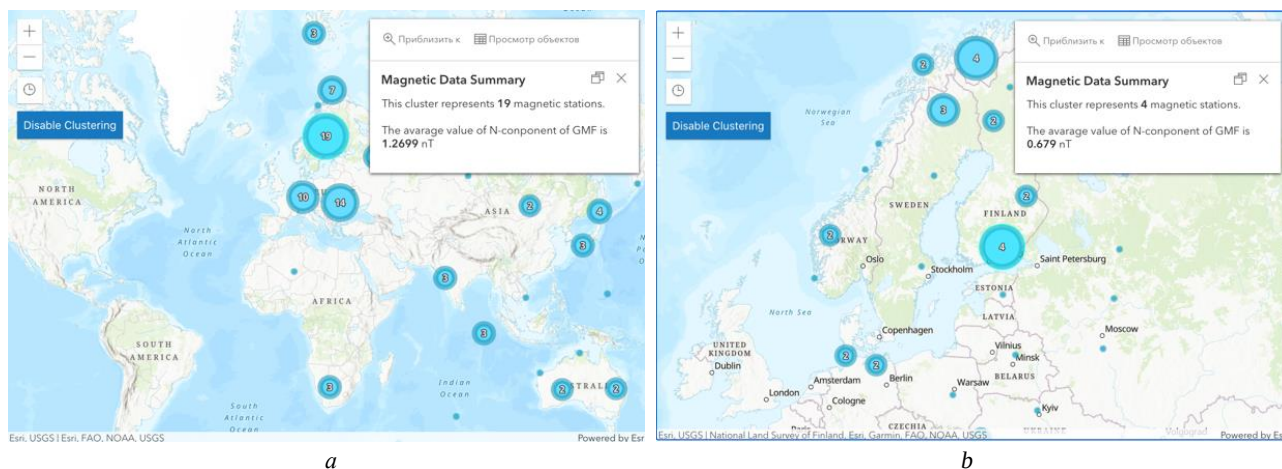


Рис. 2. Пример работы веб-приложения динамической пространственной кластеризации: *a* – пространственное изображение для масштаба по умолчанию, *b* – пространственное изображение при изменении масштаба
 Rice. 2. An example of the work of a dynamic spatial clustering web application: *a*) a spatial image for the default scale, *b*) a spatial image when the scale changes

Как и в предыдущем случае, посредством разработанного программного средства был определен индекс Моришита для анализируемых пространственных точек. Установлено, что в соответствии с построенной диаграммой Моришита сеть мониторинга может быть классифицирована как регулярная равномерная сеть [2].

Соответственно, кластеризация в данном случае не представляется возможной. Серверный сценарий, реализующий предложенный алгоритм, завершает на этом работу. На клиентскую сторону передается представленный в формате GeoJSON набор исходных пространственных точек. При этом на стороне браузера выполняется пространственный рендеринг полученных данных на заданной картографической подложке.

Заключение

Одним из магистральных направлений развития информационных технологий в настоящее время является разработка моделей, методов, алгоритмов и реализующих их программных систем, которые обеспечивают обработку, анализ и визуализацию пространственных данных. Актуальность задачи подтверждается как государственными программами поддержки соответствующих исследований, так и многочисленными запросами со стороны различных отраслей народного хозяйства.

В работе рассмотрено решение одной из задач обработки пространственной информации, связанной с оценкой пространственной анизотропии распределенных источников данных. Предложен подход, обеспечивающий геокластерный анализ источников данных на основании распределения значения индекса Моришита, с одной стороны, а также оценки информационной корреляции внутри кластера – с другой. Применение подхода позволяет выделить из множества источников данных те, которые могут рассматриваться как сходные по значениям измерений в рамках выявленного анализом плотности кластера. Это является развитием метода информационного резервирования, в рамках которого объекты, прошедшие двойную кластеризацию, могут считаться относительно сходными и даже при заданном масштабировании взаимозаменяемыми, что может быть успешно применено для восстановления соответствующих синхронно зарегистрированных временных рядов.

Предложенная в работе архитектура веб-приложений, сочетающая в себе особенности «чистой» архитектуры, с одной стороны, и паттерна MVC – с другой, реализует предложенный подход к пространственной кластеризации. При этом обеспечивается динамическая пространственная кластеризация

ция, при которой размер ячейки меняется в зависимости от действий пользователя при работе с пространственным изображением в окне браузера.

Предложенные решения реализованы в рамках геоинформационных систем, обеспечивающих кластеризацию геомагнитных данных от магнитных станций и геофизических данных в области аврорального овала. Анализ сетей мониторинга посредством разработанного программного обеспечения позволил выделить пространственные кластеры для геомагнитных данных в первом случае и подтвердить регулярность сетки для геофизических данных во втором случае. Кроме того, предложенная архитектура, реализованная в рамках указанных геоинформационных систем, обеспечивает выраженную модульность в структуре веб-приложения, отделение данных от их представления, а также независимость от программного фреймворка при реализации.

Список источников

1. Белогурова Е.Б., Воробьев В.Е., Гвоздев О.Г. и др. Пространственные данные: потребности экономики в условиях цифровизации / Фед. служба гос. регистрации, кадастра и картографии; Нац. исслед. ун-т «Высшая школа экономики»; НИИ АЭРОКОСМОС. М. : НИУ ВШЭ, 2020. 128 с.
2. Демьянов В.В., Савельева Е.А. Геоэкономистика: теория и практика. М. : Наука, 2010. 327 с.
3. Sugawara Sh., Murakami D. Spatially clustered regression // *Spatial Statistics*. 2021. V. 44. Art. 100525.
4. Lin F., Tang Y., Zhu H., Zhu Z. Spatially clustered varying coefficient model // *Journal of Multivariate Analysis*. 2022. V. 192. Art. 105023.
5. Rui Xu, Wunsch D. Survey of clustering algorithms // *IEEE Transactions on Neural Networks*. 2005. V. 16, is. 3. P. 645–678.
6. Дулин С.К., Розенберг И.Н., Уманский В.И. Методы кластеризации в исследовании массивов геоданных // *Системы и средства информации* 2009. Т. 19, № 2. С. 86–113.
7. Turton I., Openshaw S., Brunsdon C. et al. Testing spacetime and more complex hyperspace geographical analysis tools // *Innovations in GIS 7*. London : Taylor & Francis, 2000. P. 87–100.
8. Zhu H., Cui K., Lin J., Yuan Z. et al. Research on WebGL-based 3D Visualization Platform for Earthquake Rapid Reporting // *IOP Conference Series: Materials Science and Engineering*. 2019. V. 631. Art. 052023.
9. Vassilvitskii S. K-means: The advantages of careful seeding // *Proc. of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*, Society for Industrial and Applied Mathematics. 2007. P. 147–164.
10. Olea R. Revisiting the declustering of spatial data with preferential sampling // *Computers & Geosciences*. 2021. V. 157. Art. 104946.
11. Purcaru I. On the information correlation coefficient // *Studii și Cercetări Matematice*. 1976. V. 28, № 2. P. 111–121.
12. Feldman J. Mutual Information and Categorical Perception // *Psychological science*. 2021. № 32. Art. 956797621996663.
13. Воробьев А.В., Воробьева Г.Р. Методы резервирования в задачах восстановления временных рядов геомагнитных данных // *Информационные процессы*. 2018. Т. 19, № 1. С. 1–18.
14. Baroudi D., Nait-Bahloul S. Observer Patterns for Timed Properties // *International Journal of Software Innovation*. 2021. № 9. P. 1–17.
15. Gjerloev J.W. The SuperMAG data processing technique // *J. Geophys. Res.* 2012. № 117. Art. A09213.
16. Martin R.C. Clean Architecture: A Craftsman's Guide to Software Structure and Design. Upper Saddle River, NJ : Prentice-Hall Press, 2017.
17. Gjerloev J.W. A Global Ground-Based Magnetometer Initiative // *Eos Trans. AGU*. 2009. № 90 (27). P. 230–231.
18. Vorobev A.V., Pilipenko V.A., Enikeev T.A., Vorobeva G.R., Khristodulo O.I. System for dynamic visualization of geomagnetic disturbances according to the data of ground magnetic stations // *Scientific Visualization*. 2021. № 13.1. P. 162–176.
19. Newell P.T., Liou K., Zhang Y., Sotirelis T., Paxton L.J., Mitchell E.J. OVATION Prime-2013: Extension of auroral precipitation model to higher disturbance levels // *Space Weather*. 2014. V. 12, № 6. P. 368–379.
20. Vorobev A.V., Vorobeva G.R., Khristodulo O.I. An information system for spatial visualization of prognostic and retrospective data on the probability of observing auroras // *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*. 2021. V. 21, № 2. P. 225–233.

References

1. Belogurova, E.B., Vorobiev, V.E., Gvozdev, O.G. et al. (2020) *Prostranstvennyye dannye: potrebnosti ekonomiki v usloviyakh tsifrovizatsii* [Spatial data: The needs of the economy in the context of digitalization]. Moscow: HSE.
2. Demyanov, V.V. & Savelyeva, E.A. (2010) *Geostatistika: teoriya i praktika* [Geostatistics: Theory and Practice]. Moscow: Nauka.
3. Sugawara, Sh. & Murakami, D. (2021) Spatially clustered regression. *Spatial Statistics*. 44. Art. 100525. DOI: 10.1016/j.spasta.2021.100525
4. Lin, F., Tang, Y., Zhu, H. & Zhu, Z. (2022). Spatially clustered varying coefficient model. *Journal of Multivariate Analysis*. Art. 105023. DOI: 10.1016/j.jmva.2022.105023
5. Rui, Xu & Wunsch, D. (2005) Survey of clustering algorithms. *IEEE Transactions on Neural Networks*. 16(3). pp. 645–678.

6. Dulin, S.K., Rozenberg, I.N. & Umansky, V.I. (2009) Metody klasterizatsii v issledovanii massivov geodannykh [Clustering methods in the study of geodata arrays]. *Sistemy i sredstva informatsii*. 19(2). pp. 86–113.
7. Turton, I., Openshaw, S., Brunsdon C. et al. (2000) Testing spacetime and more complex hyperspace geographical analysis tools. In: *Innovations in GIS 7*. London: Taylor & Francis. pp. 87–100.
8. Zhu, H., Cui, K., Lin, J., Yuan, Z. et al. (2019) Research on WebGL-based 3D Visualization Platform for Earthquake Rapid Reporting. *IOP Conference Series: Materials Science and Engineering*. 631. Art. 052023. DOI: 10.1088/1757-899X/631/5/052023
9. Vassilvitskii, S. (2007) K-means: The advantages of careful seeding. *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms, Society for Industrial and Applied Mathematics*. pp. 147–164.
10. Olea, R. (2021) Revisiting the declustering of spatial data with preferential sampling. *Computers & Geosciences*. 157. Art. 104946. DOI: 10.1016/j.cageo.2021.104946
11. Purcaru, I. (1976) On the information correlation coefficient. *Studii și Cercetări Matematice*. 28.
12. Feldman, J. (2021) Mutual Information and Categorical Perception. *Psychological Science*. 32. Art. 956797621996663. DOI: 10.1177/0956797621996663
13. Vorobev, A.V. & Vorobeva, G.R. (2018) Redundancy Methods in Solutions of Geomagnetic Data Time Series Recovery Problem. *Informatsionnye protsessy – Information processes*. 19(1). pp. 1–18.
14. Baroudi, D. & Nait-Bahloul, S. (2021) Observer Patterns for Timed Properties. *International Journal of Software Innovation*. 9. pp. 1–17. DOI: 10.4018/IJSI.2021040101
15. Gjerloev, J.W. (2012). The SuperMAG data processing technique. *Journal of Geophysical Research*. 117. Art. A09213. DOI: 10.1029/2012JA017683.
16. Martin, R.C. (2017) *Clean Architecture: A Craftsman's Guide to Software Structure and Design*. Upper Saddle River, NJ, USA: Prentice-Hall Press.
17. Gjerloev, J.W. (2009). A Global Ground-Based Magnetometer Initiative. *Eos Trans. AGU*. 90(27). pp. 230–231. DOI: 10.1029/2009EO270002
18. Vorobev, A.V., Pilipenko, V.A., Enikeev, T.A., Vorobeva, G.R. & Khristodulo, O.I. (2021) System for dynamic visualization of geomagnetic disturbances according to the data of ground magnetic stations. *Scientific Visualization*. 13.1. pp. 162–176. DOI: 10.26583/sv.13.1.11
19. Newell, P.T., Liou, K., Zhang, Y., Sotirelis, T., Paxton, L.J. & Mitchell, E.J. (2014) Ovation Prime-2013: Extension of auroral precipitation model to higher disturbance levels. *Space Weather*. 12(6). pp. 368–379. DOI: 10.1002/2014SW001056
20. Vorobev, A.V., Vorobeva, G.R. & Khristodulo, O.I. (2021) An information system for spatial visualization of prognostic and retrospective data on the probability of observing auroras. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*. 21(2). pp. 225–233. DOI: 10.17586/2226-1494-2021-21-2-225-233

Информация об авторах:

Воробьев Андрей Владимирович – доцент, доктор технических наук, профессор кафедры геоинформационных систем Уфимского университета науки и технологий (Уфа, Россия); старший научный сотрудник ФГБУН Геофизический центр РАН (Москва, Россия). E-mail: geomagnet@list.ru

Воробьева Гульнара Равиловна – доцент, доктор технических наук, профессор кафедры вычислительной математики и кибернетики Уфимского университета науки и технологий (Уфа, Россия). E-mail: gulnara.vorobeva@gmail.com

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

Information about the authors:

Vorobev Andrei V. (Associate Professor, Doctor of Technical Sciences, Professor, Ufa University of Science and Technology, Ufa, Russian Federation); Senior Researcher, Geophysical Center of the Russian Academy of Science, Moscow, Russian Federation). E-mail: geomagnet@list.ru

Vorobeva Gulnara R. (Associate Professor, Doctor of Technical Sciences, Professor, Ufa University of Science and Technology, Ufa, Russian Federation). E-mail: gulnara.vorobeva@gmail.com

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

Поступила в редакцию 10.01.2023; принята к публикации 04.09.2023

Received 10.01.2023; accepted for publication 04.09.2023

Научная статья

УДК 519.24

doi: 10.17223/19988605/64/8

О проблемах и ошибках применения критериев согласия

Борис Юрьевич Лемешко

Новосибирский государственный технический университет, Новосибирск, Россия, lemeshko@ami.nstu.ru

Аннотация. Рассматриваются изменения распределений статистик различных критериев согласия. Показывается, начиная с каких малых объемов выборок можно пренебречь отклонением распределений статистик от асимптотических в случае проверки простых гипотез. Приводятся факторы, под влиянием которых изменяются распределения статистик критериев при проверке сложных гипотез. Показывается, что при отсутствии ошибок округления применение критериев в условиях больших выборок не вызывает проблем. В этих же условиях наличие ошибок округления приводит к отклонению распределений статистик от асимптотических. При соизмеримости ошибок округления и ошибок измерений распределения статистик критериев могут не сходиться к асимптотическим и становятся зависящими от объемов выборок и от других факторов.

Ошибки применения критериев согласия в приложениях связаны не только с незнанием свойств критериев в конкретных (нестандартных) условиях, но и с присутствием ошибочных рекомендаций в литературе, в том числе учебного характера. В популярных программных системах статистического анализа, позволяющих применять рассматриваемые критерии, реальные свойства критериев, имеющие место в конкретной ситуации, как правило, не учитываются.

Показано, что корректность статистических выводов при использовании рассматриваемых критериев в нестандартных ситуациях можно гарантировать только за счет применения методов статистического моделирования в рамках соответствующего программного обеспечения.

Ключевые слова: непараметрические критерии согласия; статистическое моделирование; ошибка 1-го рода; ошибка 2-го рода; ошибки округления.

Для цитирования: Лемешко Б.Ю. О проблемах и ошибках применения критериев согласия // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2023. № 64. С. 74–90. doi: 10.17223/19988605/64/8

Original article

doi: 10.17223/19988605/64/8

About the problems and errors in applying the goodness of fit tests

Boris Yu. Lemesko

Novosibirsk State Technical University, Novosibirsk, Russian Federation, lemeshko@ami.nstu.ru

Abstract. Changes in the distributions of statistics for various goodness-of-fit tests are considered. It is shown, starting from what small sample sizes, it is possible to neglect the deviation of statistics distributions from asymptotic ones in the case of testing simple hypotheses. Factors are given, under the influence of which the distributions of test statistics change when testing composite hypotheses. It is shown that in the absence of rounding errors, the application of the tests in conditions of large samples does not cause problems. Under the same conditions, the presence of rounding errors leads to a deviation of the distributions of statistics from asymptotic ones. When rounding errors and measurement errors are commensurable, the distributions of test statistics may not converge to asymptotic ones and become dependent on sample sizes and other factors.

Errors in the application of goodness of fit criteria in applications are associated not only with ignorance of the properties of the tests in specific (non-standard) conditions, but also with the presence of erroneous recommendations in the literature, including educational ones. In popular statistical analysis software systems that allow the application of the tests under consideration, the real properties of the tests that take place in a particular situation, as a rule, are not taken into account.

It is shown that the correctness of statistical conclusions when using the tests under consideration in non-standard situations can only be guaranteed through the use of statistical simulating methods within the framework of the appropriate software.

Keywords: nonparametric goodness of fit tests; statistical simulating; type 1 error; type 2 error; rounding errors.

For citation: Lemesko, B.Yu. (2023) About the problems and errors in applying the goodness of fit tests. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naja tehnika i informatika – Tomsk State University Journal of Control and Computer Science*. 64. pp. 74–90. doi: 10.17223/19988605/64/8

Введение

Курс математической статистики, читаемый в университетах для инженерных специальностей, содержит лишь краткое изложение основ дисциплины и не дает представления о проблемах применения соответствующих методов на практике. Статистические методы и критерии обладают известными (асимптотическими) свойствами при выполнении определенных (стандартных) предположений. В приложениях эти предположения могут не выполняться, а в нестандартных условиях свойства применяемых методов и критериев могут существенно изменяться. В литературе учебного характера внимание на нюансах такого характера практически не акцентируется. В учебниках можно встретить неверные рекомендации по использованию и примеры некорректного применения различных критериев проверки гипотез. Не всегда корректно используются критерии проверки гипотез и в популярных программных системах.

Указанные причины часто приводят к неэффективному применению методов статистического анализа в приложениях и, более того, к неверным статистическим выводам. Такие случаи формируют отрицательное мнение относительно возможностей статистических методов и сокращают потенциальную область их применения, где в результате обедняется используемый в исследованиях математический аппарат. В качестве примера можно назвать сферу метрологии, где в силу объективных и субъективных причин преобладает негативное отношение к полезности методов статистического анализа вообще и к возможностям критериев согласия в частности.

В данном случае речь пойдет об изменении свойств критериев согласия (об изменении распределений статистик) под воздействием различных факторов. В таких ситуациях использование классических результатов, касающихся распределений статистик критериев, как правило, приводит к увеличению вероятностей ошибок или первого, или второго рода. А гарантировать корректность выводов можно только за счет использования методов статистического моделирования для исследования распределений статистик критериев в нестандартных условиях, что можно делать, опираясь на соответствующее программное обеспечение.

1. Непараметрические критерии согласия при проверке простых гипотез

Критерий **Колмогорова** опирается на статистику

$$D_n = \sup_{|x| < \infty} |F_n(x) - F(x, \theta)|,$$

где $F_n(x)$ – эмпирическая функция распределения; $F(x, \theta)$ – теоретическая функция распределения; n – объем выборки. Предельное распределение этой статистики для случая проверки простой гипотезы было получено Колмогоровым в [1]. При $n \rightarrow \infty$ функция распределения статистики $\sqrt{n} \cdot D_n$ сходится равномерно к функции распределения Колмогорова

$$K(s) = \sum_{k=-\infty}^{\infty} (-1)^k e^{-2k^2 s^2}.$$

Распределение статистики $\sqrt{n} \cdot D_n$ достаточно медленно сходится к асимптотическому. Поэтому при проверке гипотез с применением критерия Колмогорова рекомендуется использовать статистику с поправкой Большева [2, 3] в форме [4]

$$S_K = \sqrt{n}D_n + \frac{1}{6\sqrt{n}} = \frac{6nD_n + 1}{6\sqrt{n}}, \quad (1)$$

где $D_n = \max(D_n^+, D_n^-)$, $D_n^+ = \max_{1 \leq i \leq n} \left\{ \frac{i}{n} - F(x_i, \theta) \right\}$, $D_n^- = \max_{1 \leq i \leq n} \left\{ F(x_i, \theta) - \frac{i-1}{n} \right\}$, n – объем выборки; $x_1, x_2, \dots, x_n$ здесь и далее – упорядоченные по возрастанию выборочные значения; $F(x, \theta)$ – функция закона распределения, согласие с которым проверяют. Распределение величины S_K при простой гипотезе в пределе подчиняется закону Колмогорова с функцией распределения $K(S)$.

Критерий Колмогорова является наиболее часто используемым из непараметрических критериев согласия. К сожалению, в подавляющем числе публикаций, в том числе учебного характера, а также в программных системах статистика критерия используется без поправки. В таких случаях отклонение реального распределения статистики $\sqrt{n} \cdot D_n$ от $K(S)$ заметно при величинах n порядка 100. Поэтому вычисление достигнутого уровня значимости p_{value} по распределению $K(S)$ при меньших значениях n приводит к увеличению вероятности ошибки второго рода β .

В ряде источников учебного характера, например [5, 6], в примерах, иллюстрирующих применение критерия Колмогорова, статистика D_n , в отличие от предусмотренного в [1], вычисляется по группированным данным. В обоснование ссылаются на программную систему STATISTICA, где предусмотрена такая возможность. Это неверная рекомендация, которая всегда приводит к меньшим значениям статистики и резко увеличивает вероятность ошибки второго рода, так как в этом случае асимптотическим распределением статистики $\sqrt{n} \cdot D_n$ уже не будет являться распределение $K(S)$.

В критерии типа ω^2 **Крамера–Мизеса–Смирнова** статистика имеет вид:

$$S_\omega = n\omega_n^2 = \frac{1}{12n} + \sum_{i=1}^n \left\{ F(x_i, \theta) - \frac{2i-1}{2n} \right\}^2, \quad (2)$$

и при справедливости простой проверяемой гипотезы в пределе подчиняется закону с функцией распределения $a1(s)$, приведенной в [4].

В критерии типа Ω^2 **Андерсона–Дарлинга** [7, 8] используется статистика

$$S_\Omega = -n - 2 \sum_{i=1}^n \left\{ \frac{2i-1}{2n} \ln F(x_i, \theta) + \left(1 - \frac{2i-1}{2n} \right) \ln(1 - F(x_i, \theta)) \right\}. \quad (3)$$

При проверке простой гипотезы в пределе эта статистика подчиняется закону с функцией распределения $a2(s)$ [4].

В критерии **Купера** [9] в качестве меры расхождения используется величина $V_n = D_n^+ + D_n^-$, где D_n^+ , D_n^- находятся так же, как и в критерии Колмогорова, а в качестве статистики используется

$$s = \sqrt{n} \cdot V_n. \quad (4)$$

Купером [9] в качестве предельного распределения $G(s|H_0)$ для случая проверки простой гипотезы дана следующая функция распределения [10]

$$G(s|H_0) = 1 - \sum_{m=1}^{\infty} 2(4m^2 s^2 - 1) e^{-2m^2 s^2}.$$

Для статистики $\sqrt{n} \cdot V_n$ достаточно долго сохраняется зависимость от объема выборки n . В [11] построена модификация статистики

$$V = V_n \left(\sqrt{n} + 0,155 + \frac{0,24}{\sqrt{n}} \right),$$

для которой зависимостью распределения от объема выборки можно пренебречь при $n \geq 20$.

В [12] предложено применять в критерии Купера статистику вида:

$$V_n^{mod} = \sqrt{n} \cdot V_n + \frac{1}{3\sqrt{n}},$$

зависимостью распределения которой от объема выборки можно пренебречь при $n \geq 30$.

Статистика критерия **Ватсона** [13, 14] используется в следующей форме:

$$U_n^2 = \sum_{i=1}^n \left(F(x_i, \theta) - \frac{i-1/2}{n} \right)^2 - n \left(\frac{1}{n} \sum_{i=1}^n F(x_i, \theta) - \frac{1}{2} \right)^2 + \frac{1}{12n}. \quad (5)$$

Асимптотическим распределением статистики U_n^2 при проверке простой гипотезы является следующее распределение:

$$G(s|H_0) = 1 - 2 \sum_{m=1}^{\infty} (-1)^{m-1} e^{-2m^2\pi^2 s}.$$

Зависимость распределения этой статистики от объема выборки выражена слабо. В [15] приведена модифицированная статистики Ватсона в форме

$$U_n^{2*} = (U_n^2 - 0.1/n + 0.1/n^2)(1 + 0.8/n),$$

отличием распределения которой от асимптотического можно пренебречь при $n \geq 20$.

На основании исследования распределений статистик критериев методами статистического моделирования в табл. 1 указаны объемы выборок, начиная с которых реальные распределения статистик (1)–(5) при справедливости простой проверяемой гипотезы практически (визуально) не отличаются от асимптотических, а также приведены объемы выборок, при которых отклонение этих распределений от асимптотических не отражается на статистических выводах. Последнее объясняется тем, что все эти критерии правосторонние, и при имеющемся отличии реального распределения статистики от асимптотического на правом «хвосте» эти распределения практически совпадают.

Таблица 1

Сходимость распределений статистик к асимптотическим

Тест	Статистика	Нет отклонения от асимптотического распределения	Можно пренебречь отклонением
Колмогорова (K)	(1)	при $n \geq 20$	при $n \geq 15$
Крамера–Мизеса–Смирнова (CMS)	(2)	при $n \geq 18$	при $n \geq 12$
Андерсона–Дарлинга (AD)	(3)	при $n \geq 10$	при $n \geq 6$
Купера (Ku)	(4)	при $n \geq 55$	при $n \geq 10$
Ватсона (W)	(5)	при $n \geq 75$	при $n \geq 10$

В диссертации **Жанга** и его последующих работах [16, 17] были предложены непараметрические критерии согласия, статистики которых имеют вид:

$$Z_K = \max_{1 \leq i \leq n} \left[(i-1/2) \log \left\{ \frac{i-1/2}{nF(x_i, \theta)} \right\} + (n-i+1/2) \log \left\{ \frac{n-i+1/2}{n\{1-F(x_i, \theta)\}} \right\} \right], \quad (6)$$

$$Z_A = - \sum_{i=1}^n \left[\frac{\log \{F(x_i, \theta)\}}{n-i+1/2} + \frac{\log \{1-F(x_i, \theta)\}}{i-1/2} \right], \quad (7)$$

$$Z_C = \sum_{i=1}^n \left[\log \left\{ \frac{[F(x_i, \theta)]^{-1} - 1}{(n-1/2)/(i-3/4) - 1} \right\} \right]^2. \quad (8)$$

Справедливость утверждений автора критериев о их более высокой мощности по сравнению с критериями Колмогорова, Крамера–Мизеса–Смирнова и Андерсона–Дарлинга была подтверждена исследованиями, проведенными в [12, 18]. Однако широкому применению критериев Жанга в приложениях препятствует зависимость распределений статистик от объема выборки n , которая, естественно, сохраняется и в ситуации проверки сложных гипотез.

2. Проблемы применения критериев при проверке сложных гипотез

При проверке сложных гипотез, когда по этой же выборке оцениваются параметры наблюдаемого закона распределения, все непараметрические критерии согласия теряют свойство «свободы от распределения» [19], и асимптотические распределения статистик $G(S|H_0)$ становятся зависящими от факторов, определяющих «сложность» гипотезы. В частности, на распределение статистики $G(S|H_0)$ критерия влияют следующие факторы:

- (a) – вид закона распределения $F(x, \theta)$, соответствующего проверяемой гипотезе H_0 ;
- (b) – тип оцениваемого параметра;
- (c) – число оцениваемых параметров;
- (d) – используемый метод оценивания параметров [20];
- (e) – в некоторых ситуациях распределения статистик зависят от конкретных значений параметров закона $F(x, \theta)$ (например, от значений параметров формы обобщенного нормального закона [21], гамма-распределения, бета-распределений и т.п.).

Хотя проблемы, связанные с применением непараметрических критериев согласия для проверки сложных гипотез, известны давно [19], с ними крайне редко знакомят в курсах лекций или освещают в учебной литературе. В то же время отличия в асимптотических распределениях тех же самых статистик при проверке простых и (различных) сложных гипотез настолько существенны, что пренебрегать этим абсолютно недопустимо. Большинство самых грубых и наиболее частых ошибок применения непараметрических критериев согласия, приводящих к существенному увеличению вероятностей ошибок 2-го рода, связано именно с незнанием этих обстоятельств.

Разрешение проблем, связанных с проверкой сложных гипотез аналитическими методами, связано с серьезными трудностями, а удачные попытки ряда авторов приводили и приводят к положительным результатам лишь частного характера. В то же время успешное использование в этих целях методов статистического моделирования, впервые продемонстрированное в [22, 23], позволяет получать полезные результаты с меньшими затратами. На базе результатов статистического моделирования распределений статистик критериев, имеющих место при проверке различных сложных гипотез, и построенных для них моделей были разработаны рекомендации Госстандарта РФ [24]. Дальнейшее развитие средств вычислительной техники и программного обеспечения исследований позволили уточнить модели и таблицы критических значений распределений статистик непараметрических критериев согласия [25–27], представленные ранее в [24], и подготовить руководство [28].

Возможность построения моделей для асимптотических распределений $G(S|H_0)$ статистик критериев ограничено множеством тех параметрических законов $F(x, \theta)$, в случае которых проявляется влияние только комбинаций факторов (a)–(d). При наличии влияния фактора (e) возможность исследования неизвестного распределения статистики $G(S|H_0)$ или оценки по этому распределению достигнутого уровня значимости p_v появляется только после нахождения оценок тех параметров, от значений которых зависит $G(S|H_0)$. В такой ситуации применение критерия и реализация процедуры интерактивного исследования $G(S|H_0)$ возможны только в рамках соответствующего программного обеспечения. Таким примером является программная система [29]. При зависимости распределений статистик $G(S|H_0)$ от объемов выборок, а это мы имеем в случае применения критериев Жанга со статистиками (6)–(8), к интерактивному моделированию распределений статистик критериев приходится прибегать при любой комбинации факторов (a)–(e), определяющих сложность гипотезы.

3. Применение критериев при малых объемах выборок

Насколько быстро распределения статистик непараметрических критериев согласия в случае проверки сложных гипотез сходятся к асимптотическим (при их существовании), можно судить по

результатам, приведенным в табл. 1. При малых n распределения статистик также несколько отличаются от асимптотических. Значения p_v , вычисляемые по асимптотическим распределениям $G(S|H_0)$, как правило, несколько превышают реальные значения. Максимальная величина этой погрешности при задаваемом уровне значимости $\alpha = 0,1$ и $n \geq 30$ не превышает 0,005. При необходимости всегда имеется возможность, как и в [29], уточнить оценку p_v , воспользовавшись методами статистического моделирования.

Проблему применения критериев согласия при малых выборках видят в другом. Обычно критерии используют, чтобы подобрать некоторую удобную параметрическую модель закона $F(x, \theta)$. При малых n , как правило, не удастся однозначно идентифицировать вид модели $F(x, \theta)$, так как не отклоняются гипотезы о принадлежности выборки целому ряду законов. Причина в низкой мощности критериев согласия при малых n . Более того, при малых n (и малых α) можно столкнуться с ситуацией смещенности критериев относительно некоторых пар конкурирующих гипотез, когда мощность критерия $1 - \beta$, где β – вероятность ошибки 2-го рода (не отклонить проверяемую гипотезу при справедливости конкурирующей), оказывается меньше α [28, 30].

4. Критерии согласия при анализе больших выборок

В настоящее время значительный интерес вызывают вопросы анализа больших данных (Big Data) и извлечения из этих данных скрытых закономерностей. В процессе анализа больших выборок, в том числе для идентификации моделей законов распределения, применяют и критерии согласия. От специалистов, особенно занимающихся анализом данных из области экономики и опирающихся на свой опыт применения, можно услышать, что проверяемая гипотеза H_0 о принадлежности выборки любому закону $F(x, \theta)$ с ростом n всегда отклоняется: критерии не работают (!). Так ли это? Рассмотрим несколько примеров.

Пример 1. Будем моделировать выборки в соответствии со стандартным нормальным законом, проверяя простую гипотезу H_0 о принадлежности этих выборок моделируемому закону. Результаты проверки (значения статистик S и достигнутые уровни значимости p_v) по пяти критериям с известными асимптотическими распределениями статистик $G(S|H_0)$ приведены в табл. 2.

Как и следовало ожидать, с ростом объемов выборок n применяемые критерии не дают никаких оснований для отклонения справедливой гипотезы H_0 .

Таблица 2

Результаты проверки простой гипотезы о принадлежности выборок, смоделированных по стандартному нормальному закону, этому же закону

Test	$n = 10^2$		$n = 10^3$		$n = 10^4$		$n = 10^5$		$n = 10^6$		$n = 10^7$	
	S	p_v	S	p_v	S	p_v	S	p_v	S	p_v	S	p_v
K	0,666	0,766	0,634	0,816	0,636	0,814	0,812	0,525	0,660	0,776	0,647	0,796
Ku	1,142	0,622	1,138	0,629	1,104	0,680	1,306	0,385	0,962	0,866	0,814	0,973
CMS	0,063	0,798	0,072	0,736	0,059	0,818	0,145	0,404	0,080	0,689	0,097	0,601
AD	0,384	0,864	0,568	0,679	0,565	0,682	1,209	0,264	0,611	0,638	0,585	0,662
W	0,053	0,668	0,067	0,523	0,059	0,610	0,068	0,512	0,030	0,926	0,034	0,887

Пример 2. Логистический закон распределения с плотностью

$$f(x) = \frac{\pi}{\theta_1 \sqrt{3}} \exp \left\{ -\frac{\pi(x - \theta_0)}{\theta_1 \sqrt{3}} \right\} / \left[1 + \exp \left\{ -\frac{\pi(x - \theta_0)}{\theta_1 \sqrt{3}} \right\} \right]^2$$

достаточно близок к нормальному, и с использованием критериев согласия при малых объемах выборок его достаточно сложно отличить от нормального.

Будем моделировать выборки в соответствии с логистическим законом (для определенности с параметрами $\theta_0 = 0$, $\theta_1 = 1$), проверяя простую или сложную гипотезу H_0 о принадлежности этих

выборки нормальному закону. В случае простой гипотезы H_0 рассматривается стандартный нормальный закон. Результаты проверки приведены в табл. 3.

Таблица 3

Результаты проверки простой и сложной гипотез о нормальности
в случае принадлежности выборок логистическому закону

	При проверке простой гипотезы						При проверке сложной гипотезы					
	$n = 10^2$		$n = 10^3$		$n = 10^4$		$n = 10^2$		$n = 10^3$		$n = 10^4$	
	S	p_v	S	p_v	S	p_v	S	p_v	S	p_v	S	p_v
K	0,701	0,710	1,293	0,071	2,293	0	0,701	0,295	1,155	0,004	2,605	0
Ku	1,164	0,589	2,340	0,001	4,499	0	1,293	0,410	1,845	0,027	5,205	0
CMS	0,090	0,638	0,436	0,058	1,915	0	0,096	0,124	0,253	0,001	2,706	0
AD	0,634	0,616	2,744	0,037	13,08	0	0,675	0,078	1,573	0,000	17,29	0
W	0,082	0,393	0,433	0,000	1,914	0	0,096	0,305	0,235	0,020	2,698	0

В данном примере гипотеза H_0 неверна. Мы видим, что в случае простой гипотезы H_0 она отклоняется по всем критериям при $\alpha > 0,071$ и $n \geq 10^3$, а в случае сложной также по всем критериям и том же n отклоняется при $\alpha > 0,027$.

Пример 3. Пусть наблюдаемый закон представляет собой смесь двух нормальных законов: $F(x, \theta) = 0,9 \times N(0, 1) + 0,1 \times N(0, 1, 1)$, где 90% составляют наблюдения, принадлежащие стандартному нормальному закону. Проверяется простая гипотеза H_0 о принадлежности моделируемых выборок различного объема n стандартному нормальному закону. Результаты проверки представлены в табл. 4.

Таблица 4

Результаты проверки простой гипотезы о принадлежности выборок
со смесями стандартному нормальному закону

Test	$n = 10^2$		$n = 10^3$		$n = 10^4$		$n = 10^5$		$n = 10^6$	
	S	p_v	S	p_v	S	p_v	S	p_v	S	p_v
K	0,706	0,701	0,550	0,923	1,015	0,254	1,698	0,006	4,188	0
Ku	1,146	0,616	1,073	0,724	1,387	0,285	1,976	0,012	4,207	0
CMS	0,0707	0,747	0,060	0,815	0,178	0,313	0,615	0,021	7,999	0
AD	0,418	0,830	0,543	0,703	1,150	0,287	2,761	0,036	40,61	0
W	0,055	0,645	0,060	0,599	0,084	0,381	0,238	0,018	1,264	0

В данном случае верна конкурирующая гипотеза H_1 , соответствующая рассматриваемой смеси и очень близкая к H_0 . Мы можем видеть, что при $n = 10^4$ гипотеза H_0 еще не отклоняется ни по одному из критериев, а при $n \geq 10^5$ и задании вероятности ошибки первого рода $\alpha = 0,04$ – отклоняется по всем.

Оценки мощности $1 - \beta$ критериев относительно H_1 приведены в табл. 5. Эти оценки подтверждают закономерность получения приводимых в табл. 4 результатов проверки.

Таблица 5

Оценки мощности критериев относительно H_1 при $\alpha = 0,05$

Test	$n = 10^2$	$n = 10^3$	$n = 10^4$	$n = 10^5$	$n = 2 \times 10^5$	$n = 10^6$
K	0,0549	0,0558	0,1376	0,7807	0,9743	1
Ku	0,0511	0,0514	0,0838	0,5319	0,8773	1
CMS	0,0532	0,0616	0,1527	0,8503	0,9886	1
AD	0,0528	0,0625	0,1587	0,8670	0,9915	1
W	0,0523	0,0540	0,0872	0,4864	0,8813	1

Подчеркнем, что в данном примере закон распределения в виде смеси, соответствующий конкурирующей гипотезе H_1 , более близок к стандартному нормальному закону, рассматриваемому в качестве H_0 при простой гипотезе, чем логистический закон в той же роли в предшествующем примере. Именно поэтому в данном случае H_0 отклоняется при существенно больших объемах выборок.

Обратим внимание, что проверка сложных гипотез о принадлежности тех же выборок со смесями нормальному закону (с вычислением по выборкам оценок максимального правдоподобия (ОМП) параметров) свидетельствует о том, что получаемые смеси очень хорошо описываются нормальными законами (табл. 6).

Таблица 6

Результаты проверки сложных гипотез о принадлежности выборок со смесями нормальным законам

Test	$n = 10^2$		$n = 10^3$		$n = 10^4$		$n = 10^5$		$n = 10^6$		$n = 10^7$	
	S	p_v	S	p_v	S	p_v	S	p_v	S	p_v	S	p_v
K	0,572	0,633	0,591	0,578	0,503	0,820	0,997	0,020	0,693	0,311	0,407	0,974
Ku	1,112	0,670	0,996	0,826	0,839	0,962	1,444	0,231	1,067	0,732	0,775	0,959
CMS	0,052	0,482	0,063	0,352	0,027	0,887	0,119	0,062	0,048	0,547	0,023	0,942
AD	0,343	0,493	0,432	0,308	0,207	0,870	0,675	0,078	0,283	0,656	0,146	0,980
W	0,051	0,695	0,052	0,683	0,024	0,970	0,105	0,256	0,043	0,790	0,023	0,931

В частности, для смоделированной выборки объемом $n = 10^7$ ОМП параметров нормального закона принимают значения $\hat{\mu} = 0,010273$, $\hat{\sigma} = 1,00063$, а достигнутые уровни значимости p_v , приведенные в последней колонке, свидетельствуют об очень хорошем согласии. В то же время результаты проверки простой гипотезы о согласии этой же выборки со стандартным нормальным законом дают по всем критериям нулевые значения p_v .

Заметим, что множество параметрических моделей законов распределения, где-либо и когда-либо использованных для описания реально наблюдаемых случайных величин, представляет собой крайне ограниченный перечень, а множество законов, описывающих эти величины, бесконечно. Вид конкретного закона для ряда (ошибок) измерений наблюдаемой величины оказывается известен крайне редко. В большинстве случаев мы имеем дело с приближенными моделями, в той или иной мере отличающимися от истинного закона. В такой ситуации с ростом n проверяемая гипотеза о согласии даже с очень хорошей моделью, естественно, будет отклоняться. В то же время справедливая гипотеза H_0 отклоняться не должна! Однако именно с этим приходится сталкиваться, если опираться на классические результаты, характеризующие свойства критериев согласия при проверке простых и сложных гипотез. В чем же причина частой некорректности выводов при использовании критериев согласия для анализа больших выборок в приложениях?

Приведенные выше примеры, в которых рассматриваются выборки моделируемых непрерывных случайных величин, показывают отсутствие проблем принципиального характера. В отличие от того, что мы наблюдаем на практике в больших данных, в выборках, смоделированных в соответствии с приводимыми примерами, отсутствовали повторяющиеся значения. В Big Data мы имеем дело с выборками очень большого объема, но данные в таких выборках представлены с некоторой ошибкой округления Δ . Поэтому в них всегда присутствует значительная доля повторяющихся значений. Наличие ошибок округления отражается на распределениях статистик непараметрических критериев согласия.

На рис. 1 (без потери общности, при $\Delta = 10^{-3}\sigma$) показано, как с ростом n меняются распределения $G(S_n|H_0)$ статистики (3) критерия Андерсона–Дарлинга при проверке простой гипотезы о принадлежности выборок нормальному закону. При таком Δ в выборках из нормального закона может присутствовать порядка 9 830 уникальных значений. Как видим, при $n = 10^3$ распределение $G(S_n|H_0)$ еще не отличается от асимптотического (предельного) $a1(S)$, а при $n = 10^4$ отклонение начинает проявляться. Если при проверке опираться на распределение $a1(S)$, то очевидно, что при больших n справедливая гипотеза H_0 будет отвергаться.

Таким образом, при проверке простых гипотез наличие ошибок округления приводит к зависимости распределений статистик критериев от объемов выборок n и от величины Δ . При проверке сложных гипотез зависимость от n и Δ накладывается на изменения распределений статистик критериев, связанные с влиянием факторов (а)–(е), определяющих различие в сложных гипотезах.

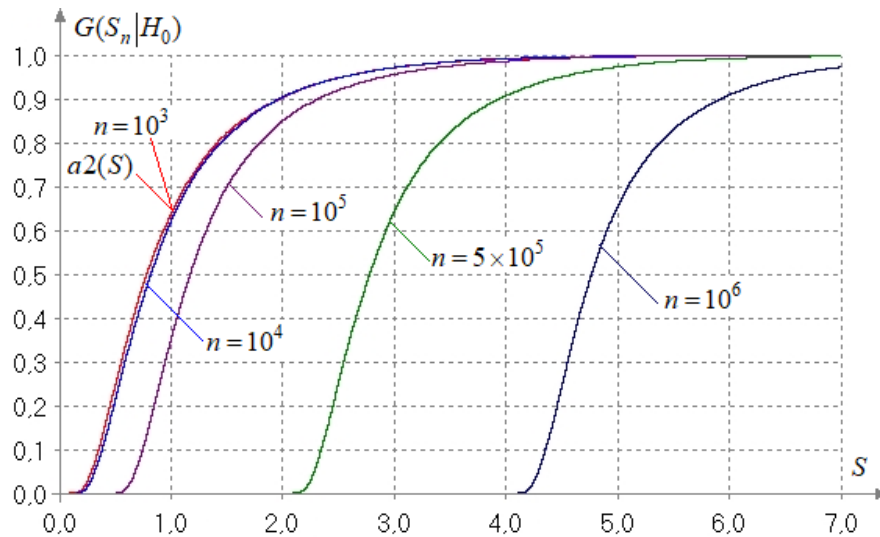


Рис. 1. Зависимость распределения статистики (3) критерия Андерсона–Дарлинга от объема выборки n при справедливости простой гипотезы H_0 о принадлежности выборки нормальному закону при $\Delta = 10^{-3}\sigma$
 Fig. 1. Dependence of the distribution of statistics (3) of the Anderson–Darling test on the sample size n under the validity of the simple hypothesis H_0 that the sample belongs to the normal law for $\Delta = 10^{-3}\sigma$

Чтобы можно было использовать классические результаты (при простых или сложных гипотезах), применяя соответствующие критерии, в [31] рекомендуется извлекать для анализа из Big Data выборки объемом не более $n_{\max}$, при котором распределение статистики $G(S_{n_{\max}}|H_0, \Delta)$ еще практически не отличается от асимптотического $G(S|H_0)$. В противном случае для обеспечения корректности выводов по результатам применения критерия потребуется нахождение реального распределения статистики $G(S_n|H_0, \Delta)$. Величина $n_{\max}$ зависит от Δ . Для ситуации, приведенной на рис. 1, в качестве $n_{\max}$ можно взять $n = 10^4$. Оценка $n_{\max}$ трудности не представляет, так как очевидно, что она не должна превышать количества уникальных значений случайной величины, которое может присутствовать в выборках из данного закона при данной величине Δ , уменьшенного в 3–4 раза.

При выполнении условия $\Delta \ll \sigma$ возможные изменения распределений статистик критериев следует предусматривать только при значительных объемах выборок, когда в выборках существенно возрастает доля повторяющихся значений.

5. Распределения статистик критериев при относительно больших ошибках округления

В приложениях различного характера достаточно часто приходится сталкиваться с выборками, полученными в том числе в результате высокоточных измерений, элементы которых отличаются друг от друга буквально единицами последнего десятичного разряда. В таких измерениях «на пределе точности» ошибка округления Δ оказывается достаточно близкой к σ ошибки измерения наблюдаемой случайной величины, а в выборках такого рода присутствует значительное количество повторяющихся значений. Как показывают исследования [32–34], распределения статистик критериев, применяемых к подобным выборкам, даже не пытаются сходиться к (существующим) асимптотическим $G(S|H_0)$, а с ростом n только удаляются от них.

В качестве примера на рис. 2 при $\Delta = 0,1\sigma$ показано изменение с ростом n распределения $G(S_n|H_0)$ статистики того же критерия Андерсона–Дарлинга при проверке сложной гипотезы о принадлежности выборок нормальному закону (при использовании оценок максимального правдоподобия (ОМП) для двух его параметров). В этом случае в выборках может присутствовать порядка 86 уникальных значений, а отклонением $G(S_n|H_0)$ от асимптотического распределения можно пре-

небрежь лишь при $n \approx 20 \div 30$. При $\Delta > 0,1\sigma$ изменения распределения $G(S_n|H_0)$ при тех же n будут еще более значительными.

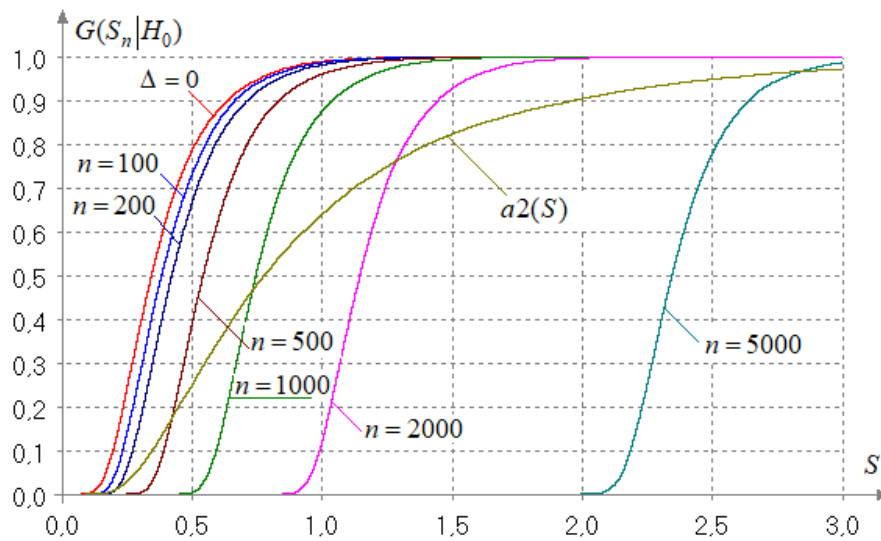


Рис. 2. Зависимость распределения статистики (3) критерия Андерсона–Дарлинга от объема выборки n при справедливости сложной гипотезы H_0 о принадлежности выборки нормальному закону при $\Delta = 0,1\sigma$

Fig. 2. Dependence of the distribution of statistics (3) of the Anderson–Darling test on the sample size n under the validity of the composite hypothesis H_0 that the sample belongs to the normal law for $\Delta = 0,1\sigma$

Аналогичные картины наблюдаются для распределений статистик всех непараметрических критериев согласия.

Таким образом, наличие влияния ошибок округления для критериев со статистиками (1)–(5) в ситуации проверки простых гипотез приводит к появлению зависимости распределений статистик критериев от n , а при заданной величине Δ – от значения параметра масштаба σ предполагаемого закона [34]. В ситуации проверки сложных гипотез эти зависимости накладываются на зависимости, определяемые факторами (а)–(е). Ясно, что подобным же изменениям подвергаются распределения статистик (6)–(8) критериев Жанга, которые и так зависят от n .

Корректность выводов по всем непараметрическим критериям согласия при Δ , близким к σ , может быть обеспечена лишь за счет интерактивного моделирования распределений $G(S_n|H_0, \Delta)$ или оценивания достигнутого уровня значимости p_v в этих условиях. Такая возможность предусмотрена в программной системе [29].

6. О применении критерия согласия χ^2

Статистика критерия согласия χ^2 Пирсона, предложенного в [35], имеет вид:

$$\chi_n^2 = n \sum_{i=1}^k \frac{(n_i / n - P_i(\theta))^2}{P_i(\theta)}, \quad (9)$$

где $P_i(\theta)$ – вероятность попадания в i -й интервал в соответствии с теоретическим законом, относительно которого проверяется гипотеза. При справедливости простой проверяемой гипотезы асимптотическим распределением статистики является χ_{k-1}^2 -распределение. При справедливости сложной гипотезы и оценивании параметров в результате минимизации статистики (9) последняя подчиняется χ_{k-r-1}^2 -распределению [36–39], где r – число оцененных по выборке параметров.

Заметим, что отказываться от χ_{k-r-1}^2 -распределения как асимптотического распределения статистики нет оснований и в случае применения ряда оценок по группированным данным, в частности

ОМП и других, обладающих такими же асимптотическими свойствами? как оценки, минимизирующие статистику (9). По крайней мере результаты статистического моделирования это подтверждают.

Дискретные распределения $G(X_n^2|H_0)$ статистики сходятся к непрерывным χ^2 -распределениям при любом способе группирования выборки. При этом целесообразно следовать рекомендации [40], разбивая выборку на интервалы таким образом, чтобы ожидаемое число наблюдений $nP_i(\theta)$, попадающих в каждый интервал, было по возможности не меньше $5 \div 10$. Иногда предлагают объединять соседние интервалы, если в каких-то из них число попавших наблюдений оказывается нулевым. Это неверная рекомендация, так как такие действия приводят к увеличению вероятностей ошибок второго рода β .

Распределение статистики $G(X_n^2|H_1)$ при справедливости конкурирующей гипотезы H_1 зависит от способа разбиения на интервалы области определения случайной величины, соответствующей H_0 . То есть мощность критерия относительно H_1 зависит от способа группирования. При этом относительно конкурирующих гипотез H_i , близких к H_0 , критерий обладает максимальной мощностью в случае использования асимптотически оптимального группирования [41, 42], при котором минимизируются потери в информации Фишера, связанные с группированием. Мощность критерия Пирсона зависит также от выбора числа интервалов [43]. Очевидно, что относительно заданной конкурирующей гипотезы H_1 за счет выбора способа группирования и числа интервалов (при заданном n) всегда можно построить критерий χ^2 Пирсона максимальной мощности.

Заметим, что при проверке простых гипотез за счет выбора способа группирования и числа интервалов критерий χ^2 Пирсона может иметь значительное преимущество в мощности перед упомянутыми выше непараметрическими критериями согласия, а при проверке сложных гипотез такое преимущество уже отсутствует [44].

Если при проверке сложной гипотезы в критерии χ^2 Пирсона использовать различные оценки по негруппированным данным, то распределение статистики (9) будет существенно отличаться от χ_{k-r-1}^2 -распределения. В таких ситуациях распределения $G(X_n^2|H_0)$ зависят не только от метода оценивания, но и от способа группирования [45], а реально лежат между χ_{k-r-1}^2 - и χ_{k-1}^2 -распределениями. Надо полагать, что на распределениях статистики (9) в ситуации оценивания параметров по негруппированным данным сказывается влияние тех же факторов (а)–(е), что имеют место в случае непараметрических критериев согласия.

В [46] было показано, что если оценки параметров закона $F(x, \theta)$ находятся по негруппированным данным с использованием метода максимального правдоподобия, то статистика (9) распределена как сумма независимых слагаемых $\chi_{k-m-1}^2 + \sum_{j=1}^m \lambda_j \xi_j^2$, где $\xi_1, \dots, \xi_m$ – стандартные нормальные случайные величины, независимые одна от другой и от случайной величины χ_{k-m-1}^2 . В этой сумме числа $\lambda_1, \dots, \lambda_m$ принимают значения между 0 и 1 и представляют собой корни уравнения $|(1-\lambda)\mathbf{J}(\theta) - \mathbf{J}_\Gamma(\theta)| = 0$, где $\mathbf{J}(\theta)$ и $\mathbf{J}_\Gamma(\theta)$ – информационные матрицы Фишера по негруппированным и группированным данным соответственно. Однако этот результат не позволяет судить о конкретном виде распределения статистики (9) и не может быть использован в процессе формирования корректного вывода о проверяемой гипотезе.

Тем не менее, по-видимому, именно вследствие работы [46] появился ряд модификаций критерия χ^2 Пирсона, отличающихся от него лишь при проверке сложных гипотез и предусматривающих использование оценок по негруппированным данным. В первую очередь следует отметить критерий Никулина–Рао–Робсона [47–49], в котором предусмотрено использование ОМП. Позже появились и

другие критерии, большинство которых перечислено в [50]. Замечательно то, что вместе с модификациями были представлены и распределения их статистик, что облегчает применение соответствующих критериев. Кроме того, критерий Никулина–Рао–Робсона, как правило, демонстрирует преимущество в мощности по сравнению с критерием Пирсона [51]. К сожалению, совокупность этих модификаций не нашла широкого применения на практике. В первую очередь этот факт объясняется «сложностью подготовки» критериев, так как конкретные выражения статистик меняются в зависимости от вида параметрического закона, относительно которого проверяется гипотеза H_0 , и от совокупности оцениваемых параметров. Но при наличии программного обеспечения, предусматривающего проверку относительно различных моделей законов с использованием этих критериев (как в [29]), проблемы применения снимаются.

Подчеркнем, что фактом отличия распределения статистики (9) от χ^2_{k-r-1} -распределения очень часто пренебрегают, в том числе в популярных программных системах. Примером может служить широко применяемая система STATISTICA, где для вычисления p_v в подобной ситуации используют χ^2_{k-r-1} -распределения. Такие действия несколько увеличивают вероятность ошибок первого рода α .

В то же время корректность выводов по критерию χ^2 Пирсона при проверке сложных гипотез с оцениванием параметров по негруппированным данным всегда может быть обеспечена за счет моделирования в этих условиях распределения статистики $G(X_n^2|H_0)$ или оценки достигнутого уровня значимости p_v , что достаточно просто реализовать [29].

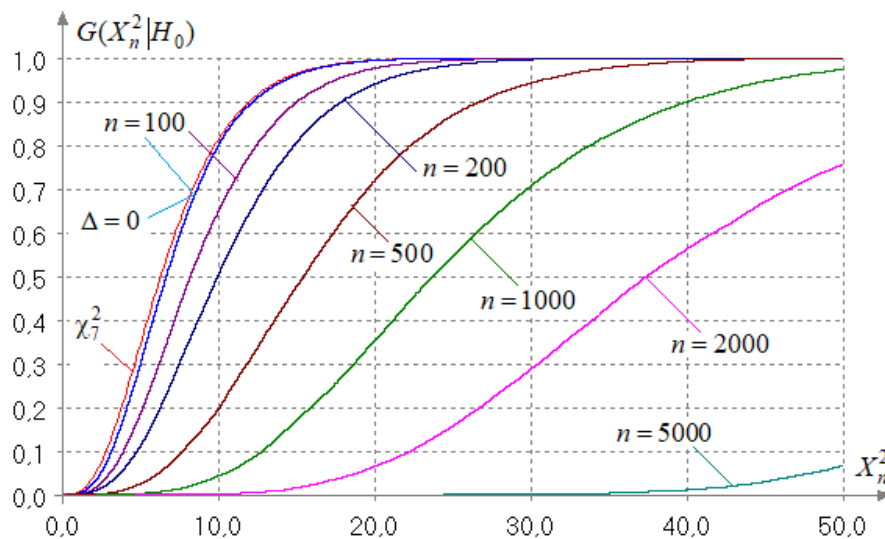


Рис. 3. Зависимость распределения статистики (9) критерия χ^2 Пирсона от объема выборки n при справедливости сложной гипотезы H_0 о принадлежности выборки нормальному закону при $\Delta = 0,1\sigma$

Fig. 3. Dependence of the distribution of statistics (9) of the χ^2 Pearson test on the sample size n under the validity of the composite hypothesis H_0 that the sample belongs to the normal law for $\Delta = 0,1\sigma$

Наличие ошибок округления Δ сказывается на распределениях $G(X_n^2|H_0)$ статистики критерия χ^2 Пирсона точно так же, как и на распределениях непараметрических критериев согласия. На рис. 3 показано, как изменяется распределение статистики $G(X_n^2|H_0)$ с ростом n при проверке сложной гипотезы о принадлежности выборок нормальному закону с вычислением ОМП двух его параметров по негруппированной выборке (число интервалов $k=10$, равновероятные интервалы, $\Delta = 0,1\sigma$).

Заключение

Таким образом, при различии всех обстоятельств, причина некорректности вывода по результатам применения любого критерия согласия оказывается всегда одной и той же. И заключается она в использовании в качестве $G(S|H_0)$ не того распределения.

Очевидно, что область использования классических результатов, касающихся проверки простых гипотез, а также частных результатов относительно проверки сложных гипотез, учитывая возможное влияние ошибок округления, представляется достаточно узкой. Следовательно, в большинстве приложений мы сталкиваемся с применением критериев в нестандартных условиях, когда распределения статистик $G(S_n|H_0, \Delta)$ в той или иной мере отличаются от асимптотических (или от известных). В современных условиях это не представляет серьезной проблемы, так как, используя методы статистического моделирования, мы в интерактивном режиме проверки можем исследовать $G(S_n|H_0, \Delta)$ и оценить достигнутый уровень значимости p_v [29].

Список источников

1. Kolmogoroff A.N. Sulla determinazione empirica di una legge di distribuzione // G. Ist. Ital. Attuar. 1933. V. 4, № 1. P. 83–91.
2. Большев Л.Н. Асимптотические пирсоновские преобразования // Теория вероятностей и ее применение. 1963. Т. 8, № 2. С. 129–155.
3. Большев Л.Н. Теория вероятностей и математическая статистика : избр. тр. / под ред. Ю.В. Прохорова. М. : Наука, 1987. 286 с.
4. Большев Л.Н., Смирнов Н.В. Таблицы математической статистики. М. : Наука, 1983. 416 с.
5. Кремер Н.Ш. Теория вероятностей и математическая статистика : учебник. 3-е изд., перераб. и доп. М. : ЮНИТИ-ДАНА, 2010. 551 с.
6. Шпаков П.С., Юнаков П.Л. Математическая обработка результатов измерений : учеб. пособие. Красноярск : Сиб. федер. ун-т, 2014. 410 с.
7. Anderson T.W., Darling D.A. Asymptotic theory of certain “Goodness of fit” criteria based on stochastic processes // The Annals of Mathematical Statistics. 1952. V. 23, № 2. P. 193–212. URL: <https://www.jstor.org/stable/2236446>
8. Anderson T.W., Darling D.A. A test of goodness of fit // Journal of the American Statistical Association. 1954. V. 29. P. 765–769.
9. Kuiper N.H. Tests concerning random points on a circle // Proc. Koninkl. Nederl. Akad. Van Wetenschappen. Series A. 1960. V. 63. P. 38–47.
10. Stephens M.A. The goodness-of-fit statistic V_N : distribution and significance points // Biometrika. 1965. V. 52, № 3/4. P. 309–321.
11. Stephens M.A. EDF statistics for goodness of fit and some comparisons // Journal of the American Statistical Association. 1974. V. 69, № 347. P. 730–737.
12. Лемешко Б.Ю., Горбунова А.А. О применении и мощности непараметрических критериев согласия Купера, Ватсона и Жанга // Измерительная техника. 2013. № 5. С. 3–9.
13. Watson G.S. Goodness-of-fit tests on a circle. I // Biometrika. 1961. V. 48, № 1/2. P. 109–114.
14. Watson G.S. Goodness-of-fit tests on a circle. II // Biometrika. 1962. V. 49, № 1/2. P. 57–63.
15. Biometrika tables for Statistics / E.S. Pearson, H.O. Hartley (eds.). Cambridge : Cambridge University Press, 1972. V. 2. 385 p.
16. Zhang J. Powerful goodness-of-fit tests based on the likelihood ratio // Journal of the Royal Statistical Society. Series B. 2002. V. 64, № 2. P. 281–294.
17. Zhang J., Wu Yu. Likelihood-ratio tests for normality // Computational Statistics & Data Analysis. 2005. V. 49, № 3. P. 709–721.
18. Лемешко Б.Ю., Горбунова А.А., Лемешко С.Б., Рогожников А.П. О решении проблем применения некоторых непараметрических критериев согласия // Автометрия. 2014. № 1. С. 26–43.
19. Кас М., Kiefer J., Wolfowitz J. On tests of normality and other tests of goodness of fit based on distance methods // The Annals of Mathematical Statistics. 1955. V. 26. P. 189–211. URL: <https://www.jstor.org/stable/2236876>
20. Лемешко Б.Ю., Постовалов С.Н. О зависимости распределений статистик непараметрических критериев и их мощности от метода оценивания параметров // Заводская лаборатория. Диагностика материалов. 2001. Т. 67, № 7. С. 62–71.
21. Лемешко Б.Ю., Маклаков А.А. Непараметрические критерии при проверке сложных гипотез о согласии с распределениями экспоненциального семейства // Автометрия. 2004. № 3. С. 3–20.
22. Lilliefors H.W. On the Kolmogorov–Smirnov test for normality with mean and variance unknown // Journal of the American Statistical Association. 1967. V. 62. P. 399–402.
23. Lilliefors H.W. On the Kolmogorov–Smirnov test for the exponential distribution with mean unknown // Journal of the American Statistical Association. 1969. V. 64. P. 387–389.

24. Р 50.1.037–2002. Рекомендации по стандартизации. Прикладная статистика. Правила проверки согласия опытного распределения с теоретическим. М. : Изд-во стандартов, 2002. Ч. II: Непараметрические критерии. 64 с.
25. Лемешко Б.Ю., Лемешко С.Б. Модели распределений статистик непараметрических критериев согласия при проверке сложных гипотез с использованием оценок максимального правдоподобия. Ч. I // Измерительная техника. 2009. № 6. С. 3–11.
26. Лемешко Б.Ю., Лемешко С.Б. Модели распределений статистик непараметрических критериев согласия при проверке сложных гипотез с использованием оценок максимального правдоподобия. Ч. II // Измерительная техника. 2009. № 8. С. 17–26.
27. Lemeszko B.Yu., Lemeszko S.B., Postovalov S. N. Statistic Distribution Models for Some Nonparametric Goodness-of-Fit Tests in Testing Composite Hypotheses // Communications in Statistics – Theory and Methods. 2010. V. 39, № 3. P. 460–471.
28. Лемешко Б.Ю. Непараметрические критерии согласия : руководство по применению. М. : ИНФРА-М, 2014. 163 с.
29. Статистический анализ интервальных наблюдений одномерных непрерывных случайных величин «Интервальная статистика 5.4» : св-во о гос. регистрации программы для ЭВМ № 2018666213 / Лемешко Б.Ю., Лемешко С.Б., Блинов П.Ю., Веретельникова И.В., Новикова А.Ю. Заявка № 20186663206; заявл. 22.11.2018; запат. 13.12.2018. URL: <https://ami.nstu.ru/~headrd/ISW.htm> (дата обращения: 03.04.2023).
30. Лемешко Б.Ю., Блинов П.Ю., Лемешко С.Б. Смещенность непараметрических критериев согласия относительно некоторых пар конкурирующих гипотез // Измерительная техника. 2016. № 5. С. 16–20.
31. Лемешко Б.Ю., Лемешко С.Б., Семёнова М.А. К вопросу статистического анализа больших данных // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2018. № 44. С. 40–49.
32. Лемешко Б. Ю., Лемешко С.Б. О влиянии ошибок округления на распределение статистик критериев согласия // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2020. № 53. С. 47–60.
33. Лемешко Б.Ю., Лемешко С.Б. Проблемы применения непараметрических критериев согласия в задачах обработки результатов измерений // Системы анализа и обработки данных. 2021. № 2 (82). С. 47–66.
34. Лемешко Б.Ю., Лемешко С.Б. Непараметрические критерии согласия при проверке нормальности в условиях округления измерений // Системы анализа и обработки данных. 2022. № 2 (86). С. 21–38.
35. Pearson K. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably to have arisen from random sampling // The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science. 1900. V. 50. P. 157–175.
36. Fisher R.A. On the interpretation of χ^2 from contingency tables, and the calculation of P // Journal of the Royal Statistical Society. 1922. V. 85, № 1. P. 87–94.
37. Fisher R.A. The conditions under which χ^2 measures the discrepancy between observation and hypothesis // Journal of the Royal Statistical Society. 1924. V. 87, № 3. P. 442–450. URL: <https://www.jstor.org/stable/2341149>
38. Neyman J., Pearson E.S. On the use and interpretation of certain test criteria for purposes of statistical inference: Part I // Biometrika. 1928. V. 20A, № 1/2. P. 175–240.
39. Neyman J., Pearson E.S. On the use and interpretation of certain test criteria for purposes of statistical inference: Part II // Biometrika. 1928. V. 20A, № 3/4. P. 263–294.
40. Mann H.B., Wald A. On the choice of the number of class intervals in the application of the chi square test // The Annals of Mathematical Statistics. 1942. V. 13, № 3. P. 306–317. URL: <https://www.jstor.org/stable/2235942>
41. Денисов В.И., Лемешко Б.Ю. Оптимальное группирование при обработке экспериментальных данных // Измерительные информационные системы. Новосибирск : Новосиб. электротехн. ин-т, 1979. С. 5–14.
42. Лемешко Б.Ю. Асимптотически оптимальное группирование наблюдений в критериях согласия // Заводская лаборатория. 1998. Т. 64, № 1. С. 56–64.
43. Лемешко Б.Ю., Чимитова Е.В. О выборе числа интервалов в критериях согласия типа χ^2 // Заводская лаборатория. Диагностика материалов. 2003. Т. 69, № 1. С. 61–67.
44. Chimitova E.V., Lemeszko B.Y. Chi-Squared Goodness-of-Fit Tests: The Optimal Choice of Grouping Intervals // Recent Advances in Systems, Control and Information Technology. SCIT 2016. Advances in Intelligent Systems and Computing / R. Szweczyk, M. Kaliczyńska (eds). Cham : Springer, 2017. V. 543. P. 760–774.
45. Лемешко Б.Ю., Постовалов С.Н. О зависимости предельных распределений статистик χ^2 Пирсона и отношения правдоподобия от способа группирования данных // Заводская лаборатория. 1998. Т. 64, № 5. С. 5–63.
46. Chernoff H., Lehmann E.L. The use of maximum likelihood estimates in χ^2 test for goodness of fit // The Annals of Mathematical Statistics. 1954. V. 25. P. 579–586.
47. Никулин М.С. Критерий хи-квадрат для непрерывных распределений с параметрами сдвига и масштаба // Теория вероятностей и ее применение. 1973. Т. 18, № 3. С. 583–591.
48. Никулин М.С. О критерии хи-квадрат для непрерывных распределений // Теория вероятностей и ее применение. 1973. Т. 18, № 3. С. 675–676.
49. Rao K.C. Robson B.S. A chi-squared statistic for goodness-of-fit tests within the exponential family // Communications in Statistics – Theory and Methods. 1974. V. 3. P. 1139–1153.
50. Greenwood P.E., Nikulin M.S. A guide to chi-squared testing. New York : John Wiley & Sons, 1996. 280 p.
51. Лемешко Б.Ю., Постовалов С.Н., Чимитова Е.В. О распределениях статистики и мощности критерия типа χ^2 Никулина // Заводская лаборатория. Диагностика материалов. 2001. Т. 67, № 3. С. 52–58.

References

1. Kolmogoroff, A.N. (1933) Sulla determinazione empirica di una legge di distribuzione. *Giornale del Istituto Italiano degli Attuari*. 4(1). pp. 83–91.
2. Bolshev, L.N. (1963) Asimptoticheskie pirsonovskie preobrazovaniya [Asymptotic Pearson transformations]. *Teoriya veroyatnostey i ee primeneniye*. 8(2). pp. 129–155.
3. Bolshev, L.N. (1987) *Teoriya veroyatnostey i matematicheskaya statistika* [Theory of Probability and Mathematical Statistics]. Moscow: Nauka.
4. Bolshev, L.N. & Smirnov, N.V. (1983) *Tablitsy matematicheskoy statistiki* [Tables for Mathematical Statistics]. Moscow: Nauka.
5. Kremer N.Sh. (2010) *Teoriya veroyatnostey i matematicheskaya statistika* [Probability Theory and Mathematical Statistics]. 3rd ed. Moscow: UNITY-DANA.
6. Shpakov, P.S. & Yunakov, P.L. (2014) *Matematicheskaya obrabotka rezul'tatov izmereniy* [Mathematical Processing of Measurement Results]. Krasnoyarsk: SFU.
7. Anderson T.W. & Darling D.A. (1952) Asymptotic theory of certain “Goodness of fit” criteria based on stochastic processes. *The Annals of Mathematical Statistics*. 23(2). pp. 193–212.
8. Anderson, T.W. & Darling, D.A. (1954) A test of goodness of fit. *Journal of the American Statistical Association*. 29. pp. 765–769.
9. Kuiper, N.H. (1960) Tests concerning random points on a circle. *Proc. Koninkl. Nederl. Akad. Van Wetenschappen. Series A*. 63. pp. 38–47.
10. Stephens, M.A. (1965) The goodness-of-fit statistic V_N : distribution and significance points. *Biometrika*. 52(3/4). pp. 309–321. DOI: 10.2307/2333685
11. Stephens, M.A. (1974) EDF statistics for goodness of fit and some comparisons. *Journal of the American Statistical Association*. 69(347). pp. 730–737.
12. Lemesenko, B.Yu. & Gorbunova, A.A. (2013) Application and Power of the Nonparametric Kuiper, Watson, and Zhang Tests of Goodness-of-Fit. *Izmeritel'naya tekhnika – Measurement Techniques*. 56(5). pp. 465–475. DOI: 10.1007/s11018-013-0229-7
13. Watson, G.S. (1961) Goodness-of-fit tests on a circle. I. *Biometrika*. 48(1/2). pp. 109–114. DOI: 10.2307/2333135
14. Watson, G.S. (1962) Goodness-of-fit tests on a circle. II. *Biometrika*. 49(1/2). pp. 57–63. DOI: 10.2307/2333467
15. Pearson, E.S. & Hartley, H.O. (eds) (1972) *Biometrika tables for Statistics*. Vol. 2. Cambridge: University Press.
16. Zhang, J. (2002) Powerful goodness-of-fit tests based on the likelihood ratio. *Journal of the Royal Statistical Society: Series B*. 64(2). pp. 281–294. DOI: 10.1111/1467-9868.00337
17. Zhang, J. & Wu, Yu. (2005) Likelihood-ratio tests for normality. *Computational Statistics & Data Analysis*. 49(3). pp. 709–721. DOI: 10.1016/j.csda.2004.05.034
18. Lemesenko, B.Yu., Gorbunova, A.A., Lemesenko, S.B. & Rogozhnikov, A.P. (2014) Solving problems of using some nonparametric goodness-of-fit tests. *Optoelectronics, Instrumentation and Data Processing*. 50(1). pp. 21–35. DOI: 10.3103/S875669901401004X
19. Kac, M., Kiefer, J. & Wolfowitz, J. (1955) On tests of normality and other tests of goodness of fit based on distance methods. *The Annals of Mathematical Statistics*. 26. pp. 189–211.
20. Lemesenko, B.Yu. & Postovalov, S.N. (2001) O zavisimosti raspredeleniy statistik neparametricheskikh kriteriyev i ikh moshchnosti ot metoda otsenivaniya parametrov [On the dependence of the distributions of statistics of non-parametric tests and their power on the parameter estimation method]. *Zavodskaya laboratoriya. Diagnostika materialov*. 67(7). pp. 62–71.
21. Lemesenko, B.Yu. & Maklakov, A.A. (2004) Nonparametric Test in Testing Composite Hypotheses on Goodness of Fit Exponential Family Distributions. *Optoelectronics, Instrumentation and Data Processing*. 40(3). pp. 3–18.
22. Lilliefors, H.W. (1967) On the Kolmogorov-Smirnov test for normality with mean and variance unknown. *Journal of the American Statistical Association*. 62. pp. 399–402. DOI: 10.1080/01621459.1967.10482916
23. Lilliefors, H.W. (1969) On the Kolmogorov-Smirnov test for the exponential distribution with mean unknown. *Journal of the American Statistical Association*. 64. pp. 387–389. DOI: 10.1080/01621459.1969.10500983
24. R 50.1.037–2002. *Rekomendatsii po standartizatsii. Prikladnaya statistika. Pravila proverki soglasiya opytnogo raspredeleniya s teoreticheskimi* [Recommendations for standardization. Applied statistics. Rules of check of experimental and theoretical distribution of the consent]. Part II. Moscow: Izd-vo standartov.
25. Lemesenko, B.Yu. & Lemesenko, S.B. (2009a) Distribution models for nonparametric tests for fit in verifying complicated hypotheses and maximum-likelihood estimators. Part I. *Izmeritel'naya tekhnika – Measurement Techniques*. 52(6). pp. 555–565. DOI: 10.1007/s11018-009-9330-3
26. Lemesenko, B.Yu. & Lemesenko, S.B. (2009) Models for statistical distributions in nonparametric fitting tests on composite hypotheses based on maximum-likelihood estimators. Part II. *Izmeritel'naya tekhnika – Measurement Techniques*. 52(8). pp. 799–812. DOI: 10.1007/s11018-009-9353-9
27. Lemesenko, B.Yu., Lemesenko, S.B. & Postovalov, S.N. (2010) Statistic Distribution Models for Some Nonparametric Goodness-of-Fit Tests in Testing Composite Hypotheses. *Communications in Statistics – Theory and Methods*. 39(3). pp. 460–471. DOI: 10.1080/03610920903140148
28. Lemesenko, B.Yu. (2014) *Neparametricheskie kriterii soglasiya: rukovodstvo po primeneniyu* [Nonparametric goodness-of-fit tests. Guide on the application]. Moscow: INFRA–M. DOI: 10.12737/11873

29. Lemesheko, B.Yu., Lemesheko, S.B., Blinov, P.Yu., Veretelnikova, I.V. & Novikova, A.Yu. (n.d.) *Statisticheskiy analiz interval'nykh nablyudeniy odnomernykh nepreryvnykh sluchaynykh velichin "Interval'naya statistika 5.4": sv-vo o gos. registratsii programmy dlya EVM № 2018666213* [Statistical analysis of interval observations of one-dimensional continuous random variables "Interval statistics 5.4": Certificate of state registration of the computer program No. 2018666213 / Application No. 2018663206; Dec. 11/22/2018; registered 12/13/2018]. [Online] Available from: <https://ami.nstu.ru/~headrd/ISW.htm> (Accessed: 4th March 2023).
30. Lemesheko, B.Yu., Blinov, P.Yu. & Lemesheko S.B. (2016) Smeshchennost' neparametricheskikh kriteriev soglasiya odnositel'no nekotorykh par konkuriruyushchikh gipotez [Bias of nonparametric goodness-of-fit tests relative to certain pairs of competing hypotheses]. *Izmeritel'naya tekhnika – Measurement Techniques*. 59(5). pp. 468–475. DOI: 10.1007/s11018-016-0992-3
31. Lemesheko, B.Yu., Lemesheko, S.B. & Semenova, M.A. (2018) To question of the statistical analysis of big data. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 44. pp. 40–49. DOI: 10.17223/19988605/44/5
32. Lemesheko, B.Yu. & Lemesheko, S.B. (2020) About the influence of rounding errors on distributions of statistics of the goodness-of-fit tests. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 53. pp. 47–60. DOI: 10.17223/19988605/53/5
33. Lemesheko, B.Yu. & Lemesheko, S.B. (2021) Problemy primeneniya neparametricheskikh kriteriev soglasiya v zadachakh obrabotki rezul'tatov izmerenii [Problems of nonparametric goodness-of-fit test application in tasks of measurement results processing]. *Sistemy analiza i obrabotki dannykh – Analysis and Data Processing Systems*. 2(82). pp. 47–66. DOI: 10.17212/2782-2001-2021-2-47-66.
34. Lemesheko, B.Yu. & Lemesheko, S.B. (2022) Neparametricheskie kriterii soglasiya pri proverke normal'nosti v usloviyakh okrugleniya izmerenii [Nonparametric goodness-of-fit tests for normality testing under rounding-off measurements]. *Sistemy analiza i obrabotki dannykh – Analysis and Data Processing Systems*. 2(86). pp. 21–38. DOI: 10.17212/2782-2001-2022-2-21-38
35. Pearson, K. (1900) On the criterion that a given system of deviations from the probable in the case of a correlated system of variables in such that it can be reasonably to have arisen from random sampling. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*. 50. pp. 157–175. DOI: 10.1080/14786440009463897
36. Fisher, R.A. (1922) On the interpretation of χ^2 from contingency tables, and the calculation of P. *Journal of the Royal Statistical Society*. 85(1). pp. 87–94. DOI: 10.2307/2340521
37. Fisher, R.A. (1924) The conditions under which χ^2 measures the discrepancy between observation and hypothesis. *Journal of the Royal Statistical Society*. 87(3). pp. 442–450.
38. Neyman, J. & Pearson, E.S. (1928) On the use and interpretation of certain test criteria for purposes of statistical inference: Part I. *Biometrika*. 20A(1/2). pp. 175–240. DOI: 10.2307/2331945
39. Neyman, J. & Pearson, E.S. (1928) On the use and interpretation of certain test criteria for purposes of statistical inference: Part II. *Biometrika*. 20A(3/4). pp. 263–294. DOI: 10.2307/2332112
40. Mann, H.B. & Wald, A. (1942) On the choice of the number of class intervals in the application of the chi square test. *The Annals of Mathematical Statistics*. 13(3). pp. 306–317.
41. Denisov, V.I. & Lemesheko, B.Yu. (1979) *Optimal'noe gruppирование pri obrabotke eksperimental'nykh dannykh* [Optimal grouping in the processing of experimental data]. Novosibirsk: Novosibirsk Electrical Engineering Institute. pp. 5–14.
42. Lemesheko, B.Yu. (1998) Asimptoticheski optimal'noe gruppирование nablyudeniy v kriteriyakh soglasiya [Asymptotically optimum grouping of observations in goodness-of-fit tests]. *Zavodskaya laboratoriya – Industrial Laboratory*. 64(1). pp. 59–67.
43. Lemesheko, B.Y. & Chimitova, E.V. (2003) O vybore chisla intervalov v kriteriyakh soglasiya tipa χ^2 [On the choice of the number of intervals in χ^2 -type goodness-of-fit tests]. *Zavodskaya laboratoriya – Industrial Laboratory*. 69(1). pp. 61–67.
44. Chimitova E. V. & Lemesheko B. Y. (2017). Chi-Squared Goodness-of-Fit Tests: The Optimal Choice of Grouping Intervals. In: Szewczyk, R. & Kaliczynska, M. (eds) *Recent Advances in Systems, Control and Information Technology. SCIT 2016. Advances in Intelligent Systems and Computing*. Vol. 543. Springer, Cham. DOI: 10.1007/978-3-319-48923-0_82
45. Lemesheko, B.Yu. & Postovalov, S.N. (1998) O zavisimosti predel'nykh raspredeleniy statistik Pirsona i otnosheniya pravdopodobiya ot sposoba gruppированиa dannykh [Limit distributions of the Pearson chi 2 and likelihood ratio statistics and their dependence on the mode of data grouping]. *Zavodskaya laboratoriya – Industrial Laboratory*. 64(5). pp. 344–351.
46. Chernoff, H. & Lehmann, E.L. (1954) The use of maximum likelihood estimates in χ^2 test for goodness of fit. *The Annals of Mathematical Statistics*. 25. pp. 579–586. DOI: 10.1007/978-1-4614-1412-4_47
47. Nikulin, M.S. (1973a) Kriteriy khi-kvadrat dlya nepreryvnykh raspredeleniy s parametrami sdviga i masshtaba [Chi-square test for continuous distributions with location and scale parameters]. *Teoriya veroyatnostey i ee primeneniye*. 18(3). pp. 583–591.
48. Nikulin, M.S. (1973b) O kriterii khi-kvadrat dlya nepreryvnykh raspredeleniy [About the Chi-square test for continuous distributions]. *Teoriya veroyatnostey i ee primeneniye*. 18(3). pp. 675–676.
49. Rao, K.C. & Robson, B.S. (1974) A chi-squared statistic for goodness-of-fit tests within the exponential family. *Communications in Statistics – Theory and Methods*. 3. pp. 1139–1153. DOI: 10.1080/03610927408827216
50. Greenwood, P.E. & Nikulin, M.S. (1996) *A guide to chi-squared testing*. New York: John Wiley & Sons.
51. Lemesheko, B., Postovalov, S. & Chimitova, E. (2001) O raspredeleniyakh statistiki i moshchnosti kriteriya tipa χ^2 Nikulina [On statistic distributions and the power of the Nikulin χ^2 test]. *Zavodskaya laboratoriya – Industrial Laboratory*. 67(3). pp. 52–58.

Информация об авторе:

Лемешко Борис Юрьевич – профессор, доктор технических наук, профессор кафедры теоретической и прикладной информатики Новосибирского государственного технического университета (Новосибирск, Россия). E-mail: lemeshko@amil.nstu.ru

Автор заявляет об отсутствии конфликта интересов.

Information about the author:

Lemeshko Boris Yu. (Professor, Doctor of Technical Sciences, Professor, Department of Theoretical and Applied Informatics, Novosibirsk State Technical University, Novosibirsk, Russian Federation). E-mail: lemeshko@amil.nstu.ru

The author declares no conflicts of interests.

Поступила в редакцию 26.04.2023; принята к публикации 04.09.2023

Received 26.04.2023; accepted for publication 04.09.2023

Научная статья

УДК 519.21

doi: 10.17223/19988605/64/9

Сравнение оценок длительности мертвого времени, полученных методом моментов и методом максимального правдоподобия, в обобщенном рекуррентном MAP-поток событий с двумя состояниями

Анастасия Владимировна Кеба¹, Людмила Алексеевна Нежелская²

^{1,2} Томский государственный университет, Томск, Россия

¹ mir.na.mig7@mail.ru

² ludne@mail.ru

Аннотация. Рассматривается задача оценки длительности мертвого времени в обобщенном MAP-поток событий (Markovian Arrival Process) с двумя состояниями методом моментов и методом максимального правдоподобия в условиях непродлевающегося мертвого времени фиксированной длительности. Оценка длительности мертвого времени приводится только в случае рекуррентного задания потока. Формулируются алгоритмы вычисления значения оценки длительности мертвого времени (периода ненаблюдаемости) для двух рассматриваемых методов. Приводятся численные результаты расчетов значений оценок, полученных с использованием имитационной модели потока, и их анализ.

Ключевые слова: обобщенный MAP-поток событий; метод моментов; метод максимального правдоподобия; оценка длительности мертвого времени; рекуррентный поток.

Для цитирования: Кеба А.В., Нежелская Л.А. Сравнение оценок длительности мертвого времени, полученных методом моментов и методом максимального правдоподобия, в обобщенном рекуррентном MAP-поток событий с двумя состояниями // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2023. № 64. С. 91–101. doi: 10.17223/19988605/64/9

Original article

doi: 10.17223/19988605/64/9

Comparison of dead time estimates obtained by the method of moments and the maximum likelihood method in generalized MAP with two states

Anastasiya V. Keba¹, Lyudmila A. Nezhel'skaya²

^{1,2} Tomsk State University, Tomsk, Russian Federation

¹ mir.na.mig7@mail.ru

² ludne@mail.ru

Abstract. We consider the problem of estimating the dead time duration in generalized MAP (Markovian Arrival Process) with two states by the method of moments and the maximum likelihood method under conditions of unextendable dead time of a fixed duration. An estimate of the dead time duration is given only in the case of a recurrent flow specification. Algorithms for calculating the value of the dead time duration estimation (period of unobservability) are formulated for the two considered methods. Numerical results of calculations of the value of estimates obtained using a flow simulation model and their analysis are given.

Keywords: generalized MAP; the method of moments; the maximum likelihood method; dead time estimate; recurrent flow.

For citation: Keba, A.V., Nezhel'skaya, L.A. (2023) Comparison of dead time estimates obtained by the method of moments and the maximum likelihood method in generalized MAP with two states. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naja tehnika i informatika – Tomsk State University Journal of Control and Computer Science*. 64. pp. 91–101. doi: 10.17223/19988605/64/9

Введение

Рассматривается обобщенный МАР-поток событий, относящийся к классу дважды стохастических потоков событий [1–3]. Изучаемый поток представляет собой адекватную математическую модель реальных потоков случайных событий в телекоммуникационных системах, спутниковых сетях связи и глобальных компьютерных сетях, где сопровождающий процесс потока является кусочно-постоянным случайным процессом с конечным числом состояний [4–7].

В реальных ситуациях параметры потока событий либо известны частично, либо вообще неизвестны, либо изменяются со временем. Поэтому при анализе дважды стохастических потоков событий выделяют два основных класса задач, базой для которых служат моменты времени наступления событий в потоке: 1) оценивание состояний потока событий [8–11]; 2) оценивание параметров потока [12–15]. Задача, решаемая в настоящей статье, относится ко второму классу задач.

При решении перечисленных задач необходимо учитывать возможные искажающие факторы, существенно влияющие на качество оценивания. Одним из таких факторов выступает мертвое время регистрирующих приборов [16–19], порождаемое каждым зарегистрированным событием, – последующие события исходного потока, наступившие в течение периода мертвого времени, недоступны для наблюдения (теряются). Данный эффект характерен для большинства реальных систем. Предполагается, что период мертвого времени имеет фиксированную длительность T .

В настоящей работе исследуется рекуррентный обобщенный МАР-поток событий с двумя состояниями в условиях неполной наблюдаемости. Приводятся алгоритмы вычисления значения оценок длительности мертвого времени методом моментов и методом максимального правдоподобия и численные результаты оценивания. Производится сравнение полученных двумя методами значений оценок. Данное исследование является непосредственным развитием работ [20–22].

1. Постановка задачи

Исследуется обобщенный МАР-поток событий (далее – поток), функционирующий в установленном (стационарном) режиме. Сопровождающий случайный процесс $\lambda(t)$ изучаемого потока представляет собой кусочно-постоянный, принципиально ненаблюдаемый процесс с двумя состояниями S_1 и S_2 . Полагается, что при $\lambda(t) = \lambda_i$ имеет место i -е состояние (S_i), $i = 1, 2$, процесса $\lambda(t)$. При этом $\lambda_1 > \lambda_2 > 0$. Наблюдаемыми являются моменты времени наступления событий потока $t_1, t_2, \dots$

Функция распределения случайной величины – длительности пребывания процесса $\lambda(t)$ в состоянии S_i – является экспоненциальной: $F_i(t) = 1 - e^{-\lambda_i t}$, $t \geq 0$, $i = 1, 2$. В момент окончания состояния S_i процесс $\lambda(t)$ переходит из состояния S_i в состояние S_j с вероятностью $P_1(\lambda_j | \lambda_i)$ с наступлением события потока или с вероятностью $P_0(\lambda_j | \lambda_i)$ без наступления события потока, $i, j = 1, 2$. Отметим, что для введенных вероятностей справедливо

$$\sum_{j=1}^2 P_0(\lambda_j | \lambda_i) + \sum_{j=1}^2 P_1(\lambda_j | \lambda_i) = 1, \quad i = 1, 2.$$

Блочная матрица инфинитезимальных характеристик [19] процесса $\lambda(t)$ имеет вид $\mathbf{D} = \|\mathbf{D}_0 | \mathbf{D}_1\|$:

$$\mathbf{D}_0 = \begin{vmatrix} -\lambda_1(1 - P_0(\lambda_1 | \lambda_1)) & \lambda_1 P_0(\lambda_2 | \lambda_1) \\ \lambda_2 P_0(\lambda_1 | \lambda_2) & -\lambda_2(1 - P_0(\lambda_2 | \lambda_2)) \end{vmatrix},$$

$$\mathbf{D}_1 = \begin{vmatrix} \lambda_1 P_1(\lambda_1 | \lambda_1) & \lambda_1 P_1(\lambda_2 | \lambda_1) \\ \lambda_2 P_1(\lambda_1 | \lambda_2) & \lambda_2 P_1(\lambda_2 | \lambda_2) \end{vmatrix}.$$

Каждое зарегистрированное событие потока создает период ненаблюдаемости фиксированной длительности T (мертвое время), в течение которого другие события исходного потока недоступны для наблюдения (теряются); кроме того, события, наступившие в течение периода мертвого времени, не вызывают его продления (непродлевающееся мертвое время). Первое наступившее после периода мертвого времени событие потока снова создает период мертвого времени длительности T и т.д.

В качестве иллюстрации на рис. 1 приведена одна из реализаций процесса $\lambda(t)$ и наблюдаемого потока, где λ_i – значение процесса $\lambda(t)$ в состоянии S_i , $i = 1, 2$; штриховкой обозначены периоды мертвого времени; события обобщенного МАР-потока, недоступные наблюдению, отмечены черными кружками; $t_1, t_2, \dots$ – моменты времени наступления событий в наблюдаемом потоке.

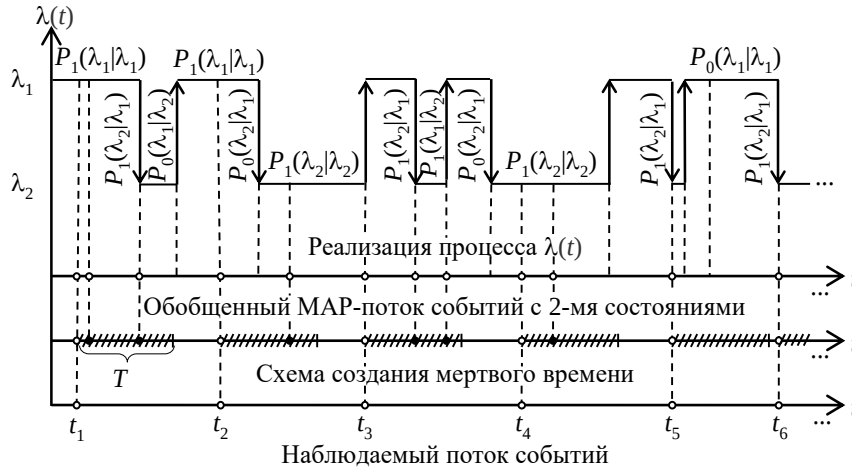


Рис. 1. Реализация обобщенного МАР-потока событий с двумя состояниями в условиях непродлевающего мертвого времени длительности T

Fig. 1. Realization of generalized MAP with 2 states under conditions of unextendible dead time of duration T

Замечание 1. Введение вероятности $P_0(\lambda_i|\lambda_i) \neq 0$, $i = 1, 2$, перехода процесса $\lambda(t)$ из состояния S_i в состояние S_i без наступления события приводит к обобщению классического МАР-потока с двумя состояниями.

Отметим, что для обобщенного МАР-потока событий процесс $\lambda(t)$ является скрытым марковским процессом.

Цель работы заключается в решении задачи оценивания периода ненаблюдаемости для рекуррентного задания обобщенного МАР-потока событий с двумя состояниями и в анализе полученных оценок. Оценка находится методом моментов и методом максимального правдоподобия на основании полученной при имитационном моделировании потока выборки последовательных моментов наступления событий $t_1, \dots, t_m$, на интервале наблюдения $(0, t)$; $0 < t_1 < \dots < t_m < t$.

2. Вид плотности вероятности длительности интервала между соседними событиями в рекуррентном обобщенном МАР-потоке при неполной наблюдаемости

Обозначим $\tau_k = t_{k+1} - t_k$, $\tau_k \geq 0$, – значение длительности k -го интервала между соседними событиями наблюдаемого потока t_k и t_{k+1} , $k = 1, 2, \dots$. Для плотности вероятности значений τ_k вследствие функционирования потока в стационарном режиме справедливо $p_T(\tau_k) = p_T(\tau)$, $\tau \geq 0$, при любом $k \geq 1$, что позволяет без ограничения общности положить момент наступления события t_k равным нулю или, что то же самое, момент наступления события есть $\tau = 0$. Тогда плотность вероятности $p_T(\tau)$ длительности интервала между соседними событиями в наблюдаемом потоке имеет вид [20]:

$$p_T(\tau) = \begin{cases} 0, & 0 \leq \tau < T, \\ \gamma(T)z_1 e^{-z_1(\tau-T)} + (1-\gamma(T))z_2 e^{-z_2(\tau-T)}, & \tau \geq T, \end{cases}$$

$$\gamma(T) = \frac{z_2 - \lambda_1 \pi_1(T) [P_1(\lambda_1|\lambda_1) + P_1(\lambda_2|\lambda_1)] - \lambda_2 \pi_2(T) [P_1(\lambda_1|\lambda_2) + P_1(\lambda_2|\lambda_2)]}{z_2 - z_1}, \quad (1)$$

$$\pi_i(T) = \pi_i - [\pi_i - \pi_i(0|T)] e^{-aT}, \quad i = 1, 2,$$

где

$$\begin{aligned}
 z_{1,2} &= \frac{\lambda_1(1 - P_0(\lambda_1|\lambda_1)) + \lambda_2(1 - P_0(\lambda_2|\lambda_2)) \mp \sqrt{D}}{2}, \\
 D &= [\lambda_1(1 - P_0(\lambda_1|\lambda_1)) - \lambda_2(1 - P_0(\lambda_2|\lambda_2))]^2 + 4\lambda_1\lambda_2P_0(\lambda_2|\lambda_1)P_0(\lambda_1|\lambda_2), \quad 0 < z_1 < z_2, \\
 \pi_1(0|T) &= \frac{p_{21} + \pi_1(p_{11} - p_{21})(1 - e^{-aT})}{p_{12} + p_{21} + (p_{22} - p_{12})(1 - e^{-aT})}, \quad \pi_2(0|T) = \frac{p_{12} + \pi_2(p_{22} - p_{12})(1 - e^{-aT})}{p_{12} + p_{21} + (p_{22} - p_{12})(1 - e^{-aT})}, \\
 P_{1i} &= \frac{P_1(\lambda_i|\lambda_1)(1 - P_0(\lambda_2|\lambda_2)) + P_0(\lambda_2|\lambda_1)P_1(\lambda_i|\lambda_2)}{(1 - P_0(\lambda_1|\lambda_1))(1 - P_0(\lambda_2|\lambda_2)) - P_0(\lambda_1|\lambda_2)P_0(\lambda_2|\lambda_1)}, \\
 p_{2i} &= \frac{P_1(\lambda_i|\lambda_2)(1 - P_0(\lambda_1|\lambda_1)) + P_0(\lambda_1|\lambda_2)P_1(\lambda_i|\lambda_1)}{(1 - P_0(\lambda_1|\lambda_1))(1 - P_0(\lambda_2|\lambda_2)) - P_0(\lambda_1|\lambda_2)P_0(\lambda_2|\lambda_1)}, \quad \sum_{j=1}^2 p_{ij} = 1, \quad i = 1, 2, \\
 \pi_1 &= \frac{\lambda_2(P_0(\lambda_1|\lambda_2) + P_1(\lambda_1|\lambda_2))}{a}, \quad \pi_2 = \frac{\lambda_1(P_0(\lambda_2|\lambda_1) + P_1(\lambda_2|\lambda_1))}{a}, \quad \sum_{i=1}^2 \pi_i = 1, \\
 a &= \lambda_1(P_0(\lambda_2|\lambda_1) + P_1(\lambda_2|\lambda_1)) + \lambda_2(P_0(\lambda_1|\lambda_2) + P_1(\lambda_1|\lambda_2)).
 \end{aligned} \tag{2}$$

Так как рассматриваемый обобщенный МАР-поток событий является в общем случае коррелированным, то на основании полученной аналитической формулы совместной плотности вероятности $p_T(\tau_1, \tau_2)$ в работе [20] получаем условие рекуррентности: $P_1(\lambda_1|\lambda_1)P_1(\lambda_2|\lambda_2) - P_1(\lambda_1|\lambda_2)P_1(\lambda_2|\lambda_1) = 0$. Тогда выражение для плотности вероятности $p_T(\tau)$ длительности интервала между соседними событиями в наблюдаемом потоке в (1) преобразуется к виду:

$$\begin{aligned}
 p_T(\tau) &= \begin{cases} 0, & 0 \leq \tau < T, \\ \gamma(T)z_1e^{-z_1(\tau-T)} + (1 - \gamma(T))z_2e^{-z_2(\tau-T)}, & \tau \geq T, \end{cases} \\
 \gamma(T) &= \frac{z_2 - \lambda_1\pi_1(T)[P_1(\lambda_1|\lambda_1) + P_1(\lambda_2|\lambda_1)] - \lambda_2\pi_2(T)[P_1(\lambda_1|\lambda_2) + P_1(\lambda_2|\lambda_2)]}{z_2 - z_1}, \\
 \pi_i(T) &= \pi_i - [\pi_i - p_{ji}]e^{-aT}, \quad i, j = 1, 2, \quad i \neq j,
 \end{aligned} \tag{3}$$

где $z_1, z_2, p_{ij}, i, j = 1, 2, \pi_i, i = 1, 2, a$ определены в (2).

3. Оценка длительности мертвого времени методом моментов

Введем в рассмотрение выборку $\tau_1, \dots, \tau_{m-1}$ из распределения $p_T(\tau) = \gamma(T)z_1e^{-z_1(\tau-T)} + (1 - \gamma(T))z_2e^{-z_2(\tau-T)}, \tau \geq 0$, зависящего от неизвестного параметра T . В силу близости теоретической и эмпирической функций распределения, т.е. неограниченного сближения их значений при $n \rightarrow \infty$, следует ожидать близости их числовых характеристик – моментов одного и того же порядка. Теоретический начальный момент 1-го порядка $M_T\tau = \int_T^\infty \tau p_T(\tau) d\tau$ близок к соответствующему выборочному моменту – статистике $C_1 = \frac{1}{m-1} \sum_{k=1}^{m-1} \tau_k$, где $\tau_k = t_{k+1} - t_k$ – значение длительности между моментами t_k и t_{k+1} наступления событий в наблюдаемом потоке событий. Таким образом, для оценки параметра T необходимо иметь уравнение моментов

$$M_T\tau = C_1.$$

Выполняя интегрирование левой части уравнения, находим

$$M_T \tau = \int_T^\infty \tau (\gamma(T) z_1 e^{-z_1(\tau-T)} + (1-\gamma(T)) z_2 e^{-z_2(\tau-T)}) d\tau = T + \frac{\gamma(T)}{z_1} + \frac{1-\gamma(T)}{z_2} = C_1, \quad (4)$$

где $\gamma(T)$ определена в (3) для рекуррентного обобщенного МАР-потока.

Подчеркнем, что решение уравнения (4) является значением $\hat{T}_{MM}$ оценки длительности мертвого времени $\hat{T}_{MM}$ методом моментов и возможно только численно; в данной работе решение уравнения моментов (4) находится с применением метода простой итерации.

При выбранных параметрах рекуррентного потока, как показывают результаты численных экспериментов, приведенные в п. 5, математическое ожидание $M_T \tau$ есть возрастающая функция переменной (параметра) T , $T > 0$. Вследствие этого уравнение моментов имеет единственное решение.

4. Оценка длительности мертвого времени методом максимального правдоподобия

Пусть $\tau_1 = t_2 - t_1$, $\tau_2 = t_3 - t_2$, ..., $\tau_{m-1} = t_m - t_{m-1}$ – последовательность измеренных в результате наблюдения за потоком в течение интервала $(0, t)$ значений длительностей интервалов между соседними событиями наблюдаемого потока.

Упорядочим величины $\tau_1, \tau_2, \dots, \tau_{m-1}$ по возрастанию: $\tau_{\min} = \tau^{(1)} < \tau^{(2)} < \dots < \tau^{(m-1)}$. В силу предположений последовательность моментов $t_1, t_2, \dots, t_k, \dots$ образует вложенную цепь Маркова по моментам наступления событий, т.е. наблюдаемый поток обладает марковским свойством, если его эволюцию рассматривать с момента t_k , $k = 1, 2, \dots$. Тогда функция правдоподобия, с учетом (3), запишется в виде:

$$L(T | \tau^{(1)}, \dots, \tau^{(m-1)}) = \begin{cases} 0, & 0 \leq \tau_{\min} < T; \\ \prod_{i=1}^{m-1} p_T(\tau^{(i)}), & \tau_{\min} \geq T. \end{cases}$$

Так как поставленная задача заключается в построении оценки длительности мертвого времени (в предположении, что остальные параметры потока известны), то, согласно методу максимального правдоподобия, ее реализацией является решение оптимизационной задачи

$$L(T | \tau^{(1)}, \dots, \tau^{(m-1)}) = \prod_{i=1}^{m-1} p_T(\tau^{(i)}) \Rightarrow \max_{0 \leq T < \tau_{\min}}, \quad (5)$$

где $p_T(\tau)$ определена в (3).

Значение T , при котором (5) достигает своего глобального максимума, является значением $\hat{T}_{ML}$ оценки максимального правдоподобия $\hat{T}_{ML}$ длительности мертвого времени. Подчеркнем, что решение задачи (5) в работе производится численно.

При выбранных параметрах рекуррентного потока, как показывают результаты численных экспериментов, приведенные в п. 5, функция правдоподобия $L(T | \tau^{(1)}, \dots, \tau^{(m-1)})$ есть возрастающая функция переменной (параметра) T , $0 \leq \tau_{\min} < T$. Вследствие этого решение оптимизационной задачи (5) единственное.

5. Результаты численных расчетов

С целью получения численных результатов разработан алгоритм вычисления значения $\hat{T}_{MM}$ оценок $\hat{T}_{MM}$ (решение уравнения моментов (4)), значения $\hat{T}_{ML}$ оценок $\hat{T}_{ML}$ (решение оптимизационной задачи (5)) длительности мертвого времени и построена имитационная модель обобщенного рекуррентного МАР-потока событий с двумя состояниями в условиях непродлевающегося мертвого

времени. Основой имитационной модели являются датчик псевдослучайных чисел и метод обратных функций [23]. На первом этапе расчета осуществляется имитационное моделирование обобщенного рекуррентного МАР-потока события с двумя состояниями при наличии мертвого времени с тем, чтобы получить выборку моментов времени наступления событий $t_1, \dots, t_m$. На втором этапе расчета вычисляются значения оценок $\hat{T}_{MM}$ и $\hat{T}_{ML}$.

Для сравнения методов максимального правдоподобия и метода моментов по качеству получаемых оценок для заданного набора параметров $\lambda_i, P_1(\lambda_j|\lambda_i), P_0(\lambda_j|\lambda_i), i, j = 1, 2, T$ в каждой v -й, $v = \overline{1, N}$, реализации находятся значения оценок $\hat{T}_{MM}$ и $\hat{T}_{ML}$, на основании которых вычисляются выборочное среднее $\hat{M}(\hat{T}) = \frac{1}{N} \sum_{v=1}^N \hat{T}^{(v)}$ и выборочная вариация $\hat{V}(\hat{T}) = \frac{1}{N-1} \sum_{v=1}^N (\hat{T}^{(v)} - T)^2$, где $\hat{T}^{(v)} = \hat{T}_{MM}, \hat{T}_{ML}$; T – значение длительности мертвого времени, известное из имитационной модели; $\hat{T}$ – оценки $\hat{T}_{MM}, \hat{T}_{ML}$.

Оценка $\hat{T}_{ML}$ лучше, если $\hat{V}(\hat{T}_{ML}) < \hat{V}(\hat{T}_{MM})$, иначе лучше оценка $\hat{T}_{MM}$.

Эксперимент 1. Фиксируются число состояний $n = 2$, количество опытов (реализаций) $N = 1\,000$, длительность мертвого времени $T = 1$, значения $\lambda_1 = 25, \lambda_2 = 10$ процесса $\lambda(t)$ и переходные вероятности, представленные в табл. 1.

Цель данного эксперимента – нахождение временного интервала установления стационарного режима функционирования рекуррентного обобщенного МАР-потока события с двумя состояниями в условиях непродлевающегося мертвого времени фиксированной длительности, или, иными словами, установление зависимостей $\hat{M}(\hat{T}), \hat{V}(\hat{T})$ от длительности времени моделирования T_m ед. времени, где $T_m \in \{100; 200; \dots; 1\,500\}$.

Результаты первого статистического эксперимента представлены в табл. 2 и на рис. 2.

Таблица 1

Значения переходных вероятностей рекуррентного обобщенного МАР-потока событий с двумя состояниями

$P_0(\lambda_1 \lambda_1) = 0,05$	$P_0(\lambda_1 \lambda_2) = 0,05$
$P_0(\lambda_2 \lambda_1) = 0,05$	$P_0(\lambda_2 \lambda_2) = 0,05$
$P_1(\lambda_1 \lambda_1) = 0,45$	$P_1(\lambda_1 \lambda_2) = 0,45$
$P_1(\lambda_2 \lambda_1) = 0,45$	$P_1(\lambda_2 \lambda_2) = 0,45$

Таблица 2

Результаты первого статистического эксперимента

T_m	$\hat{M}(\hat{T}_{MM})$	$\hat{V}(\hat{T}_{MM}) \times 10^4$	$\hat{M}(\hat{T}_{ML})$	$\hat{V}(\hat{T}_{ML}) \times 10^4$
100	0,9451	29,1964	0,9954	0,4576
200	0,9504	23,7264	0,9952	0,4799
300	0,9521	22,0135	0,9954	0,4522
400	0,9529	21,2357	0,9955	0,4418
...	...	...	...	...
1 400	0,9548	19,4938	0,9953	0,4652
1 500	0,9549	19,4513	0,9956	0,4368

Из анализа численных результатов, приведенных в табл. 2 и на рис. 2, следует, что имеет место смещение оценок относительно исходного значения оцениваемого параметра. Оценивание длительности мертвого времени методом максимального правдоподобия дает лучшие по сравнению с методом моментов результаты (в смысле выборочной вариации) вне зависимости от длительности времени моделирования. При этом выборочная вариация для каждой из предложенных оценок стремится к близкому к нулю значению, т.е. методика оценивания качественна, и полученную оценку можно принимать за истинную с достаточно малой погрешностью.

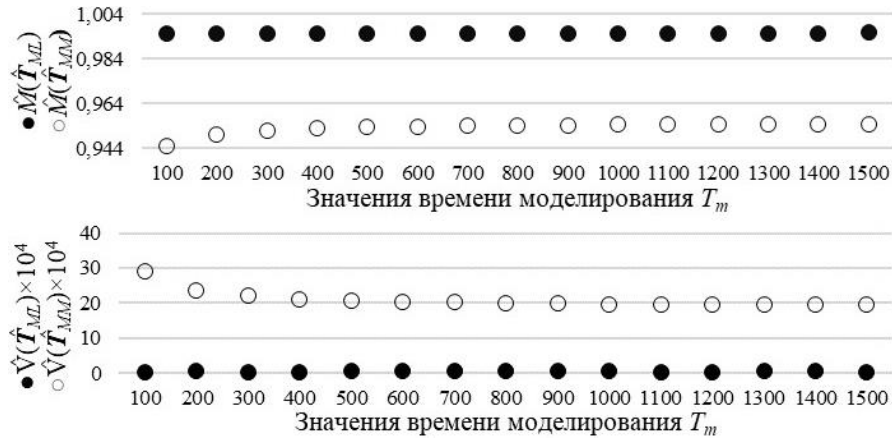


Рис. 2. Графики зависимостей $\hat{M}(\hat{T}_{ML})$, $\hat{M}(\hat{T}_{MM})$, $\hat{V}(\hat{T}_{ML}) \times 10^4$, $\hat{V}(\hat{T}_{MM}) \times 10^4$ от значения T_m

Fig. 2. $\hat{M}(\hat{T}_{ML})$, $\hat{M}(\hat{T}_{MM})$, $\hat{V}(\hat{T}_{ML}) \times 10^4$, $\hat{V}(\hat{T}_{MM}) \times 10^4$ as functions of T_m

Замечание 2. В случае рекуррентного задания обобщенного МАР-потока событий оценки $\hat{T}_{MM}$, полученные методом моментов, являются состоятельными, так как последовательность $\tau_1, \dots, \tau_{m-1}$ представляет собой взаимно независимые случайные величины, конечный теоретический момент $M_T \tau$ существует и уравнение моментов (4) имеет единственное решение [24].

Эксперимент 2. Фиксируются число состояний $n = 2$, количество опытов $N = 1\,000$, значение времени моделирования $T_m = 1\,000$ ед. времени, значения $\lambda_1 = 25$, $\lambda_2 = 10$ процесса $\lambda(t)$ и переходные вероятности, представленные в табл. 1.

В рамках эксперимента исследуются зависимости $\hat{M}(\hat{T})$, $\hat{V}(\hat{T})$ от изменения значения длительности мертвого времени T , где $T \in \{1; 2; \dots; 10\}$.

Результаты второго статистического эксперимента представлены в табл. 3 и на рис. 3.

Таблица 3

Результаты второго статистического эксперимента

T	$\hat{M}(\hat{T}_{MM})$	$\hat{V}(\hat{T}_{MM}) \times 10^4$	$\hat{M}(\hat{T}_{ML})$	$\hat{V}(\hat{T}_{ML}) \times 10^4$
1	0,9445	19,7622	0,9953	0,4662
2	1,9504	22,6564	1,9905	1,8988
3	2,9444	27,6735	2,9868	3,9532
4	3,9364	35,5709	3,9809	7,6127
...	...	...	...	...
9	8,8665	155,2961	8,9604	35,6676
10	9,8466	205,8273	9,9552	44,8211

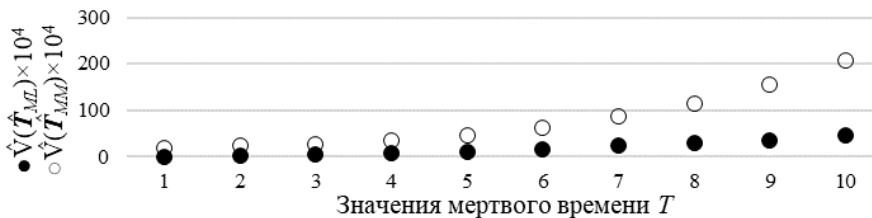


Рис. 3. Графики зависимостей $\hat{V}(\hat{T}_{ML}) \times 10^4$, $\hat{V}(\hat{T}_{MM}) \times 10^4$ от значения T

Fig. 3. $\hat{V}(\hat{T}_{ML}) \times 10^4$, $\hat{V}(\hat{T}_{MM}) \times 10^4$ as functions of T

Результаты табл. 3 и рис. 3 указывают на то, что увеличение значения длительности мертвого времени T влечет увеличение выборочной вариации полученных оценок. Это объясняется тем, что при увеличении значения T увеличивается период ненаблюдаемости потока, что приводит к увеличе-

нию числа потерянных событий и, как следствие, к ухудшению качества оценивания как методом моментов, так и методом максимального правдоподобия.

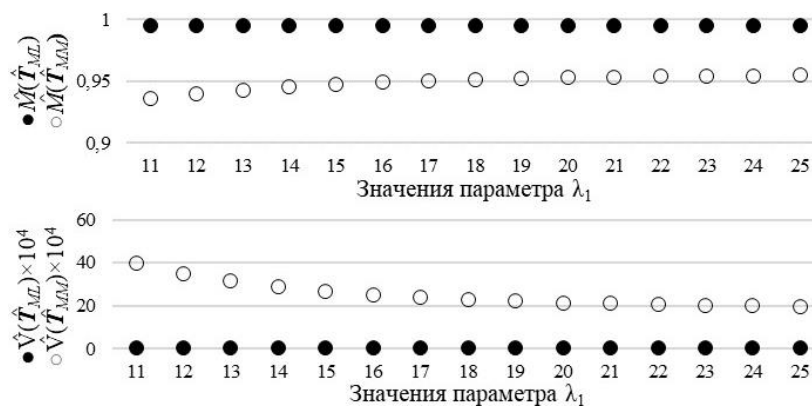
Эксперимент 3. Фиксируются число состояний $n = 2$, количество опытов $N = 1\,000$, значение времени моделирования $T_m = 1\,000$ ед. времени, длительность мертвого времени $T = 1$, значение $\lambda_2 = 10$ процесса $\lambda(t)$ и переходные вероятности, представленные в табл. 1. Будем изменять значение $\lambda_1 \in \{11; 12; \dots; 25\}$ процесса $\lambda(t)$ с целью выявления влияния близости значений λ_i , $i = 1, 2$ на качество получаемых оценок.

Результаты третьего статистического эксперимента представлены в табл. 4 и на рис. 4.

Таблица 4

Результаты 3-го статистического эксперимента

λ_1	$\hat{M}(\hat{T}_{MM})$	$\hat{V}(\hat{T}_{MM}) \times 10^4$	$\hat{M}(\hat{T}_{ML})$	$\hat{V}(\hat{T}_{ML}) \times 10^4$
11	0,9357	40,043	0,9950	0,4905
12	0,9399	34,9576	0,9953	0,4653
13	0,9453	31,5313	0,9953	0,4687
14	0,9473	28,8446	0,9951	0,4833
...	...	...	...	...
24	0,9544	19,8812	0,9951	0,4825
25	0,9545	19,7622	0,9953	0,4662

Рис. 4. Графики зависимостей $\hat{M}(\hat{T}_{ML})$, $\hat{M}(\hat{T}_{MM})$, $\hat{V}(\hat{T}_{ML}) \times 10^4$, $\hat{V}(\hat{T}_{MM}) \times 10^4$ от значения λ_1 Fig. 4. $\hat{M}(\hat{T}_{ML})$, $\hat{M}(\hat{T}_{MM})$, $\hat{V}(\hat{T}_{ML}) \times 10^4$, $\hat{V}(\hat{T}_{MM}) \times 10^4$ as functions of λ_1

Из результатов, представленных в табл. 4 и на рис. 4, следует, что при увеличении различимости состояний (при увеличении значения параметра λ_1) оценки, полученные методом моментов и методом максимального правдоподобия улучшаются, о чем свидетельствуют значения $\hat{M}(\hat{T}_{ML})$, $\hat{M}(\hat{T}_{MM})$, приближающиеся к истинному значению длительности мертвого времени T , и $\hat{V}(\hat{T}_{ML}) \times 10^4$, $\hat{V}(\hat{T}_{MM}) \times 10^4$, стремящиеся к нулю. Оценка, полученная методом максимального правдоподобия, показывает лучшие результаты.

Заключение

В данной статье рассмотрен рекуррентный обобщенный МАР-поток событий с двумя состояниями в условиях непродлевающего мертвого времени фиксированной длительности T . Приведен явный вид плотности вероятности (3) значений длительности интервала между соседними событиями рекуррентного обобщенного МАР-потока при наличии мертвого времени. Методом моментов находится значение $\hat{T}_{MM}$ оценки $\hat{T}_{MM}$ длительности мертвого времени T как решение уравнения момен-

тов (4), методом максимального правдоподобия находится значение $\hat{T}_{ML}$ оценки $\hat{T}_{ML}$ длительности мертвого времени T как решение оптимизационной задачи (5). Алгоритм имитационной модели потока и алгоритмы вычисления значения оценок реализованы на языке программирования C# в среде Visual Studio. Поставлены статистические эксперименты, указывающие на работоспособность алгоритмов оценки параметра T и приемлемое качество получаемых оценок. Проведено сравнение оценок $\hat{T}_{MM}$ и $\hat{T}_{ML}$, свидетельствующее о лучшем оценивании T методом максимального правдоподобия в смысле меньших значений выборочных вариаций.

Список источников

1. Kingman J.F.C. On doubly stochastic Poisson process // *Proceedings of the Cambridge Philosophical Society*. 1964. V. 60, is. 4. P. 923–930.
2. Cox D.R., Miller H.D. The theory of stochastic processes. New York : Wiley, 1965. 398 p.
3. Grandell J. Doubly stochastic Poisson processes. Berlin ; Heidelberg : Springer-Verlag, 1976. 240 p.
4. Башарин Г.П., Кокотушкин В.А., Наумов В.А. О методе эквивалентных замен расчета фрагментов сетей связи. Ч. 1 // *Известия АН 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 АН 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. Вишневский В.М., Дудин А.Н., Клименок В.И. Стохастические системы с коррелированными потоками. Теория и применения в телекоммуникационных сетях. М. : Техносфера, 2018. 564 с.
8. Bakholdina M.A., Gortsev A.M. Optimal estimation of the states of modulated semi-synchronous integrated flow of events in condition of its incomplete observability // *Applied Mathematical Sciences*. 2015. T. 9, № 29–32. P. 1433–1451.
9. Горцев А.М., Голофастова М.Н. Оптимальная оценка состояний модулированного синхронного дважды стохастического потока событий // *Вестник Томского государственного университета. Управление, вычислительная техника и информатика*. 2013. № 2 (23). С. 42–53.
10. Горцев А.М., Леонова М.А. Оптимальная оценка состояний обобщенного асинхронного дважды стохастического потока событий // *Вестник Томского государственного университета. Управление, вычислительная техника и информатика*. 2010. № 1 (10). С. 33–47.
11. Nezhelskaya L., Tumashkina D. Optimal state estimation of semi-synchronous event of flow of the second order under its complete observability // *Communications in Computer and Information Science*. 2018. V. 912. P. 93–105.
12. Цициашвили Г.Ш., Осипова М.А. Оценка параметров мультипликативных распределений сетей массового обслуживания // *Проблемы передачи информации*. 2009. Т. 45, № 4. С. 115–120.
13. Okamura H., Dohi T., Trivedi K.S. Markovian arrival process parameter estimation with group data // *IEEE / ACM Transactions on Networking (TON)*. 2009. V. 17 (4). P. 1326–1339.
14. Нежелская Л.А. Совместная плотность вероятностей длительности интервалов модулированного МАР-потока событий и условия рекуррентности потока // *Вестник Томского государственного университета. Управление, вычислительная техника и информатика*. 2015. № 1 (30). С. 57–67.
15. 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.
16. Normey-Rico J.E. Control of dead-time processes. London : Springer-Verlag, 2007. 462 p. (Advanced textbooks in control and signal processing).
17. Леонова М.А., Нежелская Л.А. Оценка максимального правдоподобия длительности мертвого времени в обобщенном асинхронном потоке событий // *Вестник Томского государственного университета. Управление, вычислительная техника и информатика*. 2013. № 2 (23). С. 54–63.
18. Феллер В. Введение в теорию вероятностей и ее приложения. М. : Мир, 1967. Т. 2. 752 с.
19. Назаров А.А., Терпугов А.Ф. Теория вероятностей и случайных процессов. Томск : Изд-во НТЛ, 2006. 204 с.
20. Кеба А.В., Нежелская Л.А. Совместная плотность вероятности длительностей интервалов обобщенного МАР-потока событий при его неполной наблюдаемости и условия рекуррентности потока // *Труды Томского государственного университета. Сер. физико-математическая*. Томск : Изд. Дом Том. гос. ун-та, 2022. Т. 307. С. 27–33.
21. Кеба А.В., Нежелская Л.А. Оценка длительности мертвого времени в обобщенном МАР-потоке событий с двумя состояниями // *Вестник Томского государственного университета. Управление, вычислительная техника и информатика*. 2022. № 61. С. 69–80.
22. Keba A., Nezhel'skaya L. Probability Density of the Interval Duration Between Events in the Generalized MAP with Its Incomplete Observability // *Lecture Notes in Computer Science*. 2022. V. 13766. P. 349–370.
23. Соболев И.М. Численные методы Монте-Карло. М. : Наука, 1973. 312 с.
24. Малинковский Ю.В. Теория вероятностей и математическая статистика. Гомель : ГТУ им. Ф. Скорины, 2004. Ч. 2: Математическая статистика. 146 с.

References

1. Kingman, J.F.C. (1964) On doubly stochastic Poisson process. *Proceedings of the Cambridge Philosophical Society*. 60(4). pp. 923–930. DOI: 10.1017/S030500410003838X
2. Cox, D.R. & Miller, H.D. (1965) *The Theory of Stochastic Processes*. New York: Wiley.
3. Grandell, J. (1976) *Doubly Stochastic Poisson Processes*. Berlin; Heidelberg: Springer-Verlag.
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. Ch. 2 [On the Equivalent Substitutions Method for Computing Fragments of Communication Networks. Part 2]. *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. DOI: 10.2307/3213143
7. 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.
8. Bakhholdina, M.A. & Gortsev, A.M. (2015) Optimal estimation of the states of modulated semi-synchronous integrated flow of events in condition of its incomplete observability. *Applied Mathematical Sciences*. 9. 29(32). pp. 1433–1451.
9. Gortsev, A.M. & Golofastova, M.N. (2013) The optimal state estimation of modulated synchronous twice stochastic flow of events. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 2(23). pp. 42–53.
10. Gortsev, A.M. & Leonova, M.A. (2010). The optimal evaluation of states of the integrated asynchronous doubly stochastic event flow. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 1(10). pp. 33–47.
11. 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. DOI: 10.1007/978-3-319-97595-5_8
12. Tsitsiashvili, G.Sh. & Osipova, M.A. (2009) Parameter estimation for product-form distributions of queueing networks. *Problemy peredachi informatsii – Problems of Information Transmission*. 45(4). pp. 115–120.
13. Okamura, H., Dohi, T. & Trivedi, K.S. (2009) Markovian arrival process parameter estimation with group data. *IEEE/ACM Transactions on Networking (TON)*. 17(4). pp. 1326–1339. DOI: 10.1109/TNET.2008.2008750
14. Nezhelskaya, L.A. (2015) Joint probability density of the intervals duration in modulated MAP event flows and its recurrence conditions. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 1(30). pp. 57–67.
15. Vasilevskaya, T.P., Gortsev, A.M. & Nezhelskaya, L.A. (2004) Estimation of the duration of dead time and parameters of a synchronous alternating flow with the manifestation or non-manifestation of events. *Vestnik Tomskogo gosudarstvennogo universiteta. – Tomsk State University Journal*. S9-2. pp. 129–138.
16. Normey-Rico, J.E. (2007) *Control of Dead-Time Processes*. London: Springer-Verlag.
17. Leonova, M.A. & Nezhelskaya, L.A. (2013). Maximum likelihood estimation of dead time value at a generalized asynchronous flow of events. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 2(23). pp. 54–63.
18. Feller, V. (1967) *Vvedenie v teoriyu veroyatnostey i ee prilozheniya* [Introduction to the theory of probability and its applications]. Vol. 2. Moscow: Mir.
19. Nazarov, A.A. & Terpugov, A.F. (2006) *Teoriya veroyatnostey i sluchaynykh protsessov* [Theory of Probability and Random Processes]. Tomsk: NTL.
20. Keba, A.V. & Nezhelskaya, L.A. (2022) Sovmestnaya plotnost' veroyatnosti dlitel'nostey intervalov obobshchennogo MAP-potoka sobytii pri ego nepolnoy nablyudaemosti i usloviya rekurrentnosti potoka [The joint probability density of interval durations of a generalized MAP with its incomplete observability and flow recurrence conditions]. *Trudy Tomskogo gosudarstvennogo universiteta. Ser. fiziko-matematicheskaya*. 307. pp. 27–33.
21. Keba, A.V. & Nezhelskaya, L.A. (2022) Estimation of dead time duration in generalized MAP with two states. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 61. pp. 69–80. DOI: 10.17223/19988605/61/7
22. Keba, A. & Nezhelskaya, L. (2022) Probability Density of the Interval Duration Between Events in the Generalized MAP with Its Incomplete Observability. *Lecture Notes in Computer Science*. 13766. pp. 349–370.
23. Sobol, I.M. (1973) *Chislennyye metody Monte-Karlo* [Monte Carlo Numerical Methods]. Moscow: Nauka.
24. Malinkovskiy, Yu.V. (2004) *Teoriya veroyatnostey i matematicheskaya statistika* [Theory of Probability and Mathematical

Информация об авторах:

Кеба Анастасия Владимировна – аспирант кафедры прикладной математики Института прикладной математики и компьютерных наук Национального исследовательского Томского государственного университета (Томск, Россия). E-mail: mir.na.mig7@mail.ru

Нежелская Людмила Алексеевна – доцент, 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

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

Information about the authors:

Keba Anastasia V. (Post-graduate Student, Institute of Applied Mathematics and Computer Science, National Research Tomsk State University, Tomsk, Russian Federation). E-mail: mir.na.mig7@mail.ru

Nezhel'skaya Lyudmila A. (Doctor of Physical and Mathematical Sciences, Head of the Applied Mathematics Department, Institute of Applied Mathematics and Computer Science, National Research Tomsk State University, Tomsk, Russian Federation). E-mail: ludne@mail.ru

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

Поступила в редакцию 11.05.2023; принята к публикации 04.09.2023

Received 11.05.2023; accepted for publication 04.09.2023

Научная статья

УДК 51-76

doi: 10.17223/19988605/64/10

Вейвлеты Мейера с коэффициентом масштабирования $N = 12$ для классификации сигналов по высокочастотным признакам

Полина Николаевна Подкур¹, Николай Константинович Смоленцев²

¹ Кузбасский технический университет им. Т.Ф. Горбачева, Кемерово, Россия, paulina.podkur@gmail.com

² Кемеровский государственный университет, Кемерово, Россия, smolennk@mail.ru

Аннотация. Строятся ортогональные вейвлеты Мейера с коэффициентом масштабирования $N = 12$, которые используются для изучения вопроса о классификации сигналов по высокочастотным признакам. Предложенный подход применяется для классификации сигналов электрокардиограмм (ЭКГ). Получено разложение сигналов ЭКГ на отдельные компоненты. Найдены наиболее информативные частотные признаки, которые обеспечивают точность разделения классов более 95%.

Ключевые слова: вейвлеты Мейера; вейвлет-анализ; вейвлеты с коэффициентом масштабирования N ; высокочастотные компоненты; классификация ЭКГ.

Для цитирования: Подкур П.Н., Смоленцев Н.К. Вейвлеты Мейера с коэффициентом масштабирования $N = 12$ для классификации сигналов по высокочастотным признакам // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2023. № 64. С. 102–112. doi: 10.17223/19988605/64/10

Original article

doi: 10.17223/19988605/64/10

Meyer wavelets with scaling factor $N = 12$ for classifying signals by high-frequency features

Paulina N. Podkur¹, Nikolay K. Smolentsev²

¹ T.F. Gorbachev Kuzbass State Technical University, Kemerovo, Russian Federation, paulina.podkur@gmail.com

² Kemerovo State University, Kemerovo, Russian Federation, smolennk@mail.ru

Abstract. In this paper, orthogonal Meyer wavelets are constructed with a scaling factor $N = 12$, which are used to study the problem of signal classification by high-frequency features. The application of the proposed approach for classifying electrocardiogram (ECG) signals is given. Decomposition of ECG signals into separate components is obtained. The most informative frequency features are found, which provide an accuracy of class separation of more than 95%.

Keywords: Meyer wavelets; wavelet analysis; wavelets with scaling factor N ; high-frequency components; ECG classification.

For citation: Podkur, P.N., Smolentsev, N.R. (2023) Meyer wavelets with scaling factor $N = 12$ for classifying signals by high-frequency features. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naja tehnika i informatika* – Tomsk State University Journal of Control and Computer Science. 64. pp. 102–112. doi: 10.17223/19988605/64/10

Введение

Важной проблемой диагностики является проблема классификации сигналов по высокочастотным признакам. Такая проблема является актуальной в медицине при исследовании характеристик ЭКГ. Физиологическая и патологическая информация на электрокардиограмме (ЭКГ) имеет важное значение для диагностики заболеваний сердца. В настоящее время компьютерная диагностика сигналов ЭКГ привлекает все большее внимание исследователей. Автоматический анализ ЭКГ в основном включает шумоподавление сигнала, обнаружение волн и классификацию сердцебиения. Для изучения ЭКГ используются разнообразные статистические методы, разложение в ряд Фурье и анализ спектра. Более современные методы основаны на применении вейвлет-анализа, машинного обучения и нейронных сетей. Обзор и обширный список литературы можно найти в [1, 2]. В исследованиях обычно считается, что верхняя граничная частота нормального (без нагрузки) кардиосигнала, заметно влияющая на его форму, не превышает 100 Гц. Поэтому частоты ЭКГ выше 100 Гц считаются шумовыми, и их удаляют при помощи различных фильтров. Однако в настоящее время используются электрокардиографы высокого разрешения с частотой дискретизации до 20 кГц. Ясно, что при удалении «шумовых» высокочастотных колебаний ЭКГ теряется большая часть информации, зарегистрированная кардиографом высокого разрешения. Поэтому возникает вопрос о том, несет ли эта часть сигнала ЭКГ полезную информацию. Высокие частоты ЭКГ могут включать как аппаратные шумы, так и физиологические ритмы, которые являются следствием электрической активности сердца, поскольку они регистрируются датчиками, расположенными вблизи сердца. Эффективное сглаживание (шумоподавление) сигнала ЭКГ, а также выделение высокочастотных компонент возможно с использованием вейвлет-преобразования сигнала. Для каждой высокочастотной компоненты сигнала может быть вычислено много различных статистических, частотных и стохастических характеристик (признаков).

Актуальным является вопрос о том, какие из признаков несут наибольшую информацию и могут быть использованы в задаче о классификации сигналов ЭКГ. Если анализируемые данные состоят из нескольких классов (например, ЭКГ больных и здоровых пациентов), то нужно выбрать такое подпространство наиболее информативных признаков, которое эффективно с точки зрения разделимости классов. В работе [3] для уменьшения числа признаков и эффективного разделения классов использовалась процедура приведения признаков, основанная на методах статистической теории распознавания образов [4]. Однако методы, используемые в работе [3], не показывают, какие признаки являются наиболее значимыми. В данной работе для решения задачи классификации сигналов ЭКГ используется метод опорных векторов машинного обучения. Это позволяет самостоятельно выбирать и анализировать те признаки, которые имеют наибольшее значение для задач классификации и диагностики.

Использование степеней двойки для построения теории вейвлетов и при их использовании удобно во многих отношениях, хотя и не является обязательным. Можно вместо коэффициента масштабирования 2 использовать любое целое число N , большее единицы [5] (развитие этой темы см.: [6]). В работе [7] вейвлеты с коэффициентом $N = 3$ использованы для анализа ЭЭГ. Вейвлет-анализ с коэффициентом масштабирования $N > 2$ имеет определенные преимущества, обеспечивая разделение сигнала на N частотных диапазонов уже при однократном вейвлет-разложении. Это особенно актуально, когда для записи сигнала используются приборы высокого разрешения. Например, для записи электрокардиограмм в настоящее время используются электрокардиографы с частотой дискретизации до 20 кГц. Как известно, вейвлеты Мейера выделяются тем, что они обеспечивают наиболее четкое разделение сигнала по частотным диапазонам и имеют быстро убывающие фильтры [5, 6]. В работе [8] предложено построение и использование аналогов вейвлетов Мейера с произвольным коэффициентом масштабирования $N > 2$.

В данной работе построены ортогональные вейвлеты Мейера с коэффициентом масштабирования $N = 12$, и они применены для классификации сигналов ЭКГ только по высокочастотным признакам. Результаты показывают, что вейвлет-анализ с коэффициентом масштабирования больше двух позволяет получить новые информативные числовые характеристики ЭКГ, что подтверждает перспективность использования таких вейвлетов для анализа данных.

1. Предварительные сведения

Напомним основные положения ортогонального вейвлет-анализа с масштабным коэффициентом $N \geq 2$ (развитие этой темы см.: [6]). Функция $\varphi(x) \in L^2(\mathbf{R})$ называется масштабирующей с коэффициентом N , если она удовлетворяет соотношению

$$\varphi(x) = \sqrt{N} \sum_{n \in \mathbf{Z}} h_n^0 \varphi(Nx - n), \quad (1)$$

где набор действительных чисел $\{h_n^0\}$ называется *масштабирующим фильтром* функции $\varphi(x)$. Для такой функции $\varphi(x)$ определяется *частотная функция* формулой

$$H^0(\omega) = \frac{1}{\sqrt{N}} \sum_{n \in \mathbf{Z}} h_n^0 e^{-in\omega}. \quad (2)$$

Масштабирующей функции $\varphi(x)$ соответствует $N - 1$ вейвлетов $\psi^1(x), \dots, \psi^{N-1}(x)$, определенных равенствами

$$\psi^k(x) = \sqrt{N} \sum_{n \in \mathbf{Z}} h_n^k \varphi(Nx - n), \quad k = 1, 2, \dots, N - 1, \quad (3)$$

где коэффициенты разложения $\{h_n^k\}_{n \in \mathbf{Z}}$ называются *фильтрами вейвлетов*. В частотной области вейвлеты характеризуются своими частотными функциями

$$H^k(\omega) = \frac{1}{\sqrt{N}} \sum_{n \in \mathbf{Z}} h_n^k e^{-in\omega}, \quad k = 1, 2, \dots, N - 1. \quad (4)$$

В практических задачах мы имеем дело с цифровым сигналом, который представлен массивом $X = \{x_n\}$. Тогда его вейвлет-разложение $X \rightarrow \{A_1, D^1, D^2, \dots, D^{N-1}\}$ производится по формулам [6]

$$a_{1,m} = \sum_{n \in \mathbf{Z}} \overline{h_n} x_{n+Nm}, \quad d_{1,m}^k = \sum_{n \in \mathbf{Z}} \overline{h_n^k} x_{n+Nm}, \quad k = 1, 2, \dots, N - 1. \quad (5)$$

где $\{h_n\}_{n \in \mathbf{Z}}$ и $\{h_n^k\}_{n \in \mathbf{Z}}$ – фильтры вейвлетов $\varphi(x)$ и $\psi^k(x)$, а черта сверху означает комплексное сопряжение. Таким образом, вместо $X = \{x_n\}$ мы получили N массивов: $A_1 = \{a_{1,m}\}$ – коэффициенты аппроксимации, и $D^k = \{d_{1,m}^k\}$ – детализирующие коэффициенты, соответствующие вейвлетам $\psi^k(x)$. Процедуру разложения можно повторить, применив ее к набору коэффициентов A_1 . Восстановление массива $X = \{x_n\}$ по коэффициентам вейвлет-разложения $\{A_1, D^1, D^2, \dots, D^{N-1}\}$ производится следующим образом:

$$x_n = \sum_{m \in \mathbf{Z}} h_{n-Nm} a_{1,m} + \sum_{k=1}^{N-1} \sum_{m \in \mathbf{Z}} h_{n-Nm}^k d_{1,m}^k. \quad (6)$$

2. Построение частотных функций и вейвлетов Мейера

Масштабирующая функция Мейера $\varphi(x)$ определяется [5, 6] заданием ее преобразования Фурье равенством

$$\hat{\varphi}(\omega) = \begin{cases} 1, & \omega \in \left[-\frac{2\pi}{3}, \frac{2\pi}{3}\right], \\ \cos\left(\frac{\pi}{2} v\left(\frac{3}{2\pi} |\omega| - 1\right)\right), & \frac{2\pi}{3} \leq |\omega| \leq \frac{4\pi}{3}, \\ 0, & \text{для остальных } \omega, \end{cases} \quad (7)$$

где $v(x)$ есть вспомогательная функция, удовлетворяющая трем условиям: $v(x) = 0$ при $x \leq 0$, $v(x) = 1$ при $x \geq 1$ и $v(x) + v(1-x) = 1$. Мы выбираем функцию $v(x)$ в виде полиномиальной интерполяции $v(x) = x^4(35 - 84x + 70x^2 - 20x^3)$ между 0 и 1 на промежутке $[0, 1]$. Такой выбор принят в MATLAB.

Как известно, масштабирующая функция Мейера $\varphi(x)$ является N -масштабирующей для любого натурального $N > 2$ [8]. Частотная функция $H^0(\omega)$ для $\varphi(x)$ имеет вид:

$$H^0(\omega) = \begin{cases} 1, & \omega \in \left[-\frac{2\pi}{3N}, \frac{2\pi}{3N}\right], \\ \cos\left(\frac{\pi}{2} v\left(\frac{3N}{2\pi} |\omega| - 1\right)\right), & \frac{2\pi}{3N} \leq |\omega| \leq \frac{4\pi}{3N}, \\ 0, & \text{для остальных } \omega, \end{cases} \quad (8)$$

Вне промежутка $[-\pi, \pi]$ функция $H^0(\omega)$ продолжается периодически. Коэффициенты $\{h_n^0\}$ N -масштабирующего соотношения (1) находятся из разложения в ряд Фурье функции $H^0(\omega)$.

Для построения вейвлетов Мейера с нечетным коэффициентом масштабирования N сначала находятся их частотные функции $H^k(\omega)$. При $\omega \geq 0$ функция $H^k(\omega)$, $k = 1, 2, \dots, N-2$, имеет вид [8]:

$$H_+^k(\omega) = \begin{cases} 0 & \omega < \frac{k\pi}{N} - \frac{\pi}{3N} \text{ или } \omega \geq \frac{(k+1)\pi}{N} + \frac{\pi}{3N}, \\ \sin\left(\frac{\pi}{2} v\left(\frac{3N}{2\pi} \omega - \frac{3k-1}{2}\right)\right) & \frac{k\pi}{N} - \frac{\pi}{3N} \leq \omega < \frac{k\pi}{N} + \frac{\pi}{3N}, \\ 1 & \frac{k\pi}{N} + \frac{\pi}{3N} \leq \omega < \frac{(k+1)\pi}{N} - \frac{\pi}{3N}, \\ \cos\left(\frac{\pi}{2} v\left(\frac{3N}{2\pi} \omega - \frac{3(k+1)-1}{2}\right)\right) & \frac{(k+1)\pi}{N} - \frac{\pi}{3N} \leq \omega < \frac{(k+1)\pi}{N} + \frac{\pi}{3N}. \end{cases} \quad (9)$$

На весь промежуток $[-\pi, \pi]$ функции $H_+^k(\omega)$ продолжают по четности или по нечетности в зависимости от четности k . Вне промежутка $[-\pi, \pi]$ функции $H_+^k(\omega)$ продолжают по периодичности. Последняя частотная функция $H^{N-1}_+(\omega)$ на $[0, \pi]$ определяется формулой

$$H_+^{N-1}(\omega) = \begin{cases} \sin\left(\frac{\pi}{2} v\left(\frac{3N}{2\pi} \omega - \frac{3(N-1)-1}{2}\right)\right), & \omega \in \left[\frac{(N-1)\pi}{N} - \frac{\pi}{3N}, \frac{(N-1)\pi}{N} + \frac{\pi}{3N}\right], \\ 1 & \omega \in \left[\frac{(N-1)\pi}{N} + \frac{\pi}{3N}, \pi\right], \\ 0 & \text{для остальных } \omega \in [0, \pi]. \end{cases} \quad (10)$$

Таким образом, искомые частотные функции вейвлетов $\psi^1(x)$, $\psi^2(x)$, ..., $\psi^{N-1}(x)$ имеют вид: для четных значений k функция $H^k(\omega)$ является четной и на промежутке $[0, \pi]$ определяется формулой (9), а для нечетных значений k функция $H^k(\omega)$ является нечетной и на промежутке $[0, \pi]$ определяется формулой (9). Соответственно определяется и последняя частотная функция $H^{N-1}(\omega)$.

Масштабирующий фильтр $\{h_n^0\}$ и фильтры вейвлетов $\{h_n^k\}$, $k = 1, 2, \dots, N-1$, находятся разложением в ряд Фурье частотных функций $H^0(\omega)$ и $H^k(\omega)$. Вейвлеты $\psi^k(x)$ определяются своими фильтрами $\{h_n^k\}$ по формуле (4).

В случае четного $N > 2$ можно представить число N в виде $N = N_0 \cdot 2^m$, где N_0 – нечетное, и использовать кратные коэффициенты масштабирования и классические вейвлеты Мейера с коэффициентом масштабирования 2. В частности, в данной работе мы будем использовать вейвлеты Мейера с коэффициентом $12 = 3 \cdot 2^2$.

3. Вейвлеты Мейера с коэффициентом масштабирования 12

Пусть $G^k(\omega)$, $k = 0, 1, \dots, N_0 - 1$ – частотные функции вейвлетов Мейера с нечетным N_0 и $H^l(\omega)$, $l = 0, 1, \dots, M - 1$ – частотные функции вейвлетов Мейера с четным $M = 2^m$.

Если

$$G^k(\omega) = \frac{1}{\sqrt{N_0}} \sum_{n \in \mathbb{Z}} g_n^k e^{-in\omega}, \quad k = 1, 2, \dots, N_0 - 1, \text{ и } H^l(\omega) = \frac{1}{\sqrt{M}} \sum_{n \in \mathbb{Z}} h_n^l e^{-in\omega}, \quad l = 0, 1, 2, \dots, M - 1,$$

тогда, в соответствии с результатами [8], в качестве частотных функций вейвлетов $\psi^{kl}(x)$, $l = 0, 1, \dots, M - 1$, $k = 0, 1, \dots, N_0 - 1$, с коэффициентом $N_0 M$ возьмем функции

$$H^{kl}(\omega) = G^k(M\omega)H^l(\omega).$$

Масштабирующие фильтры соответствующих вейвлетов $\psi^{kl}(x)$ находятся по формулам [6]

$$h_n^{kl} = \sum_{m \in \mathbb{Z}} g_m^k h_{n-mM}^l, \quad l = 0, 1, \dots, M - 1, \quad k = 0, 1, \dots, N_0 - 1.$$

Рассмотрим построение вейвлетов Мейера с кратным коэффициентом масштабирования, когда $M = 2^m = 4$ и $N_0 = 3$. Масштабирующая функция будет одна и та же $\varphi(x)$, указанная в формуле (8).

В случае $N_0 = 3$ масштабирующий фильтр $\{g_n^0\}$ находится из разложения в ряд Фурье функции $G^0(\omega)$, определенной по формуле (8) при $N_0 = 3$. Фильтры вейвлетов $\{g_n^1\}, \{g_n^2\}$ находятся разложением в ряд Фурье частотных функций $G^1(\omega)$ и $G^2(\omega)$, определенных на основе формул (9) и (10).

Построим Вейвлеты Мейера с коэффициентом масштабирования $M = 4$. Сначала напомним вейвлеты Мейера для коэффициента $N = 2$ [5, 6]. Пусть $H^0(\omega) = \frac{1}{\sqrt{2}} \sum_{n \in \mathbb{Z}} h_n^0 e^{-in\omega}$ – частотная функция

вейвлета Мейера $\varphi(x)$ для $N = 2$. Тогда вейвлет $\psi(x)$ находится из формулы $\hat{\psi}(\omega) = e^{i\omega/2} \overline{H^0\left(\frac{\omega}{2} + \pi\right)} \hat{\varphi}\left(\frac{\omega}{2}\right)$.

Соответствующая частотная функция имеет вид: $H^1(\omega) = e^{i\omega} \overline{H^0(\omega + \pi)}$, и ее фильтр $\{h_n^1\}$ находится по формуле $h_n^1 = (-1)^{n+1} h_{-n-1}^0$.

Частотные функции вейвлетов $\psi^{pq}(x)$, $p = 0, 1$, $q = 0, 1$, с коэффициентом масштабирования 4 определим по формуле $H^{pq}(\omega) = H^p(2\omega)H^q(\omega)$. Здесь $\psi^{00}(x) = \varphi(x)$ – масштабирующая функция Мейера. Фильтры вейвлетов получаем по формуле $h_n^{pq} = \sum_{m \in \mathbb{Z}} h_m^p h_{n-2m}^q$, $p = 0, 1$, $q = 0, 1$.

Частотные функции вейвлетов $\psi^{kpq}(x)$, $k = 0, 1, 2$, $p = 0, 1$, $q = 0, 1$, с коэффициентом масштабирования 12 определим по формуле

$$H^{kpq}(\omega) = G^k(4\omega)H^p(2\omega)H^q(\omega).$$

Все эти операции и формулы (8)–(10) легко реализовать в системе Maple. Все фильтры вейвлетов для $N = 12$ находятся непосредственным численным интегрированием полученных частотных функций также в системе Maple. Данные фильтры имеют бесконечную импульсную характеристику, поэтому для практического использования мы усекаем их, вычисляя 401 элемент:

$$H^{kpq}(\omega) \cong \frac{1}{\sqrt{12}} \sum_{n=-200}^{200} h_n^{kpq} e^{-in\omega}, \quad k = 0, 1, 2, \quad p, q = 0, 1.$$

Отбрасываемые члены имеют порядок не более 10^{-6} , что обеспечивает необходимую для нас точность прямого и обратного вейвлет-преобразования и точность вычисления признаков ЭКГ

4. Вейвлет-анализ сигнала ЭЭГ

Основная операция дискретного вейвлет-преобразования с коэффициентом масштабирования N представляет собой разложение изучаемого сигнала $X = \{x_n\}$ на N наборов $X \rightarrow \{A_1, D^1, D^2, \dots, D^{N-1}\}$ массивов при помощи фильтров вейвлетов по формулам (5)

$$a_{1,m} = \sum_{n \in \mathbb{Z}} \overline{h_n} x_{n+Nm}, \quad d_{1,m}^k = \sum_{n \in \mathbb{Z}} \overline{h_n^k} x_{n+Nm}, \quad k = 1, 2, \dots, N - 1,$$

где $A_1 = a_{1,m}$ – коэффициенты аппроксимации первого уровня разложения, и $D^k = \{d_{1,m}^k\}$ – детализирующие вейвлет-коэффициенты. При многоуровневом вейвлет-анализе процедура вейвлет-разложения

применяется многократно к массивам коэффициентов аппроксимации. Восстановление первоначального сигнала производится последовательно в обратном порядке. Если мы применим процедуру восстановления только к одному набору коэффициентов D^k , когда все остальные коэффициенты состоят из нулей, то получим часть сигнала, соответствующую одному набору коэффициентов. Будем называть такую часть *компонентой* сигнала и обозначать ее символом $RecD^k$.

Сумма всех компонент сигнала $RecA_1, RecD^1, RecD^2, \dots, RecD^{N-1}$ и равна исходному сигналу X , и мы получаем разложение X на высокочастотные компоненты и сглаженную часть $RecA_1$:

$$X = RecA_1 + RecD^1 + \dots + RecD^{N-1}.$$

Будем считать, что все компоненты здесь упорядочены по частоте, т.е. $RecA_1$ – это сглаженный сигнал, самая низкочастотная компонента, а $RecD^{N-1}$ – самая высокочастотная часть (рис. 1).

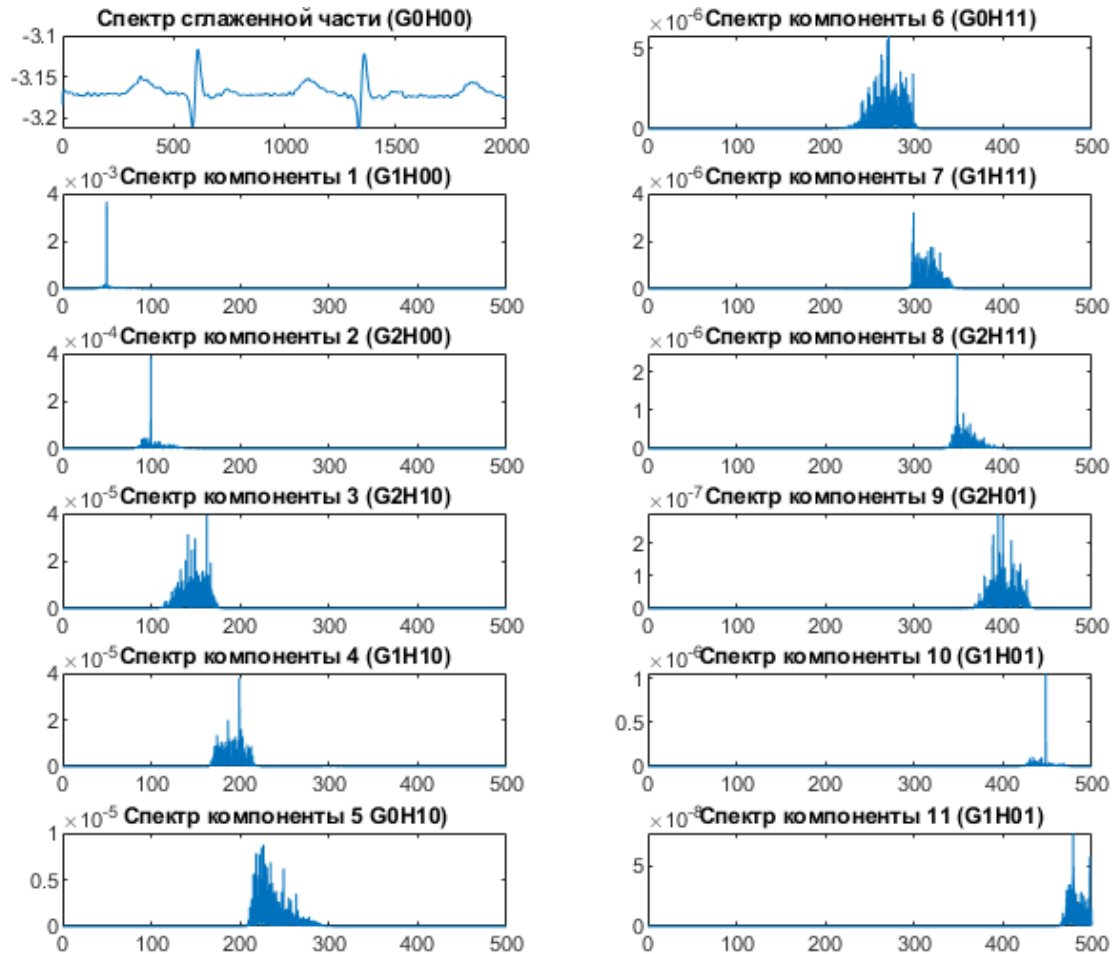


Рис. 1. График сглаженного сигнала ЭКГ и графики спектров мощности компонент сигнала ЭКГ
(по горизонтали отсчеты в Гц)

Fig. 1. Graph of the smoothed ECG signal and graphs of power spectra of the ECG signal components
(Frequency scale in Hz along the horizontal axis)

Описание признаков. Для каждой высокочастотной компоненты разложения $RecD^k$ может быть вычислено много различных статистических, частотных и стохастических характеристик: индекс Хёрста; показатели мультифрактальности; энтропия коэффициентов разложения и частотного спектра; основные частоты, их средние значения. Напомним определения перечисленных параметров.

Параметр Хёрста H сигнала $X = \{x_n\}$ представляет собой меру склонности процесса к трендам и показывает наличие в сигнале дробного броуновского движения с параметром Хёрста H . Дробное броуновское движение $x(t)$ представляет собой непрерывный гауссовский процесс, зависящий от параметра Хёрста $0 < H < 1$, когда дисперсия D приращений определяется выражением $D[x(t) - x(s)] =$

$= v|t-s|^{2H}$, где v – положительная константа. В MATLAB параметр Хёрста вычисляет функция `wfbmest1`. Она дает три разные оценки параметра Хёрста H .

Мультифрактальный анализ стал широко использоваться для обработки медико-биологических процессов после публикации в 1999 г. в журнале “Nature” работы по мультифрактальному описанию сердечного ритма [9]. Напомним основные понятия [10]. Регулярность по Гёльдеру является наиболее широко используемым понятием поточечной регулярности. Пусть $f(x)$ – локально ограниченная функция, $\alpha \geq 0$ и x_0 – некоторая точка. Будем говорить, что $f \in C^\alpha(x_0)$, если существуют константа $C > 0$ и многочлен (Тейлора) P степени $\deg(P) < \alpha$, такие что

$$|f(x) - P(x - x_0)| \leq C|x - x_0|^\alpha$$

в окрестности точки x_0 . Показатель Гёльдера функции $f(x)$ в точке x_0 равен

$$H_f(x_0) = \sup\{\alpha : f \in C^\alpha(x_0)\}.$$

Если $H_f(x_0) < 1$, то функция недифференцируема в точке x_0 и $P(x - x_0) = f(x_0)$. В этом случае показатель степени описывает негладкость функции в точке. Мультифрактальный анализ становится актуальным, когда множество точек с одинаковыми показателями Гёльдера достаточно большое. Обозначим $E_f(H) = \{x : H_f(x) = H\}$ множество точек, обладающих одинаковой регулярностью H по Гёльдеру. Такие множества образуют спектр фрактальности. Обозначим $d_f(H)$ хаусдорфову размерность множества $E_f(H)$. Если $E_f(H) \neq \emptyset$, то $0 \leq d_f(H) \leq 1$. Спектр является фрактальным, если $0 < d_f(H) < 1$ для некоторого H , и мультифрактальным, если то же самое верно для разных значений H . Значения $d_f(H)$ для разных значений H образуют спектр сингулярности. Таким образом, спектр дает нам характеристику глобального распределения показателей Гёльдера и фрактальных свойств исследуемых сигналов.

Функция MATLAB `[dH, H] = dwtleader(x)` вычисляет спектр сингулярности, dH , и показатели Гёльдера H для одномерных вещественных данных x . Спектр сингулярности и показатели Гёльдера возвращаются этой функцией как векторы из 11 элементов.

Стандартное отклонение сигнала $X = \{x_n\}$ вычисляет функция `std` MATLAB.

Энтропия Шеннона (вычисляется функцией `wentropy`) характеризует разброс значений сигнала $X = \{x_n\}$ и определяется формулой $E(X) = -\sum_i x_i^2 \ln(x_i^2)$.

Дискретное преобразование Фурье сигнала $X = \{x_n\}$ вычисляется функцией MATLAB `C = fft(X)`. Поскольку значения $\{c_m\}$ могут быть комплексными, представляет интерес спектр мощности частот (частотный спектр), вычисляемый по формуле

$$P_m = \frac{|c_m|^2}{L}, \quad m = 0, 1, \dots,$$

где L – длина массива. Вследствие симметричности график спектра мощности P_m обычно изображают для значений m в диапазоне от 0 до середины $L/2$.

Как видно на рис. 1, частотные спектры компонент сигнала ЭКГ достаточно хорошо локализованы при использовании вейвлетов Мейера и у них имеются явно выделяющиеся «основные» частоты. Мы рассматриваем как основные 10 частот с наибольшей мощностью. Среднее арифметическое этого набора частот компоненты $RecD^k$ обозначим $FrMean^k$.

Пусть $Range^k$ – набор основных (десяти) частот. Длину частотного промежутка, где расположены основные частоты компоненты $RecD^k$, будем обозначать символом $LRange^k$.

Определим также взвешенное среднее значение $FrPMean^k$ основных частот по формуле

$$FrPmean^k = \sum_{s \in Range^k} s \cdot p_s, \quad p_s = \frac{P_s}{\sum_{i \in Range^k} P_i}, \quad k = 0, 1, \dots, N-1$$

Вычисление признаков. Кардиосигнал регистрировался по восьми общепринятым каналам: L – левая рука (+) и правая рука (–), F – левая нога (+) и правая нога (–), и шесть грудных отведений, обозначаемых $C1$ – $C6$. Из данных восьми каналов L , F , $C1$, $C2$, $C3$, $C4$, $C5$, $C6$ кардиографа образуются 12 так называемых стандартных отведений: I, II, III, aVR, aVL, aVF, V1, V2, V3, V4, V5, V6 по

формулам: $I = L$, $II = F$, $III = F - L$, $aVR = -(L + F)/2$, $aVL = L - F/2$, $aVF = F - L/2$, $Vi = Ci - (L + F)/3$, $i = 1, 2, \dots, 6$.

Для разложения сигнала и восстановления его компонент созданы m -функции `dmeу12.m` и `idmeу12.m` в MATLAB в соответствии с формулами (5) и (6). Фильтры разложения и восстановления занумерованы по возрастанию частоты (см. рис. 1). Для каждого из указанных выше 12 отведений произведено вейвлет-разложение с использованием вейвлетов Мейера с коэффициентом масштабирования 12. В качестве высокочастотных взяты компоненты сигнала, соответствующие коэффициентам $D^1, D^2, \dots, D^{11}$ (см. рис. 1). Компонента $RecA_1$ соответствует сглаженной части сигнала и поэтому не будет рассматриваться, поскольку нас в данной работе интересуют только высокочастотные характеристики сигнала ЭКГ.

Для каждой высокочастотной компоненты $RecD^k$ разложения каждого канала вычислены следующие характеристики:

- индекс Хёрста исходного сигнала;
- показатели мультифрактальности исходного сигнала;
- стандартное отклонение коэффициентов разложения;
- энтропия коэффициентов разложения;
- энтропия частотного спектра компоненты;
- стандартное отклонение частотного спектра компоненты;
- средняя частота компоненты;
- средняя взвешенная частота компоненты;
- длина основного диапазона частот компоненты.

Были проанализированы ЭКГ трех групп:

- 1) ЭКГ 12 пациентов возраста от 22 до 40 лет, которые не имели жалоб на проблемы с сердцем;
- 2) ЭКГ 9 пациентов возраста от 43 до 70 лет, имеющих различные заболевания сердца (гипертрофия миокарда, рубцы з/нижней стенки VS, изменения миокарда п/перегородочной области, блокада передне-верхней ветви левой ножки пучка Гиса);
- 3) ЭКГ 13 пациентов возраста от 44 до 66 лет, недавно перенесших инфаркт миокарда (ИМ) и находящихся на лечении.

При классификации указанным группам присвоены номера классов 1, 2 и 3 соответственно. Для каждого пациента были использованы 24 записи ЭКГ длительностью по 8 с (по две 12-канальных записи), снятые на кардиографе высокого разрешения (1 028 отсчетов в секунду) «Кардиотехника – 4000, EcgShell».

В результате для каждой записи ЭКГ получается 102 значения признаков. Для классификации по группам пациентов результаты вычислений сформированы в массивы признаков и сохранены функцией `writetable` как `Data_12`, `Data_13`, `Data_23` в формате CSV и в соответствии с тем, какие группы пациентов мы хотим различить. Например, массив `Data_13` содержит признаки ЭКГ здоровых пациентов (группа 1) и перенесших инфаркт миокарда (группа 3). Его размеры 600×102 . Это числовые массивы, имеющие поля, обозначающие признаки. Первый столбец имеет поле '**Class**' и содержит числовые значения 1, 2 или 3 классов наших пациентов. Следующие столбцы содержат значения признаков ЭКГ. Например, три следующих столбца имеют поля **Hest_1**, **Hest_2**, **Hest_3** и содержат показатели трех индексов Хёрста. Для классификации можно использовать все перечисленные числовые признаки или выбрать некоторую их часть и анализировать их влияние на разделение классов.

Каждый массив признаков `Data_12`, `Data_13`, `Data_23` разбивается на два: один для обучения, другой для тестирования. Для тестирования случайным образом выбирается примерно 25% от каждой группы пациентов. Например, массив `Data_23` разбивается на `Train_23.csv` и `Test_23.csv` размерами 456×102 и 144×102 соответственно.

Классификация. Для классификации используется метод опорных векторов (SVM) пакета `scikit-learn` для машинного обучения в Python. Программа совершенно простая и легко позволяет выбирать отдельные признаки и проверять, насколько они хороши для разделения классов.

```

import pandas as pd
from sklearn import svm
predmodel=svm.LinearSVC()
train = pd.read_csv("Train_23.csv")
test = pd.read_csv("Test_23.csv")
columns_target=['Class']
columns_train=[FrP_Mean_6, FrP_Mean_7, HEST_1, Sing_3] #0,958
train_y=train[columns_target]
train_x=train[columns_train]
test_y=test[columns_target]
test_x=test[columns_train]
predmodel.fit(X_train,Y_train)
predmodel.score(X_test,Y_test)

```

Последовательно выбирались все признаки вместе, по отдельности и в комбинациях с другими признаками. В таблице представлены результаты проверки признаков на информативность для двух классов: здоровых (класс 1) и имеющих нарушения работы сердца (класс 2). Таблица состоит из трех частей. Первая часть соответствует вейвлет-разложению с коэффициентом $N = 12$. Каждая часть содержит три столбца. В первом столбце представлены результаты разделимости классов, когда выбираются все значения одного признака. Во втором столбце указано максимальное значение информативности одного признака, а в третьем – номер признака из его группы, при котором достигается это максимальное значение. Напомним, что мы имеем три значения индекса Хёрста, 22 показателя мультифрактальности (11 значений спектра сингулярности и 11 показателей Гёльдера) и по 11 высокочастотных признаков компонентов вейвлет-разложения с коэффициентом $N = 12$. Результаты вычислений показывают, что наиболее информативными являются частотные признаки для компоненты 6. Они обеспечивают точность разделения классов более 95%.

Показатели признаков разделимости групп 1 и 2

Признак	С коэффициентом 12			Пакетное, уровня 3			Пакетное, уровня 4		
	Все значения	Одно (max)	Номер признака	Все значения	Одно (max)	Номер признака	Все значения	Одно (max)	Номер признака
Индекс Хёрста	0,658	0,658	3	0,658	0,658	3	0,658	0,658	3
Спектр сингулярности	0,683	0,7	7, 8	0,692	0,683	7, 8	0,692	0,683	7, 8
Показатель Гельдера	0,758	0,7	2, 3, 4	0,742	0,7	2, 3, 4	0,742	0,7	2, 3, 4
Энтропия коэффициентов	0,575	$\leq 0,6$	все	0,667	$\leq 0,6$	все	0,783	$\leq 0,6$	все
Стандартное отклонение коэффициентов	0,591	$\leq 0,6$	все	0,742	$\leq 0,6$	все	0,81	0,65	15
Энтропия частотного спектра	0,767	0,658	2	0,633	$\leq 0,6$	все	0,4	0,692	2
Стандартное отклонение частотного спектра	0,767	$\leq 0,6$	все	0,608	$\leq 0,6$	все	0,917	0,733	2
Средняя основная частота	0,9	0,817	6	0,917	0,842	2	0,9	0,683	3
Средняя взвешенная частота	0,958	0,908	6	0,958	0,933	6	0,867	0,808	2
Длина диапазона основных частот	0,742	0,708	5	0,75	0,742	1	0,65	0,716	3, 4

Отметим результаты, получающиеся комбинациями признаков, которые показывают точность разделения классов 1 и 2 до 95,8%. Это, например, следующие группы признаков:

```

[FrP_Mean_6, FrP_Mean_7, HEST_1, Sing_3],
[FrP_Mean_7, FrP_Mean_6, HEST_1, Sing_3', Fr_Range_5],
[FrP_Mean_6, FrP_Mean_7, Fr_Mean_7, HEST_1, Sing_3, Fr_Range_5].

```

Для других пар классов пациентов результаты аналогичные. Для классов здоровых (класс 1) и пациентов, перенесших ИМ и находящихся на лечении (класс 3) наиболее оптимальным являются следующие наборы признаков, обеспечивающих точность 88,2%:

```

[FrP_Mean_6, HEST_1], [FrP_Mean_6, Fr_Mean_6, HEST_1].

```

Для классов пациентов, имеющих нарушения в работе сердца (класс 2) и пациентов, перенесших ИМ (класс 3), наиболее оптимальным является следующий набор признаков:

[FrP_Mean_4, FrP_Mean_6, FrP_Mean_7, HEST_1, Sing_3],

и он обеспечивает точность 79,2%.

Для сравнения проведены вычисление признаков и исследование их значений на разделимость групп пациентов в случае пакетного [6] разложения сигнала, т.е. при разложении в полное дерево глубины 3 и глубины 4.

Вычисление признаков при пакетном разложении 3-го уровня. Рассматриваются высокочастотные признаки компонент сигнала $RecD^k$, $k = 1, 2, \dots, 7$, соответствующих коэффициентам пакетного разложения [5, 6] в терминальных узлах дерева разложения (3,1), (3,2), (3,3), (3,4), (3,5), (3,6) и (3,7). Компонента, соответствующая узлу (3,0), представляет сглаженный сигнал, и мы ее не учитываем при вычислении признаков. Остается 7 высокочастотных компонент для каждого канала ЭКГ и 74 признака для каждого канала ЭКГ. Результаты приведены в таблице.

Наиболее информативными являются частотные признаки компонент, соответствующих узлам (3,2) и (3,6). Следующая комбинация ['FrP_Mean_2', 'FrP_Mean_6'] показывает точность 96,7% разделения классов 1 и 2.

Вычисление признаков при пакетном разложении 4-го уровня. Рассматриваются высокочастотные признаки $RecD^k$, $k = 1, 2, \dots, 15$, соответствующие коэффициентам разложения в терминальных узлах дерева разложения (4,0), (1,1), ..., (4,15). Компонента, соответствующая узлу (4,0), представляет сглаженный сигнал, и мы ее не учитываем при вычислении признаков. Остается 15 высокочастотных компонент для каждого канала ЭКГ и 130 признаков. Результаты приведены в таблице. Наиболее информативными являются частотные признаки компонент, соответствующих узлам (4,2) и (4,3).

Заключение

В данной работе построены вейвлеты Мейера с коэффициентом масштабирования $N = 12$ и показано их применение в задаче классификации сигналов. Использование кратных коэффициентов масштабирования $N > 12$ дает большую гибкость при выборе количества компонент разложения сигнала по сравнению с пакетными разложениями. В работе показана возможность классификации электрокардиограмм здоровых и больных пациентов только по высокочастотным компонентам сигнала ЭКГ. Найдены признаки высокочастотных компонент ЭКГ, наиболее информативные для задач классификации и диагностики. К таким признакам относятся значения основных частот компонент сигнала ЭКГ. Они обеспечивают точность классификации более 95% для групп здоровых и больных пациентов. Самым информативным признаком является средняя основная частота диапазона ЭКГ от 250 до 300 Гц. Данная частота для группы здоровых пациентов составляет в среднем 268,2 Гц, для пациентов с нарушениями работы сердца – 282,3 Гц, а для пациентов, перенесших инфаркт миокарда и находящихся на лечении, – 277,7 Гц. Таким образом, данные частоты и другие информативные признаки имеют диагностическое значение и могут быть использованы при классификации и диагностике как новые дополнительные характеристики ЭКГ.

Список источников

1. Li W. Wavelets for Electrocardiogram: Overview and Taxonomy // IEEE. 2019. V. 7. P. 25627–25649.
2. Павлов А.Н., Храмов А.Е., Короновский А.А., Ситникова Е.Ю., Макаров В.А., Овчинников А.А. Вейвлет-анализ в нейродинамике // Успехи физических наук. 2012. Т. 182, № 9. С. 905–939.
3. Podkur P.N., Smolentsev N.K. About classification of ECG signals based on high-frequency wavelet-components // Science Evolution. 2016. V. 1, № 1. P. 63–71.
4. Фукунага К. Введение в статистическую теорию распознавания образов. М. : Наука, 1979. 368 с.
5. Добеши И. Десять лекций по вейвлетам. М. ; Ижевск : РХД, 2001. 464 с.
6. Смоленцев Н.К. Основы теории вейвлетов. Вейвлеты в MATLAB. М. : ДМК Пресс, 2019. 628 с.
7. Подкур П.Н., Смоленцев Н.К. Вейвлет-анализ ЭЭГ с коэффициентом масштабирования 3 // Южно-Сибирский научный вестник. 2020. Вып. 6 (34). С. 56–61.

8. Подкур П.Н., Смоленцев Н.К. Вейвлеты Мейера с кратными коэффициентами масштабирования $N > 2$ // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2022. № 60. С. 84–92.
9. Ivanov P.Ch. et al, Multifractality in human heartbeat dynamics // Nature. 1999. V. 399. P. 461–465.
10. Figliola A. et al. About the Effectiveness of Different Methods for the Estimation of the Multifractal Spectrum of Natural Series // International Journal of Bifurcation and Chaos. 2010. V. 20, № 2. P. 331–339.

References

1. Li, W. (2019) Wavelets for Electrocardiogram: Overview and Taxonomy. *IEEE*. 7. pp. 25627–25649. DOI: 10.1109/ACCESS.2018.2877793
2. Pavlov, A.N., Khramov, A.E., Koronovsky, A.A., Sitnikova, E.Yu., Makarov, V.A. & Ovchinnikov, A.A. (2012) Wavelet analysis in neurodynamics. *Uspekhi fizicheskikh nauk*. 182(9). pp. 905–939.
3. Podkur, P.N. & Smolentsev, N.K. (2016) About classification of ECG signals based on high-frequency wavelet-components. *Science Evolution*. 1(1). pp. 63–71.
4. Fukunaga, K. (1972) *Introduction to Statistical Pattern Recognition*. New York: Academic Press.
5. Daubechies, I. (1992) *Ten Lectures on Wavelets*. SIAM.
6. Smolentsev, N.K. (2019) *Osnovy teorii veyvletov. Veyvlety v MATLAB* [Fundamentals of the theory of wavelets. Wavelets in MATLAB]. Moscow: DMK Press.
7. Podkur P.N. & Smolentsev N.K. (2020) Wavelet analysis of EEG with scaling factor 3. *Yuzhno-Sibirskiy nauchnyy vestnik – South-Siberian Scientific Bulletin*. 6(34). pp. 56–61. DOI: 10.25699/T7719-3385-0014-J
8. Podkur, P.N. & Smolentsev, N.R. (2022) Meyer wavelets with multiple scale factors $N > 2$. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 60. pp. 84–92. DOI: 10.17223/19988605/60/9
9. Ivanov, P.Ch. et al (1999) Multifractality in human heartbeat dynamics. *Nature*. 399. pp. 461–465.
10. Figliola, A. et al (2010) About the effectiveness of different methods for the estimation of the multifractal spectrum of natural series. *International Journal of Bifurcation and Chaos*. 20(2). pp. 331–339. DOI: 10.1142/S0218127410025788

Информация об авторах:

Подкур Полина Николаевна – кандидат физико-математических наук, доцент Кузбасского технического университета им. Т.Ф. Горбачева (Кемерово, Россия). E-mail: paulina.podkur@gmail.com

Смоленцев Николай Константинович – доктор физико-математических наук, профессор Кемеровского государственного университета (Кемерово, Россия). E-mail: smolennk@mail.ru

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

Information about the authors:

Podkur Paulina N. (Candidate of Physical and Mathematical Sciences, Associate Professor, T.F. Gorbachev Kuzbass State Technical University, Kemerovo, Russian Federation). E-mail: paulina.podkur@gmail.com

Smolentsev Nikolay K. (Doctor of Physical and Mathematical Sciences, Professor, Kemerovo State University, Kemerovo, Russian Federation). E-mail: smolennk@mail.ru

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

Поступила в редакцию 16.05.2023; принята к публикации 04.09.2023

Received 16.05.2023; accepted for publication 04.09.2023

Научная статья

УДК 681.5

doi: 10.17223/19988605/64/11

Робастная фильтрация для дискретных систем с неизвестным входом и интервальной неопределенностью

Валерий Иванович Смагин¹, Константин Станиславович Ким²

^{1,2} Томский государственный университет, Томск, Россия

¹ vsm@mail.tsu.ru

² kks93@rambler.ru

Аннотация. Рассмотрена задача робастной фильтрации для дискретного объекта с неизвестным входом и интервальными параметрами в модели объекта. Используется вероятностный подход, в основе которого лежат замена неопределенных параметров интервального типа независимыми случайными величинами с равномерным распределением. Также используются алгоритмы оценивания неизвестного входа, рекуррентные схемы калмановской фильтрации, метод наименьших квадратов и сглаживающие непараметрические процедуры. Представлены результаты моделирования.

Ключевые слова: робастная фильтрация; интервальные параметры; неизвестный вход; метод наименьших квадратов; непараметрическое сглаживание.

Для цитирования: Смагин В.И., Ким К.С. Робастная фильтрация для систем с неизвестным входом и интервальной неопределенностью // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2023. № 64. С. 113–119. doi: 10.17223/19988605/64/11

Original article

doi: 10.17223/19988605/64/11

Robust filtering in discrete systems with unknown input and interval uncertainty

Valery I. Smagin¹, Konstantin S. Kim²

^{1,2} Tomsk State University, Tomsk, Russian Federation

¹ vsm@mail.tsu.ru

² kks93@rambler.ru

Abstract. The problem of robust filtering for discrete systems with unknown input and interval parameters in the model is considered. A probabilistic approach is used, which is based on the replacement of indeterminate interval-type parameters with independent random variables with a uniform distribution. Unknown input estimation algorithms, recurrent Kalman filtering schemes, the least squares method and smoothing nonparametric procedures are also used. Simulation results are presented.

Keywords: robust filtering; interval parameters; unknown input; least squares method, nonparametric smoothing.

For citation: Smagin, V.I., Kim, K.S. (2023) Robust filtering in discrete systems with unknown input and interval uncertainty. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naja tehnika i informatika – Tomsk State University Journal of Control and Computer Science*. 64. pp. 113–119. doi: 10.17223/19988605/64/11

Введение

Задача синтеза фильтров, экстраполяторов и наблюдателей для динамических систем с неопределенными параметрами, в частности с интервальными параметрами, рассматривалась в [1–5] и др. В этих работах используются различные методы робастной обработки данных. Методы обработки

данных с использованием оценок неизвестного входа приведены в [6–9], где метод наименьших квадратов (МНК) использовался для получения оценок неизвестных входных данных. В [10–12] было предложено использовать компенсационный подход для расчета оценок неизвестного входа. В [13, 14] дополнительно использовались алгоритмы непараметрического сглаживания для повышения точности оценивания неизвестного входа.

В работе рассматривается задача робастной фильтрации в дискретных системах с аддитивными возмущениями при неизвестном входе и интервальных параметрах. В его основе лежит вероятностный подход к решению задач для модели с интервальными параметрами, заключающийся в том, что интервальный параметр заменяется равномерно распределенной случайной величиной [15]. Отметим, что применение этого подхода к задаче робастной экстраполяции рассмотрено в работах [16, 17].

1. Постановка задачи

Рассмотрим линейную дискретную систему с интервальными параметрами, описываемую разностным уравнением

$$x(k+1) = \tilde{A}x(k) + f(k) + \tilde{B}q(k), \quad x(0) = x_0, \quad (1)$$

где $x(k) \in \mathbb{R}^n$ – вектор состояния в момент времени k , $f(k)$ – неизвестный входной вектор; x_0 – случайный вектор с известным математическим ожиданием и ковариационной матрицей $N_0 = M\{(x_0 - \bar{x}_0)(x_0 - \bar{x}_0)^T\}$; $\tilde{A}$ – матрица переходов состояний с интервальной неопределенностью (с заданными нижними и верхними границами $\underline{A}$ и $\bar{A}$ соответственно), $q(k) \in \mathbb{R}^{n_1}$ – случайные возмущения со следующими характеристиками: $M\{q(k)\} = 0$, $M\{q(k)q^T(j)\} = I\delta_{kj}$, $\tilde{B}$ – матрица переходов возмущений с интервальной неопределенностью (с заданными нижними и верхними границами $\underline{B}$ и $\bar{B}$), I – единичная матрица, δ_{kj} – символ Кронекера.

Модель наблюдений определяется по формуле

$$y(k) = Sx(k) + v(k), \quad (2)$$

где $y(k) \in \mathbb{R}^l$ – вектор наблюдений, S – матрица канала наблюдений, $v(k) \in \mathbb{R}^l$ – шум наблюдений ($M\{v(k)\} = 0$, $M\{v(k)v^T(j)\} = V\delta_{kj}$).

Предполагается, что последовательности $q(k)$, $v(k)$ и x_0 взаимно независимы, система (1) наблюдаема при параметрических возмущениях матрицы динамики $\tilde{A}$. Используя информацию, доступную к моменту времени $k \in [0; T]$, необходимо построить оценку $\hat{x}(k+1)$ путем минимизации следующего критерия:

$$J(0;T) = M\left\{\sum_{k=0}^T (x(k) - \hat{x}(k))^T (x(k) - \hat{x}(k))\right\}. \quad (3)$$

2. Робастная фильтрация

Для решения задачи воспользуемся рекуррентным фильтром Калмана (ФК), при этом воспользуемся вероятностным подходом для нахождения его матрицы перехода. Суть метода заключается в том, что интервальные параметры заменяются независимыми случайными величинами, равномерно распределенными по интервалам неопределенности.

Затем, используя вероятностный подход, мы заменим неопределенные интервальные матрицы $\tilde{A}$, $\tilde{B}$ матрицами, элементы которых зависят от случайных величин

$$A(\theta) = (A + \sum_{i=1}^m A_i \theta_i), \quad B(\theta) = (B + \sum_{i=m+1}^{m+m_1} B_i \theta_i), \quad (4)$$

где θ_i – независимые равномерно распределенные случайные величины на интервале $[-1, +1]$ ($-1 \leq \theta_i \leq 1 (i = \overline{1, m+m_1})$). Будем считать, что случайные величины θ_i не зависят от x_0 , $q(k)$ и $v(k)$. В (4) матрицы $A = \frac{1}{2}(\underline{A} + \bar{A})$, и $B = \frac{1}{2}(\underline{B} + \bar{B})$ являются медианами интервальных матриц $\tilde{A}$, $\tilde{B}$. Матрицы A_i, B_i можно задать так, чтобы один элемент, соответствующий неопределенному элементу матриц $\tilde{A}, \tilde{B}$, оставался ненулевым. Его значение можно определить по ширине интервала неопределенности элементов матриц $\tilde{A}, \tilde{B}$.

В этом случае модель системы (1) принимает вид:

$$x(k+1) = A(\theta)x(k) + f(k) + B(\theta)q(k), \quad x(0) = x_0. \quad (5)$$

Для построения оценки $\hat{x}(k)$ воспользуемся рекуррентным алгоритмом (ФК):

$$\hat{x}(k+1) = A\hat{x}(k) + f(k) + K(k)(y(k+1) - S(A\hat{x}(k) + f(k))), \quad \hat{x}(0) = \bar{x}_0, \quad (6)$$

где матрица коэффициентов перехода фильтра $K(k)$ определяется на основе оптимизации критерия (3) с учетом вида распределения параметра θ и в предположении, что вектор $f(k)$ известен.

Используя свойство операции $\text{tr}(\cdot)$ ($y^T A y = \text{tr} A y y^T$) и правил дифференцирования функции $\text{tr}(\cdot)$ от произведения матриц [18]:

$$\frac{\partial \text{tr} A X B}{\partial X} = A^T B^T, \quad \frac{\partial \text{tr} A X^T B}{\partial X} = B A, \quad (7)$$

из уравнения

$$\frac{\partial J(0; T)}{\partial K} = 0 \quad (8)$$

получаем аналитическое выражение для матрицы $K(k)$

$$K(k) = (A N(k) A^T S^T + Q(k) S^T) (S A N(k) A^T S^T + S Q(k) S^T + V(k))^{-1}, \quad (9)$$

где

$$Q(k) = \frac{1}{3} \sum_{i=1}^m A_i N(k) A_i^T + \frac{1}{3} \sum_{i=1}^m A_i \hat{x}(k) \hat{x}(k)^T A_i^T + B B^T + \frac{1}{3} \sum_{i=m+1}^{m+m_1} B_i B_i^T,$$

а матрица $N(k) = M\{(x(k) - \hat{x}(k))(x(k) - \hat{x}(k))^T\}$ удовлетворяет разностному уравнению

$$N(k+1) = (A - K(k) S A) N(k) (A - K(k) S A)^T + (I - K(k) S) Q(k) (I - K(k) S)^T + K(k) V(k) K(k)^T, \quad N(0) = N_0. \quad (10)$$

Отметим, что применение в качестве матрицы динамики ФК (6) вместо неопределенной (случайной) матрицы $A(\theta)$ матрицы A (медианы интервальной матрицы $\tilde{A}$) приведет к изменению вектора неизвестного входа (этот вектор обозначается через $r(k)$):

$$r(k) = f(k) + \sum_{i=1}^m A_i \theta_i x(k), \quad -1 \leq \theta_i \leq 1 (i = \overline{1, m}), \quad (11)$$

где второй член является дополнительным неизвестным входом, возникающим из-за неопределенности матрицы перехода состояний модели (5).

В качестве алгоритма оценивания неизвестного входа $r(k)$ воспользуемся методом МНК, в этом случае оценка может быть построена на основе минимизации дополнительного критерия [6, 7]

$$I_1 = \sum_{t=1}^k \left\{ \|y(t) - S(A\hat{x}(t-1) + r(t-1))\|_C^2 + \|r(t-1)\|_D^2 \right\}. \quad (12)$$

В (12) C, D – положительно определенные весовые матрицы. МНК-оценки неизвестного входа, основанные на минимизации критерия (12), примут вид:

$$\hat{r}^{(\text{МНК})}(k) = [S^T C S + D]^{-1} S^T C [y(k+1) - S A \hat{x}(k)]. \quad (13)$$

Для повышения точности оценки неизвестного входа дополнительно будем использовать непараметрические алгоритмы [13, 14] сглаживания инновационного процесса $y(k+1) - S A \hat{x}(k)$:

$$\hat{r}^{(\text{НП})}(k) = [S^T C S + D]^{-1} S^T C \hat{\Omega}, \quad (14)$$

где j -компонента вектора вычисляется по формуле

$$\hat{\Omega}_j(k) = \frac{\sum_{t=1}^{k-1} [y(t+1) - S A \hat{x}(t)]_j G\left(\frac{k-t+1}{\mu_j}\right)}{\sum_{t=1}^k G\left(\frac{k-t+1}{\mu_j}\right)}, \quad (j = \overline{1, l}), \quad (15)$$

где $G(\cdot)$ – ядерная функция, μ_j – коэффициенты сглаживания.

Оценка робастной фильтрации в дискретных системах с интервальными параметрами определялась из рекуррентного уравнения

$$\hat{x}(k+1) = A \hat{x}(k) + \hat{r}(k) + K(k)(y(k+1) - S(A \hat{x}(k) + \hat{r}(k))), \quad \hat{x}(0) = \bar{x}_0, \quad (16)$$

где матричные коэффициенты перехода $K(k)$ рассчитывались по формулам (9), (10), а оценки $\hat{r}(k)$, использующие непараметрическое сглаживание, определялись по формулам (14), (15). Отметим, что в (13) и (15) используется медиана интервальной матрицы $\tilde{A}$. Также отметим, что в предложенном алгоритме робастной фильтрации интервальная неопределенность переходной матрицы модели $\tilde{A}$ учитывается дважды: при расчете дисперсионной матрицы $N(k)$ и при вычислении оценок неизвестного входа (в силу того, что неизвестный вход зависит от неопределенности параметра (11)).

3. Результаты моделирования

Моделирование проводилось для следующих данных ($m = 2$, $m_1 = 2$):

$$\begin{aligned} A &= \begin{pmatrix} 0,85 & 0,1 \\ -0,05 & 0,84 \end{pmatrix}, \quad S = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad A_1 = \begin{pmatrix} 0,1 & 0 \\ 0 & 0 \end{pmatrix}, \quad A_2 = \begin{pmatrix} 0 & 0 \\ 0 & 0,1 \end{pmatrix}, \\ B &= \begin{pmatrix} 0,085 & 0 \\ 0 & 0,065 \end{pmatrix}, \quad B_3 = \begin{pmatrix} 0,06 & 0 \\ 0 & 0 \end{pmatrix}, \quad B_4 = \begin{pmatrix} 0 & 0 \\ 0 & 0,04 \end{pmatrix}, \quad C = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad D = \begin{pmatrix} 0,1 & 0 \\ 0 & 0,1 \end{pmatrix}, \\ f(k) &= \begin{pmatrix} 0,2 - 0,25 \sin(0,22k) \\ -0,15 + 0,3 \sin(0,18k) \end{pmatrix}. \end{aligned} \quad (17)$$

Начальные условия:

$$x(0) = \begin{pmatrix} 2 \\ 3 \end{pmatrix}, \quad \hat{x}(0) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad N(0) = \begin{pmatrix} 0,1 & 0 \\ 0 & 0,1 \end{pmatrix}. \quad (18)$$

В (15) использовалась функция ядра Гаусса

$$G(z) = \frac{\exp\left(-\frac{z^2}{2}\right)}{\sqrt{2\pi}}.$$

Качество фильтрации оценивалось посредством сравнения стандартных ошибок отклонений оценок вектора состояния:

$$\sigma_i = \sqrt{\frac{\sum_{k=1}^N (x_i(k) - \hat{x}_i(k))^2}{N-1}}, \quad (i = \overline{1, 2}), \quad (19)$$

для следующих алгоритмов:

- ФК для систем с медианными матрицами A , B , (оценка $f(k)$ в алгоритме не используется) (ФК);
- для систем с медианными матрицами A , B был применен ФК, для вычисления оценок неизвестного входа использовался метод МНК с непараметрическим сглаживанием (формулы (14), (15)) (ФК + МНК + НПС);

• для систем с медианными матрицами A , B был применен робастный ФК (формулы (6), (9), (10)) с использованием метода МНК и непараметрического сглаживания (формулы (14), (15)) для вычисления оценок неизвестного входа (РФК).

Стандартные ошибки отклонений оценок фильтрации состояний системы σ_1 и σ_2 приведены в таблице. Результаты моделирования соответствуют 10 реализациям распределенных с равномерной плотностью значений компонент случайного вектора θ (сравнение выполнено для методов перечисленных выше) и получены для алгоритмов фильтрации ($N = 200$) с усреднением по 100 реализациям (см. таблицу).

Результаты сравнения стандартных ошибок отклонений оценок σ_1 и σ_2 вектора состояния для различных методов

№ п/п	ФК		ФК + МНК + НПС		РФК	
	σ_1	σ_2	σ_1	σ_2	σ_1	σ_2
1	0,786	0,981	0,602	0,485	0,537	0,463
2	1,029	0,999	0,710	0,475	0,612	0,461
3	1,331	1,019	0,737	0,488	0,624	0,469
4	0,824	0,981	0,613	0,512	0,549	0,482
5	1,077	1,017	0,690	0,495	0,609	0,481
6	1,004	1,097	0,660	0,478	0,575	0,460
7	0,805	0,989	0,653	0,501	0,583	0,475
8	0,592	1,025	0,528	0,476	0,482	0,459
9	0,774	0,982	0,628	0,509	0,564	0,488
10	0,570	0,966	0,512	0,486	0,468	0,470

Из таблицы видно, что алгоритм с робастной фильтрацией (РФК) имеет преимущество в точности по сравнению с другими алгоритмами. Преимущество алгоритма (РФК) в точности по сравнению с алгоритмом (ФК + МНК + НПС) составляет от 3 до 16%.

Заключение

С использованием вероятностного подхода предложен алгоритм синтеза робастного фильтра для дискретных систем с неизвестным входом и интервальными параметрами в модели объекта.

Задача решается с помощью рекуррентных алгоритмов калмановской фильтрации, метода МНК и процедур непараметрического сглаживания. Предлагаемый метод реализует уменьшение влияния неопределенностей в модели за счет замены интервальной неопределенности на вероятностную и учета оценок неизвестного входа с дополнительным сглаживанием.

Результаты моделирования показали, что совместное использование алгоритмов непараметрического сглаживания и робастного подхода позволяет повысить точность алгоритма фильтрации в условиях интервальной неопределенности в задании параметров модели объекта.

Список источников

1. Abolhasani M., Rahmani M. Robust Kalman filtering for discrete-time systems with stochastic uncertain time-varying parameters // Electronics Letters. 2017. V. 53(3). P. 146–148.
2. Ichalal D., Marx B., Maquin D., Ragot J. State estimation of system with bounded uncertain parameters: interval multi-model approach // Int. J. of Adaptive Control and Signal Processing. 2018. V. 32(3). P. 480–493.
3. Rocha K.D., Terra M.H. Robust Kalman filter for systems subject to parametric uncertainties // Systems and Control Letters. 2021. V. 157. Art. 105034.
4. Abolhasani M., Rahmani M. Robust Kalman filtering for discrete-time time-varying systems with stochastic and norm-bounded uncertainties // J. of Dynamic Systems, Measurement, and Control. 2018. V. 140 (3). Art. 030901.
5. Kim S., Deshpande V.M., Bhattacharya R. Robust Kalman filtering with probabilistic uncertainty in system parameters // IEEE Control Systems Letters. 2021. V. 5 (1). P. 295–300.

6. Janczak D., Grishin Yu. State estimation of linear dynamic system with unknown input and uncertain observation using dynamic programming // *Control and Cybernetics*. 2006. V. 4. P. 851–862.
7. Gillijns S., Moor B. Unbiased minimum-variance input and state estimation for linear discrete-time systems // *Automatica*. 2007. V. 43. P. 111–116.
8. Hsien C.-S. On the optimality of two-stage Kalman filter for systems with unknown input // *Asian J. of Control*. 2010. V. 12. № 4. P. 510–523.
9. Witczak M. Fault diagnosis and fault-tolerant control strategies for non-linear systems. Chapter 2. Unknown input observers and filters // *Lecture Notes in Electrical Engineering*. Springer International Publishing, 2014. P. 19–56.
10. Смагин В.И., Смагин С.В. Фильтрация в линейных дискретных нестационарных системах с неизвестными возмущениями // *Вестник Томского государственного университета. Управление, вычислительная техника и информатика*. 2011. № 3 (16). С. 43–51.
11. Смагин В.И. Оценивание состояний линейных дискретных систем при неизвестном входе с использованием компенсаций // *Известия вузов. Физика*. 2014. Т. 57, № 5. С. 104–110.
12. Smagin V.I., Koshkin G.M. Kalman filtering and control algorithms for systems with unknown disturbances and parameters using nonparametric technique // *Proc. 20th International Conference on Methods and Models in Automation and Robotics (MMAR 2015)*. Miedzydroje, 2015. P. 247–251.
13. Koshkin G.M., Smagin V.I. Kalman filtering and forecasting algorithms with use of nonparametric functional estimators // *Springer Proc. in Mathematics & Statistics. 2nd Conference of the International-Society-for-Nonparametric-Statistics (ISNPS)*. 2016. V. 175. P. 75–84.
14. Smagin V., Koshkin G., Udod V. State estimation for linear discrete-time systems with unknown input using nonparametric technique // *Proc. International Conference on Artificial Intelligence and Control Automation (AICA 2015)*. Bangkok : Atlantis Press, 2015. P. 675–677. doi: 10.2991/cisia-15.2015.184
15. Barmish B.R., Polyak B.T. A new approach to open robustness problems based on probabilistic predication formulae // *Proc. 13th World IFAC Congr.*, 30 June – 5 July, San Francisco, USA. 1996. V. H. P. 1–6.
16. Ким К.С., Смагин В.И. Робастная экстраполяция в дискретных системах с интервальными параметрами с использованием алгоритмов оценивания неизвестного входа // *Вестник Томского государственного университета. Управление, вычислительная техника и информатика*. 2022. № 59. С. 47–54.
17. Smagin V.I., Kim K.S. Robust extrapolation for systems with unknown input and interval uncertainty in system and observations // *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 2023. V. 62. P. 85–91.
18. Athans M. The matrix minimum principle // *Informat. and Contr.* 1968. V. 11. P. 592–606.

References

1. Abolhasani, M. & Rahmani, M. (2017) Robust Kalman filtering for discrete-time systems with stochastic uncertain time-varying parameters. *Electronics Letters*. 53(3). pp. 146–148.
2. Ichalal, D., Marx, B., Maquin, D. & Ragot, J. (2018) State estimation of system with bounded uncertain parameters: interval multi-model approach. *International Journal of Adaptive Control and Signal Processing*. 32(3). pp. 480–493.
3. Rocha, K.D.T. & Terra, M.H. (2021) Robust Kalman filter for systems subject to parametric uncertainties. *Systems and Control Letters*. 157. Art. 105034.
4. Abolhasani, M. & Rahmani, M. (2018) Robust Kalman filtering for discrete-time time-varying systems with stochastic and norm-bounded uncertainties. *Journal of Dynamic Systems, Measurement, and Control*. 140(3). Art. 030901.
5. Kim, S., Deshpande, V.M. & Bhattacharya, R. (2021) Robust Kalman filtering with probabilistic uncertainty in system parameters. *IEEE Control Systems Letters*. 5(1). pp. 295–300.
6. Janczak, D. & Grishin, Yu. (2006) State estimation of linear dynamic system with unknown input and uncertain observation using dynamic programming. *Control and Cybernetics*. 4. pp. 851–862.
7. Gillijns, S. & Moor, B. (2007) Unbiased minimum-variance input and state estimation for linear discrete-time systems. *Automatica*. 43. pp. 111–116.
8. Hsien, C.-S. (2010) On the optimality of two-stage Kalman filter for systems with unknown input. *Asian Journal of Control*. 12(4). pp. 510–523.
9. Witczak, M. (2014) Fault diagnosis and fault-tolerant control strategies for non-linear systems. Chapter 2. Unknown input observers and filters. In: *Lecture Notes in Electrical Engineering*. Springer International Publishing, Switzerland. pp. 19–56.
10. Smagin, V.I. & Smagin, S.V. (2011) Filtering for linear not stationary discrete system with unknown disturbances. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 16(3). pp. 43–50.
11. Smagin, V.I. (2014) State estimation for linear discrete systems with unknown input using compensations. *Izvestiya vuzov. Fizika – Russian Physics Journal*. 57(5). pp. 682–690. DOI: 10.1007/s11182-014-0291-x
12. Smagin, V.I. & Koshkin, G.M. (2015) Kalman filtering and control algorithms for systems with unknown disturbances and parameters using nonparametric technique. *Proceedings of the 20th International Conference on Methods and Models in Automation and Robotics (MMAR 2015)*. Miedzydroje. Poland. pp. 247–251.

13. Koshkin, G.M. & Smagin, V.I. (2016) Kalman filtering and forecasting algorithms with use of nonparametric functional estimators. *Springer Proceedings in Mathematics & Statistics. 2nd Conference of the International-Society-for-Nonparametric-Statistics (ISNPS)*. 175. pp. 75–84. DOI: 10.1007/978-3-319-41582-6_6
14. Smagin, V., Koshkin, G. & Udod, V. (2015) State estimation for linear discrete-time systems with unknown input using nonparametric technique. *Proceedings of the International Conference on Artificial Intelligence and Control Automation (AICA 2015)*. Atlantis Press, Bangkok, Thailand. pp. 675–677. DOI: 10.2991/cisia-15.2015.184
15. Barmish, B.R. & Polyak, B.T. (1996) A new approach to open robustness problems based on probabilistic predication formulae. *Proceedings 13th World IFAC Congr.* Vol. H. 30 June – 5 July, San Francisco, USA. pp. 1–6.
16. Kim, K.S. & Smagin, V.I. (2022) Robust extrapolation in discrete systems with interval parameters using algorithms for estimating unknown input. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 59. pp. 47–54. DOI: 10.17223/19988605/59/5
17. Smagin, V.I. & Kim, K.S. (2023) Robust extrapolation for systems with unknown input and interval uncertainty in system and observations. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 62. pp. 85–91. DOI: 10.17223/19988605/62/9
18. Athans, M. (1968) The matrix minimum principle. *Information and Control*. 11. pp. 592–606.

Информация об авторах:

Смагин Валерий Иванович – профессор, доктор технических наук, профессор кафедры прикладной математики Национального исследовательского Томского государственного университета (Томск, Россия). E-mail: vsm@mail.tsu.ru

Ким Константин Станиславович – кандидат физико-математических наук, доцент Высшей IT школы Национального исследовательского Томского государственного университета (Томск, Россия). E-mail: kks93@rambler.ru

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

Information about the authors:

Smagin Valery I. (Doctor of Technical Science, Professor, National Research Tomsk State University, Tomsk, Russian Federation). E-mail: vsm@mail.tsu.ru

Kim Konstantin S. (Candidate of Physical and Mathematical Sciences, National Research Tomsk State University, Tomsk, Russian Federation). E-mail: kks93@rambler.ru

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

Поступила в редакцию 18.04.2023; принята к публикации 04.09.2023

Received 18.04.2023; accepted for publication 04.09.2023

ИНФОРМАТИКА И ПРОГРАММИРОВАНИЕ

INFORMATICS AND PROGRAMMING

Научная статья

УДК 004.415.2

doi: 10.17223/19988605/64/12

**Автоматизированная система планирования и проведения экспериментов
для сбора и анализа медицинских данных в контексте открытых исследований****Яна Валерьевна Костелей¹, Дмитрий Сергеевич Жданов²,
Евгения Васильевна Голобокова³, Артём Шамильевич Буреев⁴**^{1, 2, 3, 4} Томский государственный университет, Томск, Россия^{1, 2, 3, 4} Томский научно-исследовательский институт курортологии и физиотерапии, Томск, Россия¹ Томской государственный университет систем управления и радиоэлектроники, Томск, Россия¹ kosteleyyv@gmail.com² d_s_zhdanov@mail.ru³ jane04@yandex.ru⁴ artem_bureev@mail.ru

Аннотация. Рассматривается автоматизированная система планирования и проведения экспериментов, обеспечивающая возможность выгрузки результатов проводимых экспериментов в репозитории медицинских наборов данных в соответствии с принципами FAIR. Отличительной особенностью системы является возможность определения и дополнения состава компонентов, используемых для проведения эксперимента, в зависимости от набора технического, методического и алгоритмического обеспечения, использование которых планируется для сбора и анализа данных в процессе разработки медицинских программно-аппаратных комплексов.

Ключевые слова: автоматизация проведения эксперимента; открытые исследования; хранение наборов данных; FAIR.

Благодарности: Результаты получены в рамках выполнения государственного задания Минобрнауки России, проект № FSWM-2022-0008.

Для цитирования: Костелей Я.В., Жданов Д.С., Голобокова Е.В., Буреев А.Ш. Автоматизированная система планирования и проведения экспериментов для сбора и анализа медицинских данных в контексте открытых исследований // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2023. № 64. С. 120–127. doi: 10.17223/19988605/64/12

Original article

doi: 10.17223/19988605/64/12

**Automated system for planning and conducting experiments
for the medical data collection and analysis in the open research context****Yana V. Kosteley¹, Dmitry S. Zhdanov², Evgeniya V. Golobokova³, Artem Sh. Bureev⁴**^{1, 2, 3, 4} Tomsk State University, Tomsk, Russian Federation^{1, 2, 3, 4} Tomsk Research Institute of Balneology and Physiotherapy, Tomsk, Russian Federation¹ Tomsk State University of Control Systems and Radioelectronics, Tomsk, Russian Federation¹ kosteleyyv@gmail.com

² d_s_zhdanov@mail.ru

³ jane04@yandex.ru

⁴ artem_bureev@mail.ru

Abstract. The paper considers the automated system for planning and conducting experiments, which provides the possibility of ongoing experiments results uploading to the medical datasets repository in accordance with the FAIR principle. A distinctive system feature is the ability to determine and supplement the components composition used to conduct the experiment, depending on the set of technical, methodological and algorithmic support, the use of which is planned to collect and analyze data in the developing medical software and hardware systems process.

Keywords: automation of the experiment; open research; storage of data sets; FAI.

Acknowledgments: The results were obtained as part of the implementation of the state task of the Ministry of Education and Science of Russia, project No. FSWM-2022-0008.

For citation: Kosteley, Y.V., Zhdanov, D.S., Golobokova, E.V., Bureev, A.Sh. (2023) Automated system for planning and conducting experiments for the medical data collection and analysis in the open research context. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naja tehnika i informatika – Tomsk State University Journal of Control and Computer Science*. 64. pp. 120–127. doi: 10.17223/19988605/64/12

Введение

Автоматизация процесса проведения медицинских исследований и манипуляций неразделимо связана с процессами сбора и обработки данных, которые позволяют как облегчить рутинные операции, проводимые медицинским персоналом, так и повысить качество предоставляемой услуги путем объективизации результатов, увеличения скорости, эффективности их выполнения и обеспечения удобства пациента. Трудность развития направления связана с тем, что разметка и аннотация полученных данных требуют наличия экспертных компетенций в интерпретации биологических сигналов и знания физиологических особенностей пациентов с той или иной патологией.

Создание новой медицинской технологии включает разработку аппаратной, программной и методической составляющих, при этом разработка программного компонента зависит от их реализации. Такая ситуация порождает проблему организации процесса разработки алгоритмов анализа данных тогда, когда необходимо реализовать не только итоговый программный продукт, но и промежуточные варианты, используемые для разных версий устройств и методических подходов. Зачастую в процессе разработки необходимо проводить множество экспериментов по сбору данных от устройств с различными схемами функционирования и расстановки датчиков, видеофиксацией и ведением журнала действий участников. Универсальных систем сопровождения таких процессов существовать не может, так как присутствуют особенности предметной области, конкретной методики проведения операций, используемых протоколов передачи данных от устройств, а также конечных задач.

Для соответствия концепции открытых исследований, воспроизводимости, доступности и повышения качества их результатов экспериментальные данные должны быть предоставлены в общий доступ [1, 2]. Можно отметить, что аннотация медицинских биологических сигналов – это трудоемкий процесс, требующий компетенций в выбранной предметной области, поэтому при публикации важным ресурсом являются как сами исходные сигналы, так и их аннотация. Последнее часто отсутствует в медицинских наборах данных. Кроме этого, репозитории с ними имеют свои соглашения по структуре и именованию файлов, которые исследователь должен заранее учесть, чтобы обеспечить полноту данных и их выгрузку без дополнительной подготовки. Поэтому важной задачей при разработке новых медицинских технологий является не только сбор исходных биологических сигналов, но и накопление аннотационных данных, которые включают информацию об объекте исследования, плане эксперимента, проведенной разметке и результатах анализа. Для обеспечения корректного выполнения эксперимента необходимо предоставить приложения, обеспечивающие сопровождение каждой фазы эксперимента, с детальными инструкциями.

В работе предложен вариант архитектуры автоматизированной системы планирования и проведения экспериментов для сбора и анализа данных, соответствующей решению прикладной задачи – разработке роботизированного комплекса для восстановления утраченной функции конечности.

Разработка систем управления сбором и анализом экспериментальных данных является одним из механизмов обеспечения доступности, совместимости и возможности повторного использования полученных данных. Наиболее часто встречаемые направления работ в этой области – сопровождение физических экспериментов и регистрация событий от пользователей [3, 4]. Фокус в работах часто смещен на разработку аппаратных модулей и модулей управления экспериментальными установками. Например, в работе [5] рассматривается разработка систем для сбора данных экспериментов в аэродинамических трубах на основе архитектуры систем управления технологическим процессом. Работа [6] является примером конструирования более гибкой системы с архитектурой MVC (Model-View-Controller), где в качестве модели выступают модули захвата данных устройств через аналого-цифровой преобразователь, а в виде контроллера рассматриваются механизмы миграции этих данных в базу данных (БД) и файловую систему. Общее число работ, посвященных автоматизации сбора экспериментальных данных, ограничено.

В настоящее время на рассмотрение и одобрение сообществом вносятся предложения по стандартизации сбора данных в области нейробиологии и биомедицины [7]. Международный координационный центр нейроинформатики одобрил принципы FAIR (Findable, Accessible, Interoperable and Reusable), которые направлены на повышение возможности повторного использования данных и расширение автоматического поиска [8].

В работе [9] предложена открытая веб-система для клинических и электрофизиологических экспериментов, являющаяся примером систем, поддерживающих принципы FAIR. Система оснащена модулями, позволяющими с использованием поддерживаемых протоколов собирать данные экспериментов, включая биологические сигналы и время отклика. Планирование эксперимента в системе связано с составлением его плана (характеристики экспериментальной группы, конфигурация оборудования и т.д.) и протокола (инструкций). Другие примеры такого типа систем представлены в обзоре [10].

Несмотря на то, что описанные принципы позволяют определять инструменты для сбора и унификации медицинских данных, задаче сбора данных при разработке медицинских аппаратно-программных комплексов не уделено должного внимания. Особенность такого типа систем в том, что они должны обеспечивать планирование, проведение и анализ формализованных экспериментов. Поэтому в рамках настоящей работы рассмотрена архитектура автоматизированной системы планирования и проведения экспериментов для сопровождения разработки аппаратно-программного комплекса с возможностью загрузки полученных данных в репозитории, поддерживающие принципы FAIR.

1. Предметная область системы

Разработка системы планирования и проведения экспериментов является одним из этапов создания роботизированного комплекса для восстановления утраченной функции конечности человека (например, после перенесенного инсульта). Комплекс включает в себя два типа аппаратных устройств, управляемых общим программным обеспечением. Копирующее устройство позволяет получить и записать информацию о перемещении конечности врача-реабилитолога или пациента. Исполняющее устройство осуществляет перемещение конечности пациента, аналогичное перемещению конечности врача, либо в соответствии с заранее заданной программой реабилитационных упражнений.

Программное обеспечение осуществляет формирование управляющих команд, позволяющих пациенту выполнить программу реабилитации с участием врача-реабилитолога или без него. Использование пациентом копирующего устройства дает возможность провести оценку состояния его конечности в балльной системе согласно функциональной шкале Фугл-Мейера. Для анализа используется информация о пространственной конфигурации конечности (углы поворота узлов относительно друг друга, полная кинематическая схема), а также электромиограммы, полученные с поверхности задействованных при движениях мышц.

Система планирования и проведения эксперимента должна предоставлять возможность создания четко структурированных наборов, полученных при проведении формализованных действий пациента и сопровождаемых подробной экспертной аннотацией, включающей описание периодов воз-

никновения составных действий, балл оценки по функциональной шкале и т.д. Помимо этого, система должна предоставлять возможность планирования и проведения экспериментальных исследований, позволяющих оценить эффективность версий исполнения применяемых в эксперименте аппаратных и программных средств, а также методик работы с роботизированным комплексом.

В плане эксперимента могут быть отражены типы, количество и способы расстановки датчиков, действия, проводимые при подготовке к проведению эксперимента, и другая значимая информация. Также для возможности повторной экспертной оценки проведенных в процессе эксперимента действий необходимо предусмотреть видеофиксацию его хода.

2. Архитектура системы

Архитектура предлагаемой системы включает в себя компоненты и сущности в модели данных, которые можно разделить на элементы, соответствующие предметной области, базовые элементы, которые не зависят от предметной области, а также динамически определяемые в зависимости от эксперимента компоненты. С использованием базовых элементов, определенных в системе, может производиться ее доработка на уровне кода для другой предметной области. Для части компонентов предполагается использование MVVM-архитектуры (Model-View-ViewModel), обеспечивающей связывание пользовательского интерфейса и модели данных через модель представления с обменом изменениями данных в обе стороны.

В системе определены две роли пользователей: экспериментатор и исследователь. Исследователь может как формировать план эксперимента, так и проводить тестирование новых алгоритмов, аннотацию данных с использованием полученных результатов. Для каждой из ролей определены свои компоненты. Система состоит из четырех компонент: программного обеспечения (ПО) исследователя, ПО эксперимента, ПО управления и ПО аналитика.

Центральным управляющим компонентом является ПО исследователя, функциональное назначение которого – управлять файловым хранилищем и базой наборов данных, протоколов и результатов экспериментов. Также этот компонент позволяет разрабатывать протоколы экспериментов, загружать в репозиторий результаты их проведения, производить поиск и сортировку по имеющимся наборам данных.

После разработки протокола эксперимента ПО исследователя позволяет выгрузить пакет протокола эксперимента на промежуточный файловый сервер или накопитель (рис. 1). Пакет протокола эксперимента может быть загружен в ПО эксперимента, которое используется для непосредственного проведения эксперимента. Особенность ПО эксперимента в том, что его задача состоит в выводе инструкции и управляющей информации, а сбор экспериментальных данных происходит в ПО управления.

ПО управления – это программный компонент, обеспечивающий сбор и обработку данных для определенного эксперимента. Компонент может дополняться и обновляться разработчиком. То есть ПО управления разрабатывается для работы с определенными устройствами и алгоритмами обработки данных и определяется на этапе планирования эксперимента.

ПО эксперимента в зависимости от типа эксперимента запускает необходимую версию ПО управления и управляет им посредством клиент-серверного взаимодействия по протоколу TSP (см. рис. 1).

Первично ПО эксперимента проводит запуск серверного процесса ПО управления с передачей аргумента командной строки для указания директории сохранения файлов и выполняет подключение к серверному сокету. Далее по протоколу эксперимента выполняется последовательность запросов к ПО управления: диагностика и калибровка устройств, запуск этапа эксперимента с передачей пути до файла конфигурации этапа, получение информационной строки текущего состояния проведения этапа, остановка этапа эксперимента и остановка процесса ПО управления. Серверная часть обработки запросов упакована в библиотеку и может быть использована для любой версии ПО управления.

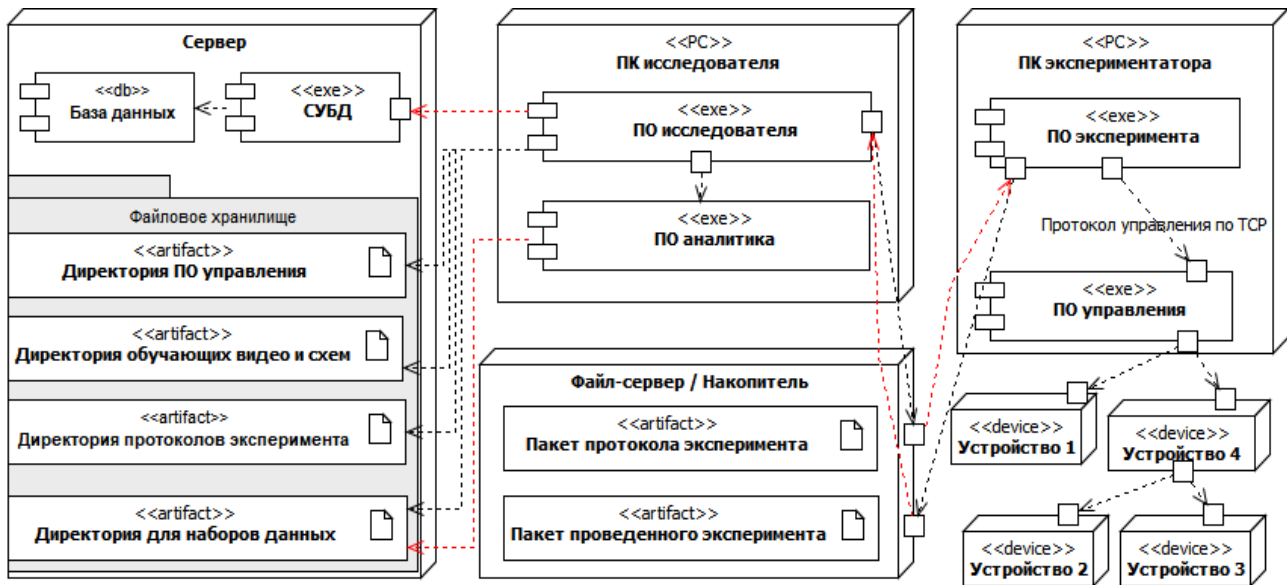


Рис. 1. Компоненты системы для сбора и обработки наборов данных
Fig. 1. The system components for collecting and processing datasets

В результате проведения эксперимента формируются файлы с сигналами и дополнительной информацией для каждого этапа эксперимента. Они сохраняются в файловой структуре, удовлетворяющей требованиям медицинского репозитория Physionet [11]. Полученные данные формируют пакет проведенного эксперимента и также через файловый сервер или накопитель выгружаются в ПО исследователя. Файловая структура пакетов протокола эксперимента и проведенного эксперимента представлена на рис. 2.

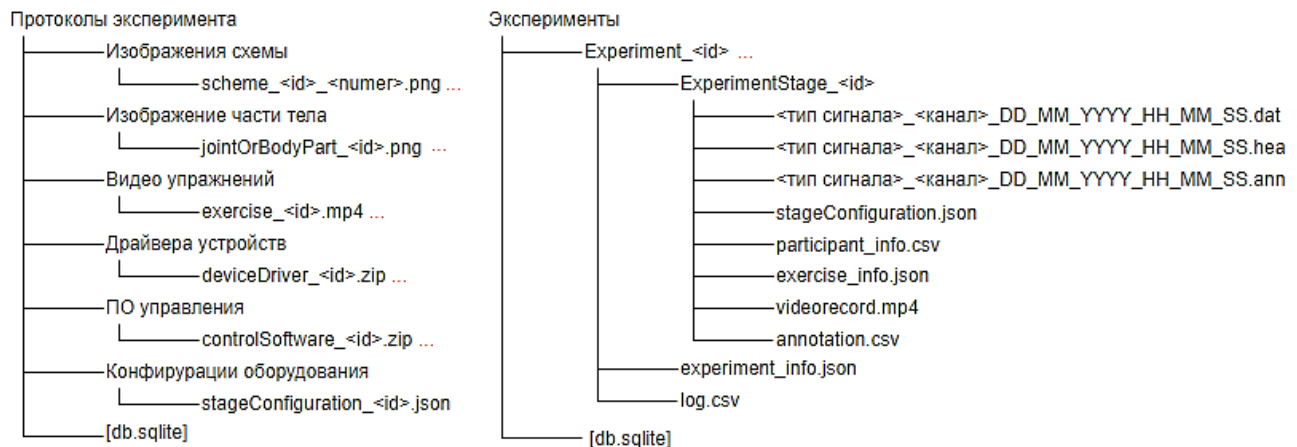


Рис. 2. Структура репозитория данных и пакетов экспериментов и их протоколов
Fig. 2. The structure of the data repository and experiments and their protocols packages

При выгрузке полученного пакета проведенного эксперимента в ПО исследователя данные перемещаются в файловое хранилище, а также частично выгружаются в БД. Для унификации структуры БД предполагается использование единой модели данных (рис. 3) для ПО исследователя и ПО эксперимента, где ПО исследователя взаимодействует с серверной БД, а при формировании пакета протокола эксперимента присоединяет к нему ее частичную файловую копию. ПО эксперимента взаимодействует с файловой версией БД, которая обновляется в процессе выгрузки пакета протокола эксперимента.

Исследователь в ПО исследователя может проводить поиск необходимых данных и запускать ПО аналитика на выбранном наборе приложение для тестирования алгоритмов и разметки сигналов. ПО аналитика, как и ПО управления, также обновляется в процессе разработки продукта.

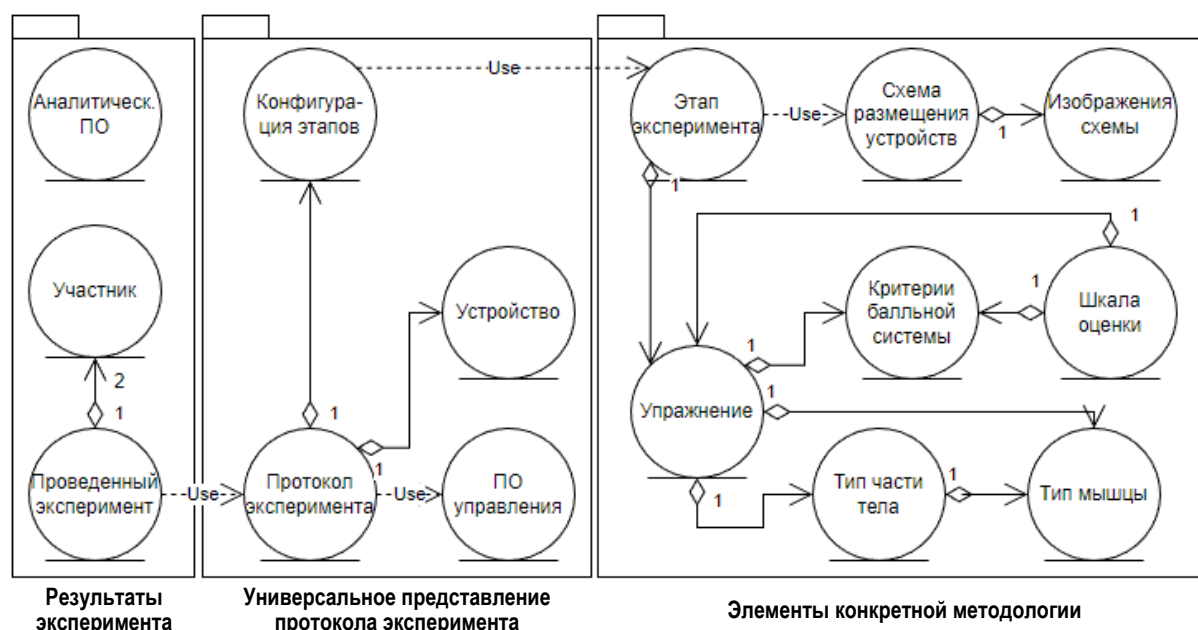


Рис. 3. Модель данных для ПО эксперимента и ПО исследователя

Fig. 3. Experiment and researcher software data model

3. Техническая реализация

В качестве фреймворка для программных компонент, определенных статически (ПО исследователя и ПО эксперимента), выбран Uno Platform [12], позволяющий создавать приложения для операционных систем Windows, Linux, MacOS, мобильных платформ и веб-сборок. Для связывания модели данных и пользовательского интерфейса применялись механизмы биндинга платформы .Net. Для реализации взаимодействия с серверной и файловой БД была использована библиотека Entity Framework Core [13]. Такой подход реализует MVVM-архитектуру и позволяет быстро адаптировать систему под другую предметную область. Для реализации серверной БД используется система управления базами данных PostgreSQL, для частичных локальных файловых копий БД – SQLite с пакетом шифрования данных.

Для создания вариантов ПО управления и ПО аналитика на данный момент разработаны приложения на языке C++ с использованием комплекта сборки MinGW, а также определена библиотека протокола взаимодействия компонент системы. В дальнейшем допускается использование компонент, разработанных с использованием других инструментов и поддерживающих заданный протокол межпроцессорного взаимодействия.

Заключение

Разработка системы, способной формировать данные, структура которых соответствует требованиям FAIR, и обладающей возможностью модификации под схожие прикладные задачи, может предоставить инструмент для сопровождения процесса разработки медицинских программно-аппаратных комплексов и информационных технологий.

Такая система должна состоять из базовых и динамически заменяемых компонентов, для того чтобы предоставлять возможность планирования и проведения экспериментов, не ограничивая исследователей в области применения такого типа систем, в том числе для сопровождения процесса разработки.

В работе представлена архитектура, технические аспекты реализации автоматизированной системы планирования и проведения экспериментов, сбора и анализа данных для сопровождения разработки функционального макета роботизированного комплекса для восстановления утраченной функ-

ции конечности. Ее использование позволило автоматизировать организацию сбора, хранения, поиска данных, а также их анализа и аннотации.

Список источников

1. What is open research? // University of Exeter. URL: <http://www.exeter.ac.uk/research/openresearch/about/explained/> (accessed: 16.04.2023).
2. Baker M. 1,500 scientists lift the lid on reproducibility // *Nature*. 2016. V. 533. P. 452–454. doi: 10.1038/533452a
3. Noldus L. The Observer: a software system for collection and analysis of observational data // *Behavior Research Methods, Instruments, & Computers*. 1991. V. 23, № 3. P. 415–429.
4. Tapp J., Wehby J., Ellis D. A multiple option observation system for experimental studies: MOOSES // *Behavior Research Methods, Instruments, & Computers*. 1995. V. 27, № 1. P. 25–31.
5. Запрягаев В.И., Гаркуша В.В., Гилев В.М. и др. Создание систем автоматизированного сбора экспериментальных данных на аэродинамических трубах // *Вычислительные технологии*. 2013. Т. 18. С. 21–28.
6. Хоченков А.Е., Галиаскаров Э.Г. Автоматизированная система сбора данных физического эксперимента // *Объектные системы*. 2014. № 1 (8). С. 57–63.
7. Abrams M.B., Bjaalie J.G., Das S. et al. A standards organization for open and FAIR neuroscience: the international neuroinformatics coordinating facility // *Neuroinformatics*. 2022. V. 20, № 1. P. 25–36.
8. Wilkinson M.D., Dumontier M., Aabersberg U.J. et al. The FAIR Guiding Principles for scientific data management and stewardship // *Scientific Data*. 2016. V. 3, № 1. P. 1–9.
9. Ruiz-Olazar M., Rocha E.S. et al. The Neuroscience Experiments System (NES)—A Software Tool to Manage Experimental Data and Its Provenance // *Frontiers in Neuroinformatics*. 2022. V. 15. P. 77.
10. Martorana M., Kuhn T., Siebes R. et al. Aligning restricted access data with FAIR: a systematic review // *PeerJ Computer Science*. 2022. V. 8. Art. e1038.
11. Sharing on PhysioNet // *Physionet*. URL: <https://physionet.org/about/publish/#guidelines> (accessed: 16.04.2023).
12. Pixel-Perfect Multi-Platform Applications with C# and WinUI // *Uno Platform*. URL: <https://platform.uno/> (accessed: 16.04.2023).
13. Entity Framework Core // *Microsoft*. URL: <https://learn.microsoft.com/en-us/ef/core/> (accessed: 16.04.2023).

References

1. University of Exeter. (n.d.) *What is open research?* [Online] Available from: <http://www.exeter.ac.uk/research/openresearch/about/explained/> (Accessed: 16th April 2023).
2. Baker, M. (2016) 1,500 scientists lift the lid on reproducibility. *Nature*. 533. pp. 452–454. DOI: 10.1038/533452a
3. Noldus, L. (1991) The Observer: a software system for collection and analysis of observational data. *Behavior Research Methods, Instruments, & Computers*. 23(3). pp. 415–429. DOI:10.3758/bf03203406
4. Tapp, J., Wehby, J. & Ellis, D. (1995) A multiple option observation system for experimental studies: MOOSES. *Behavior Research Methods, Instruments, & Computers*. 27(1). pp. 25–31. DOI: 10.3758/bf03203616
5. Zapryagaev, V.I., Garkusha, V.V., Gilev, V.M. et al. (2013) Creation of systems for automated collection of experimental data in wind tunnels. *Vychislitel'nye tekhnologii – Computing Technologies*. 18. pp. 21–28.
6. Khochenkov, A.E. & Galiaskarov, E.G. (2014) Automated system for collecting data of a physical experiment. *Ob'ektnye sistemy*. 1(8). pp. 57–63.
7. Abrams, M.B., Bjaalie, J.G., Das, S. et al. (2022) A standards organization for open and FAIR neuroscience: the international neuroinformatics coordinating facility. *Neuroinformatics*. 20(1). pp. 25–36.
8. Wilkinson, M.D., Dumontier, M., Aabersberg, U.J. et al. (2016) The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data*. 3(1). pp. 1–9. DOI: 10.1038/sdata.2016.18
9. Ruiz-Olazar, M., Rocha, E.S. et al. (2022) The neuroscience experiments system (NES)—A software tool to manage experimental data and its provenance. *Frontiers in Neuroinformatics*. 15. pp. 77.
10. Martorana, M. Kuhn, T., Siebes, R. et al. (2022) Aligning restricted access data with FAIR: a systematic review. *Peer Journal of Computer Science*. 8. Art. e1038. DOI: 10.7717/peerj-cs.1038
11. *Sharing on PhysioNet*. *Physionet*. [Online] Available from: <https://physionet.org/about/publish/#guidelines> (Accessed: 16th April 2023).
12. *Pixel-Perfect Multi-Platform Applications with C# and WinUI*. [Online] Available from: <https://platform.uno/> (Accessed: 16th April 2023).
13. *Entity Framework Core*. [Online] Available from: <https://learn.microsoft.com/en-us/ef/core/> (Accessed: 16th April 2023).

Информация об авторах:

Костелей Яна Валерьевна – кандидат технических наук, научный сотрудник лаборатории медицинского приборостроения Центра развития науки, технологий и образования в области обороны и обеспечения безопасности государства Национального исследовательского Томского государственного университета (Томск, Россия); научный сотрудник лаборатории медицинской робототехники Томского научно-исследовательского института курортологии и физиотерапии (Томск, Россия);

старший преподаватель кафедры экономической математики, информатики и статистики Томского государственного университета систем управления и радиоэлектроники (Томск, Россия). E-mail: kosteleyyv@gmail.com

Жданов Дмитрий Сергеевич – кандидат технических наук, заведующий лабораторией медицинского приборостроения Центра развития науки, технологий и образования в области обороны и обеспечения безопасности государства Национального исследовательского Томского государственного университета (Томск, Россия); научный сотрудник лаборатории медицинской робототехники Томского научно-исследовательского института курортологии и физиотерапии (Томск, Россия). E-mail: d_s_zhdanov@mail.ru

Голобокова Евгения Васильевна – научный сотрудник лаборатории медицинского приборостроения Центра развития науки, технологий и образования в области обороны и обеспечения безопасности государства Национального исследовательского Томского государственного университета; научный сотрудник лаборатории медицинской робототехники Томского научно-исследовательского института курортологии и физиотерапии (Томск, Россия). E-mail: jane04@yandex.ru

Буреев Артем Шамильевич – научный сотрудник лаборатории медицинского приборостроения Центра развития науки, технологий и образования в области обороны и обеспечения безопасности государства Национального исследовательского Томского государственного университета; заведующий лабораторией медицинской робототехники Томского научно-исследовательского института курортологии и физиотерапии (Томск, Россия). E-mail: artem_bureev@mail.ru

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

Information about the authors:

Kosteley Yana V. (Candidate of Technical Sciences, National Research Tomsk State University; Tomsk Research Institute of Balneology and Physiotherapy; Tomsk State University of Control Systems and Radioelectronics, Tomsk, Russian Federation). E-mail: kosteleyyv@gmail.com

Zhdanov Dmitry S. (Candidate of Technical Sciences, National Research Tomsk State University; Tomsk Research Institute of Balneology and Physiotherapy, Tomsk, Russian Federation). E-mail: d_s_zhdanov@mail.ru

Golobokova Evgeniya V. (National Research Tomsk State University; Tomsk Research Institute of Balneology and Physiotherapy, Tomsk, Russian Federation). E-mail: jane04@yandex.ru

Bureev Artem S. (National Research Tomsk State University; Tomsk Research Institute of Balneology and Physiotherapy, Tomsk, Russian Federation). E-mail: artem_bureev@mail.ru

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

Received 05.05.2023; accepted for publication 04.09.2023

Поступила в редакцию 05.05.2023; принята к публикации 04.09.2023

Научная статья

УДК 004.032.26, 004.932.2

doi: 10.17223/19988605/64/13

Применение закона Бенфорда в обнаружении deepfake-изображений

Светлана Павловна Никитенкова

Нижегородский государственный университет им. Н.И. Лобачевского, nikitenkova@rf.unn.ru

Аннотация. Разработка и совершенствование технологий обнаружения deepfake являются одним из приоритетных направлений обеспечения социальной и биометрической безопасности. В работе исследуются перспективы применения закона Бенфорда как инструмента обнаружения deepfake-изображений, сгенерированных нейросетями GAN. Предлагаемый подход основан на анализе спектра мощности и энтропии изображений. Эффективность предложенного метода апробировалась на датасетах, сгенерированных нейросетями StyleGAN2 и StyleGAN3. Предложенный метод не требует больших вычислительных мощностей.

Ключевые слова: AI-синтезированное изображение; закон Бенфорда; энтропия; дивергенции Кульбака–Лейблера; метод Random Forest.

Для цитирования: Никитенкова С.П. Применение закона Бенфорда в обнаружении deepfake-изображений // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2023. № 64. С. 128–137. doi: 10.17223/19988605/64/13

Original article

doi: 10.17223/19988605/64/13

Application of Benford's Law in Deepfake Image Detection

Svetlana P. Nikitenkova

Lobachevsky State University of Nizhny Novgorod, Nizhny Novgorod, Russia, nikitenkova@rf.unn.ru

Abstract. The development and improvement of deepfake detection technologies is one of the priority areas for ensuring social and biometric security. The main goal of the work is to test the application of the well-known Benford's law as a detection tool for images generated by GAN. The proposed approach is based on the analysis of power spectrum and Shannon entropy of image. The effectiveness of the proposed method was tested on datasets generated by StyleGAN2 and StyleGAN3 neural networks. The proposed method does not require large computing power.

Keywords: AI-synthesized image; Benford's law; entropy; Kullback-Leibler divergences; Random Forest method.

For citation: Nikitenkova, S.P. (2023) Application of Benford's Law in Deepfake Image Detection. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naja tehnika i informatika – Tomsk State University Journal of Control and Computer Science*. 64. pp. 128–137. doi: 10.17223/19988605/64/13

Введение

Прогресс нейросетей GAN (Generative Adversarial Network) позволил значительно улучшить качество синтетических изображений, или дипфейков. С одной стороны, это может использоваться в таких областях, как реклама, кинопроизводство, видеоигры и т.д. С другой стороны, дипфейки представляют собой серьезную угрозу для общества, политической системы и бизнеса. Миллионы сгенерированных цифровых изображений загружаются каждый день в Интернет, распространяясь

в социальных сетях, размывая границы между фактами и вымыслом. Большинство сфабрикованных новостей содержит поддельные изображения, создавая в обществе путаницу в отношении того, какие источники информации являются надежными, бросая вызов восприятию реальности [1]. Изображения искусственно сгенерированных лиц способны обмануть даже самых опытных наблюдателей и, главное, вызвать симпатию и большее доверие, чем настоящие лица [2]. Высоко реалистичные портреты фальшивых людей используются в качестве визуально убедительного инструмента для облегчения схем мошенничества и кампаний по дезинформации с потенциально катастрофическими последствиями: от репутационных рисков для знаменитостей и частных лиц до проблем социальной стабильности и национальной безопасности [3].

Закон Бенфорда широко используется в мультимедийной криминалистике для обнаружения фальсификации изображений. Как оказалось, интенсивность света на реальных изображениях при определенных ограничениях точно подчиняется закону Бенфорда, что позволяет обнаружить применение фильтров и ретушь изображений [4]. Закон Бенфорда использовался для обнаружения усиления контраста и сокрытия артефактов, возникающих в результате операции повышения контрастности [5].

Соответствие закону Бенфорда квантованных коэффициентов дискретного косинусного преобразования (DCT) изображений, сжатых в формате JPEG, неоднократно демонстрировалось как мощный инструмент для обнаружения манипуляций с изображениями, в том числе в качестве индикатора морфинга изображения лица [6]. Несоответствие закону Бенфорда распределения первых цифр квантованных коэффициентов DCT, присутствующих в файлах JPEG, использовалось в стегоанализе изображений для обнаружения скрытых сообщений [7] и изображений с двойным сжатием, для локализации несанкционированного доступа и идентификации устройства-источника [8, 9].

Исследования применения закона Бенфорда к обнаружению сгенерированных изображений пока единичны. Соответствие закону Бенфорда квантованных коэффициентов DCT использовалось в работе [10] для обнаружения изображений, сгенерированных нейросетью GAN и имеющих формат JPEG. Предложенные сценарии обнаружения базировались на знании фактора качества JPEG и типа содержимого изображения, обучении системы на изображениях, сжатых в формате JPEG.

В работе представлен новый метод обнаружения сгенерированных изображений, основанный на анализе спектра мощности и энтропии изображений в сочетании с распределением Бенфорда.

1. Закон Бенфорда

Закон Бенфорда [11], также известный как закон первой цифры, или закон значащей цифры, является эмпирическим законом. Справедливость закона Бенфорда была продемонстрирована и подтверждена в различных областях. Примерами являются распределение результатов выборов [12], суммарная длительность нот в классических музыкальных произведениях [13], фальсификация научных данных [14], данные заболеваемости и смертности от коронавирусной инфекции [15] и т.д.

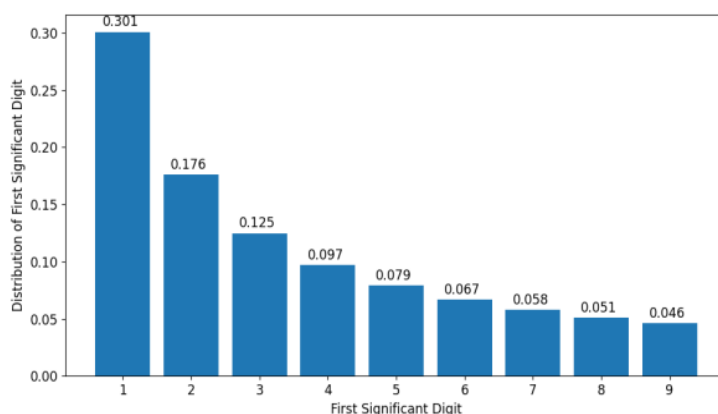


Рис. 1. Распределение Бенфорда
Fig. 1. Benford distribution

Впервые закон был обнаружен Ньюкомбом в 1881 г. [16] и переоткрыт Бенфордом в 1938 г. [17]. Статистическое объяснение этого закона предложено Хиллом в [18]. В нем говорится, что распределение вероятностей первых цифр x ($x = 1, 2, \dots, 9$) в наборе натуральных чисел является логарифмическим. В частности, если набор данных удовлетворяет закону Бенфорда, его значащие цифры будут иметь следующее распределение:

$$P(d) = \log_{10} \left(1 + \frac{1}{d} \right),$$

где $d = 1, 2, \dots, 9$. График распределения представлен на рис. 1.

2. Информационная энтропия изображения

Энтропия Шеннона успешно использовалась в оценке качества изображения. Также было показано, что различные типы искажений – JPEG-сжатие, шум, размытие и т.д. – по-разному влияют на пространственную энтропию [19].

Согласно теории энтропии Шеннона [20], энтропию изображения можно определить следующим образом. Пусть $I(x, y)$ – уровень выбранного канала изображения или уровень серого на изображении, где $x = 1, 2, \dots, M$, $y = 1, 2, \dots, N$ ($M \times N$ – размер изображения). Уровень $I(x, y) \in \{0, 1, \dots, L - 1\}$, где $L = 2^\beta$ и β – глубина изображения в пикселях (256, 65 536 и т.д.). Пусть число пикселей определенного уровня k равно n_k . Тогда энтропия изображения равна

$$E = - \sum_{k=0}^{L-1} p_k \log_2(p_k), \quad (1)$$

где $p_k = n_k / (M \times N)$ – нормированная вероятность появления каждого уровня выбранного канала изображения, который можно вычислить по гистограмме изображения.

Энтропия минимальна в однородных областях изображения и максимальна в областях, содержащих широкое распределение значений пикселей.

3. Спектральный анализ

Методы, основанные на спектральном анализе, нашли широкое применение в анализе и обработке изображений. Изображение обычно трактуется как двумерный сигнал, заданный на плоскости. Спектральные свойства изображения можно проанализировать с помощью дискретного преобразования Фурье. Для дискретного двумерного сигнала $f(x, y)$, представляющего отдельные цветовые каналы изображения размера $M \times N$, дискретное преобразование Фурье $F(k_x, k_y)$ определяется как

$$F(k_x, k_y) = \frac{1}{MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y) \exp \left(-i 2\pi \left(\frac{k_x x}{M} + \frac{k_y y}{N} \right) \right), \quad (2)$$

где x, y – позиция пикселя, $f(x, y)$ – значение пикселя выбранного канала изображения, k_x, k_y – пространственная частота. Значение $F(k_x, k_y)$ комплексное. Используя формулу Эйлера, $F(k_x, k_y)$ можно представить как

$$F(k_x, k_y) = |F(k_x, k_y)| \exp(i\varphi(k_x, k_y)), \quad (3)$$

где $|F(k_x, k_y)|$ – амплитуда, $\varphi(k_x, k_y)$ – фаза. Амплитуда изображения характеризует интенсивность различных частот в изображении и, таким образом, содержит информацию о геометрической структуре изображения в пространственной области. Фаза несет топологическую информацию о границах на изображении.

Спектр мощности изображения является важной статистической характеристикой изображения и определяется как

$$P(k_x, k_y) = |F(k_x, k_y)|^2. \quad (4)$$

После применения преобразования Фурье к изображению информация представлена в новой области, но по-прежнему содержит 2D-информацию. Размерность может быть уменьшена без существенной потери информации путем азимутального усреднения:

$$P(r) = \frac{1}{2\pi} \int_0^{2\pi} P(r, \theta) d\theta, \quad (5)$$

где

$$r = \sqrt{\frac{k_x^2 + k_y^2}{\frac{1}{4}(M^2 + N^2)}}, \quad \theta = \arctan 2(k_x, k_y).$$

Азимутальное усреднение спектра мощности позволяет преобразовать двумерное представление в одномерное. Азимутальное усреднение можно рассматривать как сжатие и усреднение схожих частотных компонент. На рис. 2 показаны графики нормализованного азимутально-усредненного спектра мощности реального изображения и изображений, сгенерированных нейросетями StyleGAN2 и StyleGAN3. Изображения взяты из датасетов [21–23].

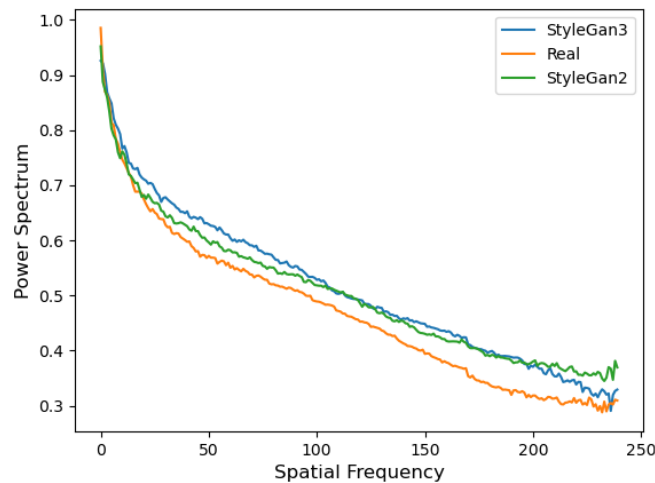


Рис. 2. Нормализованный азимутально-усредненный спектр мощности реального изображения и изображений, сгенерированных StyleGAN2 и StyleGAN3

Fig. 2. Normalized azimuth-averaged power spectrum of the real image and images generated by StyleGAN2 and StyleGAN3

4. Генеративно-сопоставительные сети StyleGAN2 и StyleGAN3

Deepfake-изображения чаще всего создаются с помощью методов, основанных на так называемых генеративно-сопоставительных сетях, или GAN (Generative Adversarial Network) [24]. Впервые представленные в 2014 г., сети GAN завоевали популярность благодаря своей способности создавать фотореалистичные изображения с нуля. Технология StyleGAN (Style Generative Adversarial Network) является расширением архитектуры GAN. Сети StyleGAN обеспечивают генерацию изображений на основе стилей, что позволяет контролировать синтез генерируемых изображений. Первый вариант технологии StyleGAN был опубликован в 2019 г. [25]. Сгенерированные StyleGAN лица содержали очевидные артефакты, в том числе артефакты, напоминающие капли воды (water droplets artifact). В 2020 г. была предложена технология StyleGAN2, позволившая добиться значительного улучшения качества изображений [26]. На основе StyleGAN2 появились многочисленные онлайн сервисы-генераторы, например известный person does not exist. В 2022 г. в официальной релиз Adobe Photoshop включены новые функции, называемые нейронными фильтрами, на основе технологии NVIDIA StyleGAN2.

При анимировании изображений сетями StyleGAN2 было обнаружено, что часть мелких деталей кажется зафиксированной. Этот дефект получил название texture sticking – «залипание текстур». Эффект связан с тем, что при генерации сети GAN усиливают алиасинг. Алиасинг – это эффект пере-

крытия частотных составляющих, возникающий из-за недостаточно большой частоты дискретизации. Алиасинг вызывает появление в амплитудно-частотном спектре частот, которых нет в исходном сигнале. Усиление алиасинга относится ко всем фильтрам, обычно используемым в глубоком обучении, и даже к высококачественным фильтрам, используемым при обработке изображений. В октябре 2021 г. компанией NVIDIA была опубликована архитектура StyleGAN3 (alias-free), главной целью которой стала адаптация технологии StyleGAN для применения в анимации и видео [27]. В StyleGAN3 для подавления артефактов алиасинга после слоя повышающей дискретизации применяется фильтр низких частот. Эксперименты показали: это приводит к значительному падению производительности и замедляет процесс обучения. Таким образом, архитектура StyleGAN2 пока остается наиболее используемой для генерации изображений.

5. Вычислительный эксперимент

Все вышесказанное мотивирует проанализировать спектр мощности сгенерированных нейросетями StyleGAN2 и StyleGAN3 изображений на соответствие распределению Бенфорда.

Рисунок 3 иллюстрирует общую архитектуру предлагаемого метода в контексте обнаружения сгенерированных цифровых изображений. Он содержит два основных блока: блок извлечения признаков и обучающий блок, в котором классификатор использует признаки, чтобы определить, настоящее изображение или сгенерированное.

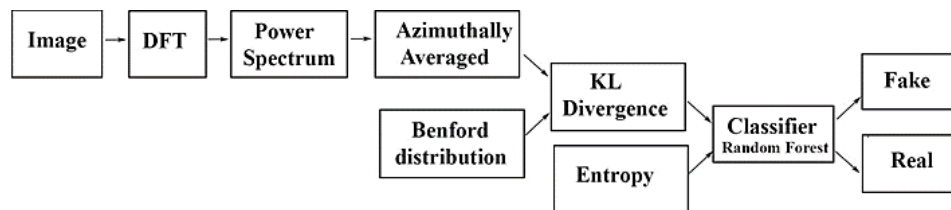


Рис 3. Общая архитектура метода на основе закона Бенфорда

Fig 3. General architecture of the method based on Benford's law

Входное изображение преобразуется в оттенки серого перед процедурой дискретного фурье-преобразования (DFT), вычисляемого по формуле (2). Далее вычисляется спектр мощности (Power Spectrum) изображения по формуле (4). Потом выполняется азимутальное усреднение (Azimuthally Averaged) спектра мощности по формуле (5). Энтропия изображения вычисляется по формуле (1). Близость значений азимутально-усредненного спектра мощности к закону Бенфорда оценивается с помощью дивергенции Кульбака–Лейблера, называемой также относительной энтропией, или величиной информационной расхожимости [28].

Дивергенция Кульбака–Лейблера определяется как

$$D_{KL}(p||q) = \sum_{i=1}^N p_i \ln \left(\frac{p_i}{q_i} \right). \quad (6)$$

В (6) q – распределение значений азимутально-усредненного спектр мощности, p – распределение (закон Бенфорда), с которым сопоставляется распределение q .

Дивергенция Кульбака–Лейблера интуитивно измеряет, насколько данное произвольное распределение отличается от истинного распределения. Если два распределения идеально совпадают, значение $D_{KL}(p||q) = 0$, в противном случае оно может принимать значения от 0 до ∞ .

Чтобы проверить применимость предложенного метода, было сформировано два датасета из произвольно выбранных изображений. Первый датасет состоит из ста изображений реальных лиц в разрешении $1\,024 \times 1\,024$ пикселя из набора Flickr-Faces-HQ Dataset (FFHQ) и ста изображений в том же разрешении, сгенерированных StyleGAN2, из набора изображений, предоставленных компанией NVIDIA. Изображения доступны для загрузки с ресурсов [21, 22]. Второй датасет состоит из ста изображений реальных лиц в разрешении $1\,024 \times 1\,024$ пикселя из набора данных Flickr-Faces-HQ

Dataset (FFHQ) и ста изображений в том же разрешении, сгенерированных StyleGAN3. Изображения доступны для загрузки с ресурсов [21, 23]. Назовем первый датасет FFHQ-StyleGAN2, второй – FFHQ-StyleGAN3. Все изображения в датасетах – изображения с близкими портретными кадрами. Как известно, генерация таких изображений удастся нейросетям наилучшим образом.

Разность между значениями азимутально-усредненного спектра мощности и закона Бенфорда каждого изображения рассматривалась как вероятностное распределение и характеризовалась дисперсией. Таким образом, каждое изображение характеризовалось 3-мерным вектором признаков: значением дивергенции Кульбака–Лейблера, значением энтропии и значением дисперсии. Полученный вектор признаков подавался на вход классификатора. Было апробировано несколько классификаторов, но наибольшую эффективность продемонстрировал классификатор RandomForest [29]. 70% изображений рассматривалось в качестве обучающего набора датасета, 30% – в качестве теста.

Чтобы проверить, являются ли энтропия и значения дивергенции Кульбака–Лейблера параметрами, по которым можно различить реальные и сгенерированные изображения, нормализованные значения энтропии Шеннона и значения дивергенции Кульбака–Лейблера были визуализированы на диаграмме рассеяния (рис. 4).

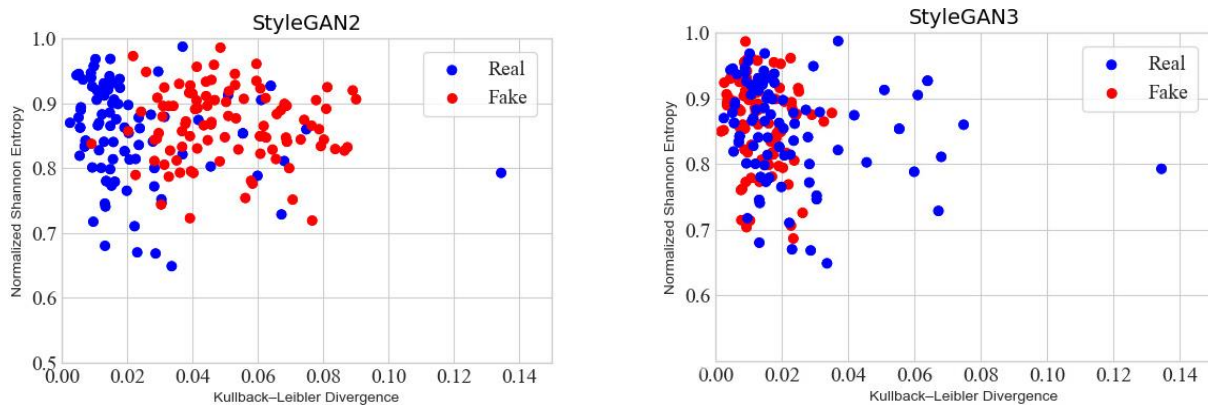


Рис. 4. Двумерный срез (нормализованные значения энтропии Шеннона и значения дивергенции Кульбака–Лейблера) реальных и сгенерированных изображений для датасетов FFHQ-StyleGAN2 (слева) и FFHQ-StyleGAN3 (справа)

Fig. 4. 2D slice (normalized Shannon entropy and Kullback-Leibler divergence values) of real and generated images for FFHQ-StyleGAN2 (left) and FFHQ-StyleGAN3 (right) datasets

Рисунок 4 показывает, что реальные и сгенерированные StyleGAN2 изображения могут быть разделены, хотя в центральной части графика все же есть явные перекрытия, в то время как реальные и сгенерированные StyleGAN3 изображения практически неразделимы.

На основании полученных данных была проведена процедура кластеризации для обнаружения сгенерированных изображений. Результаты этой процедуры можно увидеть в табл. 1. Сгенерированные StyleGAN2 изображения определялись с точностью выше 85% по метрике $F1\text{-score}$ [30]. Это превосходит результаты классификатора, базирующегося на соответствии закону Бенфорда квантованных коэффициентов DCT на том же наборе данных, точность которого заявлена авторами как 72,63% [10]. Точность предлагаемого метода, несмотря на его простоту, выше или сравнима с результатами некоторых других классификаторов, таких как Spex (основан на частотном анализе), PatchForensics (основан на анализе локальных артефактов изображений; см. обзор [31]).

Таблица 1

Метрики качества классификации датасетов

	Precision	Recall	$F1\text{-Score}$
Датасет FFHQ-StyleGAN2	0,87	0,85	0,857
Датасет FFHQ-StyleGAN3	0,59	0,59	0,596

Сгенерированные StyleGAN3 изображения определялись с точностью до ~ 60%. Опубликованные исследования применения закона Бенфорда к обнаружению изображений, сгенерированных

StyleGAN3, в настоящее время отсутствуют. Детекторы Spec, PatchForensics демонстрируют аналогичную недостаточно высокую точность (см. обзор [31]).

Чтобы проверить, было ли распределение значений азимутально-усредненного спектра мощности изображения и значений, полученных по закону Бенфорда, равным или нет, был рассчитан критерий согласия, основанный на критерии Крамера–Мизеса (Cramér–von Mises), где статистика теста измеряет расстояние между наблюдаемыми и ожидаемыми в соответствии с законом Бенфорда значениями. Согласно [32, 33] критерий Крамера–Мизеса рассчитывался как

$$W^2 = \frac{1}{N} \sum_{i=1}^9 (S_i - T_i)^2 t_i, \quad (7)$$

где $S_i = \sum_{j=1}^i q_j$ и $T_i = \sum_{j=1}^i p_j$ обозначают совокупные наблюдаемые и ожидаемые величины. Величина t_i

определяется как $t_i = (p_i + p_{i+1})/2$ ($i = 0, 1, \dots, 8$) и $t_9 = (p_9 + p_1)/2$.

Были рассмотрены те же датасеты FFHQ, StyleGAN2, StyleGAN3, каждый из $N = 100$ изображений [21–23]. Для каждого изображения была выдвинута гипотеза: подчиняется ли случайная величина значений азимутально-усредненного спектра мощности изображения распределению Бенфорда (нулевая гипотеза) или нет. Количество случаев, когда нулевая гипотеза верна, для каждого датасета представлено в табл. 2.

Таблица 2

Проверка гипотез: тест Крамера–Мизеса для $N = 100$ изображений

	Число изображений (нулевая гипотеза верна)		
	$\alpha = 0,1$	$\alpha = 0,05$	$\alpha = 0,01$
Датасет FFHQ	79	58	35
Датасет StyleGAN2	59	36	11
Датасет StyleGAN3	77	57	24

Результаты показывают, что изображения, сгенерированные StyleGAN2, часто не соответствуют закону Бенфорда, на этом основании их можно отличить от реальных изображений. Однако изображения, сгенерированные StyleGAN3, имеют практически тот же процент соответствия, что и реальные изображения на исследуемых датасетах.

Заключение

Поддельный цифровой мультимедийный контент все чаще используется для совершения киберпреступлений, разжигания социальных волнений и подрыва политических систем. И если раньше создать подделку было достаточно сложно, то в настоящее время программы для генерации изображений стали доступны обычным пользователям. В связи с этим особую актуальность приобретает разработка простых и доступных методов обнаружения deepfake-изображений.

В работе исследовалась возможность применения известного закона Бенфорда как инструмента обнаружения изображений, сгенерированных GAN. Был предложен вектор признаков, основанный на законе Бенфорда. Предложенный алгоритм имеет простую реализацию. Были проведены эксперименты на основе простого классификатора Random Forest. Рассматривались датасеты, сгенерированные нейросетями StyleGAN2 и StyleGAN3. Результаты экспериментов показали, что на основании предложенных характеристик можно отличить реальные изображения от сгенерированных нейросетью StyleGAN2 с достоверностью выше 85%. Изображения, сгенерированные StyleGAN3, труднее обнаружить предложенным способом, что мотивирует дальнейшие исследования по данной теме.

Список источников

1. Facing reality? Law enforcement and the challenge of deepfakes : an observatory report from the Europol Innovation Lab / Europol. Luxembourg : Publications Office of the European Union, 2022. 22 p. URL: <https://www.europol.europa.eu/cms/sites/default/>

- files/documents/Europol_Innovation_Lab_Facing_Reality_Law_Enforcement_And_The_Challenge_Of_Deepfakes.pdf (accessed: 22.03.2023).
2. Nightingale S., Farid H. Synthetic Faces Are More Trustworthy Than Real Faces // Journal of Vision. 2022. V. 22, № 14. Art. 3068.
3. Carlini N., Farid H. Evading deepfake-image detectors with white-and black-box attacks // Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 2020. P. 658–659.
4. Acebo E., Sbert M. Benford's law for natural and synthetic images // Proc. of the First Eurographics Conference on Computational Aesthetics in Graphics, Visualization and Imaging. 2005. P. 169–176.
5. Moin S.S., Islam S. Benford's law for detecting contrast enhancement // 2017 Fourth International Conference on Image Information Processing (ICIIP). IEEE, 2017. P. 1–4.
6. Makrushin A. et al. Generalized Benford's law for blind detection of morphed face images // Proc. of the 6th ACM Workshop on Information Hiding and Multimedia Security. 2018. P. 49–54.
7. Pevny T., Fridrich J. Detection of double-compression in jpeg images for applications in steganography // IEEE Transactions on Information Forensics and Security. 2008. V. 3, № 2. P. 247–258.
8. Milani S., Tagliasacchi M., Tubaro S. Discriminating multiple JPEG compressions using first digit features // APSIPA Transactions on Signal and Information Processing. 2014. V. 3. Art. e19.
9. Al-Bandawi H., Deng G. Classification of image distortion based on the generalized Benford's law // Multimedia Tools and Applications. 2019. V. 78. P. 25611–25628.
10. Bonettini N. et al. On the use of Benford's law to detect GAN-generated images // 2020 25th International Conference on Pattern Recognition (ICPR). IEEE, 2021. P. 5495–5502.
11. Nigrini M.J. Benford's Law: Applications for forensic accounting, auditing, and fraud detection. John Wiley & Sons, 2012. 586 p.
12. Deckert J., Myagkov M., Ordeshook P.C. Benford's Law and the detection of election fraud // Political Analysis. 2011. V. 19, № 3. P. 245–268.
13. Khosravani A., Rasinariu C. Emergence of Benford's Law in Music // arXiv preprint arXiv:1805.06506. 2018.
14. Diekmann A. Not the first digit! using benford's law to detect fraudulent scientific data // Journal of Applied Statistics. 2007. V. 34, № 3. P. 321–329.
15. Koch C., Okamura K. Benford's law and COVID-19 reporting // Economics Letters. 2020. V. 196. Art. 109573.
16. Newcomb S. Note on the frequency of use of the different digits in natural numbers // American Journal of Mathematics. 1881. V. 4, № 1. P. 39–40.
17. Benford F. The law of anomalous numbers // Proceedings of the American Philosophical Society. 1938. V. 78, № 4. P. 551–572.
18. Hill T.P. A statistical derivation of the significant-digit law // Statistical Science. 1995. V. 10, № 4. P. 354–363.
19. Liu L. et al. No-reference image quality assessment based on spatial and spectral entropies // Signal Proc.: Image Communication. 2014. V. 29, № 8. P. 856–863.
20. Shannon C.E. A mathematical theory of communication // The Bell System Technical Journal. 1948. V. 27, № 3. C. 379–423.
21. GitHub – NVlabs/FFHQ-dataset. URL: <https://github.com/NVLabs/ffhq-dataset> (accessed: 22.03.2023).
22. GitHub – NVlabs/StyleGAN2-dataset. URL: <https://github.com/NVLabs/stylegan2> (accessed: 22.03.2023).
23. Stylegan3-dataset. URL: <https://www.kaggle.com/datasets/showmik50/stylegan3-dataset> (accessed: 22.03.2023).
24. Goodfellow I. et al. Generative adversarial networks // Communications of the ACM. 2020. V. 63, № 11. P. 139–144.
25. Karras T., Laine S., Aila T. A style-based generator architecture for generative adversarial networks // Proc. of the IEEE / CVF Conference on Computer Vision and Pattern Recognition. 2019. P. 4401–4410.
26. Karras T., Laine S., Aittala M., Hellsten J., Lehtinen J., Aila T. Analyzing and Improving the Image Quality of StyleGAN // 2020 IEEE / CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2020. P. 8110–8119.
27. Karras T. et al. Alias-Free Generative Adversarial Networks // Advances in Neural Information Processing Systems. 2021. V. 34. P. 852–863.
28. Moreno P., Ho P., Vasconcelos N. A Kullback-Leibler divergence based kernel for SVM classification in multimedia applications // Advances in Neural Information Processing Systems. 2003. V. 16. P. 1–9.
29. Scikit-learn: Machine Learning in Python. URL: <http://scikit-learn.org> (accessed: 22.03.2023).
30. Ng A. Machine Learning Yearning : Technical Strategy for AI Engineers, In the Era of Deep Learning. Self-publishing, 2018. URL: <https://www.mlyearning.org> (accessed: 22.03.2023).
31. Corvi R. et al. On the detection of synthetic images generated by diffusion models // arXiv preprint arXiv:2211.00680. 2022.
32. Lesperance M. et al. Assessing conformance with Benford's law: Goodness-of-fit tests and simultaneous confidence intervals // PloS One. 2016. V. 11, № 3. Art. e0151235.
33. Шулеин В.П. Эффективные и робастные MD-оценки Крамера–Мизеса // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2010. № 3 (12). С. 107–121.

References

1. Publications Office of the European Union. (2022) *Facing reality? Law enforcement and the challenge of deepfakes, an observatory report from the Europol Innovation Lab*. [Online] Available from: https://www.europol.europa.eu/cms/sites/default/files/documents/Europol_Innovation_Lab_Facing_Reality_Law_Enforcement_And_The_Challenge_Of_Deepfakes.pdf (Accessed: 22nd March 2023).

2. Nightingale, S. & Farid, H. (2022) Synthetic Faces Are More Trustworthy Than Real Faces. *Journal of Vision*. 22(14). Art. 3068. DOI: 10.1167/jov.22.14.3068
3. Carlini, N. & Farid, H. (2020) Evading deepfake-image detectors with white-and black-box attacks. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*. pp. 658–659.
4. Acebo, E. & Sbert, M. (2005) Benford's law for natural and synthetic images. *Proceedings of the First Eurographics conference on Computational Aesthetics in Graphics. Visualization and Imaging*. pp. 169–176.
5. Moin, S. S. & Islam, S. (2017) Benford's law for detecting contrast enhancement. *2017 Fourth International Conference on Image Information Processing (ICIIP)*. pp. 1–4.
6. Makrushin, A. et al. (2018) Generalized Benford's law for blind detection of morphed face images. *Proceedings of the 6th ACM Workshop on Information Hiding and Multimedia Security*. pp. 49–54.
7. Pevny, T. & Fridrich, J. (2008) Detection of double-compression in jpeg images for applications in steganography. *IEEE Transactions on Information Forensics and Security*. 3(2). pp. 247–258.
8. Milani, S., Tagliasacchi, M. & Tubaro, S. (2014) Discriminating multiple JPEG compressions using first digit features. *APSIPA Transactions on Signal and Information Processing*. 3. Art. e19.
9. Al-Bandawi, H. & Deng, G. (2019) Classification of image distortion based on the generalized Benford's law. *Multimedia Tools and Applications*. 78. pp. 25611–25628
10. Bonettini, N. et al. (2021) On the use of Benford's law to detect GAN-generated images. *2020 25th international conference on pattern recognition (ICPR)*. IEEE. pp. 5495–5502.
11. Nigrini, M.J. (2012) *Benford's Law: Applications for forensic accounting, auditing, and fraud detection*. Vol. 586. John Wiley & Sons.
12. Deckert, J., Myagkov, M. & Ordeshook, P.C. (2011) Benford's Law and the detection of election fraud. *Political Analysis*. 19(3). pp. 245–268.
13. Khosravani, A. & Rasinariu, C. (2018) *Emergence of Benford's Law in Music*. arXiv preprint arXiv:1805.06506.
14. Diekmann, A. (2007) Not the first digit! using benford's law to detect fraudulent scientific data. *Journal of Applied Statistics*. 34(3). pp. 321–329.
15. Koch, C. & Okamura, K. (2020) Benford's law and COVID-19 reporting. *Economics Letters*. 196. Art. 109573.
16. Newcomb, S. (1881) Note on the frequency of use of the different digits in natural numbers. *American Journal of Mathematics*. 4(1). pp. 39–40.
17. Benford, F. (1938) The law of anomalous numbers. *Proceedings of the American Philosophical Society*. 78(4). pp. 551–572.
18. Hill, T.P. (1995) A statistical derivation of the significant-digit law. *Statistical Science*. 10(4). pp. 354–363.
19. Liu, L. et al. (2014) No-reference image quality assessment based on spatial and spectral entropies. *Signal processing: Image communication*. 29(8). pp. 856–863. DOI: 10.1016/j.image.2014.06.006
20. Shannon, C.E. (1948) A mathematical theory of communication. *The Bell System Technical Journal*. 27(3). pp. 379–423.
21. *GitHub – NVlabs/FFHQ-dataset*. [Online] Available from: <https://github.com/NVLabs/ffhq-dataset> (Accessed: 22nd March 2023).
22. *GitHub – NVlabs/StyleGAN2-dataset*. [Online] Available from: <https://github.com/NVLabs/stylegan2> (Accessed: 22nd March 2023).
23. *Stylegan3-dataset*. [Online] Available from: <https://www.kaggle.com/datasets/showmik50/stylegan3-dataset> (Accessed: 22nd March 2023).
24. Goodfellow, I. et al. (2020) Generative adversarial networks. *Communications of the ACM*. 63(11). pp. 139–144.
25. Karras, T., Laine, S. & Aila, T. (2019) A style-based generator architecture for generative adversarial networks. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. pp. 4401–4410.
26. Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J. & Aila, T. (2020) Analyzing and Improving the Image Quality of StyleGAN. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE. pp. 8110–8119.
27. Karras, T. et al. (2021) Alias-Free Generative Adversarial Networks. *Advances in Neural Information Processing Systems*. 34. pp. 852–863.
28. Moreno, P., Ho, P. & Vasconcelos, N. (2003) A Kullback-Leibler divergence based kernel for SVM classification in multimedia applications. *Advances in neural information processing systems*. 16. pp.1–9.
29. *Scikit-learn: Machine Learning in Python*. [Online] Available from: <http://scikit-learn.org> (Accessed: 22nd March 2023).
30. Ng, A. (2018) Machine Learning Yearning: Technical Strategy for AI Engineers, In the Era of Deep Learning. Self-publishing. [Online] Available from: <https://www.mlyearning.org>. (Accessed: 22nd March 2023).
31. Corvi, R. et al. (2022) *On the detection of synthetic images generated by diffusion models*. arXiv preprint arXiv:2211.00680.
32. Lesperance, M. et al. (2016) Assessing conformance with Benford's law: Goodness-of-fit tests and simultaneous confidence intervals. *PloS one*. 11(3). Art. e0151235.
33. Shulenin, V.P. (2010) Efficient and robust MD-estimators of Cramer – von Mises. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 3(12). pp. 107–121.

Информация об авторе:

Никитенкова Светлана Павловна – доцент кафедры «Безопасность информационных систем» радиофизического факультета Национального исследовательского Нижегородского государственного университета им. Н.И. Лобачевского (Нижний Новгород, Россия). E-mail: nikitenkova@rf.unn.ru

Автор заявляет об отсутствии конфликта интересов.

Information about the author:

Nikitenkova Svetlana P. (Candidate of Technical Sciences, Associate Professor, Lobachevsky State University of Nizhny Novgorod, Nizhny Novgorod, Russia). E-mail: nikitenkova@rf.unn.ru

The author declares no conflicts of interests.

Поступила в редакцию 10.04.2023; принята к публикации 04.09.2023

Received 10.04.2023; accepted for publication 04.09.2023

Научная статья

УДК 681.5; 004.4

doi: 10.17223/19988605/64/14

Принципы построения формального контекстно-логического конструктора запросов для точного поиска событий в больших массивах гетерогенной информации

Маис Паша оглы Фархадов¹, Екатерина Владимировна Панкратова²,
Ольга Викторовна Блинова³, Валентин Александрович Смирнов⁴

^{1,2,3} Институт проблем управления РАН им. В.А. Трапезникова, Москва, Россия

⁴ ООО «Спич Драйв», Санкт-Петербург, Россия

¹ mais.farhadov@gmail.com

² pankatya86@gmail.com

³ blinova_olga_v@mail.ru

⁴ valentin@speech-drive.ru

Аннотация. Представлен прототип формального контекстно-логического конструктора запросов, разработанного для поиска документов в базе данных гетерогенной информации. Дан краткий обзор основных проблем организации поиска по разнородным данным и существующих методов структурирования и обработки массивов таких данных. Представленный конструктор запросов в первую очередь ориентирован на работу с гетерогенными текстовыми данными для поиска в них релевантных результатов. Описаны ключевые этапы работы системы поиска документов, включая предварительную обработку текста, индексацию, создание запросов и процесс поиска. Определены основные принципы построения пользовательского интерфейса, приведены наиболее значительные преимущества прототипа для потенциальных пользователей. Предполагается внедрение конструктора для решения задач классификации документов, мониторинга информационного пространства, сбора и статистического анализа данных.

Ключевые слова: гетерогенные данные; методы текстового поиска; модели данных; контекстно-логический конструктор запросов; поисковые машины; автоматическая обработка текстовых документов.

Благодарности: Исследование выполнено в рамках научной программы Национального центра физики и математики (проект «Искусственный интеллект и большие данные в технических, промышленных, природных и социальных системах». НЦФМ-9-ИПУ-1).

Для цитирования: Фархадов М.П., Панкратова Е.В., Блинова О.В., Смирнов В.А. Принципы построения формального контекстно-логического конструктора запросов для точного поиска событий в больших массивах гетерогенной информации // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2023. № 64. С. 138–145. doi: 10.17223/19988605/64/14

Original article

doi: 10.17223/19988605/64/14

Principles of constructing a formal context-logical query constructor for accurate search of events in large arrays of heterogeneous information

Mais Pasha ogly Farkhadov¹, Ekaterina V. Pankratova², Olga V. Blinova³, Valentin A. Smirnov⁴

^{1,2,3} V.A. Trapeznikov Institute of Control Sciences of Russian Academy of Sciences, Moscow, Russian Federation

⁴ Speech Drive, LTD, Saint-Petersburg, Russian Federation

¹ mais.farhadov@gmail.com

² pankatya86@gmail.com

³ *blinova_olga_v@mail.ru*

⁴ *valentin@speech-drive.ru*

Abstract. The article presents the prototype of a formal context-logical query designer built for the document search in the database containing heterogeneous information. A concise overview is presented for the main problems in organizing the search for heterogeneous data, as well as the existing methods for structuring and processing arrays of such data. The presented query designer is primarily focused on processing heterogeneous text data, to search for relevant documents. The key components of the system are described including documents pre-processing, indexing, query build-up and the search. The main principles are defined for the query designer and the most important user benefits are highlighted. The query designer is planned to be implemented in solving the problems of document classification and in the monitoring systems of the information space, data collection and statistical analysis.

Keywords: heterogeneous data; text search methods; data models; search system; context-logical query designer; automatic processing of text documents.

Acknowledgments: The research was carried out within the framework of the scientific program of the National Center for Physics and Mathematics (the project “Research and development of artificial intelligence technologies for predictive modeling and decision support in technical, industrial, natural and social systems”).

For citation: Farkhadov, M.P., Pankratova, E.V., Blinova, O.V., Smirnov, V.A. (2023) Principles of constructing a formal context-logical query constructor for accurate search of events in large arrays of heterogeneous information. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naja tehnika i informatika – Tomsk State University Journal of Control and Computer Science*. 64. pp. 138–145. doi: 10.17223/19988605/64/14

Введение

Существующие на данный момент автоматизированные интеллектуальные системы обработки данных не позволяют в полной мере охватить и качественно проанализировать накопленные большие массивы гетерогенной информации. В контексте организации структур для хранения и обработки гетерогенных данных возможно различное понимание свойства гетерогенности. Физическая гетерогенность означает различные типы информации, форматы данных, подходы к цифровому представлению или используемым моделям, например, типов баз данных. Неоднородность представления данных обусловлена постоянным накоплением информационного фонда из различных источников и намерением последующего его использования для решения кардинально разных задач. В качестве гетерогенных данных могут выступать текстовые файлы, аудио-, видеоматериалы, графики, рисунки, таблицы, коды программ, диаграммы, схемы и т.д. Также гетерогенность может быть семантической, подразумевающей главные различия в названиях, значениях, смысловой наполненности информации [1].

При анализе больших массивов информации надо учитывать оба вида гетерогенности данных. В различных практических задачах в рамках одной базы данных могут иметь значение как один, так и оба вида гетерогенности. В случае физической гетерогенности в первую очередь необходимо решить вопрос о способах обработки и методах поиска, при семантической – о формулировке запросов, методах выборки обрабатываемых массивов, способах индексации.

Механизмы работы поисковых роботов отличаются от логики человеческого мышления, в этой связи формулировка запросов, при выполнении которой находится релевантная информация, – это сложная задача, требующая для своего решения специальных инструментов [2]. Контекстно-логические конструкторы запросов показывают хорошие результаты при обработке гетерогенных данных, так как являются инструментом составления частично формализованных запросов, который сохраняет интуитивно понятный интерфейс и естественную логику [3].

1. Методы структурирования, хранения и анализа гетерогенной информации

Основная задача структурирования и разработки методов обработки гетерогенных массивов данных – это обеспечение возможности их интеграции. Интеграция означает возможность работать с информацией как с единым массивом данных, обладающим целостностью, реализацию возможности для пользователя работать с данными как с единым информационным пространством [4].

Методы обработки гетерогенной информации можно разделить на две большие группы. К первой относится обработка запросов по каждой категории информации отдельно. Это позволяет использовать стандартные системы обработки данных в соответствии с их видом. Однако серьезным недостатком такого подхода являются трудности с объединением результатов обработки по подсистемам. Вторая группа методов – преобразование гетерогенных потоков к общему виду и последующая единая обработка по всей системе. Основными недостатками такого подхода являются ресурсоемкость преобразования информации и сложность учета особенностей каждого из рассматриваемых видов информации [5]. При использовании разрабатываемого контекстно-логического конструктора запросов используется второй метод: все поступающие данные в различных форматах преобразуются в единый текстовый формат.

2. Принципы построения контекстно-логического конструктора запросов

Разработанный прототип контекстно-логического конструктора универсален и может быть применен в различных системах анализа и обработки информации: мониторинге информационного пространства, анализе записей аудио- и видеоразговоров, анализе научных документов, стенограмм конференций, обработке архивов и т.д. Для обеспечения такой универсальности сформулированы ключевые принципы работы конструктора: модульность, инкапсуляция и эргономичность.

Модульность обеспечивает возможность неограниченного масштабирования сложности поисковых запросов путем комбинации атомарных поисковых запросов с использованием логических операторов. При обработке большого количества разнородных элементов иногда необходимо объединение большого количества правил для получения именно тех результатов, которые необходимы пользователю. При формировании длинных запросов их можно сохранять как шаблоны и потом изменять только необходимую часть, например для поиска по определенным критериям разных фраз из уже отобранной ранее категории документов.

Инкапсуляция позволяет обеспечить несколько уровней вложенности при создании элементов поисковых запросов; например, внутри одного из элементов можно использовать несколько различных контекстных правил, объединенных различными логическими операторами. Такие запросы часто формируются при поэтапном приближении к релевантным результатам путем уточнения запросов и поиска по найденным в предыдущих итерациях результатам.

Эргономичность облегчает создание и редактирование запросов через графический интерфейс, когда поисковый запрос формируется «на лету» в несколько кликов без необходимости программирования или сложных ручных операций. На практике эргономичность является одним из важнейших критериев для практического применения новых разработок, так как пользователи чаще всего не смогут оценить качество работы поискового робота, если будут испытывать сложности при использовании графического интерфейса.

Конструктор запросов предназначен для создания точных поисковых запросов к базе данных текстов. С помощью конструктора эксперт может выделить в массиве текстов интересующую его комбинацию слов и терминов, а также получить статистические данные о доле текстов, соответствующих поисковому критерию, в общей базе.

3. Этапы обработки гетерогенной информации для работы с конструктором запросов

Обработка документов состоит из нескольких этапов (рис. 1), только на последнем этапе пользователь непосредственно работает с прототипом конструктора запросов, все остальные этапы являются подготовительными.

Первый этап. Осуществляется загрузка коллекции документов из источников данных. Это могут быть как структурированные базы данных или разделы баз данных, так и неструктурированные коллекции документов, в том числе полученные в результате автоматизированного распознавания речи или изображений. Производится проверка времени изменения документов. На первом этапе ра-

боты алгоритма текстовые документы индексируются и сохраняются на сервере в виде бинарных документов, оптимизированных для последующего парсинга. Метаданные документов сохраняются в таблицу внутри базы данных с указанием уникального идентификатора документа и ключевых метаданных (автор документа, дата, источник, ключевые слова и др.).

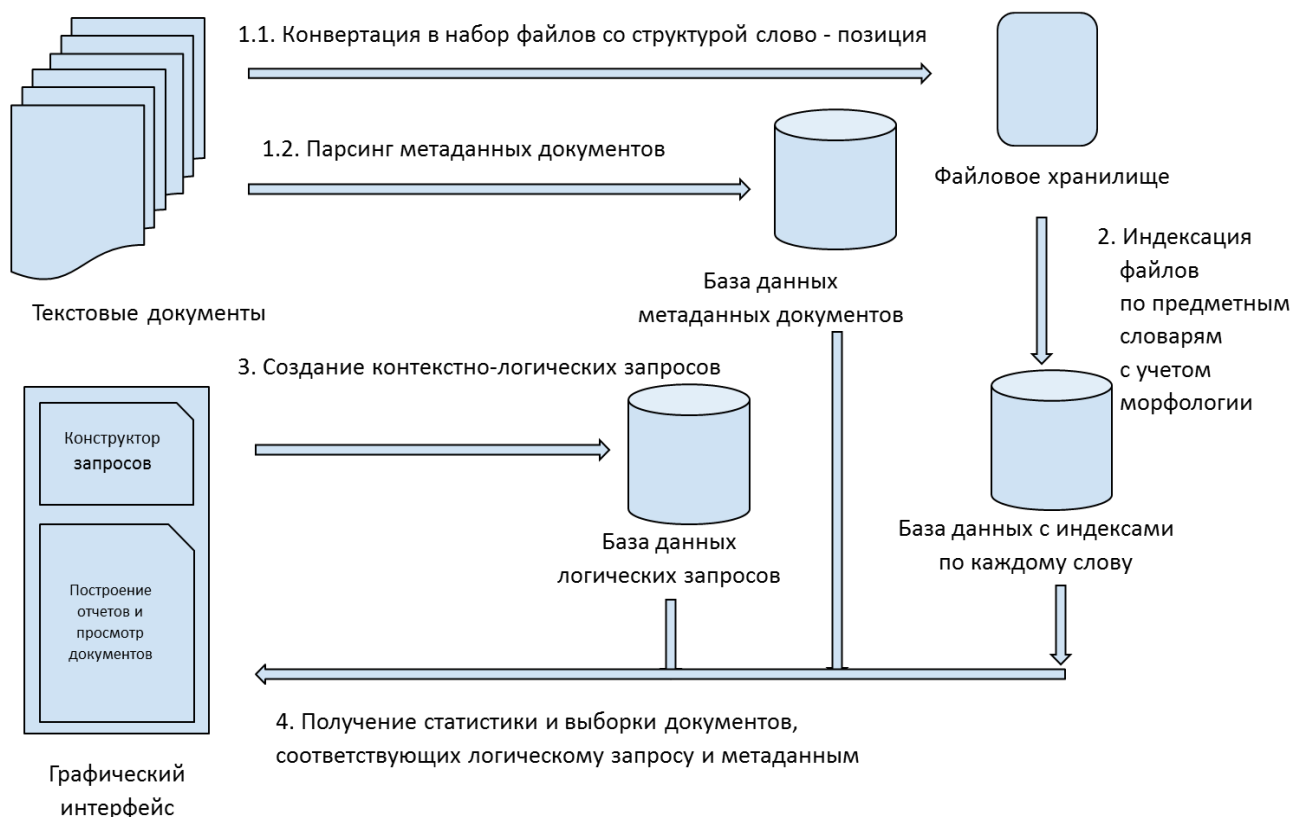


Рис. 1. Общая схема системы обработки документов с использованием конструктора запросов
Fig. 1. The general scheme of the document processing system using the query constructor

Второй этап. Все документы проходят подготовительную обработку. Цель такой обработки – очищение документа от данных, не нуждающихся в обработке, стемминг слов, т.е. поиск основы слова независимо от того, совпадает ли она с морфологическим корнем. Формирование структуры слово–позиция. Данный процесс также необходим для сокращения объема поискового индекса и способствует увеличению скорости поиска. На этом же этапе эксперт определяет словари: список слов и фраз, которые ему интересны с точки зрения содержания документов. Данные словари могут быть импортированы из уже существующего глоссария или базы знаний.

На третьем этапе тексты, загруженные в систему, анализируются на предмет наличия в них слов и фраз, входящих в созданные экспертом словари, с учетом словоформ. Так, например, если в словарь входит слово «атом», то в текстах будут найдены все формы данного слова («атома», «атомы» и др.). В случае обнаружения слова в тексте информация об этом сохраняется в отдельную таблицу, представляющую собой индекс для дальнейшего быстрого поиска по документам. При добавлении в словарь новых слов в фоновом режиме автоматически строится исторический индекс. Глубина индекса определяется настройками системы.

Четвертый, завершающий этап представляет собой непосредственно работу пользователя с конструктором запросов. При тестировании поиска в качестве запросов использовались следующие типы: информационный (поиск общей информации независимо от тематической направленности); транзакционный (поиск группы документов по определенным критериям); навигационный (поиск конкретного ресурса, обладающего заданными характеристиками). Подробнее работа пользователя с формальным контекстно-логическим конструктором запросов описана в следующем разделе [6].

4. Пользовательский интерфейс

Интерфейс прототипа конструктора запросов представляет собой набор взаимосвязанных форм, на которых отражены основные функции, операторы и правила. Запросы формируются в наглядной форме, все элементы форм предназначены для быстрого и удобного выбора существующих запросов или для создания новых.

В первой форме пользователю предлагается выбрать запрос из наиболее часто используемых в библиотеке запросов (рис. 2.). Предполагается, что при регулярной работе с системой в большинстве случаев использовать готовые запросы или формировать новые на основе существующих удобнее и быстрее, чем создавать запросы заново. Необходимый запрос можно выбрать в один клик. Выбранный запрос выделяется цветом. Также есть возможность перейти к полному списку доступных пользователю запросов.

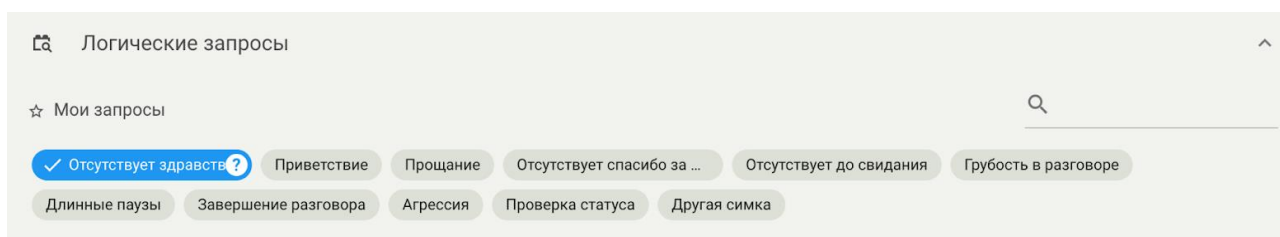


Рис. 2. Меню выбора часто используемых запросов

Fig. 2. Frequently used query selection menu

При создании новых запросов или модернизации существующих ключевой функцией является добавление новых блоков; каждый блок включает в себя указание логической связки (И / ИЛИ), наличие или отсутствие отрицания, а также конкретное контекстное правило (ПРИСУТСТВУЕТ, ОТСУТСТВУЕТ, ОКОЛО, НАЧАЛО, ОКОНЧАНИЕ) (рис. 3.).

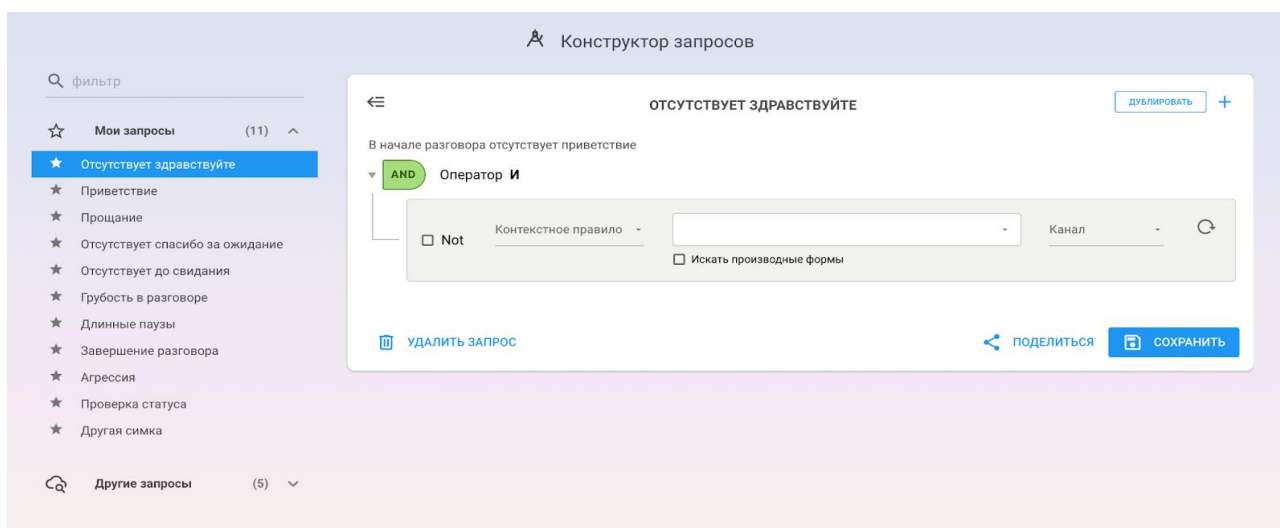


Рис. 3. Интерфейс создания нового запроса

Fig. 3. Interface for creating new request

В рамках запроса на первом этапе необходимо выбрать одно из контекстных правил, после чего пользователь вводит параметры для конкретного правила, а также выбирает, активировать или нет оператор отрицания. Активация отрицания в запросе происходит путем отметки в чек-боксе. Для добавления правила или оператора пользователь использует контекстное меню, которое также позволяет производить базовые операции с уже имеющимися в запросе правилами (вырезать, вставить, удалить). Одним из наиболее эффективных контекстных правил является правило ОКОЛО, позволяющее

искать в текстах события, связанные друг с другом. Расстояние между словами может указываться, например, в секундах (для аудиозаписи разговора) или в количестве слов (для текстовых документов) (рис. 4.).

Рис. 4. Использование оператора ОКОЛО

Fig. 4. Using the NEAR operator

Остальные правила в прототипе проверяют наличие или отсутствие слов. На рис 5 ниже приведен интерфейс использования логических правил И / ИЛИ. С их помощью можно указать взаимозаменяемые синонимичные конструкции, в случае если пользователь не знает точной формулировки фраз в искомых документах. Согласно указанным принципам запросы можно объединять.

Рис. 5. Интерфейс использования операторов И / ИЛИ

Fig. 5. Interface for using operators AND / OR

Прототип предполагает возможность поиска по словарям, а также по формам слов. В случае диалоговой речи (например, чат) прототип также предполагает опцию поиска в тексте конкретного участника диалога. Сформированный запрос сохраняется в каталог запросов. По результатам выполнения любого поискового запроса пользователь может просмотреть список документов, в которых найдены участки текста по соответствующим запросам, сами участки текста, собрать статистические данные по документам, соответствующим поисковым критериям, например число таких документов по отношению к общему числу документов в категории.

Заключение

Рассмотренные особенности работы с гетерогенными данными требуют разработки специфических инструментов для систематизации множества информационных объектов и поиска релевантных документов в массиве информации. Использование контекстно-логического конструктора запросов позволит учитывать морфологическую и синтаксическую вариативность текстов, упростить поиск по документам, в значительной степени автоматизировать классификацию и тем самым повысить скорость и качество работы эксперта с базой. В статье рассмотрены этапы работы системы обработки документов с использованием прототипа формального контекстно-логического конструктора запросов для точного поиска событий в тексте, общая схема обработки документов и создания хранилища метаданных. Прототип строится на основе принципов модульности, эргономичности и инкапсуляции, что обеспечивает его гибкость и адаптивность к большому числу практических задач. Рассмотрены основные структурные и интерфейсные особенности разработанного прототипа.

Список источников

1. Кашников А., Лядова Л. Интеграция гетерогенных источников данных на основе рекурсивной декомпозиции // *Information Technologies and Knowledge*. 2011. Т. 3, № 5. С. 274–284.
2. Обухова О.Л., Бирюкова Т.К., Гершкович М.М., Соловьев И.В., Чочиа А.П. Конструктор запросов интеллектуального поиска // *Электронные библиотеки: перспективные методы и технологии, электронные коллекции* : тр. XII Всерос. науч. конф. RCDL`2010, Казань, Россия 13–17 октября 2010 г. Казань : Казан. ун-т, 2010. С. 557–559.
3. Жижимов О.Л., Федотов А.М., Шокин Ю.И. Технологическая платформа массовой интеграции гетерогенных данных // *Вестник НГУ. Сер. Информационные технологии*. 2013. № 1. С. 24–41.
4. Бочаров А.В. Визуальный конструктор пользовательских поисковых запросов на основе комбинаторных тематических тезаурусов: реализация идеи // *Историческая информатика. Информационные технологии и математические методы в исторических исследованиях и образовании*. 2014. № 2-3. С. 10–20.
5. Куликов Г.Г., Шилина М.А., Бармин А.А., Старцев Г.В., Шамиданов Д.Г. Метод структурирования контента гетерогенного информационного пространства на основе формализованной модели предметной области для решения задач интеллектуального поиска // *Вестник ЮУрГУ. Сер. Компьютерные технологии, управление, радиоэлектроника*. 2018. № 1. С. 5–16.
6. Высочкин А.В., Квач А.И., Портнов Е.М. Разработка алгоритма поиска и ранжирования неструктурированной информации // *Инженерный вестник Дона*. 2019. № 1 (52). С. 57–61.

References

1. Kashnikov, A. & Lyadova, L. (2011) Integration of heterogeneous data sources based on recursive decomposition. *Information Technologies and Knowledge*. 3(5). pp. 274–284.
2. Obukhova, O.L., Biryukova, T.K., Gershkovich, M.M., Solovyov, I.V. & Chochia, A.P. (2010) Konstruktor zaprosov intellektual'nogo poiska [Constructor of intelligent search queries]. *Elektronnye biblioteki: perspektivnye metody i tekhnologii, elektronnye kolleksii* [Electronic libraries: promising methods and technologies, electronic collections]. Proc. of the 12th Conference RCDL`2010. Kazan', October 13–17, 2010. pp. 557–559.
3. Zhizhimov, O.L., Fedotov, A.M. & Shokin, Yu.I. (2013) Technology platform for mass integration of heterogeneous data. *Vestnik NGU. Ser. Informatsionnye tekhnologii*. 1. pp. 24–41.
4. Bocharov, A.V. (2014). Visual designer of user search queries based on combinatorial thematic thesauri: implementation of the idea. *Istoricheskaya informatika. Informatsionnye tekhnologii i matematicheskie metody v istoricheskikh issledovaniyakh i obrazovanii*. 2-3. pp. 10–20.
5. Kulikov, G.G., Shilina, M.A., Barmin, A.A., Startsev, G.V. & Shamidanov, D.V. (2018) Method of structuring the content of the heterogeneous information space based on the formalized model of the subject domain for solving the problems of intellectual search. *Vestnik YuUrGU. Ser. Komp'yuternye tekhnologii, upravlenie, radioelektronika*. 18(1). pp. 5–16.
6. Visochkin, A.V., Kvach, A.I. & Portnov, E.M. (2019) Development of an algorithm for searching and ranking unstructured information. *Inzhenernyy vestnik Dona*. 1(52). pp. 57–61.

Информация об авторах:

Фархадов Маис Паша оглы – доктор технических наук, заведующий лабораторией Института проблем управления им. В.А. Трапезникова РАН (Москва, Россия). E-mail: mais.farhadov@gmail.com

Панкратова Екатерина Владимировна – кандидат физико-математических наук, старший научный сотрудник Института проблем управления им. В.А. Трапезникова РАН (Москва, Россия). E-mail: pankatya86@gmail.com

Блинова Ольга Викторовна – научный сотрудник Института проблем управления им. В.А. Трапезникова РАН (Москва, Россия). E-mail: blinova_olga_v@mail.ru

Смирнов Валентин Александрович – кандидат технических наук, генеральный директор ООО «Спич Драйв» (Санкт-Петербург, Россия). E-mail: valentin@speech-drive.ru

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

Information about the authors:

Farkhadov Mais Pasha ogly (Doctor of Technical Sciences, Chief Researcher, V.A. Trapeznikov Institute of Control Sciences of Russian Academy of Science, Moscow, Russian Federation). E-mail: mais.farkhadov@gmail.com

Pankratova Ekaterina Vladimirovna (Candidate of Physical and Mathematical Sciences, Senior Researcher, V.A. Trapeznikov Institute of Control Sciences of Russian Academy of Science, Moscow, Russian Federation). E-mail: pankatya86@gmail.com

Blinova Olga Viktorovna (Researcher, V.A. Trapeznikov Institute of Control Sciences of Russian Academy of Science, Moscow, Russian Federation). E-mail: blinova_olga_v@mail.ru

Smirnov Valentin Aleksandrovich (Candidate of Technical Sciences, Director of Speech Drive, LTD, Saint-Petersburg, Russia). E-mail: valentin@speech-drive.ru

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

Received 15.03.2023; accepted for publication 04.09.2023

Поступила в редакцию 15.03.2023; принята к публикации 04.09.2023

Научный журнал

**ВЕСТНИК
ТОМСКОГО
ГОСУДАРСТВЕННОГО
УНИВЕРСИТЕТА**

**УПРАВЛЕНИЕ,
ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА
И ИНФОРМАТИКА**

**TOMSK STATE UNIVERSITY
JOURNAL OF CONTROL AND COMPUTER SCIENCE**

2023. № 64

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 29.09.2023 г. Формат 60х84¹/₈.
Гарнитура Times. Усл. печ. л. 17,2.
Тираж 250 экз. Заказ № 5606. Цена свободная.

Дата выхода в свет 23.10.2023 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