

**ВЕСТНИК
ТОМСКОГО
ГОСУДАРСТВЕННОГО
УНИВЕРСИТЕТА**

**УПРАВЛЕНИЕ,
ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА
И ИНФОРМАТИКА**

**TOMSK STATE UNIVERSITY
JOURNAL OF CONTROL AND COMPUTER SCIENCE**

Научный журнал

2022

№ 61

**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

Founder – Tomsk State University

EDITORIAL BOARD

Alexander Gortsev – Editor-in-Chief, Doctor of Sc., Prof., Head of the Applied Mathematics Department Tomsk State University. Tel: +73822529485

Valery Smagin – Deputy Editor-in-Chief, Doctor of Sc., Prof. of the Applied Mathematics Department Tomsk State University. Tel: +73822529485

Lyudmila Nezhelskaya – Executive Editor, Doctor. of Sc., Prof. of the Applied Mathematics Department Tomsk State University.
E-mail: vestnik_uvti@mail.tsu.ru . +73822529485

Sergey Vorobeychikov – Doctor of Sc., Prof. of the System Analysis and Mathematical Modeling Department Tomsk State University

Vladimir Vishnevsky – Doctor of Sc., Prof. Head of the laboratory Institute of Control Sciences of Russian Academy of Sciences (Moscow, Russia).

Gennady Koshkin – Doctor of Sc., Prof. of the System Analysis and Mathematical Modeling Department Tomsk State University

Yury Kostyuk – Doctor of Sc., Prof. of the Theoretical Informatics Department Tomsk State University

Anjela Matrosova – Doctor of Sc., Prof. of the Department of Computer Security Tomsk State University

Anatoly Nazarov – Doctor of Sc., Prof. of the Probability Theory and Mathematical Statistics Department Tomsk State University

Konstantin Samouylov – Doctor of Sc., Prof., Head of the Applied Probability and Informatics Department RUDN University (Moscow, Russia)

Eugene Semenkina – Doctor of Sc., Prof. System Analysis and Operations Research Department Reshetnev Siberian State University of Science and Technology (Krasnoyarsk, Russia)

Sergey Sushchenko – Doctor of Sc., Prof., Head of the Applied of Information Department Tomsk State University

Mais Farkhadov – Doctor of Sc., Head of the laboratory Institute of Control Sciences of Russian Academy of Sciences (Moscow, Russia).

Gurami Tsitsiasvili – Doctor of Sc., Prof., Chief researcher Institute for Applied Mathematics Far Eastern Branch of RAS, Prof. Far Eastern Federal University (Vladivostok, Russia)

Editorial address:

Institute of Applied Mathematics and Computer Science,
unit of Applied Mathematics

National Research Tomsk State University

36 Lenina Avenue, Tomsk, 634050

Telephone / fax: +73822529485

E-mail: vestnik_uvti@mail.tsu.ru

EDITORIAL COUNCIL

Ana Rosa Cavalli

PhD, Prof.
University VII Paris, France

Alexander Dudin

Doctor of Sc., Prof.
Belarusian State University

Minsk, Republic Belorussia

Enco Orzinger

PhD, Prof.
University of Rome, Italy

Paolo Prinetto

Prof. Politechnic Institute,
Torino, Italy

Raimund Ubar

Doctor of Sc., Prof.
University of Technology

Tallinn, Estonia

Nina Yevtushenko

Doctor of Sc., Prof.
Ivannikov V.P. ISP RAS

Moscow, Russia

Yervant Zorian

PhD, Fellow & Chief Architect,
Synopsys, Mountain View,
CA, USA

Учредитель – Томский государственный университет

РЕДАКЦИОННАЯ КОЛЛЕГИЯ

Горцев Александр Михайлович – гл. редактор, проф., д-р техн. наук, зав. кафедрой прикладной математики ТГУ. Тел. +73822529485

Смагин Валерий Иванович – зам. гл. редактора, проф., д-р техн. наук, проф. кафедры прикладной математики ТГУ. Тел. +73822529485

Нежельская Людмила Алексеевна – ответственный секретарь, доц., д-р физ.-мат. наук, проф. кафедры прикладной математики ТГУ
E-mail: vestnik_uvti@mail.tsu.ru . +73822529485

Воробейчиков Сергей Эрикович – д-р физ.-мат. наук, проф. кафедры системного анализа и математического моделирования ТГУ

Вишневский Владимир Миронович – проф., д-р техн. наук, зав. лабораторией Института проблем управления РАН (г. Москва)

Кожкин Геннадий Михайлович – проф., д-р физ.-мат. наук, проф. кафедры системного анализа и математического моделирования ТГУ

Костюк Юрий Леонидович – проф., д-р техн. наук, проф. кафедры теоретической информатики ТГУ

Матросова Анжела Юрьевна – проф., д-р техн. наук, проф. кафедры компьютерной безопасности ТГУ

Назаров Анатолий Андреевич – проф., д-р техн. наук, проф. кафедры теории вероятностей и математической статистики ТГУ

Самуйлов Константин Евгеньевич – проф., д-р техн. наук, зав. кафедрой прикладной информатики и теории вероятностей РУДН (г. Москва)

Семенин Евгений Станиславович – проф., д-р техн. наук, проф. каф. системного анализа и исследования операций, СибГУ им. акад. М.Ф. Решетнева (г. Красноярск)

Сушенко Сергей Петрович – проф., д-р техн. наук, зав. кафедрой прикладной информатики ТГУ

Фархадов Манс Паша Оглы – д-р техн. наук, зав. лабораторией Института проблем управления РАН (г. Москва)

Цициашвили Гурами Шалвович – проф., д-р физ.-мат. наук, гл. науч. сотр. Института прикладной математики ДВО РАН, проф. ДВФУ (г. Владивосток)

Адрес редакции и издателя: 634050, Томск, пр. Ленина, 36

Национальный исследовательский Томский государственный университет,
Институт прикладной математики и компьютерных наук,

отделение прикладной математики

Телефон / факс: +73822529485

E-mail: vestnik_uvti@mail.tsu.ru

РЕДАКЦИОННЫЙ СОВЕТ

Ана Роза Ковали

д-р. философии, проф.
Университет VII Париж, Франция

Александр Дудин

д.ф.-м.н., проф.
БГУ, Минск,
Республика Беларусь

Енцо Орзингер

д-р. философии, проф.
Римский университет, Италия

Паоло Принетто

проф. Политехнический институт
Турин, Италия

Раймонд Убар

д-р. проф.
Технологический университет,
Таллинн, Эстония

Нина Евтушенко

д-р. техн. наук, проф.
ИСП РАН им. Иванникова В.П.,
Москва, Россия

Ервант Зорян

д-р. философии,
гл. научный сотр.
Фирмы «Синописис»,
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

CONTENTS

УПРАВЛЕНИЕ ДИНАМИЧЕСКИМИ СИСТЕМАМИ

CONTROL OF DYNAMICAL SYSTEMS

Чырахова М.У.
 Об одной дискретной задаче управления, описываемой разностным уравнением типа Вольтерра и негладким критерием качества 4

Chirakhova M.U.
 On the one discrete control problem described by Volterra type difference equation and non-smooth quality criterion 4

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ

MATHEMATICAL MODELING

Михеев П.А., Поддубный В.В., Приступа П.В., Сущенко С.П., Карим П.Х.
 Быстродействие транспортного протокола с селективным режимом отказа в условиях соперничества за полосу пропускания тракта передачи данных 12

Mikheev P.A., Poddubny V.V., Pristupa P.V., Sushchenko S.P., Karim P.Kh.
 Performance of a selective failure mode transport protocol under conditions of contention for the bandwidth of the data transmission path 12

ОБРАБОТКА ИНФОРМАЦИИ

DATA PROCESSING

Abdushukurov A.A., Sayfulloyeva G.S.
 Asymptotic properties of modified empirical Kac processes under general random censorship model 26

Abdushukurov A.A., Sayfulloyeva G.S.
 Asymptotic properties of modified empirical Kac processes under general random censorship model 26

Воробьева Г.Р., Воробьев А.В.
 Подход к восстановлению временных рядов пространственных данных на основе информационно-корреляционного и геостатистического анализа 37

Vorobeva G.R., Vorobev A.V.
 An approach to the recovery of spatial data time series based on information-correlation and geostatistical analysis 37

Горцев А.М., Веткина А.В.
 Оценивание методом моментов параметра равномерного распределения длительности непродлевающегося случайного мертвого времени в рекуррентном полусинхронном потоке событий в общем и особом случаях 47

Gortsev A.M., Vetkina A.V.
 Estimation by method of moments of the parameter of the uniform distribution of the duration of unextendable random dead time in recurrent semi-synchronous events flow in general and special case 47

Добронев Б.С., Попова О.А.
 Распределения второго порядка: построение, операции, приложения 61

Dobronets B.S., Popova O.A.
 Second-order distributions: construction, operations, applications 61

Кеба А.В., Нежелская Л.А.
 Оценка длительности мертвого времени в обобщенном MAP-потоке событий с двумя состояниями 69

Keba A.V., Nezhel'skaya L.A.
 Estimation of dead time duration in generalized MAP with two states 69

Rouban A.I.
 The sensitivity coefficients for dynamic systems described by nonlinear difference generalized equations with the distributed memory and characterized by generalized functionals 81

Rouban A.I.
 The sensitivity coefficients for dynamic systems described by nonlinear difference generalized equations with the distributed memory and characterized by generalized functionals 81

Харченко И.К., Боровской И.Г., Шельмина Е.А.
 Использование ансамбля сверточных нейронных сетей для распознавания дорожных знаков 88

Kharchenko I.K., Borovskoy I.G., Shelmina E.A.
 Usage of convolutional neural network ensemble for traffic sign recognition 88

ИНФОРМАТИКА И ПРОГРАММИРОВАНИЕ

INFORMATICS AND PROGRAMMING

Григорьев Ю.Д.
 Перечисление симметричных ожерелий 97

Grigoriev Yu.D.
 Enumeration of symmetrical necklaces 97

Melnikov T.Y., Chuprov S.S., Lazarev E.A., Gataullin R.I., Viksnin I.I.
 Improving reputation and trust-based approach with reliability indicators for autonomous vehicles intergroup communication 108

Melnikov T.Y., Chuprov S.S., Lazarev E.A., Gataullin R.I., Viksnin I.I.
 Improving reputation and trust-based approach with reliability indicators for autonomous vehicles intergroup communication 108

Шашев Д.В., Шатравин В.В.
 Реализация сигмоидной функции активации с помощью концепции перестраиваемых вычислительных сред 117

Shashev D.V., Shatravin V.V.
 Implementation of the sigmoid activation function using the reconfigurable computing environments 117

ДИСКРЕТНЫЕ ФУНКЦИИ И АВТОМАТЫ

DISCRETE FUNCTION AND AUTOMATONS

Провкин В.А., Матросова А.Ю.
 Графовые представления множеств всех достижимых реакций комбинационной схемы 128

Provkin V.A., Matrosova A.Yu.
 Graph representations of the sets of all reachable reactions of the combinational circuit 128

УПРАВЛЕНИЕ ДИНАМИЧЕСКИМИ СИСТЕМАМИ

CONTROL OF DYNAMICAL SYSTEMS

Научная статья

УДК 517.977.56

doi: 10.17223/19988605/61/1

Об одной дискретной задаче управления, описываемой разностным уравнением типа Вольтерра и негладким критерием качества

Махнура Узейр кызы Чырахова

Институт систем управления НАН Азербайджана, Баку, Азербайджан, mahnuraciraqova@gmail.com

Аннотация. Рассматривается задача оптимального управления процессами, описываемая нелинейным разностным уравнением типа Вольтерра и негладким критерием качества. Доказаны необходимые условия оптимальности в терминах производных по направлениям. Отдельно изучена задача на минимум. Получены дискретные аналоги принципа максимума и линеаризованного принципа максимума.

Ключевые слова: разностное уравнение, негладкий функционал, производная по направлению, необходимое условие оптимальности, допустимое управление, задача на минимум, принцип максимума

Для цитирования: Чырахова М.У. Об одной дискретной задаче управления, описываемой разностным уравнением типа Вольтерра и негладким критерием качества // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2022. № 61. С. 4–11. doi: 10.17223/19988605/61/1

Original article

doi: 10.17223/19988605/61/1

On the one discrete control problem described by Volterra type difference equation and non-smooth quality criterion

Mahnura U. Chirakhova

National Academy of Sciences of Azerbaijan, Institute of Control Systems, Baku, Azerbaijan mahnuraciraqova@gmail.com

Abstract. The work devoted to study of one optimal control problem described by a system of nonlinear Volterra type difference equations, with a non-smooth quality functional assuming that the quality functional satisfies the Lipschitz condition and has directional derivatives.

Using method based on linearization of equation of process under several of assumptions necessary conditions of optimality in terms of derivatives in direction are established.

The minimax control problems has been separately considered. That is that problem of finding the minimum value of the functional of the type maximum (problem by minimax). Using the derivative formula in a direction of the function of maximum type under the assumption of the convexity of the analogue of the set of permissible velocities, the necessary condition of optimality of the type of the maximum principle is proved. In the case of the convexity of the control domain, an analogue of the linearized principle of maximum is proved.

Keywords: difference equation; non-smooth functional; directional derivative; necessary optimality condition; admissible control; problem of minimax; maximin principle

For citation: Chirakhova, M.U. (2022) On the one discrete control problem described by Volterra type difference equation and non-smooth quality criterion. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naja tehnika i informatika – Tomsk State University Journal of Control and Computer Science*. 61. pp. 4–11. doi: 10.17223/19988605/61/1

В работе [1] изучена одна задача оптимального управления, описываемая системой разностных уравнений типа Вольтерра с гладким терминальным критерием качества, доказан ряд необходимых условий оптимальности типа принципа максимума Л.С. Понтрягина и также исследован случай их вырождения [2–5].

В работе [6] установлен ряд необходимых условий для существования седловой точки.

В предлагаемой работе изучается случай негладкого функционала качества. Доказаны необходимые условия оптимальности в терминах производных по направлениям.

Отдельно изучена задача на минимакс.

1. Постановка задачи

Пусть $T = \{t_0, t_0 + 1, \dots, t_1\}$ – заданный дискретный отрезок времени, $U \subset R^r$ – заданное непустое и ограниченное множество, $u(t)$ – r -мерная дискретная управляющая функция, удовлетворяющая ограничению

$$u(t) \in U, \quad t \in T, \quad (1)$$

которую назовем допустимым управлением, а $\Phi(x)$ – заданная скалярная функция, удовлетворяющая условию Липшица и имеющая производные по любому направлению.

Предположим, что управляемый процесс описывается системой нелинейных разностных уравнений типа Вольтерра

$$x(t) = \sum_{\tau=t_0}^t f(t, \tau, x(\tau), u(\tau)), \quad t \in T. \quad (2)$$

Здесь $f(t, \tau, x, u)$ – заданная n -мерная вектор-функция, непрерывная по (x, u) с частными производными по x при всех (t, τ) .

Задача оптимального управления заключается в нахождении минимального значения терминального функционала

$$S(u) = \Phi(x(t_1)) \quad (3)$$

при ограничениях (1), (2).

Допустимое управление $u(t)$, доставляющее минимальное значение функционалу (3) при ограничениях (1), (2), называется оптимальным управлением.

2. Необходимые условия оптимальности в терминах производных по направлениям

Пусть $(u(t), x(t))$ – фиксированный допустимый процесс, множество

$$f(t, \tau, x(\tau), U) = \{\alpha : \alpha = f(t, \tau, x(\tau), v), v \in U\} \quad (4)$$

выпукло при всех (t, τ) , а $\varepsilon \in [0, 1)$ – произвольное число.

Через $x(t; \varepsilon)$ обозначим произвольное допустимое управление, такое что

$$x(t; \varepsilon) = \sum_{\tau=t_0}^t f(t, \tau, x(\tau; \varepsilon), u(\tau; \varepsilon)) = \sum_{\tau=t_0}^t [(1 - \varepsilon) f(t, \tau, x(\tau; \varepsilon), u(\tau))] + \sum_{\tau=t_0}^t \varepsilon f(t, \tau, x(\tau; \varepsilon), v(\tau)), \quad (5)$$

где $v(t)$ – произвольное допустимое управление.

Это возможно в силу выпуклости множества (4).

Введем обозначение:

$$y(t) = \left. \frac{\partial x(t; \varepsilon)}{\partial \varepsilon} \right|_{\varepsilon=0}. \quad (6)$$

Из (5), в силу условий гладкости, наложенных на правую часть уравнения (2), следует, что вектор-функция $y(t)$, определяемая формулой (6), является решением уравнения в вариациях

$$y(t) = \sum_{\tau=t_0}^t \left[f_x(t, \tau, x(\tau), u(\tau)) y(\tau) + (f(t, \tau, x(\tau), v(\tau)) - f(t, \tau, x(\tau), u(\tau))) \right]. \quad (7)$$

Из (6) ясно, что

$$x(t; \varepsilon) = x(t) + \varepsilon y(t) + o(\varepsilon; t). \quad (8)$$

Вычислим специальное приращение функционала качества (3), соответствующее допустимым управлениям $u(t; \varepsilon)$, $u(t)$.

Имеем

$$\begin{aligned} S(u(t; \varepsilon)) - S(u(t)) &= \Phi(x(t_1; \varepsilon)) - \Phi(x(t_1)) = \left[\Phi(x(t_1) + \varepsilon y(t_1) + o(t_1; \varepsilon)) - \right. \\ &\quad \left. - \Phi(x(t_1) + \varepsilon y(t_1)) \right] - \left[\Phi(x(t_1) + \varepsilon y(t_1)) - \Phi(x(t_1)) \right]. \end{aligned} \quad (9)$$

По предположению функция $\Phi(x)$ удовлетворяет условию Липшица. Поэтому получаем, что

$$\left| \Phi(x(t_1) + \varepsilon y(t_1) + o(\varepsilon)) - \Phi(x(t_1) + \varepsilon y(t_1)) \right| \leq o_1(\varepsilon). \quad (10)$$

Далее, используя определение производной по направлениям, получаем, что

$$\Phi(x(t_1) + \varepsilon y(t_1)) - \Phi(x(t_1)) = \varepsilon \frac{\partial \Phi(x(t_1))}{\partial y(t_1)} + o_2(\varepsilon). \quad (11)$$

Если предполагать, что допустимое управление $u(t)$ оптимальное, то из (11), учитывая соотношения (10), (11), получаем

$$\frac{\partial \Phi(x(t_1))}{\partial y(t_1)} + o_3(\varepsilon) \geq 0. \quad (12)$$

Из этого неравенства следует

Теорема 1. Если множество (4) выпуклое, то для оптимальности допустимого управления $u(t)$ необходимо, чтобы неравенство

$$\frac{\partial \Phi(x(t_1))}{\partial y(t_1)} \geq 0 \quad (13)$$

выполнялось для всех $v(t) \in U$, $t \in T$.

Неравенство (13) является общим необходимым условием оптимальности и носит неявный характер.

Перейдем к конкретизации полученного необходимого условия оптимальности.

Уравнение в вариациях является линейным неоднородным разностным уравнением. Из результатов работ [7, 8] следует, что решение $y(t)$ уравнения в вариациях допускает представление

$$\begin{aligned} y(t) &= \sum_{\tau=t_0}^t (f(t, \tau, x(\tau), v(\tau)) - f(t, \tau, x(\tau), u(\tau))) - \\ &\quad - \sum_{\tau=t_0}^t \sum_{s=t_0}^{\tau} R(t, \tau) (f(t, s, x(s), v(s)) - f(t, s, x(s), u(s))), \end{aligned} \quad (14)$$

где $R(\tau, t)$ ($n \times n$) – матричная функция, являющаяся решением матричного разностного уравнения

$$R(t, \tau) = \sum_{s=t}^{\tau} R(t, s) (f_x(s, t, x(t), u(t)) - f_x(\tau, t, x(t), u(t))), t_0 \leq t \leq \tau.$$

По аналогии с [7, 8] можно показать, что матричная функция $R(\tau, t)$ является также решением уравнения

$$R(t, \tau) = \sum_{s=t}^{\tau} (f_x(\tau, s, x(s), u(s))R(s, \tau) - f_x(\tau, t, x(t), u(t))). \quad (15)$$

С помощью дискретного аналога теоремы Фубини (см., напр.: [8]) представление (14) преобразуется к виду:

$$y(t) = \sum_{\tau=t_0}^t \left[(f(t, \tau, x(\tau), v(\tau)) - f(t, \tau, x(\tau), u(\tau))) - \sum_{s=t}^{\tau} R(t, s) (f(s, \tau, x(\tau), v(\tau)) - f(s, \tau, x(\tau), u(\tau))) \right]. \quad (16)$$

Положим

$$L_1(v) = \sum_{\tau=t_0}^{t_1} (f(t_1, \tau, x(\tau), v(\tau)) - f(t_1, \tau, x(\tau), u(\tau))) - \sum_{\tau=t_0}^{t_1} R(t_1, \tau) (f(\tau, t, x(t), v(t)) - f(\tau, t, x(t), u(t))). \quad (17)$$

С учетом (13), (17) теорема 1 может быть сформулирована в виде:

Теорема 2. Если множество (4) выпуклое, то для оптимальности допустимого управления $u(t)$ в рассматриваемой задаче необходимо, чтобы неравенство

$$\frac{\partial \Phi(x(t_1))}{\partial L_1(v)} \geq 0 \quad (18)$$

выполнялось для всех $v(t) \in U, t \in T$.

3. Линеаризованное необходимое условие оптимальности

Предположим, что вектор-функция $f(t, \tau, x, u)$ непрерывно-дифференцируема по (x, u) , а множество U выпукло. В силу сделанных предположений можно написать «возмущенную» систему вида:

$$x(t; \mu) = \sum_{\tau=t_0}^t f(t, \tau, x(\tau; \mu), u(\tau; \mu)) \equiv \sum_{\tau=t_0}^t (1 - \varepsilon) f(t, \tau, x(\tau; \mu), \mu v(\tau) + (1 - \mu)u(\tau)), \quad (19)$$

где $\mu \in [0, 1)$ – произвольное число, а $v(t)$ – произвольное допустимое управление.

Предположим

$$z(t, v) = \left. \frac{\partial x(t; \mu)}{\partial \mu} \right|_{\mu=0}. \quad (20)$$

Из (19) следует, что $z(t, v)$, определяемое формулой (20), является решением следующего уравнения в вариациях:

$$z(t, v) = \sum_{\tau=t_0}^t [f_x(t, \tau, x(\tau), v(\tau))z(\tau, v) + f_u(t, \tau, x(\tau), u(\tau))(v(\tau) - u(\tau))]. \quad (21)$$

Запишем специальное приращение критерия качества, соответствующее допустимым управлениям $u(t; \mu), u(t)$.

Имеем

$$S(u(t; \mu)) - S(u(t)) = [\Phi(x(t_1) + \mu z(t_1, v) + o(\mu)) - \Phi(x(t_1) + \mu z(t_1, v))] = \mu \frac{\partial \Phi(x(t_1))}{\partial z(t_1, v)} + o(\mu). \quad (22)$$

Из полученного разложения следует

Теорема 3. Если множество U выпукло, то для оптимальности допустимого управления $u(t)$ в рассматриваемой задаче необходимо, чтобы неравенство

$$\frac{\partial \Phi(x(t_1))}{\partial z(t_1, v)} \geq 0 \quad (23)$$

выполнялось для всех $v(t) \in U, t \in T$.

Конкретизируем полученное необходимое условие оптимальности.

Решение $z(t, v)$ линеаризованного уравнения (21) представимо в виде

$$z(t, v) = \sum_{\tau=t_0}^t f_u(t, \tau, x(\tau), u(\tau))(v(\tau) - u(\tau)) - \sum_{\tau=t_0}^t \sum_{s=\tau}^{\tau} [R(t, s) f_u(s, \tau, x(\tau), u(\tau))](v(\tau) - u(\tau)), \quad (24)$$

где $R(\tau, t)$ решение уравнения (15).

Введя обозначение

$$Q(t, \tau) = f_u(t, \tau, x(\tau), u(\tau)) + \sum_{s=\tau}^t [R(t, s) f_u(s, \tau, x(\tau), u(\tau))],$$

из (24) получаем, что

$$z(t, v) = \sum_{\tau=t_0}^t Q(t, \tau)(v(\tau) - u(\tau)).$$

Следовательно,

$$z(t_1, v) = \sum_{\tau=t_0}^{t_1} Q(t_1, \tau)(v(\tau) - u(\tau)).$$

Положим

$$L_2(v) = \sum_{\tau=t_0}^{t_1} Q(t_1, \tau)(v(\tau) - u(\tau)).$$

Тогда теорема 3 может быть сформулирована в виде:

Теорема 4. Если множество U выпукло, то для оптимальности допустимого управления $u(t)$ в рассматриваемой задаче необходимо, чтобы неравенство

$$\frac{\partial \Phi(x(t_1))}{\partial L_2(v)} \geq 0 \quad (25)$$

выполнялось для всех $v(t) \in U, t \in T$.

4. Необходимое условие оптимальности в задаче на минимакс

Рассмотрим задачу о минимуме функционала типа максимум

$$S(u) = \max_{a \in A} \varphi(x(t_1), a), \quad (26)$$

при ограничениях (1)–(5), где $\varphi(x, a)$ – заданная, непрерывно-дифференцируемая по x скалярная функция, $A \in R^m$ – заданное непустое и ограниченное множество.

Задачи оптимального управления с функционалом типа (26) обычно называются задачами на минимакс. В работах [2, 9, 10] при различных предположениях изучен ряд свойств функций типа максимум или же минимум. В частности, установлена формула для производной по направлению функций типа максимум.

Используя дифференциальные свойства функций типа максимум, получим необходимое условие оптимальности в задаче на минимакс.

Пусть $u(t)$ в задаче (1)–(5) является оптимальным управлением. Тогда из теоремы 2 получаем, что вдоль процесса $(u(t), x(t))$

$$\frac{\partial}{\partial L_1(v)} \left(\max_{a \in A} \varphi(x(t_1), a) \right) \geq 0.$$

Пусть $A(x)$ – множество максимумов функции $\varphi(x, a)$, т.е.

$$A(x(t_1)) = \left\{ a \in A : \varphi(x(t_1), a) = \max_{\bar{a} \in A} \varphi(x(t_1), \bar{a}) \right\} \max_{a \in A} \varphi(x(t_1), a).$$

Учитывая известную формулу о производной по направлению функции типа максимум (см., напр.: [9, 10]), из неравенства (18) получаем, что

$$\frac{\partial}{\partial L_1(v)} \left(\max_{a \in A} \varphi(x(t_1), a) \right) = \max_{a \in A(x(t_1))} \frac{\partial \varphi'(x(t_1), a)}{\partial x} L_1(v) \geq 0. \quad (27)$$

Принимая во внимание выражение $L_1(v)$ из (17), неравенство (27) записывается в виде:

$$\begin{aligned} \max_{a \in A(x(t_1))} \frac{\partial \varphi'(x(t_1), a)}{\partial x} \left[\sum_{\tau=t_0}^{t_1} (f(t_1, t, x(t), v(t)) - f(t_1, t, x(t), u(t))) - \right. \\ \left. - \sum_{\tau=t_0}^{t_1} R(t_1, \tau) (f(\tau, t, x(t), v(t)) - f(\tau, t, x(t), u(t))) \right] \geq 0. \end{aligned} \quad (28)$$

Введем обозначения

$$\psi_a(t) = \frac{\partial \varphi'(x(t_1), a)}{\partial x} R(t_1, \tau),$$

$$H(t, x(t), u(t), \psi_a(t)) = -\varphi'_x(x(t_1), a) f(t_1, t, x(t), u(t)) + \sum_{\tau=t_0}^{t_1} \psi_a(\tau) f(\tau, t, x(t), u(t)).$$

Тогда неравенство (28) записывается в виде:

$$\min_{a \in A(x(t_1))} (H(t_1, x(t_1), v(t_1), \psi_a(t_1)) - H(t_1, x(t_1), u(t_1), \psi_a(t_1))) \leq 0. \quad (29)$$

Сформулируем полученный результат:

Теорема 5. Для оптимальности допустимого управления $u(t)$ в рассматриваемой задаче (1), (2), (26) необходимо, чтобы неравенство (29) выполнялось для всех $v(t) \in U$, $t \in T$.

Перейдем теперь к доказательству линеаризованного условия максимина при выполнении предположений теоремы 3.

При выполнении условий теоремы 3 получаем, что вдоль оптимального управления $u(t)$ выполняется неравенство

$$\max_{a \in A(x(t_1))} \frac{\partial \varphi'(x(t_1), a)}{\partial x} L_2(v) \geq 0.$$

Учитывая выражение $L_2(v)$, последнее неравенство записывается в виде:

$$\max_{a \in A(x(t_1))} \frac{\partial \varphi'(x(t_1), a)}{\partial x} \sum_{\tau=t_0}^{t_1} Q(t_1, \tau) (v(t) - u(t)) \geq 0$$

для всех $v(t) \in U$, $t \in T$.

Следовательно, учитывая вид функции Гамильтона Понтрягина, приходим к следующему утверждению:

Теорема 6. Если множество U выпуклое, а $f(t, \tau, x, u)$ имеет также по u непрерывную производную, то для оптимальности допустимого управления $u(t)$ в задаче (1)–(2), (26) необходимо, чтобы неравенство

$$\min_{a \in A(x(t_1))} H'_u(t, x(t), u(t), \psi_a(t)) (v(t) - u(t)) \leq 0 \quad (30)$$

выполнялось для всех $v(t) \in U$, $t \in T$.

Неравенство (30) является аналогом линеаризованного принципа максимина, доказанный в работе [11], для задачи оптимального управления обыкновенными дифференциальными уравнениями.

Заключение

В работе рассматривается дискретная задача оптимального управления, описываемая системой разностных уравнений Вольтерра и недифференцируемым функционалом качества. Получены общие необходимые условия оптимальности, которые охватывают задачу в минимакс, также используя дифференциальное свойство функций максимума, доказаны необходимые условия оптимальности в форме принципа максимина и линеаризованного принципа максимина.

Список источников

1. Мансимов К.Б., Чырахова М.У. Об одной задаче управления процессами описываемыми разностными уравнениями Вольтерра // Вестник Бакинского государственного университета. Сер. физико-математических наук. 2019. № 1. С. 66–89.
2. Габасов Р., Кириллова Ф.М., Альсевич В.В. Методы оптимизации. Минск : Четыре четверти, 2011. 472 с.
3. Габасов Р., Кириллова Ф.М. Особые оптимальные управления. М. : Либроком, 2011, 256 с.
4. Мансимов К.Б. Дискретные системы. Баку : Изд-во Бакинского гос. ун-та, 2013. 151 с.
5. Васильев Ф.П. Методы оптимизации. М. : Факториал пресс. 2002. 814 с.
6. Мансимов К.Б., Чырахова М.У. О необходимых условиях существования седловой точки в задаче оптимального управления для систем разностных уравнений типа Вольтерра // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2022. № 59. С. 4–12.
7. Колмановский В.Б. Устойчивость дискретных уравнений Вольтерра // Доклады РАН. 1996. Т. 349, № 5. С. 40–61.
8. Колмановский В.Б. Об асимптотических свойствах решений некоторых нелинейных систем Вольтерра // Автоматика и телемеханика. 2000. № 4. С. 42–50.
9. Демьянов В.Ф., Рубинов А.М. Основы негладкого анализа и квазидифференциальное исчисление. М. : Наука, 1990. 432 с.
10. Демьянов В.Ф., Малоземов В.Н. Введение в минимакс. М. : Наука, 1972. 368 с.
11. Габасов Р., Кириллова Ф.М. Принцип максимума в теории оптимального управления. М. : Либроком, 2011. 272 с.

References

1. Mansimov, K.B. & Chirakhova, M.U. (2019) On one problem of process control described by Wroltterra difference equations. *Vestnik Bakinskogo gosudarstvennogo universiteta. Ser. fiziko-matematicheskikh nauk – Bulletin of the Baku State University. Univer. ser. Phys.-Math. Sciences.* 1. pp. 66–89.
2. Gabasov, R., Kirillova, F.M. & Alseovich, V.V. (2011) *Metody optimizatsii* [Optimization methods]. Minsk: Chetyre chetverti.
3. Gabasov, R. & Kirillova, F.M. (2011) *Osobyie optimal'nye upravleniya* [Singular optimal controls]. Moscow: Librokom.
4. Mansimov, K.B. (2013) *Diskretnye sistemy* [Discrete Systems]. Baku: Baku State University.
5. Vasilyev, F.P. (2002) *Metody optimizatsii* [Optimization methods]. Moscow: Faktorial press.
6. Mansimov, K.B. & Chyrahkova, M.U. (2022) On necessary conditions for the existence of a saddle point in the problem of optimal control for systems of difference equations of Volterra type. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science.* 59. pp. 4–12. DOI: 10.17223/19988605/59/2
7. Kolmanovskiy, V.B. (1996) Ustoychivost' diskretnykh uravneniy Vol'terra [Stability of discrete Volterra equations]. *Dokl. RAN.* 349(5). pp. 40–61.
8. Kolmanovskiy, V.B. (2000) On the asymptotic properties of solutions of some nonlinear volterra equations. *Avtomatika i telemekhanika.* 4. pp. 42–50.
9. Demyanov, V.F. & Rubinov, A.M. (1990) *Osnovy negladkogo analiza i kvazidifferentsial'noe ischislenie* [Fundamentals of non-smooth analysis and quasi-differential calculus]. Moscow: Nauka.
10. Demyanov, V.F. & Malozemov, V.N. (1972) *Vvedenie v minimaks* [Introduction to minimax]. Moscow: Nauka.
11. Gabasov, R. & Kirillova, F.M. (2011) *Printsip maksimuma v teorii optimal'nogo upravleniya* [The maximum principle in the theory of optimal control]. Moscow: Librokom.

Информация об авторе:

Чырахова Махнура Узейр кызы – докторант Института систем управления НАН Азербайджана (Баку, Азербайджан).
E-mail: mahnuraciraqova@gmail.com

Автор заявляет об отсутствии конфликта интересов.

Information about the author:

Chirakhova Mahnura U. (Post-graduate Student of the Institute of Control System of the National Academy of Sciences of Azerbaijan, Baku, Azerbaijan). E-mail: mahnuraciraqova@gmail.com

The author declares no conflicts of interests.

Поступила в редакцию 04.06.2022; принята к публикации 29.11.2022

Received 04.06.2022; accepted for publication 29.11.2022

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ

MATHEMATICAL MODELING

Научная статья

УДК 681.324

doi: 10.17223/19988605/61/2

Быстродействие транспортного протокола с селективным режимом отказа в условиях соперничества за полосу пропускания тракта передачи данных**Павел Андреевич Михеев¹, Василий Васильевич Поддубный², Павел Викторович Приступа³, Сергей Петрович Сущенко⁴, Пешанг Хасан Карим⁵***^{1, 2, 3, 4, 5} Томский государственный университет, Томск, Россия**⁴ Томский филиал Федерального исследовательского центра информационных и вычислительных технологий, Томск, Россия**¹ doka.patrick@gmail.com**² vvpoddubny@gmail.com**³ pristupa@gmail.com**⁴ ssp.inf.tsu@gmail.com**⁵ peshangkarimov@gmail.com*

Аннотация. Предложена индикаторная модель транспортного соединения для режима селективного отказа в условиях конкуренции между различными абонентскими соединениями за полосу пропускания передающего тракта. Функционирование транспортного соединения описывается двумерной цепью Маркова с дискретным временем и конечным числом состояний. Индикатором соперничества являются очереди конкурентных потоков данных в транзитных узлах транспортного соединения с заданными параметрами. Проведен анализ доступной полосы пропускания в различных условиях соперничества.

Ключевые слова: транспортный протокол; селективный режим отказа; соперничество за ресурсы; полоса пропускания; протокольные параметры; круговая задержка; математическая модель; цепь Маркова

Для цитирования: Михеев П.А., Поддубный В.В., Приступа П.В., Сущенко С.П., Карим П.Х. Быстродействие транспортного протокола с селективным режимом отказа в условиях соперничества за полосу пропускания тракта передачи данных // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2022. № 61. С. 12–25. doi: 10.17223/19988605/61/2

Original article

doi: 10.17223/19988605/61/2

Performance of a selective failure mode transport protocol under conditions of contention for the bandwidth of the data transmission path**Pavel A. Mikheev¹, Vasiliy V. Poddubny², Pavel V. Pristupa³,
Sergey P. Sushchenko⁴, Peshang Kh. Karim⁵***^{1, 2, 3, 4, 5} Tomsk State University, Tomsk, Russian Federation**⁴ Tomsk Branch of the Federal Research Center for Information and Computational Technologies, Tomsk, Russian Federation**¹ doka.patrick@gmail.com**² vvpoddubny@gmail.com**³ pristupa@gmail.com*

⁴ ssp.inf.tsu@gmail.com

⁵ peshangkarimov@gmail.com

Abstract. An indicator model of a transport connection is proposed for the selective failure mode under conditions of competition between different subscriber connections for the bandwidth of the transmitting path. The operation of a transport connection is described by a two-dimensional Markov chain with discrete time and a finite number of states. The contention indicator is the queues of competitive data flows in the transit nodes of the transport connection with the given parameters. The analysis of the available bandwidth in various conditions of rivalry is carried out.

Keywords: transport protocol; selective reject mode; competition for resources; bandwidth; protocol parameters; round-trip delay; mathematical model; Markov chain

For citation: Mikheev, P.A., Poddubny, V.V., Pristupa, P.V., Sushchenko, S.P., Karim, P.Kh. (2022) Performance of a selective failure mode transport protocol under conditions of contention for the bandwidth of the data transmission path. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naja tehnika i informatika – Tomsk State University Journal of Control and Computer Science*. 61. pp. 12–25. doi: 10.17223/19988605/61/2

Важнейшей операционной характеристикой абонентского соединения, управляемого транспортным протоколом компьютерной сети, является его пропускная способность. Данный показатель в значительной мере определяется интенсивностью внешних по отношению к данному соединению потоков, имеющих с ним хотя бы часть общего маршрута. Традиционной моделью многозвенного транспортного соединения являются сети систем массового обслуживания (СМО). Однако аналитическое исследование сетей СМО возможно лишь при существенных ограничениях на входной поток, дисциплины обслуживания требований и конкурентный трафик от других абонентов, имеющих хотя бы часть общего маршрута с исследуемым соединением, и буферную емкость транзитных узлов. Кроме того, модель транспортного соединения в виде сети СМО не позволяет учесть особенности управляющих протокольных процедур. Основным индикатором «внешней» нагрузки на тракт, в котором проложено исследуемое транспортное соединение, являются размеры очередей перед протокольными блоками данных рассматриваемого соединения в транзитных узлах. Мониторинг такого индикатора позволяет оценить распределение длин очередей в транзитных узлах от внешних по отношению к анализируемому соединению сетевых потоков и использовать при расчете операционных характеристик соединения и выборе протокольных параметров на время сеанса связи между заданной парой абонентов.

Известные модели асинхронных управляющих процедур отдельного звена передачи данных и транспортного протокола [1–8] ориентированы на анализ операционных характеристик при монопольном использовании каналов связи и не позволяют учитывать нагрузку на разделяемые сетевые ресурсы, обеспечиваемую соседством с другими виртуальными соединениями, агрегируемыми на различных участках пути в отдельных звеньях маршрута заданного абонентского соединения, и проявляющуюся в виде «внешних» очередей в транзитных узлах. Не исследовано влияние предельно малых длительностей тайм-аута при переменной круговой задержке, обусловленной переменной нагрузкой на транспортном соединении. Анализ процессов управления параметрами транспортного протокола в нагруженном транспортном соединении [9–11] выполнен при существенных ограничениях на значения протокольных параметров и характеристик тракта передачи данных.

В данной работе предложена математическая модель транспортного соединения, управляемого транспортным протоколом в режиме селективного отказа, учитывающая кроме фактора искажений в прямом и обратном трактах передачи данных и механизмов повторных передач, обусловленных искажениями и истечением тайм-аута неприема ответа от получателя потока информации, еще и очереди ненулевой длины от «внешних» межабонентских соединений для длительностей сквозного тайм-аута с интервальными ограничениями и ограничениями снизу.

1. Индикаторная модель тракта передачи данных

Рассмотрим обмен между абонентами, соединенными многозвенным трактом передачи данных. Предположим, что выполняются следующие допущения. Узлы тракта соединены дуплексными кана-

лами связи, имеющими одинаковое быстродействие в обоих направлениях. Длина тракта, выраженная в количестве участков переприема, равна D_n . Обратный канал, по которому доставляются подтверждения отправителю о корректности приема последовательности сегментов данных, имеет длину D_o . Заданы вероятности искажения сегмента в канале связи для прямого $R_n(d), d = \overline{1, D_n}$ и обратного $R_o(d), d = \overline{1, D_o}$ направлений передачи каждого участка переприема. Тогда достоверности передачи сегментов данных вдоль тракта от источника до адресата и обратно составят

$$F_n = \prod_{d=1}^{D_n} (1 - R_n(d)); \quad F_o = \prod_{d=1}^{D_o} (1 - R_o(d)).$$

Время обработки сегментов в узлах тракта одинаково. Взаимодействующие абоненты имеют неограниченный поток сегментов для передачи, а обмен выполняется сегментами одинаковой длины. Подтверждения получателя о корректности приема принимаемых данных переносятся в сегментах встречного потока. Полагаем, что повторная передача сегментов организована в соответствии с селективной процедурой отказа [1]. Считаем, кроме того, что потеря сегментов из-за отсутствия буферной памяти в узлах тракта не происходит.

Задана функция вероятностей $b_n, n = \overline{0, N}$ того, что каждый сегмент из потока анализируемого соединения в транзитных узлах встретит очередь размера $n \leq N$, где N – максимальный размер очереди, определяемый емкостью буферных пулов транзитных узлов. Будем называть тактом время t , необходимое для вывода сегмента в линию. Такт определяется суммой времени вывода сегмента в линию, времени распространения сигнала в канале связи и времени обработки сегмента принимающим узлом. Тайм-аут S , выраженный в длительностях t , запускается перед началом передачи первого сегмента последовательности и фиксируется для всех сегментов в пределах ширины окна. Будем считать, что размер окна управляющего протокола определяется величиной W , а $S > W$ задает длительность тайм-аута ожидания подтверждения корректности доставки данных. Очевидно, что сумму длин прямого и обратного трактов передачи данных $D = D_n + D_o$ можно интерпретировать как длительность круговой задержки в ненагруженном монополюсно используемом тракте, выраженную в тактах t . После передачи очередного сегмента протокол копирует его в очередь переданных, но не подтвержденных данных и запускает тайм-аут. Как только размер очереди становится равным ширине окна W , управляющий протокол приостанавливает передачу в ожидании получения квитанции или истечения тайм-аута ожидания подтверждения S . При получении подтверждения из очереди удаляются сегменты, дошедшие до адресата без искажений. По истечении тайм-аута S соответствующий сегмент передается повторно, и тайм-аут запускается вновь. Тогда время получения отправителем сквозной квитанции распределено по геометрическому закону с параметром F_o и длительностью такта дискретизации t .

Функционирование виртуального соединения, управляемого транспортным протоколом, в нагруженном многозвенном тракте передачи данных с очередями сегментов перед отправляемыми данными или подтверждениями может быть описано марковизированным процессом динамики очереди переданных, но не подтвержденных сегментов, в котором размер очереди перед прямым или обратным потоком данных исследуемого соединения является дополнительной переменной Марковского процесса. В состоянии цепи Маркова (i, n) источник отправил последовательность размера $i-n$ сегментов, которая в процессе переноса в одном из звеньев встретила очередь длиной n сегментов. Значениям координаты $i = \overline{0, W+n}, n = \overline{0, N}$ состояний цепи Маркова соответствуют количество переданных, но не подтвержденных получателем сегментов и время от начала передачи последовательности, а значениям $i = \overline{W+n+1, S-1}, n = \overline{0, N}$ – время, в течение которого отправитель не активен и ожидает получения квитанции о корректности приема переданной последовательности из W сегментов. Обозначим через $P(i, n), i = \overline{0, S-1}, n = \overline{0, N}$, – вероятности состояний цепи Маркова. Тогда последовательность переданных, но не подтвержденных сегментов данных рассматриваемого виртуального соединения при очереди нулевой длины растет до состояния цепи Маркова с координатами $(D-1, 0)$ с вероятностью

стью b_0 . Дальнейший рост размера этой последовательности происходит с вероятностью $b_0(1 - F_o)$. В состояниях (i, n) , $i = \overline{D-1+n, S-1}$, $n = \overline{0, N}$, возможно получение отправителем квитанции, и в зависимости от результатов доставки отправитель передает новые сегменты (при положительной квитанции) либо повторно – искаженные. Поскольку отправленная последовательность сегментов исследуемого виртуального соединения может встретить очередь ненулевой длины в любой момент процесса передачи (на пути последовательности до адресата или при переносе подтверждения отправителю информационного потока), то переход из состояния $(i, 0)$, $i = \overline{0, S-2}$, в состояние (i, n) , $i = \overline{0, S-2}$, $n = \overline{1, N}$, происходит с вероятностью b_n .

2. Вероятности состояний цепи Маркова

Обозначим через π_{in}^{jm} переходные вероятности цепи Маркова, где (i, n) – координаты исходного, а (j, m) – измененного состояния цепи. Тогда динамику процесса передачи информационного потока в режиме селективного отказа в нагруженном тракте передачи данных можно задать следующими значениями переходных вероятностей:

$$\pi_{in}^{jm} = \begin{cases} b_0, i = \overline{0, D-2}, n = 0; j = i+1, m = 0; \\ b_0(1 - F_o), i = \overline{D-1, S-2}, n = 0; j = i+1, m = 0; \\ b_m, i = \overline{0, S-2}, n = 0; j = i, m = \overline{1, N}; \\ b_0 F_o, i = \overline{D-1, W-1}, n = 0; j = D-1, m = 0; \\ b_0 F_o, i = \overline{W, W+D-2}, n = 0; j = W+D-2-i, m = 0; \\ b_0 F_o, i = \overline{W+D-1, S-2}, n = 0; j = 0, m = 0; \\ 1, i = S-1, n = \overline{0, N}; j = 0, m = 0; \\ 1, i = \overline{0, D-2+n}, n = \overline{1, N}; j = i+1, m = n; \\ 1 - F_o, i = \overline{D-1+n, S-2}, n = \overline{1, N}; j = i+1, m = n. \\ F_o, i = \overline{D-1+n, W-1+n}, n = \overline{1, N}; j = D-1, m = 0; \\ F_o, i = \overline{W+n, W+n+D-2}, n = \overline{1, N}; j = W+n+D-2-i, m = 0; \\ F_o, i = \overline{W+n+D-1, S-2}, n = \overline{1, N}; j = 0, m = 0. \end{cases} \quad (1)$$

Разнообразие вида решения системы уравнений равновесия для вероятностей состояний цепи Маркова определяется соотношениями между протокольными параметрами W , S , общей длиной тракта (круговой задержкой) D и максимальным размером длин очередей N . Поскольку длительность тайм-аута должна превышать ширину окна, быть не короче круговой задержки ($S \geq D$), а также превышать время ожидания в очередях из протокольных блоков данных сопутствующего трафика до начала передачи в транзитных узлах, то выделяется широкое разнообразие вариантов решения для различных областей изменения значений протокольных параметров и длин очередей. Анализ процесса передачи в аналитическом виде для произвольных значений протокольных параметров в условиях соперничества за сетевые ресурсы возможен только в предположении, что «внешние» очереди имеют ненулевую длину ($b_0 = 0$).

3. Анализ процесса передачи с ограничениями снизу на длительность тайм-аута

Рассмотрим процесс передачи для протокольных параметров, связанных с общей длиной тракта и максимальным размером очереди неравенствами вида $W \geq D$, $S \geq D+W+N-1$. Система уравнений равновесия при этом записывается следующим образом:

$$P(0,0) = F_o \sum_{n=1}^N \sum_{i=D+W+n-2}^{S-2} P(i,n) + \sum_{n=0}^N P(S-1,n); \quad (2)$$

$$P(i,0) = F_o \sum_{n=1}^N P(D+W+n-2-i,n), i = \overline{1, D-2}; \quad (3)$$

$$P(D-1,0) = \sum_{n=1}^N \sum_{i=D+n-1}^{W+n-1} F_o P(i,n); \quad (4)$$

$$P(0,n) = b_n P(0,0), n = \overline{1, \bar{N}}; \quad (5)$$

$$P(i,n) = P(i-1,0) + b_n P(i,0), i = \overline{1, D-1}, n = \overline{1, \bar{N}}; \quad (6)$$

$$P(i,n) = P(i-1,n), i = \overline{D, D+n-1}, n = \overline{1, \bar{N}}; \quad (7)$$

$$P(i,n) = (1-F_o)P(i-1,n), i = \overline{D+n, S-1}, n = \overline{1, \bar{N}}. \quad (8)$$

Найдем решение данной системы уравнений. Согласно уравнению (7) получаем: $P(i,n) = P(D-1,n)$, $i = \overline{D, D+n-1}$, $n = \overline{1, \bar{N}}$, а из (8) имеем: $P(i,n) = (1-F_o)^{i-D-n+1} P(D-1,n)$, $i = \overline{D+n, S-1}, n = \overline{1, \bar{N}}$. С учетом данных соотношений из (3), (4) для $i = \overline{1, D-1}$ находим $P(i,0) = F_o (1-F_o)^{W-i-1} \sum_{m=1}^N P(D-1,m)$, $i = \overline{1, D-2}$, $P(D-1,0) = \left(1 - (1-F_o)^{W-D+1}\right) \sum_{m=1}^N P(D-1,m)$. Подставляя найденные соотношения в (6), с учетом (5) получаем

$$P(i,n) = b_n \left[P(0,0) + (1-F_o)^{W-i-1} \left(1 - (1-F_o)^i\right) \sum_{m=1}^N P(D-1,m) \right], i = \overline{1, D-2}, n = \overline{1, \bar{N}},$$

$$P(D-1,n) = b_n \left[P(0,0) + \left(1 - (1-F_o)^{W-1}\right) \sum_{m=1}^N P(D-1,m) \right], n = \overline{1, \bar{N}}.$$

Отсюда последовательно для $n = \overline{1, \bar{N}}$ выражаем $P(D-1,n)$ через вероятности состояний $P(D-1,m)$, $m = \overline{n+1, N}$:

$$P(D-1,n) = \frac{b_n}{1 - \left(1 - (1-F_o)^{W-1}\right) \sum_{m=1}^n b_m} \left[P(0,0) + \left(1 - (1-F_o)^{W-1}\right) \sum_{m=n+1}^N P(D-1,m) \right], n = \overline{1, \bar{N}}. \quad (9)$$

При $n = N$ отсюда приходим к $P(D-1,N) = \frac{b_N P(0,0)}{(1-F_o)^{W-1}}$. Подставляя данное соотношение в (9), для значений n от $N-1$ до 1 рекурсивно находим функциональные выражения для вероятностей состояний $P(D-1,n)$ через $P(0,0)$: $P(D-1,n) = \frac{b_n P(0,0)}{(1-F_o)^{W-1}}$, $n = \overline{1, \bar{N}}$. Отсюда из найденных ранее соотношений окончательно получаем распределение вероятностей состояний цепи Маркова

$$P(i,0) = \frac{F_o P(0,0)}{(1-F_o)^i}, i = \overline{1, D-2}; P(D-1,0) = \frac{\left(1 - (1-F_o)^{W-D+1}\right) P(0,0)}{(1-F_o)^{W-1}};$$

$$P(i,n) = \frac{b_n P(0,0)}{(1-F_o)^i}, i = \overline{0, D-2}, n = \overline{1, \bar{N}}; P(i,n) = \frac{b_n P(0,0)}{(1-F_o)^{W-1}}, i = \overline{D-1, D+n-1}, n = \overline{1, \bar{N}};$$

$$P(i,n) = \frac{b_n (1-F_o)^{i-D-n+1} P(0,0)}{(1-F_o)^{W-1}}, i = \overline{D+n, S-1}, n = \overline{1, \bar{N}},$$

а из условия нормировки находим вероятность начального состояния

$$P(0,0) = \frac{F_o (1-F_o)^{W-1}}{1 + F_o (1 + \bar{N}) - (1-F_o)^W + (1-F_o)^{W-D+1} \left[1 - \sum_{m=1}^N b_m (1-F_o)^{S-W-m} \right]}, \text{ где } \bar{N} = \sum_{n=1}^N n b_n.$$

Рассмотрим найденное решение в ряде частных случаев. Для детерминированного обратного тракта ($F_o = 1$) пространство значимых состояний (i, n) образует плоскость равнобедренного по координатам i и n треугольника $i = \overline{D-1, D-1+n}, n = \overline{0, \bar{N}}$: $P(D-1, 0) = \frac{1}{2 + \bar{N}}$, $P(i, n) = \frac{b_n}{2 + \bar{N}}$, $i = \overline{D-1, D-1+n}$, $n = \overline{1, \bar{N}}$.

При неограниченной ширине окна ($W = \infty$) состояния (i, n) , $i = \overline{0, D-2}$, $n = \overline{0, \bar{N}}$, являются невозвратными ($P(i, n) = 0$), и вероятности состояния цепи Маркова принимают вид:

$$P(D-1, 0) = \frac{F_o}{1 + F_o(1 + \bar{N})}; P(i, n) = \frac{b_n F_o}{1 + F_o(1 + \bar{N})}, i = \overline{D-1, D+n-1}, n = \overline{1, \bar{N}};$$

$$P(i, n) = \frac{b_n F_o (1 - F_o)^{i-D-n+1}}{1 + F_o(1 + \bar{N})}, i \geq D+n, n = \overline{1, \bar{N}}.$$

Рассмотрим процесс передачи данных в условиях недогруженного соединения, когда ширина окна не превышает длительности круговой задержки ($1 \leq W < D$), а размер тайм-аута ограничен снизу ($S \geq D+W+N-1$). Согласно (1) система уравнений равновесия, приведенная выше, изменится следующим образом. Уравнения (2), (5), (8) останутся без изменений, (3) – справедливо при $i = \overline{1, W-1}$, уравнение (4) примет вид: $P(D-1, 0) = 0$, уравнение (6) – справедливо для $i = \overline{1, W-1}$, $n = \overline{1, \bar{N}}$, уравнение (7) – при $i = \overline{W, D+n-1}$, $n = \overline{1, \bar{N}}$. Решение системы уравнений равновесия имеет вид:

$$P(i, 0) = \frac{F_o P(0, 0)}{(1 - F_o)^i}, i = \overline{1, W-1}; P(i, n) = \frac{b_n P(0, 0)}{(1 - F_o)^i}, i = \overline{0, W-1}, n = \overline{1, \bar{N}};$$

$$P(i, n) = \frac{b_n P(0, 0)}{(1 - F_o)^{W-1}}, i = \overline{W-1, D+n-1}, n = \overline{1, \bar{N}}; P(i, n) = \frac{b_n (1 - F_o)^{i-D-n+1} P(0, 0)}{(1 - F_o)^{W-1}}, i = \overline{D+n, S-1}, n = \overline{1, \bar{N}},$$

а из условия нормировки получаем вероятность начального состояния

$$P(0, 0) = \frac{F_o (1 - F_o)^{W-1}}{2 + F_o(D - W + \bar{N}) - (1 - F_o)^W - \sum_{m=1}^N b_m (1 - F_o)^{S-D+1-m}}. \quad (10)$$

При $F_o = 1$ значимыми будут только состояния $P(W-1, 0) = \frac{1}{D-W+\bar{N}+2}$, $P(i, n) = \frac{b_n}{D-W+\bar{N}+2}$, $i = \overline{W-1, D+n-1}, n = \overline{1, \bar{N}}$. Неограниченная длительность тайм-аута ($S \rightarrow \infty$) приводит к вероятности начального состояния следующего вида: $P(0, 0) = \frac{F_o (1 - F_o)^{W-1}}{2 + F_o(D - W + \bar{N}) - (1 - F_o)^W}$. Для старт-стопного протокола ($W = 1$) распределение принимает вид:

$$P(i, n) = b_n P(0, 0), i = \overline{0, D+n-1}, n = \overline{1, \bar{N}}, P(i, n) = b_n (1 - F_o)^{i-D-n+1} P(0, 0), i = \overline{D+n, S-1}, n = \overline{1, \bar{N}},$$

$$P(0, 0) = \frac{F_o}{1 + F_o(D + \bar{N}) - \sum_{m=1}^N b_m (1 - F_o)^{S-D+1-m}}.$$

4. Анализ процесса передачи при интервальных ограничениях на длительность тайм-аута

Рассмотрим функционирование транспортного соединения при интервальных ограничениях на протокольные параметры и максимальный размер очереди вида: $W \geq D$, $D+W-1 \leq S \leq D+W+N-1$, $1 \leq N \leq D-2$ либо $W \geq D$, $W+N+1 \leq S \leq D+W+N-1$, $D-2 \leq N$. При данных ограничениях уравнения (2)–(3) исходной системы уравнений равновесия (2)–(8) преобразуются к виду:

$$P(0,0) = F_o \sum_{n=1}^{S-D-W+1} \sum_{i=D+W+n-2}^{S-2} P(i,n) + \sum_{n=0}^N P(S-1,n), \quad (11)$$

$$P(i,0) = F_o \sum_{n=1}^{S-D-W+i} P(D+W+n-2-i,n), i = \overline{1, D+W+N-S-1}, \quad (12)$$

$$P(i,0) = F_o \sum_{n=1}^N P(D+W+n-2-i,n), i = \overline{D+W+N-S, D-2}. \quad (13)$$

Из уравнений (7), (8), (12), (13), (4) находим

$$P(i,n) = P(D-1,n), i = \overline{D, D+n-1}, n = \overline{1, N}, \quad P(i,n) = (1-F_o)^{i-D-n+1} P(D-1,n), \quad i = \overline{D+n, S-1}, n = \overline{1, N},$$

$$P(i,0) = F_o (1-F_o)^{W-i-1} \sum_{m=1}^{S-D-W+i} P(D-1,m), i = \overline{1, D+W+N-S-1},$$

$$P(i,0) = F_o (1-F_o)^{W-i-1} \sum_{m=1}^N P(D-1,m), i = \overline{D+W+N-S, D-2},$$

$$P(D-1,0) = \left(1 - (1-F_o)^{W-D+1}\right) \sum_{m=1}^N P(D-1,m).$$

Подставляя данные зависимости в (6), с учетом (5) получаем

$$P(i,n) = b_n \left[P(0,0) + (1-F_o)^{W-i-1} \sum_{m=1}^{S-D-W+i} P(D-1,m) - (1-F_o)^{W-1} \sum_{m=1}^{S-D-W} P(D-1,m) - \sum_{m=S-D-W+1}^{S-D-W+i} P(D-1,m) (1-F_o)^{S-D-m} \right], \quad i = \overline{1, D+W+N-S-1}, n = \overline{1, N}.$$

$$P(i,n) = b_n \left[P(0,0) + (1-F_o)^{W-i-1} \sum_{m=1}^N P(D-1,m) - (1-F_o)^{W-1} \sum_{m=1}^{S-D-W} P(D-1,m) - \sum_{m=S-D-W+1}^N P(D-1,m) (1-F_o)^{S-D-m} \right], \quad i = \overline{D+W+N-S, D-2}, n = \overline{1, N}.$$

$$P(D-1,n) = b_n \left[P(0,0) + \left(1 - (1-F_o)^{W-1}\right) \sum_{m=1}^{S-D-W} P(D-1,m) + \sum_{m=S-D-W+1}^N P(D-1,m) \left(1 - (1-F_o)^{S-D-m}\right) \right], \quad n = \overline{1, N}.$$

Далее последовательно для $n = \overline{1, S-D-W}$ выражаем $P(D-1,n)$ через вероятности состояний $P(D-1,m)$, $m = \overline{n+1, N}$ и переписываем данное уравнение в виде:

$$P(D-1,n) = b_n \frac{P(0,0) + \left(1 - (1-F_o)^{W-1}\right) \sum_{m=n+1}^{S-D-W} P(D-1,m) + \sum_{m=S-D-W+1}^N P(D-1,m) \left(1 - (1-F_o)^{S-D-m}\right)}{1 - \left(1 - (1-F_o)^{W-1}\right) \sum_{m=1}^n b_m}, \quad (14)$$

$$n = \overline{1, N}.$$

При $n = S-D-W$ отсюда приходим к

$$P(D-1, S-D-W) = \frac{b_{S-D-W}}{1 - (1-F_o)^{W-1} \sum_{m=1}^{S-D-W} b_m} \left[P(0,0) + \sum_{m=S-D-W+1}^N P(D-1,m) \left(1 - (1-F_o)^{S-D-m}\right) \right].$$

Подставляя данное соотношение в (14), далее для значений n от $S-D-W-1$ до 1 находим функциональные выражения для вероятностей состояний $P(D-1,n)$ через $P(0,0)$ и $P(D-1,m)$, $m = \overline{S-D-W+1, N}$ и упрощаем уравнение (14) до

$$P(D-1, n) = \frac{b_n \left[P(0, 0) + \sum_{m=S-D-W+1}^N P(D-1, m) (1 - (1 - F_o)^{S-D-m}) \right]}{1 - \left(1 - (1 - F_o)^{W-1} \right) \sum_{m=1}^{S-D-W} b_m}, \quad (15)$$

$n = \overline{1, N}$.

Отсюда последовательно для $n = \overline{S-D-W+1, N}$ получаем

$$P(D-1, n) = \frac{b_n \left[P(0, 0) + \sum_{m=n+1}^N P(D-1, m) (1 - (1 - F_o)^{S-D-m}) \right]}{1 - \left(1 - (1 - F_o)^{W-1} \right) \sum_{m=1}^{S-D-W} b_m - \sum_{m=S-D-W+1}^n b_m (1 - (1 - F_o)^{S-D-m})}.$$

Из данного соотношения последовательно для n от N до $S-D-W+1$ с учетом (15) окончательно выражаем $P(D-1, n)$ через вероятность начального состояния $P(0, 0)$ и согласно ранее найденным зависимостям получаем вероятности состояний цепи Маркова:

$$P(i, 0) = \frac{P(0, 0)}{E} F_o (1 - F_o)^{W-i-1} \sum_{m=1}^{S-D-W+i} b_m, \quad E = (1 - F_o)^{W-1} \sum_{m=1}^{S-D-W} b_m + \sum_{m=S-D-W+1}^N b_m (1 - F_o)^{S-D-m}, \quad (16)$$

$i = \overline{1, D+W+N-S-1}$.

$$P(i, 0) = \frac{P(0, 0)}{E} F_o (1 - F_o)^{W-i-1}, \quad i = \overline{D+W+N-S, D-2}, \quad (17)$$

$$P(D-1, 0) = \frac{P(0, 0)}{E} \left(1 - (1 - F_o)^{W-D+1} \right),$$

$$P(0, n) = P(0, 0) b_n, \quad n = \overline{1, N}, \quad (18)$$

$$P(i, n) = \frac{P(0, 0)}{E} b_n \left[(1 - F_o)^{W-i-1} \sum_{m=1}^{S-D-W+i} b_m + \sum_{m=S-D-W+i+1}^N b_m (1 - F_o)^{S-D-m} \right], \quad (19)$$

$i = \overline{1, D+W+N-S-1}, \quad n = \overline{1, N}$,

$$P(i, n) = \frac{P(0, 0)}{E} b_n (1 - F_o)^{W-i-1}, \quad i = \overline{D+W+N-S, D-2}, \quad n = \overline{1, N}, \quad (20)$$

$$P(i, n) = \frac{P(0, 0)}{E} b_n, \quad i = \overline{D-1, D+n-1}, \quad n = \overline{1, N}, \quad (21)$$

$$P(i, n) = \frac{P(0, 0)}{E} b_n (1 - F_o)^{i-D-n+1}, \quad i = \overline{D+n, S-1}, \quad n = \overline{1, N}. \quad (22)$$

Вероятность начального состояния, найденная из условия нормировки, имеет вид:

$$P(0, 0) = F_o E / \left[1 + F_o (1 + \bar{N}) + (1 - F_o)^{W-D+1} - \sum_{m=1}^N b_m (1 - F_o)^{S-D+1-m} - (1 - F_o)^W \sum_{m=1}^{S-D-W} b_m - (1 - F_o (D+W+N-S)) \sum_{m=S-D-W+1}^N b_m (1 - F_o)^{S-D-m} \right].$$

Нетрудно убедиться в том, что данное распределение совпадает с полученным ранее распределением для ограничений снизу на длительность тайм-аута при $S = D+W+N-1$.

Проанализируем процесс информационного переноса в транспортном соединении при размере скользящего окна, не превышающем длительность круговой задержки ($W \leq D$), и интервальных ограничениях на длительность тайм-аута и максимальный размер очереди вида $D+W-1 \leq S \leq D+W+N-1$, $1 \leq N \leq W-2$. При данных условиях уравнение (2) исходной системы уравнений равновесия (2)–(8)

преобразуются к $P(0, 0) = F_o \sum_{n=1}^{S-D-W+1} \sum_{i=D+W+n-2}^{S-2} P(i, n) + \sum_{n=0}^N P(S-1, n)$. Уравнение (3) переопределяется соотношениями (12) и (13). При этом уравнение (13) справедливо для множества индексов

$i = \overline{D+W+N-S, W-1}$. Уравнение (4) принимает вид $P(D-1, 0) = 0$, уравнение (6) выполняется для $i = \overline{1, W-1}$, $n = \overline{1, N}$, уравнение (7) – для $i = \overline{W, D+n-1}$, $n = \overline{1, N}$. Стационарные вероятности состояний цепи Маркова, описываемой данными уравнениями, с точностью до вероятности начального состояния принимают вид (16)–(22), но выражения (17) и (20) справедливы для индексов $i = \overline{D+W+N-S, W-1}$, а (21) – для $i = \overline{W-1, D+n-1}$. Согласно условию нормировки начальное состояние определяется соотношением

$$P(0, 0) = F_o E \left/ \left[2 + F_o(D-W+\bar{N}) - \sum_{m=1}^N b_m(1-F_o)^{S-D+1-m} - (1-F_o)^W \sum_{m=1}^{S-D-W} b_m - \sum_{m=S-D-W+1}^N b_m(1-F_o)^{S-D-m+1} \right] \right.$$

При $S = D+W+N-1$ данное соотношение для $P(0, 0)$ совпадает с (10).

Найдем вероятности состояний цепи Маркова при ограничениях $W \geq D$, $D+W-1 \leq S \leq W+N+1$, $D-2 \leq N \leq W-2$ либо $W \geq D$, $D+N+1 \leq S \leq W+N+1$, $W-2 \leq N$. В данных условиях уравнение (2) исходной системы уравнений локального равновесия принимает вид (11), уравнение (3) – вид (12), но для множества состояний $i = \overline{1, D-2}$, а уравнение (4) – вид $P(D-1, 0) = \sum_{n=1}^{S-W-1} \sum_{i=D+n-1}^{W+n-1} F_o P(i, n) + \sum_{n=S-W}^N \sum_{i=D+n-1}^{S-2} F_o P(i, n)$. Принцип поиска решения полученной системы уравнений повторяет последовательность действий для рассмотренных ранее ограничений $W \geq D$, $D+W-1 \leq S \leq D+W+N-1$, $1 \leq N \leq D-2$. Тогда вероятности состояний цепи Маркова принимают вид:

$$P(i, 0) = \frac{P(0, 0)}{E} F_o (1-F_o)^{W-i-1} \sum_{m=1}^{S-D-W+i} b_m, \quad i = \overline{1, D-2};$$

$$P(D-1, 0) = \frac{P(0, 0)}{E} \left[1 - (1-F_o)^{W-D+1} \sum_{m=1}^{S-W-1} b_m - \sum_{m=S-W}^N b_m(1-F_o)^{S-D-m} \right];$$

$$P(i, n) = \frac{P(0, 0)}{E} b_n \left[(1-F_o)^{W-i-1} \sum_{m=1}^{S-D-W+i} b_m + \sum_{m=S-D-W+i+1}^N b_m(1-F_o)^{S-D-m} \right], \quad i = \overline{0, D-2}, \quad n = \overline{1, N};$$

$$P(i, n) = \frac{P(0, 0)}{E} b_n, \quad i = \overline{D-1, D+n-1}, \quad n = \overline{1, N};$$

$$P(i, n) = \frac{P(0, 0)}{E} b_n (1-F_o)^{i-D-n+1}, \quad i = \overline{D+n, S-1}, \quad n = \overline{1, N};$$

$$\begin{aligned}
 P(0, 0) = F_o E \left/ \left[1 + F_o(1+\bar{N}) + (1-F_o)^{W-D+1} \sum_{m=1}^{S-W-2} b_m - (1-F_o)^W \sum_{m=1}^{S-D-W} b_m - \sum_{m=1}^N b_m(1-F_o)^{S-D-m} - \right. \right. \\
 \left. \left. - \sum_{m=S-D-W+1}^{S-W-2} b_m(1-F_o)^{S-D-m} + F_o \left((D-1) \sum_{m=S-W}^N b_m(1-F_o)^{S-D-m} + \sum_{m=S-D-W+2}^{S-W-1} (D+W-S-m) b_m(1-F_o)^{S-D-m} + \right. \right. \right. \\
 \left. \left. \left. + \sum_{m=1}^N b_m(1-F_o)^{S-D-m} + b_m(1-F_o)^{W-1} \right) \right] \right|.
 \end{aligned}$$

Перейдем к поиску вероятностей состояний цепи Маркова при ограничениях $W \geq D$, $D+W-1 \leq S \leq D+N+1$, $W-2 \leq N$. При этом уравнение (2) исходной системы (2)–(8) примет вид (11), уравнение (3) – вид (12) для множества индексов $i = \overline{1, D-2}$, а уравнения (4), (7), (8) преобразуются соответственно к виду:

$$P(D-1, 0) = \sum_{n=1}^{S-W-1} \sum_{i=D+n-1}^{W+n-1} F_o P(i, n) + \sum_{n=S-W}^{S-D-1} \sum_{i=D+n-1}^{S-2} F_o P(i, n);$$

$$P(i, n) = P(i-1, n), \quad i = \overline{D, D+n-1}, \quad n = \overline{1, S-D-1}; \quad i = \overline{D, S-1}, \quad n = \overline{S-D, N};$$

$$P(i, n) = (1-F_o)P(i-1, n), \quad i = \overline{D+n, S-1}, \quad n = \overline{1, S-D-1}.$$

Выполняя последовательность ранее рассмотренных преобразований для вероятностей состояний цепи Маркова, получаем

$$\begin{aligned}
 P(i,0) &= \frac{P(0,0)}{e} F_o (1-F_o)^{W-i-1} \sum_{m=1}^{S-D-W+i} b_m, \quad i = \overline{1, D-2}; \\
 e &= 1 - \sum_{m=1}^{S-D-1} b_m + (1-F_o)^{W-1} \sum_{m=1}^{S-D-W} b_m + \sum_{m=S-D-W+1}^{S-D-1} b_m (1-F_o)^{S-D-m}; \\
 P(D-1,0) &= \frac{P(0,0)}{e} \left[\sum_{m=1}^{S-D-1} b_m - (1-F_o)^{W-D+1} \sum_{m=1}^{S-W-1} b_m - \sum_{m=S-W}^{S-D-1} b_m (1-F_o)^{S-D-m} \right]; \\
 P(i,n) &= \frac{P(0,0)}{e} b_n \left[1 - \sum_{m=1}^{S-D-1} b_m + (1-F_o)^{W-i-1} \sum_{m=1}^{S-D-W+i} b_m + \sum_{m=S-D-W+i+1}^{S-D-1} b_m (1-F_o)^{S-D-m} \right], \\
 &\quad i = \overline{0, D-2}, \quad n = \overline{1, N}; \\
 P(i,n) &= \frac{P(0,0)}{e} b_n, \quad i = \overline{D-1, D+n-1}, \quad n = \overline{1, S-D-1}; \quad i = \overline{D-1, S-1}, \quad n = \overline{S-D, N}; \\
 P(i,n) &= \frac{P(0,0)}{e} b_n (1-F_o)^{i-D-n+1}, \quad i = \overline{D+n, S-1}, \quad n = \overline{1, S-D-1}.
 \end{aligned}$$

Вероятность начального состояния не приводим из-за ее громоздкости.

5. Доступная полоса пропускания транспортного соединения

Пропускная способность транспортного соединения в условиях соперничества потоков различных корреспондирующих абонентов за полосу пропускания тракта передачи данных определяется как отношение среднего объема данных, передаваемых между двумя последовательными получениями квитанций, к среднему времени получения квитанции [4, 5, 8]. Вклад в быстродействие виртуального соединения дают те состояния цепи Маркова, для которых возможно получение квитанции. Нормированная на единицу пропускная способность виртуального соединения в нагруженном тракте определяется отношением среднего количества сегментов данных, передаваемых отправителем между поступлениями двух последовательных квитанций, к среднему времени между поступлениями квитанций, выраженному в количестве интервалов длительности t : $Z(W, S) = \bar{V} / \bar{T}$. Поскольку квитанции переносятся в каждом сегменте независимо и поступают к отправителю каждый такт t при условии, что они не искажены на пути длины D от получателя до отправителя информационного потока, то среднее время между приходами квитанций распределено по геометрическому закону с параметром F_o и составит: $\bar{T} = 1/F_o$. Средний объем передаваемых между поступлениями квитанций

данных с учетом того, что каждый сегмент исследуемого соединения с вероятностью $b_n, n = \overline{1, N}$ встречает очередь размера n и дает вклад в объем переданной информации, обратно пропорциональный величине $n+1$, задается обобщением соотношения, приведенного в работе [11]:

$$\bar{V} = \sum_{n=0}^N \frac{1}{n+1} \left[\sum_{l=2D-1+n}^{W+2D-2+n} l P(l, n) + \sum_{l=W+2D-1+n}^{S-1} \bar{W} P(l, n) \right].$$

Величины \bar{T} и \bar{W} определяются средним количеством сегментов, достигших адресата при селективной процедуре повтора искаженных сегментов: $\bar{T} = (l-D-n+2)F_n$, $\bar{W} = WF_n$. Тогда зависимость пропускной способности виртуального соединения от протокольных параметров (W, S) , характеристик передающего тракта (D, F_n, F_o) и нагрузочных параметров на соединительный путь порождаемых соперничающими абонентами $(b_n, n = \overline{1, N})$ примет

$$\text{вид } Z(W, S) = F_n F_o \times \sum_{n=0}^N \frac{1}{n+1} \left[\sum_{l=2D-1+n}^{W+2D-2+n} (l-2D+2-n) P(l, n) + W \sum_{l=W+2D-1+n}^{S-1} P(l, n) \right].$$

Отсюда для произвольной ширины окна при $S \geq D+W+N-1$ окончательно получаем

$$Z_c(W, S) = \begin{cases} F_n \frac{\sum_{n=1}^N \frac{b_n}{n+1} [1 - (1 - F_o)^W - W F_o (1 - F_o)^{S-D-n+1}]}{2 + F_o (D - W + \bar{N}) - (1 - F_o)^W - \sum_{n=1}^N b_n (1 - F_o)^{S-D-n+1}}, & W < D; \\ F_n \frac{F_o^2 (1 - (1 - F_o)^{W-D+1}) + \sum_{n=1}^N \frac{b_n}{n+1} [1 - (1 - F_o)^W - W F_o (1 - F_o)^{S-D-n+1}]}{1 + F_o (\bar{N} + 1) + (1 - F_o)^{W-D+1} - (1 - F_o)^W - \sum_{n=1}^N b_n (1 - F_o)^{S-D-n+1}}, & W \geq D. \end{cases}$$

Для интервальных ограничений на длительность тайм-аута $D + W - 1 \leq S \leq D + W + N - 1$ и размера очередей соперников $1 \leq N \leq D - 2$ быстродействие транспортного соединения в конкурентной среде передачи данных составит

$$Z_c(W, S) = F_n \left\{ F_o^2 (1 - (1 - F_o)^{W-D+1}) + \sum_{n=1}^N b_n - \sum_{n=1}^{S-D-W+1} \frac{b_n}{n+1} [(1 - F_o)^W + W F_o (1 - F_o)^{S-D-n+1}] - \sum_{n=S-D-W+2}^N \frac{b_n}{n+1} (1 - F_o)^{S-D-n+1} [1 + F_o (S - D - n + 1)] \right\} / \left\{ 1 + F_o (1 + \bar{N}) + (1 - F_o)^{W-D+1} - (1 - F_o)^W - \sum_{n=1}^N b_n (1 - F_o)^{S-D-n+1} \right\}.$$

В случае абсолютно надежного обратного канала ($F_o = 1$) доступная полоса пропускания транспортного соединения при $W \leq D$ в значительной мере определяется близостью ширины окна к длительности круговой задержки $Z_c(W, S) = \frac{F_n}{2 + D - W + \bar{N}} \sum_{n=1}^N \frac{b_n}{n+1}$, а для $W > D$ – инвариантна к D $Z_c(W, S) = \frac{F_n}{2 + \bar{N}} \left[1 + \sum_{n=1}^N \frac{b_n}{n+1} \right]$. Неограниченная длительность тайм-аута ($S \rightarrow \infty$) при $W < D$ приводит к зависимости вида:

$$Z_c(W, \infty) = \frac{F_n (1 - (1 - F_o)^W) \sum_{n=1}^N \frac{b_n}{n+1}}{2 + F_o (D - W + \bar{N}) - (1 - F_o)^W},$$

а для неограниченно возрастающей ширины окна получаем $Z_c(\infty, \infty) = \frac{F_n}{1 + F_o (\bar{N} + 1)} \left[F_o^2 + \sum_{n=1}^N \frac{b_n}{n+1} \right]$.

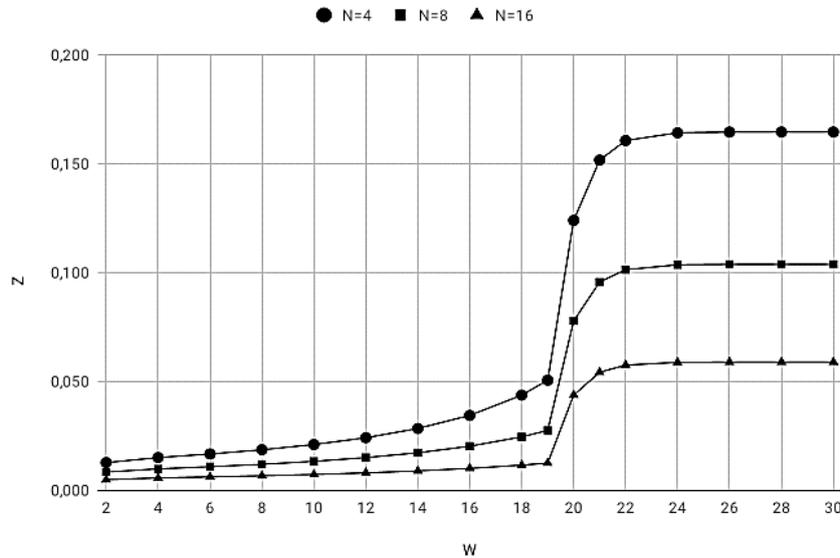


Рис. 1. Зависимость доступной полосы пропускания от ширины окна при равномерном распределении длины очереди для различных значений N , $D = 20$, $F_o = F_n = 0.7$ и неограниченном размере тайм-аута

Fig. 1. Dependence of the available bandwidth on the window width with a uniform distribution of the queue length for different values of N , $D = 20$, $F_o = F_n = 0.7$ and an unlimited timeout size

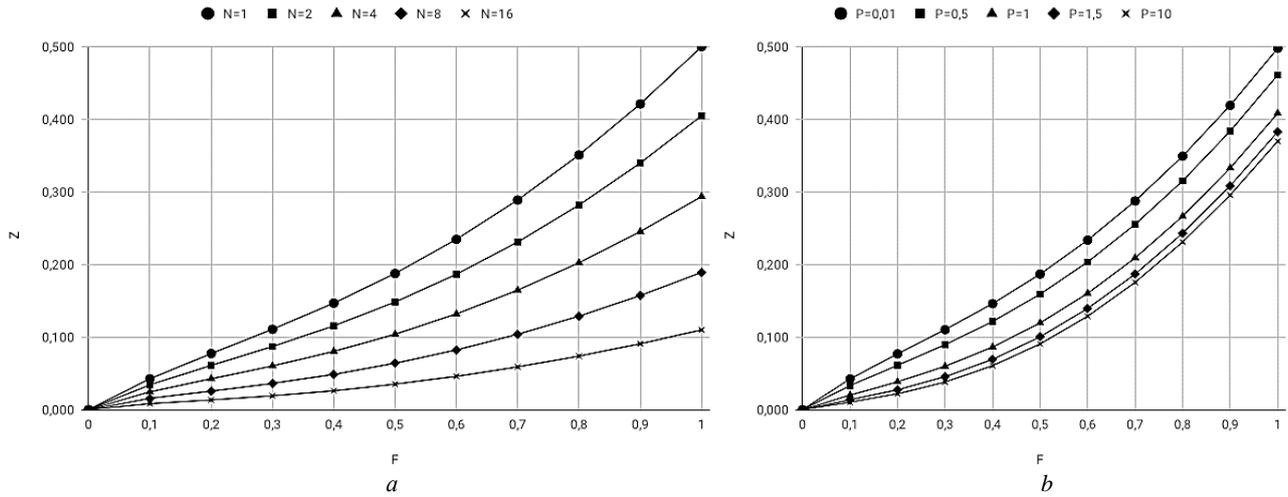


Рис. 2. Зависимость доступной полосы пропускания от достоверности передачи данных при неограниченной ширине окна, $D = 20, F_o = F_n = F$, равномерном распределении (a) и различных значениях параметра P усеченного геометрического распределения длины очереди для $N = 8$ (b)

Fig. 2. Dependence of the available bandwidth on the reliability of data transmission with an unlimited window width, $D = 20, F_o = F_n = F$, uniform distribution (a) and various values of the parameter P of the truncated geometric distribution of the queue length for $N = 8$ (b)

Численный анализ показывает, что доступная транспортному соединению полоса пропускания для $W \geq D$ практически инвариантна к длительности круговой задержки, ощутимо снижаясь от области насыщения при $W = D$ и $F_o < 1$. В случае $W < D$ доступная полоса пропускания недогружена и эффективная скорость передачи данных значительно снижается (рис. 1). С ростом конкуренции между абонентами за полосу пропускания тракта передачи данных средний размер очереди увеличивается и скорость информационного переноса быстро падает (рис. 2).

6. Выбор значений протокольных параметров

Поскольку показатель пропускной способности при неограниченном росте протокольных параметров размера окна (W) и длительности тайм-аута ожидания квитанции (S) имеет зависимость в виде кривой с насыщением, то будем искать их рациональные значения из условия заданного уровня потенциальной пропускной способности в два этапа. В силу того, что протокольные параметры связаны неравенством $S > W$, на первом этапе определяется рациональный размер окна W_o из условия $Z(W_o, \infty) = y_w Z(\infty, \infty)$, а на втором этапе – рациональная длительность тайм-аута S_o из условия $Z(W_o, S_o) = y_s Z(W_o, \infty) = y_w y_s Z(\infty, \infty)$. Здесь $y_w < 1$ и $y_s < 1$ – заданные уровни пропускной способности по координатам W и S соответственно. При $W \geq D$, $S \geq D + W + N - 1$ аналитические соотношения для рациональных значений параметров принимают вид:

$$W_o = \left| \frac{1}{\ln(1-F_o)} \ln \frac{(1-y_w)(1+F_o(1+\bar{N})) \left(F_o^2 + \sum_{n=1}^N \frac{b_n}{n+1} \right) (1-F_o)^{D-1}}{F_o^2 (1+y_w(1-F_o^{D-1}) + F_o(1+\bar{N})) + (y_w + (1-F_o)^{D-1} (1-y_w + F_o(1+\bar{N}))) \sum_{n=1}^N \frac{b_n}{n+1}} \right|,$$

$$S_o = \left| \frac{1}{\ln(1-F_o)} \ln(1-F_o)^{D-1} \left\{ y \left[1 + F_o(1+N) + (1-F_o)^{W-D+1} - (1-F_o)^{W_o} \right] \left(F_o^2 + \sum_{n=1}^N \frac{b_n}{n+1} \right) - \right. \right.$$

$$\left. - (1+F_o(1+\bar{N})) (1 - (1-F_o)^{W_o}) \sum_{n=1}^N \frac{b_n}{n+1} - F_o^2 (1 - (1-F_o)^{W_o-D+1}) \right\} / \left\{ W_o F_o (1+F_o(1+\bar{N})) \sum_{n=1}^N \frac{b_n}{(n+1)F_o^n} - \right.$$

$$\left. - y \left(F_o^2 + \sum_{n=1}^N \frac{b_n}{n+1} \right) \sum_{n=1}^N \frac{b_n}{F_o^n} \right\} \right|,$$

где $\lceil \dots \rceil$ означает округление до большего целого, $y = y_w y_s$. Для недогруженного транспортного соединения ($W < D$) из условия $Z(W, S_o) = y_s Z(W, \infty)$ определяется рациональная длительность тайм-аута

$$S_o = \left\lceil \frac{1}{\ln(1 - F_o)} \ln \frac{(1 - y_s)(1 - F_o)^{D-1} \left(2 + F_o(D - W + \bar{N}) - (1 - F_o)^w\right) \sum_{n=1}^N \frac{b_n}{n+1}}{WF_o \left(2 + F_o(D - W + \bar{N}) - (1 - F_o)^w\right) \sum_{n=1}^N \frac{b_n}{(n+1)(1 - F_o)^n} - y_s \left(1 - (1 - F_o)^w\right) \sum_{n=1}^N \frac{b_n}{n+1} \sum_{n=1}^N \frac{b_n}{(1 - F_o)^n}} \right\rceil.$$

Заключение

Проведен анализ процесса соперничества информационных потоков различных межабонентских соединений за полосу пропускания разделяемых участков пути. Предложена индикаторная модель транспортного соединения, конкурирующего за полосу пропускания отдельных участков маршрута, в виде двумерной цепи Маркова с дискретным временем, описывающей динамику очереди отправленных, но не подтвержденных протокольных блоков данных. Получено распределение состояний цепи Маркова при различных условиях функционирования транспортного соединения. Найдены аналитические зависимости быстродействия транспортного соединения для различных соотношений между параметрами транспортного протокола, характеристиками сетевых каналов и нагрузочными параметрами. Численные исследования доступной полосы пропускания транспортного соединения в селективном режиме повторной передачи показали, что скорость передачи между абонентами определяется достоверностью передачи данных, распределением длин очередей протокольных блоков в транзитных узлах и соотношением между длительностью круговой задержки и шириной окна. Из условия заданного уровня потенциальной пропускной способности найдены аналитические зависимости рациональных значений протокольных параметров. Направлением дальнейших исследований необходимо выделить задачу анализа эффективности применения процедур прямой коррекции ошибок на уровне транспортного протокола при конкурентном использовании сетевых каналов связи.

Список источников

1. Богуславский Л.Б. Управление потоками данных в сетях ЭВМ. М. : Энергоатомиздат, 1984. 168 с.
2. Gelenbe E., Labetoulle J., Pugolle G. Performance evaluation of the HDLC protocol // Comput. Networks. 1978. V. 2. P. 409–415.
3. Боровихин Е.А., Коротаев И.А. Анализ функционирования и оптимизация протокола HDLC // Автоматика и вычислительная техника. 1993. № 2. С. 47–51.
4. Сущенко С.П. Аналитические модели асинхронных процедур управления звеном передачи данных // Автоматика и вычислительная техника. 1988. № 2. С. 32–40.
5. Кокшенев В.В., Сущенко С.П. Анализ быстродействия асинхронной процедуры управления звеном передачи данных // Вычислительные технологии. 2008. Т. 13, № 5. С. 61–65.
6. Herrero R. Modeling and comparative analysis of Forward Error Correction in the context of multipath redundancy // Telecommunication Systems. Modelling, Analysis, Design and Management. 2017. V. 65 (4). P. 783–794.
7. Приступа П.В., Сущенко С.П., Михеев П.А., Моисеев А.Н. Анализ эффективности прямой коррекции ошибок на уровне транспортного протокола // Автоматика и телемеханика. 2022. № 7. С. 59–81.
8. Кокшенев В.В., Михеев П.А., Сущенко С.П. Сравнительный анализ быстродействия селективного и группового режимов повтора транспортного протокола // Автоматика и телемеханика. 2017. № 2. С. 65–81.
9. Ewald N.L., Kemp A.H. Analytical Model of TCP New Reno through CTMC // Computer Performance Engineering. Proc. 6th European Performance Engineering Workshop. EPEW 2009. London, UK, July 9–10. 2009. P. 183–196.
10. Padhye J., Firoiu V., Towsley D.F., Kurose J.F. Modeling TCP Reno Performance: A Simple Model and Its Empirical Validation // IEEE/ACM Transactions on Networking. 2000. V. 8 (2). P. 133–145.
11. Кокшенев В.В., Михеев П.А., Сущенко С.П. Анализ селективного режима транспортного протокола в нагруженном тракте передачи данных // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2013. № 3 (24). С. 78–94.

References

1. Boguslavsky, L.B. (1984) *Upravlenie potokami dannykh v setyakh EVM* [Control of Data Flows in Computer Networks]. Moscow: Energoatomizdat.

2. Gelenbe, E., Labetoulle, J. & Pugolle, G. (1978) Performance evaluation of the HDLC protocol. *Computer Networks*. 2. pp. 409–415.
3. Borovikhin, E.A. & Korotaev, I.A. (1993) Analiz funktsionirovaniya i optimizatsiya protokola HDLC [Analysis of the functioning and optimization of the HDLC protocol]. *Avtomatika i vychislitel'naya tekhnika – Automation and Computer Science*. 2. pp. 47–51.
4. Sushchenko, S.P. (1988) Analiticheskie modeli asinkhronnykh protsedur upravleniya zvenom peredachi dannykh [Analytical models of asynchronous control procedures for a data transmission link]. *Avtomatika i vychislitel'naya tekhnika – Automation and Computer Science*. 2. pp. 32–40.
5. Kokshenev, V.V. & Sushchenko, S.P. (2008) Analiz bystrodeystviya asinkhronnoy protsedury upravleniya zvenom peredachi dannykh [Analysis of the speed of an asynchronous control procedure for a data transmission link]. *Computational Technologies*. 13(5). pp. 61–65.
6. Herrero, R. (2017) Modeling and comparative analysis of Forward Error Correction in the context of multipath redundancy. *Telecommunication Systems. Modelling, Analysis, Design and Management*. 65(4). pp. 783–794. DOI: 10.1007/s11235-016-0267-y
7. Pristupa, P.V., Sushchenko, S.P., Mikheev, P.A. & Moiseev, A.N. (2022) Efficiency Analysis of the Forward Error Correction at the Transport Protocol Level. *Automation and Remote Control*. 83(7). pp. 1059–1077. DOI: 10.1134/S0005117922070049
8. Kokshenev, V.V., Mikheev, P.A. & Sushchenko, S.P. (2017) Comparative analysis of the performance of selective and group repeat transmission models in a transport protocol. *Automation and Remote Control*. 78(2). pp. 247–261. DOI: 10.1134/S0005117917020059
9. Ewald, N.L., Andrew, H. & Kemp, A.H. (2009) Analytical Model of TCP New Reno through CTMC. *Computer Performance Engineering*. Proc. of the 6th European Performance Engineering Workshop, EPEW-2009. London, UK, July 9–10. pp. 183–196.
10. Padhye, J., Firoiu, V., Towsley, D.F. & Kurose, J.F. (2000) Modeling TCP Reno Performance: A Simple Model and Its Empirical Validation. *IEEE/ACM Transactions on Networking*. 8(2). April. pp. 133–145.
11. Kokshenev, V.V., Mikheev, P.A. & Sushchenko, S.P. (2013) Transport protocol selective acknowledgements analysis in loaded transmission data path. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 3(24). pp. 78–94.

Информация об авторах:

Михеев Павел Андреевич – кандидат технических наук, ведущий программист кафедры прикладной информатики Института прикладной математики и компьютерных наук Национального исследовательского Томского государственного университета (Томск, Россия). E-mail: doka.patrick@gmail.com

Поддубный Василий Васильевич – профессор, доктор технических наук, профессор кафедры прикладной информатики Института прикладной математики и компьютерных наук Национального исследовательского Томского государственного университета (Томск, Россия). E-mail: vvpoddubny@gmail.com

Приступа Павел Викторович – ассистент кафедры прикладной информатики Института прикладной математики и компьютерных наук Национального исследовательского Томского государственного университета (Томск, Россия). E-mail: pristupa@gmail.com

Сущенко Сергей Петрович – профессор, доктор технических наук, заведующий кафедрой прикладной информатики Института прикладной математики и компьютерных наук Национального исследовательского Томского государственного университета; ведущий научный сотрудник Томского филиала Федерального исследовательского центра информационных и вычислительных технологий (Томск, Россия). E-mail: ssp.inf.tsu@gmail.com

Карим Пешанг Хасан – аспирант кафедры прикладной информатики Института прикладной математики и компьютерных наук Национального исследовательского Томского государственного университета (Томск, Россия). E-mail: peshangkarimov@gmail.com

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

Information about the authors:

Mikheev Pavel A. (Candidate of Technical Sciences, National Research Tomsk State University, Tomsk, Russian Federation). E-mail: doka.patrick@gmail.com

Poddubny Vasilii V. (Doctor of Technical Sciences, Professor, National Research Tomsk State University, Tomsk, Russian Federation). E-mail: vvpoddubny@gmail.com

Pristupa Pavel V. (Assistant, National Research Tomsk State University, Tomsk, Russian Federation). E-mail: pristupa@gmail.com

Suschenko Sergey P. (Doctor of Technical Sciences, Professor, National Research Tomsk State University, Tomsk Branch of the Federal Research Center for Information and Computational Technologies, Tomsk, Russian Federation). E-mail: ssp.inf.tsu@gmail.com

Karim Peshang Kh. (Post-graduate Student, National Research Tomsk State University, Tomsk, Russian Federation). E-mail: peshangkarimov@gmail.com

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

Поступила в редакцию 16.08.2022; принята к публикации 29.11.2022

Received 16.08.2022; accepted for publication 29.11.2022

ОБРАБОТКА ИНФОРМАЦИИ

DATA PROCESSING

Original article

doi: 10.17223/19988605/61/3

**Asymptotic properties of modified empirical Kac processes
under general random censorship model****Abdurakhim A. Abdushukurov¹, Gulnoz S. Sayfulloyeva²**¹ *Moscow State University named after M.V.Lomonosov, Tashkent Branch, Tashkent, Uzbekistan, a_abdushukurov@rambler.ru*² *Navoi State Pedagogical Institute, Navoi, Uzbekistan, sayfullayevagulnoz@gmail.com*

Abstract. In this paper, we consider a general random censorship model and prove asymptotic properties of modified empirical Kac processes. This model includes such important special cases as random censorship on the right and competing risks model. Our results use strong approximation method. Cumulative hazard processes also investigated in a similar manner in the general setting.

Keywords: censored data; empirical estimates; Kac estimate; strong approximation; Gaussian process

For citation: Abdushukurov, A.A., Sayfulloyeva, G.S. (2022) Asymptotic properties of modified empirical Kac processes under general random censorship model. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naja tehnik i informatika – Tomsk State University Journal of Control and Computer Science*. 61. pp. 26–36. doi: 10.17223/19988605/61/3

Научная статья

УДК 519.2

doi: 10.17223/19988605/61/3

**Асимптотические свойства модифицированных эмпирических процессов Каца
в общей модели случайного цензурирования****Абдурахим Ахмедович Абдушукуров¹, Гулноз Сайфуллоевна Сайфуллоева²**¹ *Ташкентский филиал Московского государственного университета им. М.В. Ломоносова, Ташкент, Узбекистан, a_abdushukurov@rambler.ru*² *Навуйский государственный педагогический институт, Навои, Узбекистан, sayfullayevagulnoz@gmail.com*

Аннотация. Рассматривается общая модель случайного цензурирования, включающая в себя модели случайного цензурирования справа и конкурирующих рисков. В ней определены эмпирические процессы Каца и определены их модификации. Также исследованы кумулятивные процессы риска с доказательством свойств сильной аппроксимации последовательностями гауссовских процессов.

Ключевые слова: цензурированные данные; эмпирические оценки; оценка Каца; сильная аппроксимация; гауссовские процессы

Для цитирования: Абдушукуров А.А., Сайфуллоева Г.С. Асимптотические свойства модифицированных эмпирических процессов Каца в общей модели случайного цензурирования // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2022. № 61. С. 26–36. doi: 10.17223/19988605/61/3

The empirical distribution function has been widely used as an estimator for the distribution function of the elements of a random sample. It is not, however, appropriate when the observations are incomplete. Developing the corresponding theory of convergence of considered empirical and concentrated processes to a Gaussian process has been obtained by many scientists. A generalization of these results for the case of competing risks or when present various types of censorship considered by authors [see, for example, [1–3]]. These results have numerous statistical application in areas such as medical follow-up studies, life testing, actuarial sciences and demography (see, also, [4–6]). A general scheme of random censorship was considered by authors includes an competing risks model and random censoring from both sides.

1. Mathematical model

Let Z be a real random variable (r.v.) with distribution function (d.f.) $H(x) = P(Z \leq x)$, $x \in \mathbb{R}$. For a fixed integer $k \geq 1$ let $A^{(1)}, \dots, A^{(k)}$ be pairwise disjoint random events, and define the subdistribution functions $H(x; i) = P(Z \leq x, A^{(i)})$, $i \in \mathfrak{S} = \{1, \dots, k\}$. Suppose that when observing Z we are interested in the joint behaviour of the pairs $(Z, A^{(i)})$, $i \in \mathfrak{S}$. Let $\{(Z_j, A_j^{(1)}, \dots, A_j^{(k)}), j \geq 1\}$ be a sequence of independent replicas of the $(Z, A^{(1)}, \dots, A^{(k)})$ defined on some probability space $\{\Omega, \mathcal{A}, P\}$. We assume throughout that the functions $H(x), H(x; 1), \dots, H(x; k)$ are continuous. Let $H_n(x)$ denote the ordinary empirical d.f. of Z_1, \dots, Z_n and introduce the empirical sub d.f. $H_n(x; i)$, $i \in \mathfrak{S}$

$$H_n(x; i) = \frac{1}{n} \sum_{j=1}^n \delta_j^{(i)} I(Z_j \leq x), \quad (x; i) \in \bar{\mathbb{R}} \times \mathfrak{S},$$

where $\bar{\mathbb{R}} = [-\infty; \infty]$, $\delta_j^{(i)} = I(A_j^{(i)})$ is an indicator of event $A_j^{(i)}$ and

$$H_n(x; 1) + \dots + H_n(x; k) = \frac{1}{n} \sum_{j=1}^n I(Z_j \leq x) = H_n(x), \quad x \in \bar{\mathbb{R}},$$

is the ordinary empirical d.f. Properties of many biometrical estimates depends on limit behaviours of proposed empirical statistics.

The following results are a straightforward consequences of exponential inequality of Dvoretzky-Kiefer-Wolfowitz with exactly constant $D = 2$ from [7, 8]:

For all $n = 1, 2, \dots$ and $\varepsilon > 0$:

$$P \left(\sup_{|x| < \infty} |H_n(x) - H(x)| > \left(\frac{(1 + \varepsilon)}{2} \cdot \frac{\log n}{n} \right)^{1/2} \right) \leq 2n^{-(1 + \varepsilon)}, \quad (1)$$

and

$$P \left(\sup_{|x| < \infty} |H_n(x; i) - H(x; i)| > 2 \left(\frac{(1 + \varepsilon)}{2} \frac{\log n}{n} \right)^{1/2} \right) \leq 4n^{-(1 + \varepsilon)}. \quad (2)$$

A crucial role is played the vector-valued empirical process

$$\left\{ a_n(t) = \left(a_n^{(0)}(t_0), a_n^{(1)}(t_1), \dots, a_n^{(k)}(t_k) \right), t = (t_0, \dots, t_k) \in \bar{\mathbb{R}}^{k+1} \right\};$$

where

$$a_n^{(0)}(x) = \sqrt{n} (H_n(x) - H(x)), \quad a_n^{(i)}(x) = \sqrt{n} (H_n(x; i) - H(x; i)), \quad i \in \mathfrak{S}.$$

The results of our approximation theorems presented here is, quite naturally, the approximation theorems of Komlós–Major–Tusnády’s, for the ordinary empirical process with the approximation with the rate of order $n^{-1/2} \log n$. We will construct the approximation Gaussian processes in terms of Wiener sequences. The following theorem of Burke-Csörgő-Horváth [9, 10] is an extended analogue of Komlós–Major–Tusnády’s result [11, 12].

Theorem A [9, 10]. If the underlying probability space $\{\Omega, A, P\}$ is rich enough, then one can define $k+1$ sequences of Gaussian processes $B_n^{(0)}(x), B_n^{(1)}(x), \dots, B_n^{(k)}(x)$ such that for $a_n(t)$ and $B_n(t) = (B_n^{(0)}(x_0), B_n^{(1)}(x_1), \dots, B_n^{(k)}(x_k))$, $t = (t_0, \dots, t_k)$, we have

$$P \left\{ \sup_{t \in \overline{\mathbb{R}}^{k+1}} \|a_n(t) - B_n(t)\|^{(k+1)} > n^{-1/2} (M(\log n) + z) \right\} \leq K \exp(-\lambda z), \quad (3)$$

for all real z , where $M = (2k+1)A_1$, $K = (2k+1)A_2$ and $\lambda = A_3/(2k+1)$ with A_1, A_2 and A_3 are absolute constants. Moreover, B_n itself is a $(k+1)$ dimensional vector-valued Gaussian process having the same covariance structure as the vector $a_n(t)$, namely $EB_n^{(i)}(x) = 0$, $(x, i) \in \overline{\mathbb{R}} \times \overline{\mathfrak{S}} = \overline{\mathfrak{S}} \cup \{0\}$ and for any $i, j \in \overline{\mathfrak{S}}$, $i \neq j$, $x, y \in \overline{\mathbb{R}}$:

$$\begin{aligned} EB_n^{(0)}(x)B_n^{(0)}(y) &= \min\{H(x), H(y)\} - H(x) \cdot H(y), \\ EB_n^{(i)}(x)B_n^{(i)}(y) &= \min\{H(x; i), H(y; i)\} - H(x; i) \cdot H(y; i), \\ EB_n^{(i)}(x)B_n^{(j)}(y) &= -H(x; i) \cdot H(y; j), \\ EB_n^{(0)}(x)B_n^{(i)}(y) &= \min\{H(x; i), H(y; j)\} - H(x) \cdot H(y; i). \end{aligned} \quad (4)$$

Note that in proving of theorem A (theorem 3.1 in [10]) authors constructed sequence of two – parametrical Gaussian processes $\mathbb{Q}^{(0)}(x; n), \mathbb{Q}^{(2)}(x; n), \dots, \mathbb{Q}^{(k)}(x; n)$ such that for $a_n(t)$ and $\mathbb{Q}(t; n) = (\mathbb{Q}^{(0)}(x; n), \dots, \mathbb{Q}^{(k)}(x; n))$, $t \in \overline{\mathbb{R}}^{k+1}$ the following its Borel-Cantelly consequence of approximation have used

$$\sup_{t \in \overline{\mathbb{R}}^{k+1}} \|a_n(t) - n^{-1/2} \mathbb{Q}(t; n)\|^{(k+1)} \stackrel{a.s.}{=} O\left(n^{-1/2} \log^2 n\right),$$

where $\mathbb{Q}(t; n)$ is the $(k+1)$ dimensional vector-valued Gaussian process that $\mathbb{Q}(t; n) \stackrel{D}{=} n^{1/2} a_n(t)$. Hence

$$E\mathbb{Q}^{(i)}(x; n) = 0, \quad (x, i) \in \overline{\mathbb{R}} \times \overline{\mathfrak{S}}$$

and for any $i, j \in \overline{\mathfrak{S}}$, $i \neq j$, $x, y \in \overline{\mathbb{R}}$:

$$\begin{aligned} E\mathbb{Q}^{(0)}(x; n)\mathbb{Q}^{(0)}(y; m) &= \min(n, m) \{ \min\{H(x), H(y)\} - H(x)H(y) \}, \\ E\mathbb{Q}^{(0)}(x; n)\mathbb{Q}^{(i)}(y; m) &= \min(n, m) \{ \min\{H(x), H(y; i)\} - H(x)H(y; i) \}, \\ E\mathbb{Q}^{(i)}(x; n)\mathbb{Q}^{(i)}(y; m) &= \min(n, m) \{ \min\{H(x; i), H(y; i)\} - H(x; i)H(y; i) \}, \\ E\mathbb{Q}^{(i)}(x; n)\mathbb{Q}^{(j)}(y; m) &= -\min(n, m) H(x; i) \cdot H(y; j). \end{aligned} \quad (5)$$

Observe that $\{\mathbb{Q}^{(i)}, i \in \overline{\mathfrak{S}}\}$ are Kiefer processes and they satisfying the distributional equality

$$\mathbb{Q}^{(i)}(x; n) \stackrel{D}{=} W^{(i)}(H(x; i); n) - H(x; i)W^{(i)}(1; n), \quad (6)$$

where $\{W^{(i)}(y; n), 0 \leq y \leq 1, n \geq 1, i \in \overline{\mathfrak{S}}\}$ itself are two-parametrical Wiener processes with $EW^{(i)}(y; n) = 0$ and

$$EW^{(i)}(y;n)W^{(i)}(u;m) = \min(n,m)\min(y,u), \quad i \in \mathfrak{I}.$$

It is important to note that though Kiefer processes $\{\mathbb{Q}^{(i)}, i \in \mathfrak{I}\}$ are dependent processes, but corresponding Wiener processes are independent. Indeed, from proof of theorem A are follows that

$$\begin{aligned} \mathbb{Q}^{(1)}(x;n) &\stackrel{D}{=} K(H(x;1);n), \\ \mathbb{Q}^{(2)}(x;n) &\stackrel{D}{=} K(H(x;2) - H(+\infty;1);n) - K(H(+\infty;1);n), \\ &\dots \end{aligned}$$

$$\mathbb{Q}^{(i)}(x;n) \stackrel{D}{=} K(H(x;i) + H(+\infty;1) + \dots + H(+\infty;i-1);n) - K(H(+\infty;1) + \dots + H(+\infty;i-1);n), \quad i \in \mathfrak{I},$$

where $H(+\infty;i) = \lim_{x \uparrow +\infty} H(x;i)$, $H(+\infty;1) + \dots + H(+\infty;k) = 1$.

The Kiefer processes $\{K(y;n), 0 \leq y \leq 1, n \geq 1\}$ are represented through two-parametrical Wiener processes $\{W(y;n), 0 \leq y \leq 1, n \geq 1\}$ by distributional equality

$$\{K(y;n), 0 \leq y \leq 1, n \geq 1\} \stackrel{D}{=} \{W(y;n) - yW(1;n), 0 \leq y \leq 1, n \geq 1\}. \quad (7)$$

Consequently, in view of (6) and (7) the Wiener process $\{W^{(i)}, i \in \mathfrak{I}\}$ also admits representations for all $(x;i) \in \overline{\mathbb{R}} \times \mathfrak{I}$:

$$\begin{aligned} W^{(1)}(H(x;1);n) &\stackrel{D}{=} W(H(x;1);n), \\ W^{(2)}(H(x;2);n) &\stackrel{D}{=} W(H(x;2) + H(+\infty;1);n) - W(H(+\infty;1);n), \\ &\dots \\ W^{(i)}(H(x;i);n) &\stackrel{D}{=} W(H(x;i) + H(+\infty;i-1);n) - W(H(+\infty;1) + \dots + H(+\infty;i-1);n). \end{aligned}$$

Now by directly calculations of covariance of processes $\{W^{(i)}, i \in \mathfrak{I}\}$ it is easy to believing on its independency.

This paper further structured as follows. In section 1 we introduce the classical Kac processes analogues and their modifications. For its we prove approximation results. Then in section 2 we propose corresponding estimators of hazard functions. For them we also prove approximation results.

2. Kac processes under general censoring

Authors [9] proved the general theorems to obtain approximation for the usual empirical and corresponding cumulative hazard estimates by Gaussian processes for the competing risk generalizations. We prove these results for a corresponding Kac-type processes.

Following of [12] we introduce the modified empirical d.f. of Kac by the following way. Along with sequence $\{Z_j, j \geq 1\}$ on a probability space $\{\Omega, A, P\}$ consider also a sequence $\{v_n, n \geq 1\}$ of r.v.-s having Poisson distribution with parameter $Ev_n = n$, $n = 1, 2, \dots$. Assume throughout that the two sequences $\{Z_j, j \geq 1\}$ and $\{v_n, n \geq 1\}$ are independent. Kac's empirical d.f. is

$$H_n^*(x) = \begin{cases} \frac{1}{n} \sum_{j=1}^{v_n} I(Z_j \leq x), & \text{if } v_n \geq 1 \text{ a.s.}, \\ 0, & \text{if } v_n = 0 \text{ a.s.}, \end{cases}$$

while the empirical sub-d.f. one is

$$H_n^*(x; i) = \begin{cases} \frac{1}{n} \sum_{j=1}^{v_n} I(Z_j \leq x, A_j^{(i)}), & i \in \mathfrak{I}, \text{ if } v_n \geq 1 \text{ a.s.}, \\ 0, & i \in \mathfrak{I} \text{ if } v_n = 0 \text{ a.s.}, \end{cases}$$

with $H_n^*(x; 1) + \dots + H_n^*(x; k) = H_n^*(x)$ for all $x \in \bar{\mathbb{R}}$. Here we suppose that sequence $\{v_n, n \geq 1\}$ is independent of random vectors $\left\{ \left(Z_j, \delta_j^{(1)}, \dots, \delta_j^{(k)} \right), j \geq 1 \right\}$, where $\delta_j^{(i)} = I(A_j^{(i)})$. Note that statistics $H_n^*(x; i)$ (consequently also $H_n^*(x)$) are unbiased estimators of $H(x; i)$, $i \in \mathfrak{I}$ (consequently also of $H(x)$):

$$\begin{aligned} E(H_n^*(x; i)) &= \frac{1}{n} E \left\{ \sum_{m=1}^{\infty} E \left[\sum_{k=1}^n \delta_k^{(i)} \cdot I(Z_k \leq x) \right], v_n = m \right\} = \\ &= \frac{1}{n} E \left\{ \sum_{m=1}^{\infty} E \left[\sum_{k=1}^n \delta_k^{(i)} \cdot I(Z_k \leq x) / v_n = m \right] \cdot P(v_n = m) \right\} = \\ &= \frac{1}{n} \sum_{m=1}^{\infty} H(x; i) m P(v_n = m) = \frac{1}{n} H(x; i) \sum_{m=1}^{\infty} m \cdot \frac{n^m e^{-n}}{m!} = \\ &= H(x; i) e^{-n} \sum_{m=0}^{\infty} \frac{n^m}{m!} = H(x; i), \quad (x; i) \in \bar{\mathbb{R}} \times \mathfrak{I}. \end{aligned}$$

Consequently,

$$E[H_n^*(x)] = \sum_{i=1}^k E[H_n^*(x; i)] = \sum_{i=1}^k H(x; i) = H(x), \quad x \in \bar{\mathbb{R}}.$$

Let's define $a_n^{(i)*}(x) = \sqrt{n}(H_n^*(x; i) - H(x; i))$, $i \in \mathfrak{I}$ and $a_n^{(0)*}(x) = \sqrt{n}(H_n^*(x) - H(x))$ the empirical Kac processes.

Theorem 1. If the underlying probability space $\{\Omega, A, P\}$ is rich enough, then one can define $k+1$ sequences of Gaussian processes $W_n^{(0)}(x), W_n^{(1)}(x), \dots, W_n^{(k)}(x)$ such that for $a_n^*(t) = (a_n^{(0)*}(t_0), a_n^{(1)*}(t_1), \dots, a_n^{(k)*}(t_k))$ and $W_n^*(t) = (W_n^{(0)}(t_0), W_n^{(1)}(t_1), \dots, W_n^{(k)}(t_k))$, $t = (t_0, t_1, \dots, t_k)$, we have

$$P \left\{ \sup_{t \in \bar{\mathbb{R}}^{k+1}} \|a_n^*(t) - W_n^*(t)\|^{(k+1)} > C^* n^{-1/2} \log n \right\} \leq K^* n^{-r}, \quad (8)$$

where $r \geq 2$ is an arbitrary integer, $C^* = C^*(r)$ -depends only on r and K^* is an absolute constant. Moreover, $W_n^*(t)$ itself is a $(k+1)$ -dimensional vector-valued Gaussian process with expectation $EW_n^{(i)}(x) = 0$, $(x, i) \in \bar{\mathbb{R}} \times \mathfrak{I}$ and for any $i, j \in \mathfrak{I}$, $i \neq j$, $x, y \in \bar{\mathbb{R}}$:

$$\begin{aligned} EW_n^{(0)}(x)W_n^{(0)}(y) &= \min\{H(x), H(y)\}, \\ EW_n^{(i)}(x)W_n^{(j)}(y) &= \min\{H(x; i), H(y; j)\}, \\ EW_n^{(i)}(x)W_n^{(0)}(y) &= \min\{H(x; i), H(y)\}. \end{aligned} \quad (9)$$

The basic relation between $a_n(t)$ and $a_n^*(t)$ is the following easily checked identity

$$a_n^*(x) = \sqrt{\frac{v_n}{n}} a_{v_n}^{(i)}(x) + H(x; i) \frac{(v_n - n)}{\sqrt{n}}, \quad i \in \mathfrak{I} \quad (10)$$

Hence the approximating sequence have respectively the form

$$W_n^{(i)}(x) = B_{v_n}^{(i)}(x) + H(x; i) \frac{W_n^*(n)}{\sqrt{n}}, \quad i \in \mathfrak{I},$$

where $B_{v_n}^{(i)}(x)$ is a Poisson indexed Brownian bridge type process of Teorem A and $\{W^{(*)}(x), x \geq 0\}$ is a Wiener process. Easy to verify that $\{W_n^{(i)}(x), (x;i) \in \mathbb{R} \times \overline{\mathfrak{I}}\} \stackrel{D}{=} \{W^*(H(x;i)), (x,i) \in \overline{\mathbb{R}} \times \overline{\mathfrak{I}}\}$. The proof of Teorem 1 is coincides with the proof of theorem 1 of Stute in [13] hence it is omitted.

In so far as $\lim_{x \uparrow +\infty} H_n^*(x) = H_n^*(+\infty) = \frac{V_n}{n}$, then by Stirlings formula

$$P(v_n = n) = P(H_n^*(+\infty) = 1) = \frac{n^n e^{-n}}{n!} = \frac{1}{\sqrt{2\pi n}}(1 + o(1)), \quad n \rightarrow \infty,$$

and

$$P(H_n^*(+\infty) > 1) = P(v_n > n) = \sum_{k=n+1}^{\infty} \frac{n^k e^{-n}}{k!} = o(1), \quad n \rightarrow \infty.$$

Thus $H_n^*(x)$ with positive probability t_0 be greater than 1. In order to avoid these undesirable properties, we propose following modifications of Kac statistics:

$$\begin{aligned} H_n(x) &= 1 - (1 - H_n^*(x))I(H_n^*(x) < 1), \quad x \in \overline{\mathbb{R}}, \\ H_n(x;i) &= 1 - (1 - H_n^*(x;i))I(H_n^*(x;i) < 1), \quad (x;i) \in \overline{\mathbb{R}} \times \overline{\mathfrak{I}}. \end{aligned} \quad (11)$$

The following inequalities are useful in investigating of Kac processes.

Theorem 2. Let $\{v_n, n \geq 1\}$ be a sequence of Poisson r.v.-s with $E v_n = n$. Then for any $\varepsilon > 0$ such that

$$\frac{n}{\log n} \geq \frac{\varepsilon}{8(1+e/3)^2}, \quad (12)$$

we have

$$P\left(|v_n - n| > \frac{1}{2}\left(\frac{\varepsilon}{2}n \log n\right)^{1/2}\right) \leq 2n^{-\varepsilon w}, \quad (13)$$

$$P\left(\sup_{|x| < \infty} |H_n^*(x;i) - H(x;i)| > 2\left(\frac{\varepsilon \log n}{2n}\right)^{1/2}\right) \leq 4n^{-4\varepsilon w}, \quad i \in \overline{\mathfrak{I}}, \quad (14)$$

$$P\left(\sup_{|x| < \infty} |H_n(x;i) - H(x;i)| > 2\left(\frac{\varepsilon \log n}{2n}\right)^{1/2}\right) \leq 4n^{-4\varepsilon w}, \quad i \in \overline{\mathfrak{I}}, \quad (15)$$

where $w = [16(1+e/3)]^{-1}$.

Proof. Let $\gamma_1, \gamma_2, \dots$ be a sequence of Poisson r.v.-s with $E \gamma_k = 1$ for all $k = 1, 2, \dots$. Then

$$S_n = v_n - v = \sum_{k=1}^n (\gamma_k - 1) = \sum_{k=1}^n \xi_k \quad \text{and} \quad E \exp(t \xi_k) = e^{-t} \exp(t \gamma_1) = \exp(-(t+1)) \sum_{k=0}^{\infty} \frac{(e^t)^k}{k!} = \exp\{e^t - (t+1)\}.$$

Using Taylor expansion for e^t , we get

$$E \exp(t \xi_k) = \exp\left\{1 + t + \frac{t^2}{2} + \psi(t) - (t+1)\right\} = \exp\left\{\frac{t^2}{2} + \psi(t)\right\},$$

where $\psi(t) = \frac{t^3}{6} \exp(\theta t)$, $0 < \theta < 1$. Estimate $\psi(t)$ taking into account that $t^3 \leq t^2$ under $0 \leq t \leq 1$:

$$\psi(t) \leq \frac{t^3}{6} e \leq e \frac{t^2}{6}. \quad \text{Thus, } E \exp(t \xi_k) = \exp\left\{\frac{t^2}{2}\left(1 + \frac{e}{3}\right)\right\}, \quad 0 \leq t \leq 1.$$

The following result (theorem from [14]) is necessary for our further investigations.

Lemma 1 [14]. Let $\{\xi_n, n \geq 1\}$ be a sequence of independent r.v.-s with $E \xi_n = 0$, $n = 1, 2, \dots$. Suppose that $U, \lambda_1, \dots, \lambda_n$ positive real numbers such that

$$E \exp(t\xi_k) \leq \exp\left(\frac{1}{2}\lambda_k t^2\right) \text{ for } k=1,2,\dots,n \quad |t| \leq U. \quad (16)$$

Let $\Lambda = \lambda_1 + \dots + \lambda_n$. Then

$$P(|\xi_1 + \dots + \xi_k| \geq z) \leq \begin{cases} 2 \exp\left(-\frac{z^2}{2\Lambda}\right), & \text{if } 0 \leq z \leq \Lambda U, \\ 2 \exp\left(-\frac{Uz}{2}\right), & \text{if } z > \Lambda U. \end{cases}$$

Let in lemma 1 $\lambda_k = 1 + e/3$, $U = 1$, $z = \frac{1}{2}\left(\frac{\varepsilon}{2}n \log n\right)^{1/2}$, then we obtain (13). Here $0 \leq z = \frac{1}{2}\left(\frac{\varepsilon}{2}n \log n\right)^{1/2} \leq (1 + e/3)n = \Lambda U$. Consider probability in (14). By total probability formula

$$\begin{aligned} & P\left(\sup_{|x| < \infty} |H_n^*(x; i) - H(x; i)| > 2\left(\frac{\varepsilon \log n}{2n}\right)^{1/2}\right) = \\ & = P\left(\sup_{|x| < \infty} \left|H_n(x; i) - H(x; i) + \frac{1}{n} \sum_{k=v_n+1}^{v_n} \delta_k^{(i)} I(Z_k \leq x)\right| > 2\left(\frac{\varepsilon \log n}{2n}\right)^{1/2} \middle/ v_n > n\right) \cdot P(v_n > n) + \\ & + P\left(\sup_{|x| < \infty} \left|H(x; i) - H(x; i) - \frac{1}{n} \sum_{k=v_n+1}^n \delta_k^{(i)} I(Z_k \leq x)\right| > 2\left(\frac{\varepsilon \log n}{2n}\right)^{1/2} \middle/ v_n \leq n\right) \cdot P(v_n \leq n) \leq \\ & \leq P\left(\sup_{|x| < \infty} |H_n(x; i) - H(x; i)| > \left(\frac{\varepsilon \log n}{2n}\right)^{1/2}\right) + P\left(\sup_{|x| < \infty} \left|\frac{1}{n} \sum_{k=\min(n, v_n)+1}^{\max(n, v_n)} \delta_k^{(i)} I(Z_k \leq x)\right| > \left(\frac{\varepsilon \log n}{2n}\right)^{1/2}\right) \leq \\ & \leq 2n^{-4\varepsilon} + P\left(\left|\frac{v_n - n}{n}\right| > \left(\frac{\varepsilon \log n}{2n}\right)^{1/2}\right) \leq 2n^{-4\varepsilon} + 2n^{-4w\varepsilon} \leq 4n^{-4w\varepsilon}, \quad i \in \mathfrak{I}, \end{aligned}$$

where we applied (2) and (13) that proves (14). Let's define $T_n^{(i)} = \inf\{x: H_n(x; i) = 1\}$, $i \in \mathfrak{I}$. If $x \geq T_n^{(i)}$ and $v_n > n$, then $H_n(x; i) = 1$ and $H_n^*(x; i) - H(x; i) \geq H_n^*(x; i) - H_n(x; i) \geq 0$. Then assuming $v_n > n$, we obtain

$$\begin{aligned} & \sup_{|x| < \infty} |H_n(x; i) - H(x; i)| = \left\{ \max \left[\sup_{x < T_n^{(i)}} |H_n^*(x; i) - H(x; i)|, \sup_{x \geq T_n^{(i)}} |H_n(x; i) - H(x; i)| \right] \right\} \leq \\ & \leq \left\{ \max \left[\sup_{x < T_n^{(i)}} |H_n^*(x; i) - H(x; i)|, \sup_{x \geq T_n^{(i)}} |H_n^*(x; i) - H(x; i)| \right] \right\} = \sup_{|x| < \infty} |H_n^*(x; i) - H(x; i)|, \quad i \in \mathfrak{I}. \end{aligned} \quad (17)$$

Under $v_n \leq n$, it is obvious that $H_n(x; i) = H_n^*(x; i)$, for all $(x; i) \in \overline{\mathbb{R}} \times \mathfrak{I}$.

Now taking into account last two relations, total probability formula and (14) we obtain (15). Theorem 2 is proved.

Let $a_n(t) = (a_n^{(0)}(t_0), a_n^{(1)}(t_1), \dots, a_n^{(k)}(t_k))$, where $a_n^{(0)}(x) = \sqrt{n}(H_n(x) - H(x))$, $a_n^{(i)}(x) = \sqrt{n}(H_n(x; i) - H(x; i))$, $(x; i) \in \overline{\mathbb{R}} \times \mathfrak{I}$. We shall prove an approximation theorem of the vector-valued modified empirical Kac process $a_n(t)$ by the appropriate Gaussian vector-valued process $W_n^*(t)$, $t \in \overline{\mathbb{R}}^{k+1}$ from theorem 2.

Theorem 3. Let $\{T_n, n \geq 1\}$ be a numerical sequence satisfying, for each n , the condition $T_n < T_H = \inf\{x: H(x) = 1\} \leq \infty$ such that

$$\min_{i \in \mathfrak{I}} \left\{ P(A^{(i)} - H(T_n, i)) \right\} \geq 1 - H(T_n) \geq 2\left(\frac{r \log n}{2wn}\right)^{1/2}. \quad (18)$$

If for any $\varepsilon > 0$ condition (12) hold, then on a probability space of theorem 2 one can define $k + 1$ sequences of mean zero Gaussian processes $W_n^{(0)}(x), W_n^{(1)}(x), \dots, W_n^{(k)}(x)$ with the covariance structure (9) such that for $a_n(t)$ and $W_n^*(t) = (W_n^{(0)}(t_0), W_n^{(1)}(t_1), \dots, W_n^{(k)}(t_k))$ we have

$$P \left\{ \sup_{t \in (-\infty; T_n]^{k+1}} \|a_n(t) - W_n^*(t)\|^{(k+1)} > Cn^{-1/2} \log n \right\} \leq Kn^{-\beta}, \quad (19)$$

where K is an absolute constant, $C = C(\varepsilon)$ and $\beta = \min(r, \varepsilon w)$ for any $\varepsilon > 0$.

Proof. It is easy to seen that probability in (19) can be estimated by sum

$$P \left\{ \sup_{x \leq T_n} |a_n^{(0)}(x) - W_n^{(0)}(x)| > Cn^{-1/2} \log n \right\} + \sum_{i=1}^k P \left\{ \sup_{x \leq T_n} |a_n^{(i)}(x) - W_n^{(i)}(x)| > Cn^{-1/2} \log n \right\} = q_{1n} + q_{2n}. \quad (20)$$

Taking into account that for any $x \leq T_n$, $H_n^*(x) \leq H_n^*(T_n)$ and if $H_n^*(T_n) \leq 1$, then $a_n^{(0)}(x) = a_n^{(0)*}(x)$ and by formula of total probability

$$\begin{aligned} q_{1n} &\leq P \left(\sup_{x \leq T_n} |a_n^{(0)}(x) - W_n^{(0)}(x)| > Cn^{-1/2} \log n / H_n^*(T_n) \leq 1 \right) + P(H_n^*(T_n) > 1) \leq \\ &\leq P \left(\sup_{x \leq T_n} |a_n^{(0)*}(x) - W_n^{(0)}(x)| > Cn^{-1/2} \log n \right) + P(H_n^*(T_n) > 1) \leq \\ &\leq Kn^{-r} + P(H_n^*(T_n) - H(T_n) > 1 - H(T_n)) \leq \\ &\leq Kn^{-r} + P \left(\sup_{|x| < \infty} |H_n^*(x) - H(x)| > \left(\frac{r \log n}{2wn} \right)^{1/2} \right) \leq Ln^{-r}, \end{aligned} \quad (21)$$

where we have used theorem 1 and analogue of (14) for $H_n^* - H$, $L = K^* + 4$. Analogously,

$$\begin{aligned} q_{2n} &\leq \sum_{i=1}^k P \left(\sup_{x \leq T_n} |a_n^{(i)}(x) - W_n^{(i)}(x)| > Cn^{-1/2} \log n \right) + \sum_{i=1}^k P(H_n^*(T_n; i) > P(A^{(i)})) \leq \\ &\leq \sum_{i=1}^k P \left(\sup_{x \leq T_n} |a_n^{(i)*}(x) - W_n^{(i)}(x)| > Cn^{-1/2} \log n \right) + \sum_{i=1}^k P \left(\sup_{|x| < \infty} |a_n^{(i)*}(x) - W_n^{(i)}(x)| > Cn^{-1/2} \log n \right) + \\ &\quad + kP \left(\frac{|V_n - \nu|}{n} > \frac{1}{2} \left(\frac{4r \log n}{2wn} \right)^{1/2} \right) \leq kLn^{-r} + 2kn^{-4r}, \end{aligned} \quad (22)$$

where we also have used inequalities (13), (15) and theorem 1. Now from (21) and (22) follows (19). Theorem 3 is proved.

3. Estimation of hazard function

In many practical situations, when we are interested in the joint behaviors of the pairs $\{(Z, A^{(i)}), i \in \mathfrak{I}\}$, a crucial role is played by the so-called cumulative hazard functions

$\{S^{(i)}(x) = \exp(-\Lambda^{(i)}(x)), i \in \mathfrak{I}\}$, where $\Lambda^{(i)}(x)$ is the i -th hazard function $\left(\int_{-\infty}^x = \int_{(-\infty; x]} \right)$:

$$\Lambda^{(i)}(x) = \int_{-\infty}^x \frac{dH(u; i)}{1 - H(u)}, \quad i \in \mathfrak{I},$$

with $\Lambda^{(1)}(x) + \dots + \Lambda^{(k)}(x) = \Lambda(x) = \int_{-\infty}^x \frac{dH(u)}{1 - H(u)}$ is the corresponding hazard function of d.f. $H(x)$.

Consider two important special cases of considered generalized censorship model:

a) Let $\{X_1, X_2, \dots\}$ be a sequence of independent r.v.-s with common continuous d.f. F . These are censored on the right by $\{Y_1, Y_2, \dots\}$ a sequence of independent r.v.-s, independent of the X – sequence, with common continuous d.f. G . One can only observe the sequence of pairs $\{(Z_k, \delta_k), k = \overline{1, n}\}$, where $Z_j = \min(X_j, Y_j)$ and $\delta_j = \delta_j^{(1)}$ is the indicator of event $A_j = A_j^{(1)} = \{Z_j = X_j\}$. In this case $k = 2$, $1 - H(x) = (1 - F(x))(1 - G(x))$, $H(x; 1) = \int_{-\infty}^x (1 - G(u)) dF(u)$, thus $S^{(1)}(x) = S(x) = 1 - F(x)$. The useful special case when $1 - G(x) = (1 - F(x))^\beta$, $\beta > 0$, which corresponds to independence of r.v.-s Z_j and $\delta_j, j \geq 1$.

b) For $k > 1$ consider independent sequences $\{Y_1^{(i)}, Y_2^{(i)}, \dots\}$ ($i = 1, \dots, k$) of independent r.v.-s with common continuous d.f. F and let $Z_j = \min(Y_j^{(1)}, \dots, Y_j^{(k)})$. One observes the sequences $\{(Z_j, \delta_j^{(i)}), i = \overline{1, k}\}_{j=1}^n$, where $\delta_j^{(i)}$ is the indicator of the event $A_j^{(i)} = \{Z_j = Y_j^{(i)}\}$. This is the competing risks model with $S^{(i)}(x) = 1 - F^{(i)}(x)$, $i \in \mathfrak{I}$.

Define the natural Kac-type estimator

$$\Lambda_n^{(i)}(x) = \int_{-\infty}^x \frac{dH(u; i)}{1 - H_n(u)}, \quad i \in \mathfrak{I},$$

of $\Lambda^{(i)}(x)$, $i \in \mathfrak{I}$. Let $w_n^{(i)}(x) = \sqrt{n}(\Lambda_n^{(i)}(x) - \Lambda^{(i)}(x))$, $i \in \mathfrak{I}$ is an Kac-type hazard process and $w_n(t) = (w_n^{(1)}(t_1), \dots, w_n^{(k)}(t_k))$, $t = (t_1, \dots, t_k)$, $Y_n(t) = (Y_n^{(1)}(t_1), \dots, Y_n^{(k)}(t_k))$ corresponding vector process with

$$Y_n^{(i)}(x) = \int_{-\infty}^x \frac{W_n^{(0)}(u) dH(u; i)}{(1 - H(u))^2} + \frac{W_n^{(i)}(x)}{1 - H(x)} - \int_{-\infty}^x \frac{W_n^{(i)}(u) dH(u)}{(1 - H(u))^2}, \quad i \in \mathfrak{I}$$

and $\{W_n^{(0)}(x), W_n^{(1)}(x), \dots, W_n^{(k)}(x)\}$ are Wiener processes with the covariance structure (9). Then for $i \in \mathfrak{I}$, $EY_n^{(i)}(x) = 0$ and

$$EY_n^{(i)}(x)Y_n^{(i)}(y) = C(x, y),$$

where $x, y \leq T_H = \inf\{x: H(x) = 1\} \leq \infty$.

Theorem 4. Let $\{T_n, n \geq 1\}$ be a numerical sequence satisfying for each n , the condition $T_n < T_H$ such that

$$\frac{n}{\log n} \geq \max\left\{32\varepsilon w^2, \frac{2rb_n^2}{w}, \frac{2\varepsilon b_n^2}{w}\right\}, \quad (23)$$

where $b_n = (1 - H(T_n))^{-1}$, $\varepsilon > 0$, $r \geq 2$. Then on a probability space of theorem 2

$$P\left(\sup_{t \in (-\infty; T_n]^{(k)}} \|w_n(t) - Y_n(t)\|^{(k)} > r(n)\right) \leq k\Phi_1 n^{-\beta}, \quad (24)$$

where $r(n) = \Phi_0 b_n^2 n^{-1/2} \log n$, $\Phi_0 = \Phi_0(\varepsilon, r)$, Φ_1 – (absolute) constants.

Proof. It is enough to prove that for each $i \in \mathfrak{I}$

$$P\left(\sup_{x \leq T_n} (w_n^{(i)}(x) - Y_n^{(i)}(x)) > r(n)\right) \leq \Phi_1 n^{-\beta}. \quad (25)$$

For difference we have representation for each $i \in \mathfrak{I}$:

$$\begin{aligned} w_n^{(i)}(x) - Y_n^{(i)}(x) &= \int_{-\infty}^x \frac{(a_n^{(0)}(u) - W_n^{(0)}(u)) dH(u; i)}{(1 - H(u))^2} + \frac{a_n^{(i)}(x) - W_n^{(i)}(x)}{1 - H(x)} - \\ &- \int_{-\infty}^x \frac{(a_n^{(i)}(u) - W_n^{(i)}(u)) dH(u)}{(1 - H(u))^2} + n^{-1/2} \int_{-\infty}^x \frac{(a_n^{(0)}(u))^2 dH(u; i)}{(1 - H(u))^2 (1 - H_n(u))} + \\ &+ n^{-1/2} \int_{-\infty}^x \frac{a_n^{(0)}(u) da_n^{(i)}(u)}{(1 - H(u))(1 - H_n(u))} = \sum_{m=1}^4 R_{mn}^{(i)}(x). \end{aligned}$$

For sum $R_{1n}^{(i)}(x) + R_{2n}^{(i)}(x) + R_{3n}^{(i)}(x)$ using (15) and (19) we have

$$P\left(\sup_{x \leq T_n} \left| \sum_{m=1}^4 R_{mn}^{(i)}(x) \right| > 3Cn^{-1/2} \log n + \varepsilon n^{-1/2} b_n^3 \log n\right) \leq 3Kn^{-\beta} + 2Ln^{-w\varepsilon} \leq (3K + 2L)n^{-\beta}, \quad i \in \mathfrak{I}. \quad (26)$$

Rewrite $R_{4n}^{(i)}$ as

$$R_{4n}^{(i)}(x) = n^{-1/2} \int_{-\infty}^x \frac{(a_n^{(0)}(u))^2 d(H(u; i) - H(u; i))}{(1 - H(u))^2 (1 - H_n(u))} + n^{-1/2} \int_{-\infty}^x \frac{a_n^{(0)}(u) da_n^{(i)}(u)}{(1 - H(u))^2} = \bar{R}_{4n}^{(i)}(x) + \overline{\bar{R}}_{4n}^{(i)}(x). \quad (27)$$

Then by (15) for $i \in \mathfrak{I}$

$$P\left(\sup_{x \leq T_n} \left| \bar{R}_{4n}^{(i)}(x) \right| > 2\varepsilon n^{-1/2} b_n^3 \log n\right) \leq 2Ln^{-w\varepsilon} \leq 2Ln^{-\beta}. \quad (28)$$

There exists an absolute constant A such that

$$\begin{aligned} P\left(\sup_{x \leq T_n} \left| \overline{\bar{R}}_{4n}^{(i)}(x) \right| > 3An^{-1/2} b_n^2 \log n\right) &\leq P(H_n^*(T_n) > 1) + \\ &+ P\left(\sup_{x \leq T_n} n^{-1/2} \left| \int_{-\infty}^x \frac{a_n^{(0)*}(u) da_n^{(i)*}(u)}{(1 - H(u))^2} \right| > 3An^{-1/2} b_n^2 \log n\right) \leq Ln^{-r} + p_n, \end{aligned} \quad (29)$$

so that for any $x \leq T_n$, $H_n^*(x) \leq H_n^*(T_n)$ and if $H_n^*(T_n) \leq 1$, then $H_n^*(x; i) \leq H_n^*(T_n)$ and hence

$a_n^{(i)}(x) = a_n^{(i)*}(x)$ for $i \in \mathfrak{I}$. It is enough to estimate probability p_n . According to proof of theorem 1 in [13], supposing $a_{v_n}^{(0)}(x) = \sqrt{v_n}(H_{v_n}^*(x) - H(x))$, $a_{v_n}^{(i)}(x) = \sqrt{v_n}(H_{v_n}^*(x; i) - H(x; i))$, $i \in \mathfrak{I}$ and using representation (10), we have proved the theorem 4.

Conclusion

We consider Kac processes in a general censorship scheme, including competing risks model and random censoring from both sides. Our results uses strong approximation method. Cumulative hazard processes also investigated in a similar manner in the general setting. In paper we obtain corresponding approximation results for ordinary empirical processes, for a Kac processes and their modifications and for hazard processes. All results are new and have approximation rates of order $n^{-1/2} \log n$.

References

1. Abdushukurov, A.A. (1998) Nonparametric estimation of the distribution function based on relative risk function. *Communications in Statistics – Theory and Methods*. 27(8). pp. 1991–2012.

2. Abdushukurov, A.A. (1999) On nonparametric estimation of reliability indices by censored samples. *Theory of Probability and Its Applications*. 43(1). pp. 3–11.
3. Csörgő, S. & Horváth, L. (1982) On random censorship from the right. *Acta Scientiarum Mathematicarum*. 44. pp. 23–34.
4. Shulenin, V.P. (2010) Asymptotic and nonasymptotic properties of Hodges – Lehmann adaptive estimators. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 2(11). pp. 96–112.
5. Shulenin, V.P. (2021) Asymptotic properties and robustness of trimmed versions estimates of standard deviation and mean absolute deviations. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 55. pp. 91–102. DOI: 10.17223/19988605/55/11
6. Zenkova, J.N. (2011) Using of information about Sf-equal-shoulder symmetry at the processing of censored data. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 2(15). pp. 32–45.
7. Dvoretzky, A., Kiefer, J. & Wolfowitz, J. (1956) Asymptotic minimax character of the sample distribution function and of the multinomial estimator. *The Annals of Mathematical Statistics*. 27. pp. 642–669.
8. Massart, P. (1990) The tight constant in the Dvoretzky-Kiefer-Wolfowitz inequality. *The Annals of Probability*. 18(3). pp. 1269–1283.
9. Abdushukurov, A.A. (2009) *Statistics of Incomplete Observations*. Tashkent: University Press.
10. Burke, M.D., Csörgő, S. & Horváth, L. (1981) Strong approximations of some biometric estimates under random censorship. *Z. Wahrschein. Verw. Gebiete*. 56. pp. 87–112.
11. Mason, D.M. (2015) Classical empirical process theory and wighted approximation. *Comunicaciones del CIMAT*. N.I-15-03/09-11-2015/ (PE/CIMAT).
12. Mason, D.M. (2017) Selected definition and results from modern empirical theory. *Comunicaciones del CIMAT*. N.I-17-01, 16.03.2017. (PE).
13. Csörgő, S. (1980) Strong approximation of empirical Kac processes. *Carleton Math. Lect. Note*. 26. pp. 71–86.
14. Petrov, W.W. (1987) *Limit theorems for sum of random variables*. Moscow: Nauka.

Information about the authors:

Abdushukurov Abdurakhim A. (Doctor of Physical and Mathematical Sciences, Professor, Branch of Moscow State University in Tashkent, Tashkent, Uzbekistan). E-mail: a_abdushukurov@rambler.ru

Sayfulloeva Gulnoz S. (Post-graduate Student of the Department of Probability Theory and Mathematical Statistics of the Navoi State Pedagogical Institute, Navoi, Uzbekistan). E-mail: sayfullayevagulnoz@gmail.com

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

Информация об авторах:

Абдушукуров Абдурахим Ахмедович – профессор, доктор физико-математических наук, профессор Ташкентского филиала Московского государственного университета им. М.В. Ломоносова (Ташкент, Республика Узбекистан). E-mail: a_abdushukurov@rambler.ru

Сайфуллоева Гулноз Сайфуллоевна – докторант кафедры теории вероятностей и математической статистики Навоийского государственного педагогического института (Навои, Республика Узбекистан). E-mail: sayfullayevagulnoz@gmail.com

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

Поступила в редакцию 04.03.2022; принята к публикации 29.11.2022

Received 04.03.2022; accepted for publication 29.11.2022

Научная статья

УДК 004.91

doi: 10.17223/19988605/61/4

Подход к восстановлению временных рядов пространственных данных на основе информационно-корреляционного и геостатистического анализа

Гульнара Равилевна Воробьева¹, Андрей Владимирович Воробьев²

^{1,2} Уфимский государственный авиационный технический университет, Уфа, Россия

¹ gulnara.vorobeveva@gmail.com

² cpu16bit@gmail.com

Аннотация. Обсуждаются вопросы повышения эффективности процессов обработки и анализа пространственных данных в контексте решения задачи восстановления соответствующих временных рядов с точностью, не превышающей заданной допустимой величины. Предложен подход, основанный на резервировании распределенных источников пространственных данных в соответствии с результатами их теоретико-информационного анализа, с одной стороны, и геостатистического анализа – с другой. На примере результатов мониторинга параметров геомагнитного поля показана эффективность применения предложенного подхода.

Ключевые слова: пространственные данные; обработка данных; теоретико-информационный анализ; геостатистический анализ; восстановление временных рядов

Благодарности: Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 20-07-00011.

Для цитирования: Воробьева Г.Р., Воробьев А.В. Подход к восстановлению временных рядов пространственных данных на основе информационно-корреляционного и геостатистического анализа // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2022. № 61. С. 37–46. doi: 10.17223/19988605/61/4

Original article

doi: 10.17223/19988605/61/4

An approach to the recovery of spatial data time series based on information-correlation and geostatistical analysis

Gulnara R. Vorobeveva¹, Andrei V. Vorobev²

^{1,2} Ufa State Aviation Technical University, Ufa, Russian Federation

¹ gulnara.vorobeveva@gmail.com

² cpu16bit@gmail.com

Abstract. The issues of increasing the efficiency of processing and analysis of spatial data in the context of solving the problem of recovering the corresponding time series with an accuracy not exceeding a given allowable value are discussed. An approach based on redundancy of distributed sources of spatial data in accordance with the results of their information-theoretic analysis, on the one hand, and geostatistical analysis, on the other, is proposed. Using the results of monitoring the parameters of the geomagnetic field as an example, the effectiveness of the proposed approach is shown.

Keywords: spatial data; data processing; theoretical information analysis; geostatistical analysis; recovery of time series

Acknowledgments: The research was carried out with the financial support of the Russian Foundation for Basic Research № 20-07-00011.

For citation: Vorobeva, G.R., Vorobev, A.V. (2022) An approach to the recovery of spatial data time series based on information-correlation and geostatistical analysis. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naja tehnika i informatika – Tomsk State University Journal of Control and Computer Science*. 61. pp. 37–46. doi: 10.17223/19988605/61/4

Одним из магистральных направлений развития технологий обработки пространственных данных является применение методов геостатистического анализа для оценки их пространственно-временной анизотропии в различных прикладных областях науки и техники. Вместе с тем, к примеру, детерминистические модели и методы пространственной интерполяции, крикинга и кокрикинга, моделирования пространственной неопределенности предполагают непрерывность временных рядов данных, синхронно регистрируемых пространственно распределенными источниками данных (геодезическими опорными пунктами) [1. С. 50]. Однако несовершенство используемого для регистрации пространственных данных информационно-измерительного оборудования, сбои в каналах передачи информации, обусловленные причинами различной природы, а также ошибки, вызванные человеческим фактором, приводят к тому, что в настоящее время практически не существует источников данных, гарантирующих их бесперебойную регистрацию.

Ярким примером обозначенной проблемы являются наземные магнитные обсерватории и вариационные станции, которые в режиме реального времени осуществляют регистрацию параметров магнитного поля Земли и его вариаций. При этом проведенные авторами исследования показали, что годовые архивы соответствующих геомагнитных наблюдений содержат от 22,36 до 35,12% пропущенных значений, что крайне негативно сказывается на работе информационных систем, функционирование которых основано на применении указанных геопространственных данных [2. Р. 627].

Представляется целесообразным отметить, что в силу ряда причин (преимущественно экономического характера) решить проблему совершенствованием технического обеспечения измерительного (и регистрационного) процесса не представляется возможным, так же как и нивелировать негативное воздействие человеческого фактора. В этой связи необходима и актуальна разработка подхода, позволяющего восстановить временные ряды пространственных данных с точностью, значение которой должно быть максимально приближено к значению, регламентированному действующими в соответствующей прикладной области деятельности стандартами, спецификациями и рекомендациями. Так, к примеру, в случае геомагнитных данных в качестве такого документа может быть использована спецификация IAGA (International Association of Geomagnetism and Aeronomy – Международная ассоциация геомагнетизма и аэронавтики) [3. Р. 9], которая определяет допустимую ошибку при регистрации параметров геомагнитного поля и его вариаций не более 1 нТл.

В настоящее время широко известны методы восстановления временных рядов данных, обеспечивающие с некоторой точностью импутацию недостающих единичных значений и целых фрагментов. К примеру, для единичных значений часто применяется сглаживание временного ряда методом скользящей средней, предполагающее замену отсутствующего фрагмента данных усредненным значением соседних по отношению к нему элементов [4. Р. 137]. При тех же условиях может быть применена линейная интерполяция, основанная на подборе заданного уравнением прямой полинома первой степени на базе известных значений уровня временного ряда геомагнитных данных [4. Р. 137]. Для более длинных фрагментов известны решения, основанные на модели авторегрессии (AR) первого порядка, которая обеспечивает возможность прогнозирования отсутствующих фрагментов временного ряда на основании предшествующих ему значений [4. Р. 141]. Аналогично для длительных фрагментов может быть применена интегрированная модель авторегрессии – скользящего среднего (ARIMA), которая обеспечивает лучшую по сравнению с AR метрику качества прогнозирования пропущенного фрагмента за счет гибкой параметризации обработки данных [4. Р. 142].

Вместе с тем анализ эффективности перечисленных методов восстановления временных рядов геомагнитных данных (был проанализирован 10-минутный фрагмент) показал невозможность их применения в соответствующих информационных системах ввиду возникновения ошибки, превышающей допустимое спецификациями значение в 1 нТл [5. Р. 1062]. К примеру, метод скользящей

средней обеспечивает среднеквадратическую ошибку восстановления данных 1,3 нТл. Метод линейной интерполяции хорошо показал себя при восстановлении единичных пропусков (среднеквадратическая ошибка составила 0,03 нТл), но при увеличении восстанавливаемого фрагмента до 10 значений величина среднеквадратической ошибки превысила 1,43 нТл. Модели AR и ARIMA также показали хорошие результаты на небольших фрагментах, обеспечив восстановление 5-минутного фрагмента с приемлемой точностью в 0,7 нТл, но при увеличении импутируемого фрагмента ошибка возросла до 1,8 нТл.

При этом важно отметить, что известные методы восстановления временных рядов пространственных данных никак не учитывают характерные для процессов особенности статистических и энтропийных характеристик, без понимания которых затруднительна оценка соответствующих параметров в контексте внешних для них факторов. В частности, при восстановлении геомагнитных данных целесообразно учитывать недетерминированную зависимость характера изменения параметров магнитного поля от состояния магнитосферы в соответствующий момент времени. Кроме того, важной характеристикой является пространственная анизотропия параметров геомагнитного поля и его вариаций, что также является одним из критериев, которые необходимо учитывать в процессе восстановления временных рядов.

1. Кластеризация источников пространственных данных на основе их корреляционных и энтропийных характеристик

В основе предлагаемого подхода к восстановлению временных рядов пространственных данных лежит оценка взаимного расположения их источников. В этой связи предлагается объединять в группы (кластеры) источники данных, которые по своим параметрам пространственной корреляции, а также взаимной информации могут считаться условно близкими.

На начальном этапе представим множество источников данных посредством выражений в соответствии с системой аксиом Цермело–Франкеля [6]:

$$A = \langle a_1, a_2, \dots, a_n \rangle, \quad (1)$$

где в интенциональном описании a_i – группа, объединяющая источники пространственных данных (для заданной предметной области, например, геомагнитные данные) по всему миру.

Группы источников данных на начальном этапе делятся на глобальную, региональные и локальные в зависимости от конкретной прикладной задачи, для них применяются соответствующие данные:

$$A = \langle A^g, A^r, A^l \rangle, A^g = \langle a_1^g, \dots, a_k^g \rangle, A^r = \langle a_1^r, \dots, a_m^r \rangle, A^l = \langle a_1^l, \dots, a_p^l \rangle, \quad (2)$$

где A^g, A^r, A^l – глобальная, региональная и локальная группы источников данных соответственно.

Группы источников данных более низкого уровня могут представлять собой подмножества одной или нескольких групп более высокого уровня, что в терминах дополнений множеств может быть представлено как

$$A = A^l \cup (A^l/A^r) \cup (A^r/A^g), \quad (3)$$

где дополнение (A^l/A^r) задано элементами множества A^r , не входящими во множество A^l , а дополнение (A^r/A^g) задано элементами множества A^g , не входящими во множество A^r . При этом $A^g \supset A^r, A^r \supset A^l$.

На последующем этапе источники данных, относящиеся к одной и той же пространственной группе (кластеру), подвергаются дополнительному агрегированию на основании соответствующих теоретико-информационных характеристик. Для этого анализируется пространственное распределение информационных характеристик соответствующих пространственных данных для оценки степени связности временных рядов данных с различной географической привязкой. Ключевой характеристикой здесь является информационная энтропия (энтропия Шеннона) $H(X)$, определяемая согласно выражению [7. Р. 104]:

$$H(X) = \sum_{i=1}^n p_i \log_2 p_i, \quad (4)$$

где p_i – вероятность наблюдения значения x_i , X – временной ряд пространственных данных, представленный совокупностью значений x_1, \dots, x_n .

Согласно выражению (4) энтропия является усредненной характеристикой сообщения – математическим ожиданием распределения случайной величины i_1, i_2, \dots, i_n , где i_k – итерация наблюдений соответствующего параметра (содержательная / атрибутивная составляющая анализируемых пространственных данных). Являясь в определенной степени мерой рассеяния, информационная энтропия подобна дисперсии, но не зависит от типа распределения и характеризуется универсальностью и аддитивностью. Кроме того, энтропия, в отличие, к примеру, от корреляции, характеризует любую, в том числе нелинейную, связь переменных.

Вообще говоря, информационная энтропия Шеннона, по определению [7. Р. 104], количественно характеризует произвольное распределение какого-либо параметра процесса. Если во внешней среде или в самой исследуемой системе происходят какие-либо изменения, то это приводит к изменению распределения ее параметров. В этом смысле информационная энтропия Шеннона может рассматриваться как функция состояния системы, поскольку количественно описывает меру неопределенности значений параметров, характеризующих систему [7. Р. 104].

Анализ взаимных энтропийных характеристик ставит своей целью выявление статистически значимых зависимостей произвольного вида между приращениями пар временных рядов. Чем более отдалены друг от друга в пространстве временные ряды, тем менее выражены их взаимные теоретико-информационные характеристики.

При этом основной количественной мерой взаимных энтропийных характеристик является взаимная информация (Mutual Information), которая, согласно [8. Р. 31], является мерой взаимной зависимости между двумя переменными. Она квантифицирует количество информации (по мере Хартли [8. Р. 31]), полученное от одной величины при наблюдении другой:

$$I(X, Y) = \sum_{i=1}^n \sum_{j=1}^n p(x_i, y_j) \log_2 \frac{p(x_i, y_j)}{p(x_i)p(y_j)}, \quad (6)$$

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 значений x_i в точке X и значения y_j в точке Y

Вероятностные составляющие информационной энтропии могут быть интерпретированы таким образом, что чем меньше вероятность появления какого-либо значения исследуемого атрибутивного параметра, тем большую неопределенность снимает регистрирующее его появление сообщения и тем больше информации оно несет. Иными словами, по формуле Бриллюэна $H + I = 1$.

При этом, согласно [7. Р. 104; 8. Р. 31], для статистически независимых выборок условные энтропии $H(X/Y)$ и $H(Y/X)$ не пересекаются. В противном случае площадь их пересечения характеризует меру взаимной информации между ними. Чем больше величина взаимной информации, тем теснее связь и меньше величина условной энтропии $H(X/Y)$.

По результатам анализа величины взаимной информации внутри пространственного кластера могут выделяться дополнительные группы, представленные источниками данных с наиболее сильной связью. В итоге осуществляется декомпозиция исходного кластера на дочерние пространственные кластеры с соответствующими источниками данных.

На последующем этапе продолжается анализ выделенных пространственных кластеров источников данных на предмет оценки связывающего их корреляционного параметра – коэффициента информационной корреляции, который определяется согласно выражению

$$R(X, Y) = \sqrt{1 - e^{-2I(X, Y)}}, \quad (7)$$

где $I(X, Y)$ – мера взаимной информации для выборок данных, синхронно зарегистрированных источниками X и Y .

Свойства взаимной информации для источников данных X и Y полностью определяют свойства информационного коэффициента корреляции, показывая связь между соответствующим источниками данных. Так, к примеру, для независимых источников данных (независимых как друг от друга, так и от некоторых внешних факторов) $R(X, Y) = 0$.

Здесь представляется целесообразным отметить, что значения и информационной энтропии (на основании которой рассчитывается взаимная информация), и информационной корреляции для физически специфических величин являются основанием для формулирования суждений о пространственной однородности физических процессов, данные о которых синхронно зарегистрированы источниками X и Y . Кроме того, перечисленные параметры в совокупности могут быть использованы в качестве индикатора, свидетельствующего о сопоставимом изменении величин X и Y в результате одинакового воздействия экзогенных факторов.

Полученные значения коэффициента информационной корреляции используются для последующей декомпозиции каждого пространственного кластера. В каждую группу дополнительно выделяются источники данных с наиболее сильной связью с точки зрения попарной информационной корреляции.

Предполагается, что источники данных, представленные в одном пространственном кластере, могут быть использованы для частичного взаимного информационного резервирования, а также применены в составе известных статистических моделей для восстановления недостающих значений соответствующих временных рядов. Однако известно, что проблема пространственной автокорреляции, указывающая на пространственную зависимость источников данных, может внести существенные искажения в процедуру применения стандартных статистических методов, предполагающих зависимость между наблюдениями [9. С. 330]. Так, к примеру, не компенсирующий пространственную зависимость регрессионный анализ может сопровождаться неустойчивыми оценками параметров, что способно привести к недостоверным результатам проводимых тестов значимости [9. С. 339].

Для исключения пространственной автокорреляции необходимо ввести расчет дополнительного геостатистического параметра, который показывает, носят ли случайный характер внешние процессы / факторы, определяющие данные, регистрируемые пространственно распределенными источниками. С этой целью выдвигается нулевая гипотеза, согласно которой пространственные процессы, обуславливающие анализируемые пространственные данные, носят случайный характер. Для подтверждения / опровержения данной нулевой гипотезы для каждой пары пространственных объектов (источников данных) рассчитывается значение глобального индекса Морана.

В этой связи при обработке геопространственных данных необходимо учитывать параметры пространственной неоднородности и пространственной зависимости источников данных [10]:

$$I_G = \frac{\sum_i \sum_j w_{ij} (x_i - \mu)(x_j - \mu)}{\sum_i (x_i - \mu)^2}, \quad (8)$$

где I_G – индекс Морана, x_i, x_j – значения параметра x в пространственных точках i и j , μ – среднее значение параметра x , w – экспертный весовой коэффициент.

В зависимости от соотношения значений I_G и $I(E)$ (где для n точек $I(E) = -1/(n - 1)$) возможно определить, являются ли значения в соседних пространственных регионах подобными ($I_G > I(E)$), отличающимися ($I_G < I(E)$) или расположенными случайным образом ($I_G = I(E)$), и оценить зависимость описываемого процесса от экзогенных факторов

Значение индекса Морана может быть использовано для уточнения полученных на предшествующих этапах пространственных кластеров. Данные, синхронно регистрируемые территориально распределенными источниками, традиционно измеряются с помощью интервальных иколичественных шкал. Принимая во внимание их геопространственную привязку, теснота связи между ними может быть определена на основании анализа пространственной неоднородности и пространственной зависимости, что, в свою очередь, предполагает наличие пространственной корреляции (положительной или отрицательной) между пространственными наблюдениями, что и показывает индекс Морана.

Целесообразно заметить, что для корректного расчета коэффициента пространственной автокорреляции, а также перечисленных теоретико-информационных характеристик необходимо, чтобы количество значений, синхронно зарегистрированных в исследуемых распределенных источниках данных X и Y , было одинаковым, исследуемые временные ряды от X и Y были распределены нормально и измерены в интервальной шкале или шкале отношений [10]. Установлено, что совокупное

применение теоретико-информационных и геостатистических характеристик источников данных позволяет исключить ложную пространственную автокорреляцию, которая может внести существенные искажения в сформулированные на их основе выводы.

2. Описание предлагаемого подхода

Резюмируя описание приведенных выше методов и моделей теоретико-информационного и геостатистического анализа, представляется целесообразным обобщенное представление предлагаемого подхода к восстановлению временных рядов пространственных данных. Процедуре импутации пропущенных фрагментов временных рядов пространственных данных предшествует составление так называемого доверительного перечня резервных геодезических пунктов. Представленные в составе последнего источники данных являются основой для восстановления временного ряда пространственных данных, относящихся к анализируемому источнику данных. В случае сильной корреляционной связи наиболее близкий по доверительному списку источник данных может использоваться в качестве резервного для анализируемого. Замена пропусков исходного временного ряда нормализованными данными резервного источника данных на основе доверительного списка выполняется посредством сопоставления временных индексов и установления соответствия между ними. Выбранный фрагмент резервного временного ряда копируется под соответствующие временные индексы восстанавливаемого ряда, заменяя в нем обнаруженные пропуски.

Важно отметить, что полученные в ходе применения метода информационного резервирования результаты восстановления данных, как правило, являются смещенными и поэтому должны быть аппроксимированы относительно известных соседних пропущенному фрагменту значений уровней временного ряда. Нормализация данных выполняется посредством метода наименьших квадратов и предполагает вычисление значений коэффициентов линейной зависимости двух массивов [11. Р. 56].

Составление доверительного списка источников данных – многоэтапный процесс последовательной пространственной кластеризации опорных геодезических пунктов [12, 13]. На начальном этапе пространственные кластеры формируются в соответствии со спецификой предметной области или прикладной задачи, для которой используются соответствующие временные ряды пространственных данных. Так, распределенные источники данных могут быть изначально кластеризованы, к примеру, по своей ведомственной принадлежности, используемой информационно-измерительной технике, территориальному распределению и пр. При этом допускается наличие как единственного глобального пространственного кластера, который представлен всеми доступными источниками данных, так и множества локальных и региональных пространственных кластеров, каждый из которых семантически отличается от остальных.

На последующем этапе отдельно рассматривается каждый из выделенных ранее пространственных кластеров. Осуществляется декомпозиция пространственного кластера на вложенные кластеры, каждый из которых, в свою очередь, представлен источниками данных, наиболее близкими по своим теоретико-информационным характеристикам. Для этого попарно для всех источников данных в соответствующем пространственном кластере последовательно рассчитываются значения информационной энтропии Шеннона и взаимной информации. Для источников данных с наилучшим показателями по взаимной информации формируются новые включающие их пространственные кластеры. При этом, как и в случае предшествующих этапов кластеризации, возможны ситуации, когда кластер представлен единственным источником данных, для которого не были обнаружены «близкие» к нему иные геодезические пункты.

Далее для выделенных пространственных кластеров осуществляется попарный анализ коэффициента информационной корреляции, в том числе на основании ранее рассчитанного значения по взаимной информации между теми же источниками данных. Результаты такого анализа также позволяют внутри кластеров выделить источники данных с наиболее сильной корреляционной связью, что, в свою очередь, является основой для формирования новых пространственных кластеров при декомпозиции исходного.

На следующем шаге попарно анализируются параметры пространственной неоднородности и пространственной зависимости источников данных внутри кластеров, сформированных на предшествующем этапе, для чего рассчитывается значение индекса Морана по соответствующим временным рядам пространственных данных. При этом выявляются те опорные геодезические пункты из исходного множества, вариации в данных которых определяются одними и теми же внешними факторами. В результате выполняется корректировка полученных ранее пространственных кластеров. Непосредственно при восстановлении временных рядов определяется принадлежность анализируемого источника данных соответствующему пространственному кластеру, на основании чего как непосредственно идентифицируется доверительный список опорных геодезических пунктов с учетом обозначенных энтропийных и геостатистических характеристик, так и определяется тот источник данных, который может быть рассмотрен в качестве резервного.

3. Анализ эффективности предложенного подхода

Оценка эффективности предложенного подхода восстановления временных рядов пространственных данных была выполнена на примере годовых архивов геомагнитных данных, регистрируемых пространственно распределенными магнитными обсерваториями и вариационными станциями. Рассматриваются фрагменты одномерного временного ряда пространственных данных, которые характеризуют минутные наблюдения вариаций горизонтальной составляющей вектора геомагнитного поля (ВН, нТл). В качестве анализируемого источника данных была выбрана магнитная обсерватория DOU (Dourbes) глобальной сети магнитных обсерваторий INTERMAGNET (архив свободно распространяется и доступен по адресу <https://intermagnet.org>) [14. P. 2; 15. P. 110].

Множество обсерваторий рассматриваемой сети было подвергнуто процедуры многоэтапной пространственной кластеризации. Для этого на начальном этапе определяются взаимные энтропийные характеристики итерационного попарного сравнения основного источника данных с каждым доступным в глобальной группе источником данных, на основании которого анализируются информационная энтропия Шеннона, условная энтропия и основной показатель – взаимная информация. В результате для глобального множества источников данных, сформированного на предыдущем шаге, формируется множество, которое, с одной стороны, включает в себя анализируемый опорный геодезический пункт DOU, а с другой – содержит источники данных, показавшие максимальные или близкие к ним значения энтропийных характеристик (преимущественно рассматриваются значения взаимной информации).

Например, для геомагнитных данных в доверительный список попадают магнитные обсерватории и вариационные станции с близкими к единице показателями взаимной информации и условной энтропии, со значениями информационной энтропии, отличающимися не более чем на 0,01, а также относящиеся к одной пространственной группе и описываемые, как правило, одним статистическим законом распределения.

Примечательно, что расчет информационной энтропии Шеннона для каждого набора геомагнитных данных показал, что наименьшая неопределенность наблюдается в районе средних широт ($46,9^\circ$ N). При этом энтропия возрастает по направлению к низким широтам рис. 1). Аналогичные исследования, проведенные для других значений широт, показали повторяемость полученных зависимостей и подтвердили результаты анализа. Физическое объяснение полученных результатов теоретико-информационного анализа заключается в том, что значения параметров геомагнитного поля и его вариаций на средних широтах зависят от наименьшего количества факторов. При этом по мере приближения к экватору и полюсам число факторов, влияющих на результаты наблюдений, возрастает, что приводит к росту неопределенности данных о состоянии геомагнитного поля и его вариаций, но повышает информативность каждой итерации геомагнитного мониторинга. При этом установлено, что в пределах одного широтного диапазона энтропийные характеристики магнитных обсерваторий почти неизменны. Физический смысл полученных результатов можно интерпретировать таким образом, что на значения параметров геомагнитного поля и его вариаций в пределах одного широтного

диапазона оказывают влияние одни и те же внешние факторы (или их совокупность), что приводит к относительной неизменности информационной энтропии с изменением географической долготы.

Кроме того, при анализе значения условной энтропии было установлено его увеличение по мере пространственного удаления анализируемого источника данных от основного (в данном случае DOU). Аналогичная тенденция была обнаружена и при анализе значений взаимной информации для пар источников геомагнитных данных.

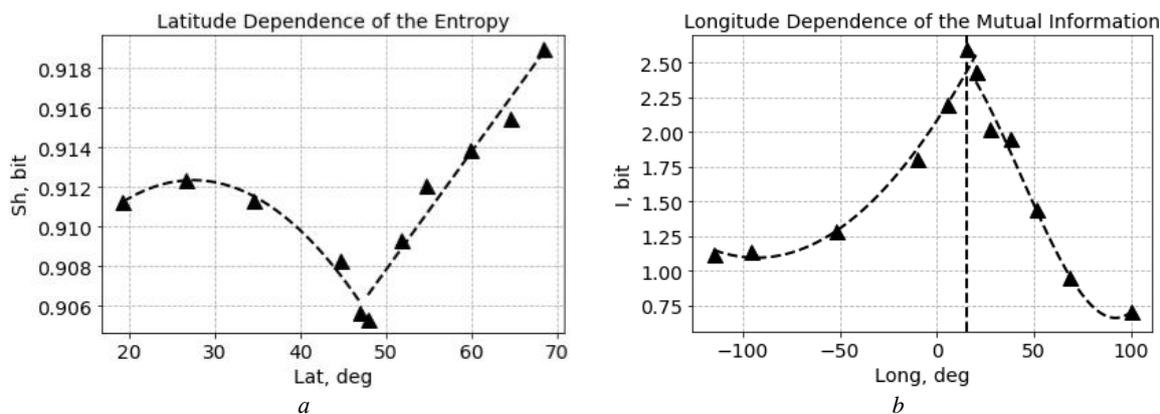


Рис. 1. Широтная и долготная зависимости энтропийных характеристик геомагнитных данных (a – информационная энтропия Шеннона, b – взаимная информация)

Fig. 1. Latitudinal and longitude dependences of entropy characteristics of geomagnetic data (a – Shannon information entropy, b – mutual information)

На последующем этапе для сформированного пространственного кластера с обсерваторией DOU выполняется анализ информационной корреляции. При этом осуществляется итерационная обработка данных посредством попарного сравнения входящих в кластер опорных геодезических пунктов и того источника данных, для которого необходимо выполнить восстановление временного ряда (DOU). В результате получено подмножество в составе исходного пространственного кластера, для которого были достигнуты наилучшие показатели информационной корреляции. Анализ пространственной однородности соответствующих источников данных показал, что сформированный кластер является корректным.

Представленные в окончательно сформированном кластере источники данных образуют доверительный список для DOU. Данные в списке ранжируются по своим энтропийным и корреляционным характеристикам. Проведенные вычислительные эксперименты показали, что данные, зарегистрированные станцией DOUrbes, наилучшим образом коррелируют с наблюдениями обсерватории MAV (за исследуемый год). Это позволяет сделать вывод, что для реконструкции искомого временного ряда геомагнитных данных обсерваторию MAV можно назначить резервной станцией (при условии наличия данных за соответствующий временной интервал).

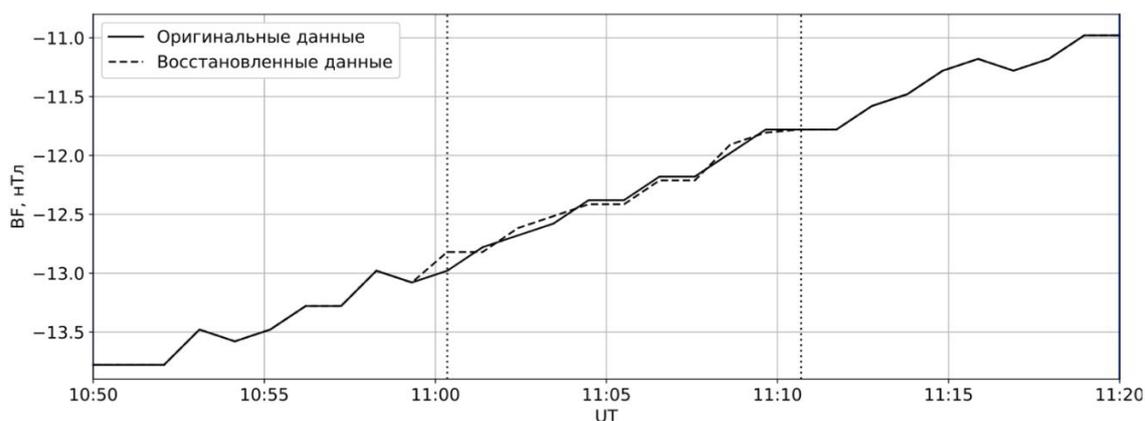


Рис. 2. Результаты восстановления 10-минутного фрагмента временного ряда геомагнитных данных
Fig. 2. Results of recovery of a 10-minute fragment of the time series of geomagnetic data

Дальнейший анализ показал, что применение метода информационного резервирования для восстановления 10-минутного пропуска в экспериментальном временном ряду характеризуется среднеквадратической ошибкой со значением $\sim 0,036$ нТл, что в целом меньше регламентированной спецификацией / стандартом IAGA-2002 погрешности для измерений параметров геомагнитного поля и его вариаций (~ 1 нТл) (рис. 2).

Заключение

В настоящей работе предлагается повысить эффективность используемых источниками данных технических систем регистрации с помощью информационного резервирования, которое предполагает использование дополнительных данных, выступающих в качестве вспомогательных. Выбор в пользу метода информационного резервирования обусловлен в наибольшей степени тем, что в большинстве известных случаев, когда отказ в работе технической системы приводит к потере или искажению информации (обрабатываемой локально или передаваемой по каналам связи), повышение надежности достигается преимущественно посредством информационного резервирования. В настоящее время метод информационного резервирования активно используется в системах управления и связи, информационных, измерительных и вычислительных системах сбора и обработки информации и позволяет повысить надежность технической системы регистрации данных при недостаточной надежности носителей информации, невозможности возобновления информации с помощью первичных источников и т.д.

На основании перечисленных характеристик и особенностей метода информационного резервирования как способа повышения надежности технических систем, обеспечивающих данными системы поддержки принятия решений, в работе предложен метод восстановления временных рядов. Метод предполагает определение наиболее вероятных значений посредством формирования доверительного списка источников данных на основании оценки пространственной гетерогенности и зависимости синхронно регистрируемых данных, а также сравнительной оценки соответствующих фрагментов временных рядов геомагнитных данных, зарегистрированных в момент времени, предшествующий восстанавливаемому.

В ходе проведенных вычислительных экспериментов для геомагнитных данных было установлено, что применение предложенного подхода позволяет восстанавливать временные ряды с точностью $0,01-0,5$ нТл, что не превышает допустимой величины ошибки геомагнитных измерений.

Список источников

1. Демьянов В.В., Савельева Е.А. Геоestatистика: теория и практика. М. : Наука, 2010. 327 с.
2. Vorobev A.V., Vorobeva G.R. Approach to Assessment of the Relative Informational Efficiency of Intermagnet Magnetic Observatories // *Geomagnetism and Aeronomy*. 2018. V. 58 (5). P. 625–628.
3. St-Louis B.J. Intermagnet Technical Reference Manual. Version 4. 1999. 156 с.
4. Mantas J. Statistical methods // *Studies in Health Technology and Informatics*. 2002. V. 65. P. 136–147.
5. Vorobeva G.R. Approach to the recovery of geomagnetic data by comparing daily fragments of a time series with equal geomagnetic activity // *Computer Optics*. 2019. № 43. P. 1053–1063.
6. Swan A. A class of higher inductive types in Zermelo-Fraenkel set theory // *Mathematical Logic Quarterly*. 2022. № 1. P. 118–127. doi: 10.1002/malq.202100040
7. Ricci L., Perinelli A., Castelluzzo M. Estimating the variance of Shannon entropy // *Physical Review*. 2021. V. 104 (2-1). Art. 024220. doi: 10.1103/PhysRevE.104.024220.
8. Carrara N., Ernst J. On the Estimation of Mutual Information // *Proceedings*. 2020. № 33. P. 31. doi: 10.3390/proceedings2019033031.
9. Гудчайлд М.Ф. Пространственный аналитическая перспектива на географических информационных системах // *Международный журнал географических информационных систем*. 1987. № 1 (4). С. 327–344.
10. Pinheiro P. [et al.] Lacunarity exponent and Moran index: A complementary methodology to analyze AFM images and its application to chitosan films // *Physica A: Statistical Mechanics and its Applications*. 2021. V. 581. Art. 126192.
11. Liu Q., Li Ch., Wei Y. Condition numbers of multidimensional mixed least squares-total least squares problems // *Applied Numerical Mathematics*. 2022. № 178. doi: 10.1016/j.apnum.2022.03.014
12. Sugawara Sh., Murakami D. Spatially clustered regression // *Spatial Statistics*. 2021. № 44. Art. 100525.
13. Ilmi Nasution B., Kurniawan R., Caraka R. Nature-Inspired Spatial Clustering // *The R Journal*. 2021. March. P. 1–33. doi: 10.1109/CITSM.2014.7042178
14. Lin J.-W. Real-time Magnetic Observatory Network: A Review // *European Journal of Environment and Earth Sciences*. 2021. № 2. P. 1–2.

15. Kim J.-H., Chang H.-Y. Geomagnetic field variations observed by INTERMAGNET during 4 total solar eclipses // Journal of Atmospheric and Solar-Terrestrial Physics. 2018. V. 172. P. 107–116.

References

1. Demiyarov, V.V. & Savelieva, E.A. (2010) *Geostatistika: teoriya i praktika* [Geostatistics: Theory and Practice]. Moscow: Nauka.
2. Vorobev, A.V. & Vorobeva, G.R. (2018) Approach to Assessment of the Relative Informational Efficiency of Intermagnet Magnetic Observatories. *Geomagnetism and Aeronomy*. 58(5). pp. 625–628.
3. St-Louis, B.J. (1999). *Intermaget Technical Reference Manual*. Version 4.
4. Mantas, J. (2002) Statistical methods. *Studies in Health Technology and Informatics*. 65. pp. 136–147. DOI: 10.3233/978-1-60750-909-7-136.
5. Vorobeva, G.R. (2019) Approach to the recovery of geomagnetic data by comparing daily fragments of a time series with equal geomagnetic activity. *Computer Optics*. 43. pp. 1053–1063.
6. Swan, A. (2022) A class of higher inductive types in Zermelo-Fraenkel set theory. *Mathematical Logic Quarterly*. 1. pp. 118–127. doi:10.1002/malq.202100040.
7. Ricci, L., Perinelli, A. & Castelluzzo, M. (2021) Estimating the variance of Shannon entropy. *Physical Review*. 104 (2-1). Art. 024220. DOI: 10.1103/PhysRevE.104.024220
8. Carrara, N. & Ernst, J. (2020) On the Estimation of Mutual Information. *Proceedings*. 33. p. 31. DOI: 10.3390/proceedings2019033031.
9. Gudchild, M.F. (1987) Prostranstvennyy analiticheskaya perspektiva na geograficheskikh informatsionnykh sistemakh [Spatial Analytical Perspective on Geographic Information Systems]. *Mezhdunarodnyy zhurnal geograficheskikh informatsionnykh sistem International – Journal of Geographic Information Systems*. 1(4). pp. 327–344.
10. Pinheiro, P. et al. (2021) Lacunarity exponent and Moran index: A complementary methodology to analyze AFM images and its application to chitosan films. *Physica A: Statistical Mechanics and its Applications*. 581. Art. 126192. DOI: 10.1016/j.physa.2021.126192
11. Liu, Q., Li, Ch. & Wei, Y. (2022) Condition numbers of multidimensional mixed least squares-total least squares problems. *Applied Numerical Mathematics*. 178. DOI: 10.1016/j.apnum.2022.03.014.
12. Sugawara, Sh. & Murakami, D. (2021) Spatially clustered regression. *Spatial Statistics*. 44. Art. 100525. DOI: 10.1016/j.spasta.2021.100525
13. Ilmi Nasution, B., Kurniawan, R. & Caraka, R. (2021) Nature-Inspired Spatial Clustering. *The R Journal*. March. pp. 1–33. DOI: 10.1109/CITSM.2014.7042178.
14. Lin, J.-W. (2021) Real-time Magnetic Observatory Network: A Review. *European Journal of Environment and Earth Sciences*. 2. pp. 1–2. DOI: 10.24018/ejgeo.2021.2.5.177.
15. Kim, J.-H. & Chang, H.-Y. (2018) Geomagnetic field variations observed by INTERMAGNET during 4 total solar eclipses. *Journal of Atmospheric and Solar-Terrestrial Physics*. 172. pp. 107–116. DOI: 10.1016/j.jastp.2018.03.023.

Информация об авторах:

Воробьева Гульнара Равиелевна – доцент, доктор технических наук, профессор кафедры вычислительной математики и кибернетики Уфимского государственного авиационного технического университета. E-mail: gulnara.vorobeva@gmail.com

Воробьев Андрей Владимирович – доцент, кандидат технических наук, доцент кафедры геоинформационных систем Уфимского государственного авиационного технического университета. E-mail: cpu16bit@gmail.com

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

Information about the authors:

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

Vorobev Andrei V. (Associate Professor, Candidate of Technical Sciences, Associate Professor, Ufa State Aviation Technical University, Ufa, Russian Federation). E-mail: cpu16bit@gmail.com

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

Поступила в редакцию 11.05.2022; принята к публикации 29.11.2022

Received 11.05.2022; accepted for publication 29.11.2022

Научная статья

УДК 519.2

doi: 10.17223/19988605/61/5

Оценивание методом моментов параметра равномерного распределения длительности непродлевающегося случайного мертвого времени в рекуррентном полусинхронном потоке событий в общем и особом случаях

Александр Михайлович Горцев¹, Анна Васильевна Веткина²

^{1, 2} *Томский государственный университет, Томск, Россия*

¹ *a-gortsev@mail.ru*

² *anyavetkina@gmail.com*

Аннотация. Исследуется полусинхронный поток событий, относящийся к классу дважды стохастических потоков событий. Функционирование потока рассматривается в условиях непродлевающегося случайного мертвого времени, распределенного по равномерному закону на отрезке $[0, T^*]$. Рассматривается частный случай функционирования полусинхронного потока событий, когда такой поток является рекуррентным, в общем и особом случаях соотношения его параметров. Производится оценивание параметра T^* равномерного распределения длительности непродлевающегося случайного мертвого времени методом моментов. Приводятся результаты статистических экспериментов, показывающие адекватность полученных оценок.

Ключевые слова: рекуррентный полусинхронный поток событий; непродлевающееся случайное мертвое время; оценка параметра; метод моментов

Для цитирования: Горцев А.М., Веткина А.В. Оценивание методом моментов параметра равномерного распределения длительности непродлевающегося случайного мертвого времени в рекуррентном полусинхронном потоке событий в общем и особом случаях // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2022. № 61. С. 47–60. doi: 10.17223/19988605/61/5

Original article

doi: 10.17223/19988605/61/5

Estimation by method of moments of the parameter of the uniform distribution of the duration of unextendable random dead time in recurrent semi-synchronous events flow in general and special case

Alexander M. Gortsev¹, Anna V. Vetkina²

^{1, 2} *Tomsk State University, Tomsk, Russian Federation*

¹ *a-gortsev@mail.ru*

² *anyavetkina@gmail.com*

Abstract. The paper describes recurrent semi-synchronous events flow with $p = 1$ that is a common mathematical model of information flows of messages operating in telecommunication and information-computing networks, and that belongs to the class of doubly stochastic event flows. A general and special cases are considered. Operation of the flow is considered with random unextendable dead time that has uniform distribution on the interval $[0, T^*]$. Parameter T^* of the dead time is estimated using the method of moments. Results of statistical experiments are presented.

Mathematical expectation of the duration of τ – the interval between adjacent events of the observed flow – is given for the general case by the formula

$$M(\tau | T^*) = \frac{T^*}{2} + \frac{1}{\lambda_2 + \alpha_2} + \frac{\alpha_2}{\lambda_1(\lambda_1 + \alpha_2)} + \frac{\alpha_2(\lambda_1 - \lambda_2)}{\lambda_1(\lambda_1 + \alpha_2)^2(\lambda_2 + \alpha_2)} \frac{1}{T^*} \left(1 - e^{-(\lambda_1 + \alpha_2)T^*}\right), T^* > 0;$$

and for special case:

$$M(\tau|T^*) = \frac{T^*}{2} + \frac{1}{\lambda_1} + \frac{\alpha_2}{\lambda_1(\lambda_1 + \alpha_2)} + \left[\frac{\alpha_2}{\lambda_1(\lambda_1 + \alpha_2)} \right]^2 \frac{1}{T^*} (1 - e^{-(\lambda_1 + \alpha_2)T^*}), \quad T^* \geq 0.$$

The estimate \hat{T}^* is found numerically from the equation of moments $M(\tau|T^*) = C$, $C = (1/n) \sum_{k=1}^n (t_{k+1} - t_k)$, where t_1, t_2, \dots, t_{n+1} are moments of occurrence of events in the observed flow. Value C is found using simulation modeling of the observed flow. Analysis of the numerical results shows that in the sense of the introduced criterion $\hat{V}(\hat{T}^*)$ (sample variance of the estimate \hat{T}^*), an increase in the parameter T^* has a negative effect on the quality of estimates \hat{T}^* , that is quite natural: increasing the parameter T^* leads to an increase in the number of lost events of the initial flow.

According to the results of the research, there are the following conclusions: 1) it is shown analytically that the equation of moments has a unique solution; 2) the results of simulation modelling show that the quality of the estimates in the sense of the introduced criterion (sample variance of the estimate \hat{T}^*) is quite satisfactory, and the bias of estimates \hat{T}^* relative to the true value of parameter T^* doesn't exceed hundredths of values.

Keywords: recurrent semi-synchronous events flow; unextendable random dead time; uniform distribution; estimation of the parameter; method of moments

For citation: Gortsev, A.M., Vetkina, A.V. (2022) Estimation by method of moments of the parameter of the uniform distribution of the duration of unextendable random dead time in recurrent semi-synchronous events flow in general and special case. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naja tehnika i informatika – Tomsk State University Journal of Control and Computer Science*. 61. pp. 47–60. doi: 10.17223/19988605/61/5

В современных телекоммуникационных системах [1, 2] входящие потоки событий наиболее адекватно описывают дважды стохастические потоки событий – потоки, у которых случайными являются моменты наступления событий и интенсивность. Поэтому исследование дважды стохастических потоков является актуальной задачей. В общем случае дважды стохастические потоки событий являются коррелированными потоками [2].

Дважды стохастические потоки делятся на два класса: первый класс составляют потоки, сопровождающий процесс (интенсивность) которых есть непрерывный случайный процесс [3, 4]; второй – потоки, сопровождающий процесс (интенсивность) которых есть кусочно-постоянный случайный процесс с конечным (произвольным) числом состояний [5, 6].

В зависимости от того, каким образом происходит переход интенсивности из состояния в состояние, выделяется три типа дважды стохастических потоков: 1) синхронные потоки (потоки, у которых состояние сопровождающего процесса меняется в случайные моменты времени, являющиеся моментами наступления событий) [7–10]; 2) асинхронные потоки (потоки, у которых переход из состояния в состояние сопровождающего процесса происходит в случайные моменты времени и не зависит от моментов наступления событий) [11–14]; 3) полусинхронные потоки (потоки, у которых одна часть состояний сопровождающего процесса меняется в моменты наступления событий потока, другая часть состояний сопровождающего процесса меняется в произвольные моменты времени, не связанные с моментами наступления событий потока) [15–18].

На практике часто приходится иметь дело с потоками, у которых не все события доступны наблюдению. Как правило, причиной ненаблюдаемости служит мертвое время регистрирующих приборов [19], порождаемое зарегистрированным событием, так что другие события, наступившие в этот период, теряются. Регистрирующие приборы при этом делятся на два вида: с непродлевающимся мертвым временем и продлевающимся. Кроме того, длительность мертвого времени может быть как детерминированной величиной, одинаковой для всех событий, так и случайной с тем или иным законом распределения. В реальных регистрирующих устройствах величина и характер мертвого времени зависят от многих факторов, причем регистрирующие приборы обладают значением длительности мертвого времени, ограниченным сверху некоторой величиной. Переходя к случайному мертвому времени [19], вполне естественно рассматривать его распределение как равномерное на некотором отрезке.

В данной работе исследуется полусинхронный дважды стохастический поток событий с интенсивностью, являющейся кусочно-постоянным случайным процессом с двумя состояниями. На параметр потока накладывается условие, такое что исходный коррелированный поток вырождается в рекуррентный. Рассматривается два случая функционирования такого потока: общий и особый. Производится оценивание параметра длительности случайного мертвого времени, распределенного по равномерному закону. Для этого выводится аналитическая формула математического ожидания длительности интервала между соседними событиями наблюдаемого потока и находится оценка параметра равномерного распределения длительности непродлевающегося случайного мертвого времени с использованием уравнения моментов. С помощью построенной имитационной модели наблюдаемого потока реализуются статистические эксперименты для получения численных результатов оценивания.

1. Математическая модель наблюдаемого потока. Постановка задачи

Рассматривается стационарный режим функционирования полусинхронного дважды стохастического потока событий, сопровождающий процесс (интенсивность) которого есть кусочно-постоянный стационарный случайный процесс $\lambda(t)$ с двумя состояниями S_1 и S_2 . Будем говорить, что имеет место первое состояние процесса (потока) S_1 , если $\lambda(t) = \lambda_1$, и, наоборот, имеет место второе состояние процесса (потока) S_2 , если $\lambda(t) = \lambda_2$ ($\lambda_1 > \lambda_2 \geq 0$). Если имеет место первое состояние процесса S_1 , то в течение временного интервала, когда $\lambda(t) = \lambda_1$, поступает пуассоновский поток событий с интенсивностью λ_1 . Если имеет место второе состояние процесса S_2 , то в течение временного интервала, когда $\lambda(t) = \lambda_2$, поступает пуассоновский поток событий с интенсивностью λ_2 . Переход из состояния S_1 процесса $\lambda(t)$ в состояние S_2 возможен только в момент наступления события (свойство синхронности потока), при этом этот переход осуществляется с вероятностью p (с вероятностью $1 - p$ процесс $\lambda(t)$ остается в состоянии S_1). Переход из состояния S_2 процесса $\lambda(t)$ в состояние S_1 может осуществляться в произвольный момент времени, не связанный с моментом наступления события (свойство асинхронности потока). При этом длительность пребывания процесса $\lambda(t)$ во втором состоянии есть случайная величина, распределенная по экспоненциальному закону $F(t) = 1 - e^{-\alpha_2 t}$, $t \geq 0$, где α_2 – интенсивность смены состояния S_2 на S_1 . Так как переход из второго состояния в первое не привязан к моменту наступления события во втором состоянии, то поток называется полусинхронным дважды стохастическим потоком событий. В сделанных предположениях $\lambda(t)$ – скрытый марковский процесс ($\lambda(t)$ – принципиально ненаблюдаемый процесс; наблюдаемыми являются только моменты наступления событий потока).

После каждого зарегистрированного события в момент времени t_k наступает период мертвого времени случайной длительности, который порождается этим событием, так что другие события исходного потока, наступившие в течение этого периода мертвого времени, недоступны наблюдению и не вызывают его продления (непродлевающееся мертвое время). Принимается, что случайная длительность мертвого времени распределена по равномерному закону с плотностью вероятности $p(T) = 1/T^*$, где T – значение длительности мертвого времени, $0 \leq T \leq T^*$.

Исследуется частный случай функционирования полусинхронного потока событий, когда $p = 1$, т.е. такой поток, который при каждом наступлении события в первом состоянии мгновенно переходит во второе. При выполнении данного условия исходный поток, действующий в условиях детерминированного мертвого времени, становится рекуррентным потоком:

$$p(\tau_1, \tau_2 | T) = p(\tau_1 | T) p(\tau_2 | T), \quad \tau_1 \geq T, \quad \tau_2 \geq T,$$

где $p(\tau | T)$ – плотность вероятности значений длительности интервала между соседними событиями в полусинхронном потоке, $p(\tau_1, \tau_2 | T)$ – совместная плотность вероятности [20. С. 254].

Вместе с тем рассматриваются общий и особый случаи соотношения параметров данного потока: когда $\lambda_1 - \lambda_2 - \alpha_2 \neq 0$ и когда $\lambda_1 - \lambda_2 - \alpha_2 = 0$.

Возможный вариант возникающей ситуации приведен на рис. 1, где S_1 и S_2 – состояния случайного процесса $\lambda(t)$; временная ось $(0, t)$ – ось моментов наступления наблюдаемых событий в моменты времени t_1, t_2, \dots ; временная ось $(0, t^{(1)})$ – ось наступления событий в моменты времени $t_1^{(1)}, t_2^{(1)}, \dots$ в первом (S_1) состоянии процесса $\lambda(t)$, на которой также указаны значения длительностей $T_1^{(1)}, T_2^{(1)}, \dots$ мертвых времен, порождаемых наблюдаемыми событиями потока; аналогично для временной оси $(0, t^{(2)})$; белыми кружками обозначены наблюдаемые события, черными – ненаблюдаемые, штриховкой – периоды мертвого времени; траектория процесса $\lambda(t)$ привязана к временной оси $(0, t^{(1)})$.

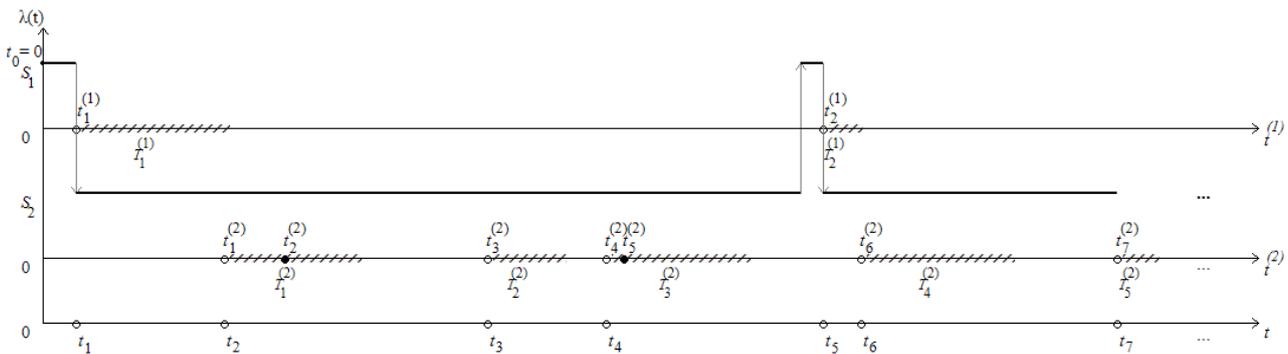


Рис. 1. Формирование наблюдаемого потока событий
Fig. 1. Formation of the observed event flow

Цели данной работы:

1. На основании выборки моментов наступления событий t_1, t_2, \dots, t_n наблюдаемых потоков в общем и особом случаях на временном интервале $(0, T_m)$, где T_m – время наблюдения за потоком ($t_n < T_m$), оценить параметр равномерного распределения длительности непродлевающегося случайного мертвого времени T^* .

2. Исследовать оценку \hat{T}^* для общего и особого случаев рассматриваемого потока. Для этого провести статистические эксперименты, устанавливающие стационарный режим и определяющие свойства полученных оценок.

2. Уравнение моментов для оценивания параметра T^*

Введем $\tau_k = t_{k+1} - t_k$, $k = 1, 2, \dots$, – значение длительности k -го интервала между соседними событиями наблюдаемого потока ($\tau_k \geq 0$). Так как рассматривается стационарный режим функционирования наблюдаемого потока, то плотность вероятности значений длительности k -го интервала есть $p(\tau_k) = p(\tau)$, $\tau \geq 0$, для любого k , т.е. момент наступления события есть $\tau = 0$.

Для оценки неизвестного параметра T^* равномерного распределения длительности случайного непродлевающегося мертвого времени используется метод моментов [21]. Для этого находится теоретический момент – математическое ожидание случайной величины τ (длительность интервала между соседними событиями в наблюдаемом потоке) $M(\tau | T^*)$, после чего ММ-оценка параметра T^* находится численно из уравнения моментов $M(\tau | T^*) = C$, где C – выборочное среднее,

$$C = (1/n) \sum_{k=1}^n \tau_k, \tau_k = t_{k+1} - t_k > 0, \text{ являющееся оценкой математического ожидания } M(\tau | T^*).$$

Из работы [20] имеем, что плотность вероятности значений длительности интервала между соседними событиями в полусинхронном потоке, функционирующем в условиях непродлевающегося мертвого времени фиксированной длительности T , в общем случае имеет вид:

$$p(\tau|T) = \begin{cases} 0, & 0 \leq \tau < T, \\ \gamma(T)\lambda_1 e^{-\lambda_1(\tau-T)} + (1-\gamma(T))(\alpha_2 + \lambda_2) e^{-(\alpha_2 + \lambda_2)(\tau-T)}, & \tau \geq T, \end{cases} \quad (1)$$

где $\gamma(T) = \frac{\alpha_2(\lambda_1 - \lambda_2 - \alpha_2 - \lambda_1 p)}{(\lambda_1 p + \alpha_2)(\lambda_1 - \lambda_2 - \alpha_2)} \left[1 - \frac{p(\lambda_1 - \lambda_2)}{(1-p)\lambda_2 - (\alpha_2 + \lambda_2)e^{(\lambda_1 p + \alpha_2)T}} \right]$, $\lambda_1 - \lambda_2 - \alpha_2 \neq 0$; τ_1, τ_2 – смежные интервалы между наблюдаемыми событиями. Для обобщенного полусинхронного потока событий с непродлевающимся мертвым временем фиксированной длительности T в особом случае $\lambda_1 - \lambda_2 - \alpha_2 = 0$ справедливы формулы

$$p(\tau|T) = \begin{cases} 0, & 0 \leq \tau < T, \\ \left[\lambda_1 - \alpha_2(1-\delta)\pi_2(T)(1-\lambda_1(\tau-T)) \right] e^{-\lambda_1(\tau-T)}, & \tau \geq T, \end{cases} \quad (2)$$

где $\pi_2(T) = \pi_2 - [\pi_2 - \pi_2(0|T)]e^{-(\alpha_2 + p\lambda_1)T}$, $\pi_2 = \frac{p\lambda_1}{\alpha_2 + p\lambda_1}$, $\pi_2(0|T) = \frac{p\lambda_1 - \pi_2[\alpha_2(1-p+\delta p) - \lambda_1(1-p)] [1 - e^{-(\alpha_2 + p\lambda_1)T}]}{\lambda_1 + [\alpha_2(1-p+\delta p) - \lambda_1(1-p)] e^{-(\alpha_2 + p\lambda_1)T}}$,

δ – вероятность иницирования дополнительного события при переходе процесса $\lambda(t)$ из второго состояния в первое ($0 \leq \delta \leq 1$).

Подчеркнем, что внесение непродлевающегося случайного мертвого времени в математическую модель полусинхронного потока событий может только изменить (в меньшую или большую сторону) корреляцию в потоке по сравнению с ситуацией отсутствия мертвого времени ($T^* = 0$) либо с ситуацией наличия детерминированного мертвого времени ($T > 0$), но не устранить ее полностью.

Тогда искомая плотность вероятности $p(\tau)$ примет вид

$$p(\tau) = \int_{(T)} p(\tau, T) dT = \int_{(T)} p(T) p(\tau|T) dT,$$

где $p(\tau, T)$ – совместная плотность вероятности значений τ и T ; условная плотность вероятности $p(\tau|T)$ определяется выражениями (1) и (2) для общего и особого случаев соответственно при подстановке параметров $p = 1$ и $\delta = 0$; равномерная плотность $p(T)$ определена в разделе 1; (T) – область интегрирования значений случайной величины – длительности непродлевающегося случайного мертвого времени. Отметим, что внесение непродлевающегося случайного мертвого времени в математическую модель для случая рекуррентного потока, когда $p = 1$ и $\delta = 0$, оставляет наблюдаемый поток в классе рекуррентных потоков.

Область значений случайной величины мертвого времени представляет собой объединение двух областей, когда $0 \leq \tau < T^*$ и когда $\tau \geq T^*$, поэтому выражение для плотности $p(\tau)$ имеет следующий вид:

$$p(\tau) = \begin{cases} p_1(\tau) = \int_0^\tau p(T) p(\tau|T) dT, & 0 \leq \tau < T^*, \\ p_2(\tau) = \int_0^{T^*} p(T) p(\tau|T) dT, & \tau \geq T^*. \end{cases} \quad (3)$$

Подставляя выражение (1) при $p = 1$ в (3) и учитывая, что $p(T) = 1/T^*$, $0 \leq T < T^*$ для общего случая $\lambda_1 - \lambda_2 - \alpha_2 \neq 0$, находим

$$p_1(\tau) = \frac{1}{T^*} \left\{ 1 - e^{-\lambda_1 \tau} - e^{-(\lambda_1 + \alpha_2)\tau} + e^{-(\lambda_2 + \alpha_2)\tau} \right\}, \quad 0 \leq \tau < T^*, \quad (4)$$

$$p_2(\tau) = \frac{1}{T^*} \left\{ e^{-\lambda_1 \tau} \left[-1 + C_3 e^{\lambda_1 T^*} + C_4 e^{-\alpha_2 T^*} \right] + e^{-(\lambda_2 + \alpha_2)\tau} \left[-1 + C_3 e^{-(\lambda_1 - \lambda_2)T^*} + C_4 e^{(\lambda_2 + \alpha_2)T^*} \right] \right\}, \quad \tau \geq T^*, \quad (5)$$

где $C_3 = -\frac{\alpha_2(\lambda_2 + \alpha_2)}{(\lambda_1 + \alpha_2)(\lambda_1 - \lambda_2 - \alpha_2)}$, $C_4 = \frac{\lambda_1(\lambda_1 - \lambda_2)}{(\lambda_1 + \alpha_2)(\lambda_1 - \lambda_2 - \alpha_2)}$.

Аналогично подставляя выражение (2) в (3) при $p = 1$, $\delta = 0$ для особого случая $\lambda_1 - \lambda_2 - \alpha_2 = 0$, находим

$$p_1(\tau) = \frac{1}{T^*} \left\{ 1 - 2e^{-\lambda_1 \tau} + e^{-(\lambda_1 + \alpha_2)\tau} \right\}, \quad 0 \leq \tau < T^*, \quad (6)$$

$$p_2(\tau) = \frac{1}{T^*} e^{-\lambda_1 \tau} \left\{ -2 + \left[1 + \frac{\lambda_1 \alpha_2}{\lambda_1 + \alpha_2} (\tau - T^*) \right] e^{\lambda_1 T^*} + \left[1 - \frac{\lambda_1 \alpha_2}{\lambda_1 + \alpha_2} (\tau - T^*) \right] e^{-\alpha_2 T^*} \right\}, \quad \tau \geq T^*. \quad (7)$$

Отметим, что и для общего, и для особого случая рассматриваемого потока в точке $\tau = T^*$ выполняется равенство двух плотностей $p_1(T^*) = p_2(T^*)$ и неравенство их производных $p_1'(T^*) \neq p_2'(T^*)$, т.е. функция $p(\tau)$ является непрерывной и в точке $\tau = T^*$ имеет излом.

По определению математического ожидания случайной величины τ – длительности интервала между двумя соседними событиями наблюдаемого потока – с учетом формулы (3) имеем $M(\tau|T^*) = \int_0^{T^*} \tau p_1(\tau) d\tau + \int_{T^*}^{\infty} \tau p_2(\tau) d\tau$. Подставляя сюда плотности вероятности $p_1(\tau)$, $p_2(\tau)$, определенные в (4), (5) для общего случая, находим

$$M(\tau|T^*) = \frac{T^*}{2} + \frac{1}{\lambda_2 + \alpha_2} + \frac{\alpha_2}{\lambda_1(\lambda_1 + \alpha_2)} + \frac{\alpha_2(\lambda_1 - \lambda_2)}{\lambda_1(\lambda_1 + \alpha_2)^2(\lambda_2 + \alpha_2)} \frac{1}{T^*} \left(1 - e^{-(\lambda_1 + \alpha_2)T^*} \right), \quad T^* > 0. \quad (8)$$

Далее, подставляя формулы (6), (7) для особого случая, находим

$$M(\tau|T^*) = \frac{T^*}{2} + \frac{1}{\lambda_1} + \frac{\alpha_2}{\lambda_1(\lambda_1 + \alpha_2)} + \left[\frac{\alpha_2}{\lambda_1(\lambda_1 + \alpha_2)} \right]^2 \frac{1}{T^*} \left(1 - e^{-(\lambda_1 + \alpha_2)T^*} \right), \quad T^* \geq 0. \quad (9)$$

Аналитически доказано, что математическое ожидание $M(\tau|T^*)$, определяемое формулой (8) для общего случая и формулой (9) для особого случая, является возрастающей функцией переменной (параметра) T^* ($T^* \geq 0$). Из этого следует, что уравнение моментов $M(\tau|T^*) = C$ имеет единственное решение. Уравнение моментов может не иметь решения только в одном единственном случае, когда $C < M(\tau|T^* = 0)$, тогда принимается $\hat{T}^* = 0$. При этом оценки, получаемые методом моментов, являются состоятельными (выполнены условия теоремы о состоятельности оценок [21]).

3. Результаты статистических экспериментов для наблюдаемого потока

Для установления стационарного режима и определения свойств найденных оценок проведены статистические эксперименты.

Первый статистический эксперимент (установление стационарного режима). Было получено 100 реализаций ($N = 100$) имитационной модели наблюдаемого потока для общего и особого случаев при $T^* = 1; 3$ и $T_m = 50, 100, \dots, 1\,500$. Далее для каждого набора параметров было получено 100 решений уравнения моментов методом простой итерации при $\varepsilon = 0,0001$, $\Delta T^* = 0,001$. Каждое i -е решение есть значение оценки \hat{T}_i^* , $i = 1, \dots, 100$, параметра T^* . Заданный набор параметров для рекуррентного полусинхронного потока в общем случае: $\lambda_1 = 2, \lambda_2 = 1, \alpha_2 = 0,2$; в особом случае: $\lambda_1 = 2, \lambda_2 = 1,8, \alpha_2 = 0,2$.

На основании полученных данных вычислялись выборочное среднее искомым оценок $\hat{M}(\hat{T}^*) = \frac{1}{N} \sum_{i=1}^N \hat{T}_i^*$ и их выборочная вариация $\hat{V}(\hat{T}^*) = \frac{1}{N} \sum_{i=1}^N (\hat{T}_i^* - T^*)^2$, где T^* – известное из имитационной модели значение параметра.

В табл. 1 приведены результаты для общего случая наблюдаемого потока при $T^* = 1$. В первой строке таблицы указано время моделирования T_m (время наблюдения за потоком) ($T_m = 50, 100, \dots, 2\,000$ ед. времени); во второй и третьей строках – выборочное среднее $\hat{M}(\hat{T}^*)$ и выборочная вариация $\hat{V}(\hat{T}^*)$ для $T^* = 1$.

Таблица 1

Численные результаты первого статистического эксперимента для $T^* = 1$, общий случай

T_m	50	100	150	200	250	300	350	400	450	500
$\hat{M}(T^*)$	0,9628	1,1062	1,108	1,2059	1,1562	1,0810	1,1997	0,9151	1,1594	1,0282
$\hat{V}(T^*)$	9,8829	1,7811	0,5607	0,5945	0,5632	0,6084	0,6523	0,3512	0,3633	0,4652
550	600	650	700	750	800	850	900	950	1 000	1 050
0,9341	1,0573	1,0133	1,0789	1,0381	1,0684	0,9962	0,9714	1,0331	1,0255	0,9910
0,4155	0,2409	0,2246	0,2367	0,1766	0,1648	0,1976	0,1366	0,1361	0,1256	0,1571
1 100	1 150	1 200	1 250	1 300	1 350	1 400	1 450	1 500	1 550	1 600
1,0189	1,0316	1,0252	1,0204	1,0178	1,0246	1,0394	0,9830	0,9875	1,0169	0,9991
0,1034	0,1100	0,1342	0,0964	0,1271	0,0946	0,1037	0,1179	0,1085	0,0997	0,0796
	1 650	1 700	1 750	1 800	1 850	1 900	1 950	2 000		
	1,0007	1,0256	0,9533	1,0344	1,0140	1,0293	1,0153	0,9892		
	0,0642	0,0661	0,0840	0,0770	0,0807	0,0635	0,0909	0,0696		

Для наглядности на рис. 2 и 3 приведены графики зависимостей $\hat{M}(T^*)$ и $\hat{V}(T^*)$ от значения времени моделирования T_m для $T^* = 1$, построенные по данным табл. 1.

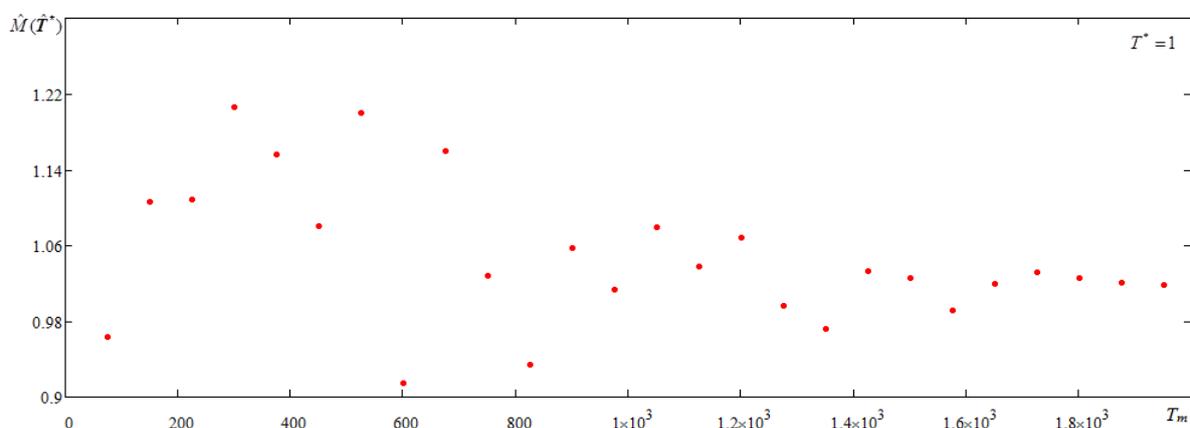


Рис. 2. График зависимости $\hat{M}(T^*)$ от T_m при $T^* = 1$ в общем случае

Fig. 2. Plot of $\hat{M}(T^*)$ versus T_m with $T^* = 1$ in general case

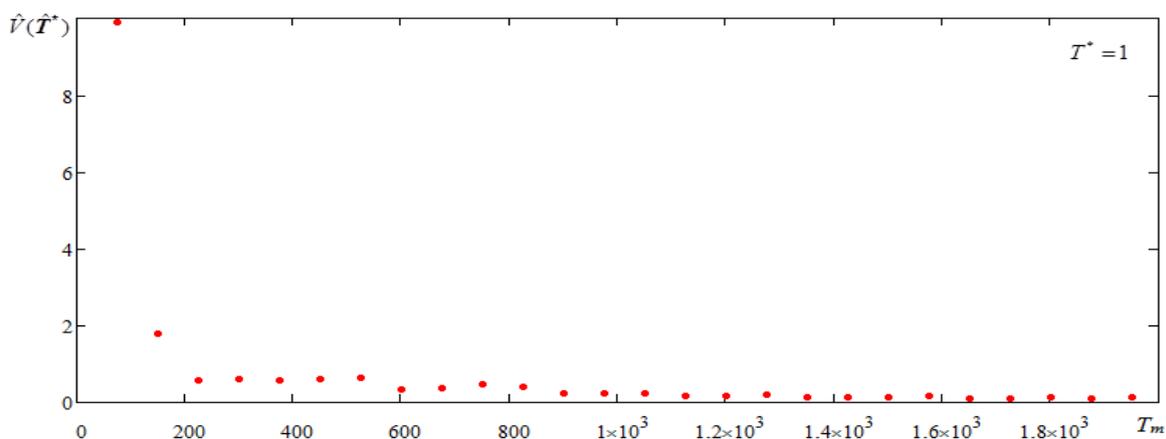


Рис. 3. График зависимости $\hat{V}(T^*)$ от T_m при $T^* = 1$ в общем случае

Fig. 3. Plot of $\hat{V}(T^*)$ versus T_m with $T^* = 1$ in general case

В табл. 2 приведены результаты для общего случая наблюдаемого потока при $T^* = 3$. Структура табл. 2 аналогична структуре табл. 1.

Численные результаты первого статистического эксперимента для $T^* = 3$, общий случай

T_m	50	100	150	200	250	300	350	400	450	500
$\hat{M}(T^*)$	3,1570	2,8336	3,3744	3,0769	3,0023	2,8782	3,1557	3,0098	2,9714	2,9456
$\hat{V}(T^*)$	23,089	6,4750	2,5065	2,3764	1,6844	1,0933	1,6212	1,2113	0,8877	0,8757
550	600	650	700	750	800	850	900	950	1 000	1 050
2,9310	3,1351	3,0106	2,9751	2,9633	2,9780	3,1098	2,9862	3,0876	3,0449	2,9822
0,5920	0,6786	0,7178	0,5904	0,4972	0,4203	0,5318	0,5243	0,3918	0,3930	0,4549
1 100	1 150	1 200	1 250	1 300	1 350	1 400	1 450	1 500	1 550	1 600
2,9933	2,9860	2,9937	2,9608	2,9516	3,0190	3,0104	2,9546	2,9979	3,0287	2,9543
0,3848	0,3413	0,4107	0,3657	0,3227	0,3035	0,2829	0,2885	0,2738	0,2367	0,2568
1 650	1 700	1 750	1 800	1 850	1 900	1 950	2 000			
2,9897	2,9656	3,0635	2,9404	3,0245	2,9945	2,9154	2,9995			
0,2090	0,2899	0,2852	0,2508	0,2480	0,1882	0,1711	0,1858			

На рис. 4 и 5 приведены графики зависимостей $\hat{M}(T^*)$ и $\hat{V}(T^*)$ от времени моделирования T_m для $T^* = 3$, построенные по данным табл. 2.

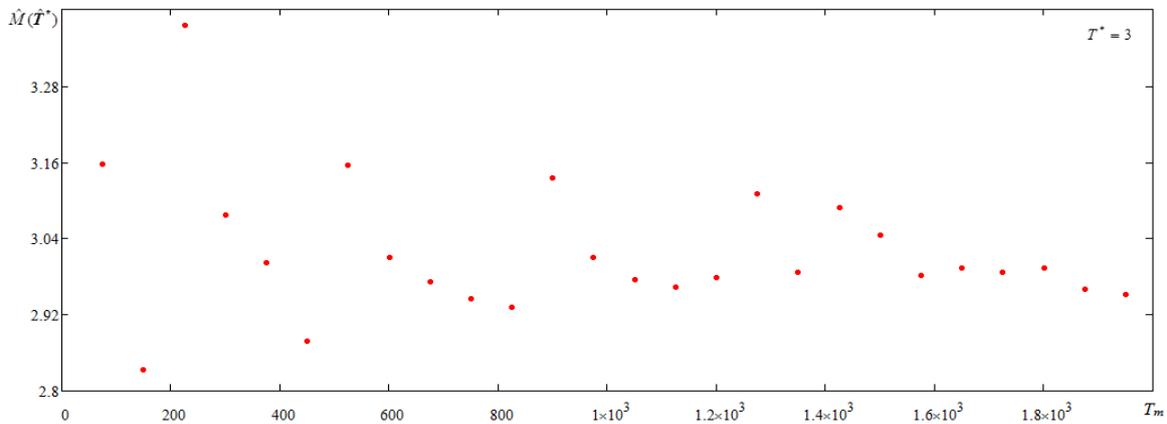


Рис. 4. График зависимости $\hat{M}(T^*)$ от T_m при $T^* = 3$ в общем случае

Fig. 4. Plot of $\hat{M}(T^*)$ versus T_m with $T^* = 3$ in general case

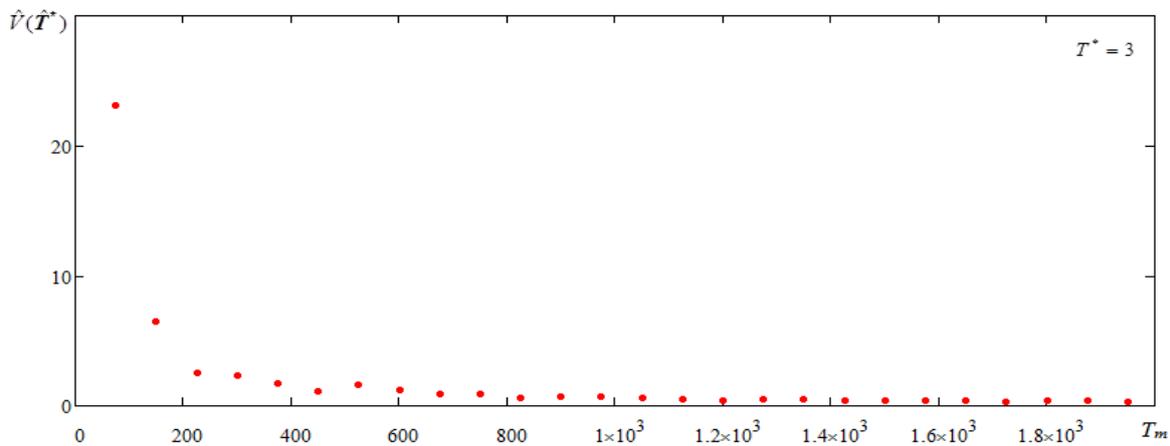


Рис.5. График зависимости $\hat{V}(T^*)$ от T_m при $T^* = 3$ в общем случае

Fig. 5. Plot of $\hat{V}(T^*)$ versus T_m with $T^* = 3$ in general case

Из анализа результатов первого статистического эксперимента следует, что для рекуррентного полусинхронного потока событий в общем случае справедливо:

1) стационарный режим функционирования потока устанавливается при $T_m \geq 1600$ ед. времени, так как выборочное среднее $\hat{M}(\hat{T}^*)$ стремится к постоянному значению, когда $T_m \geq 1600$ ед. времени;

2) оценка \hat{T}^* является смещенной оценкой; причиной смещения оценки \hat{T}^* ($\hat{T}^* < T^*$) относительно истинного T^* (известного из имитационной модели) является то, что значения случайного мертвого времени T сосредоточены около теоретического среднего ($T^*/2$).

Результаты первого статистического эксперимента для особого случая приведены в табл. 3, 4 для $T^* = 1$ и $T^* = 3$ соответственно, а также на рис. 6, 7 и 8, 9 для $T^* = 1$ и $T^* = 3$ соответственно.

Таблица 3

Численные результаты эксперимента для $T^* = 1$, особый случай

T_m	50	100	150	200	250	300	350	400	450	500
$\hat{M}(\hat{T}^*)$	0,9133	0,9025	0,9058	0,8939	0,9135	0,9120	0,9102	0,8962	0,9076	0,9105
$\hat{V}(\hat{T}^*)$	0,0370	0,0236	0,0176	0,0139	0,0140	0,0144	0,0120	0,0114	0,012	0,0128
550	600	650	700	750	800	850	900	950	1 000	1 050
0,8989	0,9081	0,911	0,9011	0,9073	0,9012	0,9050	0,9046	0,9019	0,9043	0,9040
0,0115	0,0131	0,0100	0,0108	0,0106	0,0114	0,0100	0,0108	0,0104	0,0103	0,0105
1 100	1 150	1 200	1 250	1 300	1 350	1 400	1 450	1 500	1 550	1 600
0,9031	0,9048	0,9055	0,9043	0,8792	0,9039	0,9048	0,9049	0,9040	0,9029	0,9046
0,0106	0,0091	0,0105	0,0105	0,0184	0,0101	0,0100	0,0099	0,0098	0,0099	0,0101
1 650	1 700	1 750	1 800	1 850	1 900	1 950	2 000			
0,9032	0,9038	0,9042	0,9050	0,9044	0,9042	0,9030	0,9053			
0,0098	0,0095	0,0099	0,0103	0,0098	0,0097	0,0099	0,0101			

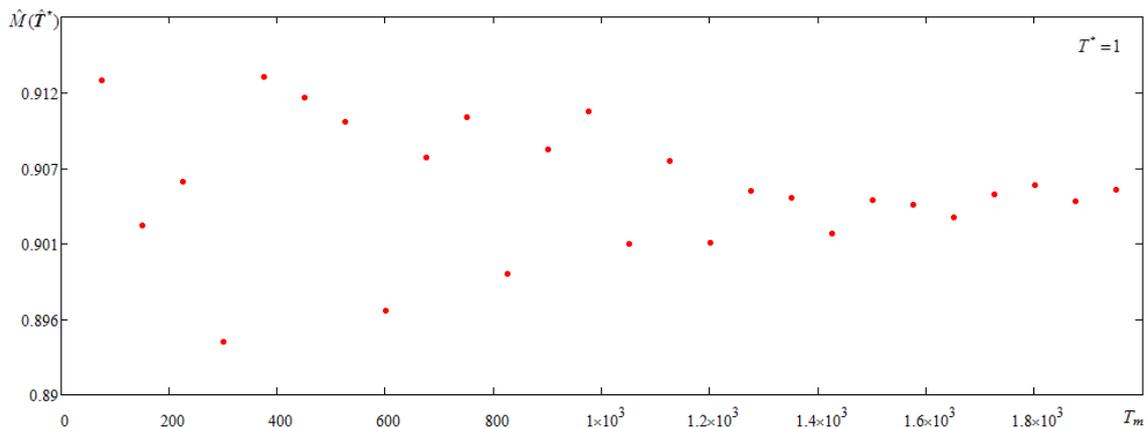


Рис. 6. График зависимости $\hat{M}(\hat{T}^*)$ от T_m при $T^* = 1$ в особом случае

Fig. 6. Plot of $\hat{M}(\hat{T}^*)$ versus T_m with $T^* = 1$ in special case

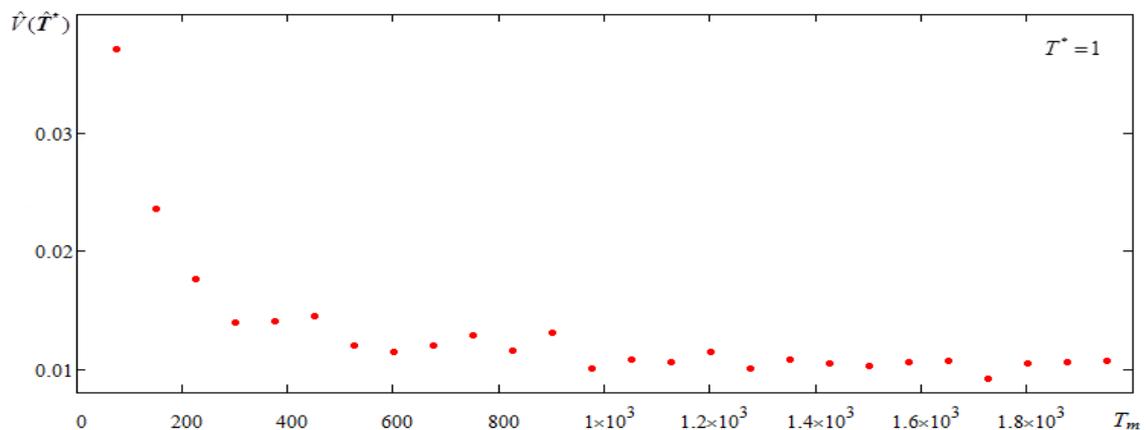


Рис. 7. График зависимости $\hat{V}(\hat{T}^*)$ от T_m при $T^* = 1$ в особом случае

Fig. 7. Plot of $\hat{V}(\hat{T}^*)$ versus T_m with $T^* = 1$ in special case

Численные результаты эксперимента для $T^* = 3$, особый случай

T_m	50	100	150	200	250	300	350	400	450	500
$\hat{M}(\hat{T}^*)$	2,9409	2,9383	2,9173	2,8994	2,9050	2,9087	2,9050	2,9146	2,9075	2,9096
$\hat{V}(\hat{T}^*)$	0,1629	0,0890	0,0537	0,0484	0,0359	0,0350	0,0356	0,0283	0,0186	0,0262
550	600	650	700	750	800	850	900	950	1 000	1 050
2,9019	2,8999	2,9134	2,9057	2,9028	2,9063	2,9125	2,9011	2,9013	2,9032	2,9070
0,0218	0,0198	0,0219	0,0185	0,0218	0,0181	0,0190	0,0162	0,0175	0,0163	0,0155
1 100	1 150	1 200	1 250	1 300	1 350	1 400	1 450	1 500	1 550	1 600
2,9070	2,9056	2,9038	2,9107	2,8707	2,9058	2,9043	2,9023	2,9037	2,9013	2,9058
0,0151	0,0144	0,0163	0,0164	0,0244	0,0144	0,0138	0,0129	0,0147	0,0131	0,0163
1 650	1 700	1 750	1 800	1 850	1 900	1 950	2 000			
2,9049	2,9109	2,9028	2,9065	2,9033	2,9048	2,9070	2,9101			
0,0125	0,0144	0,0138	0,0118	0,0128	0,0128	0,0130	0,0121			

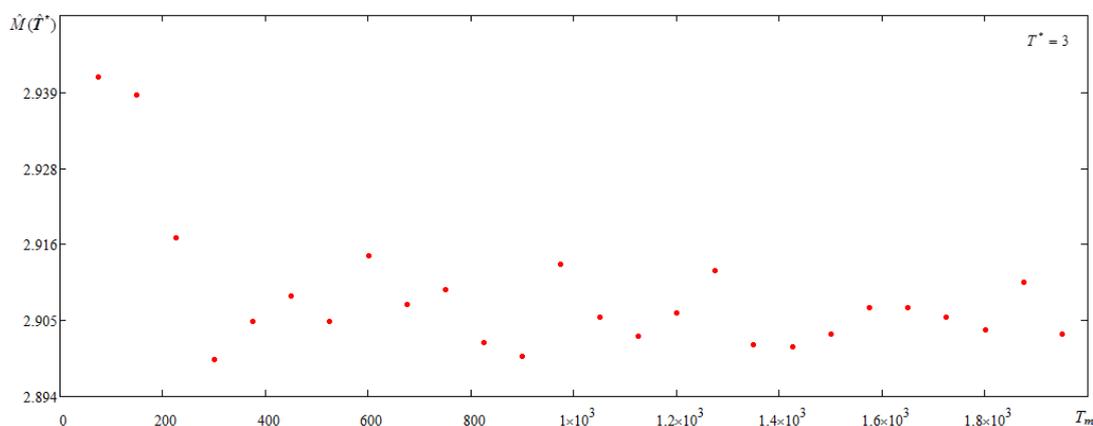


Рис. 8. График зависимости $\hat{M}(\hat{T}^*)$ от T_m при $T^* = 3$ в особом случае

Fig. 8. Plot of $\hat{M}(\hat{T}^*)$ versus T_m with $T^* = 3$ in special case

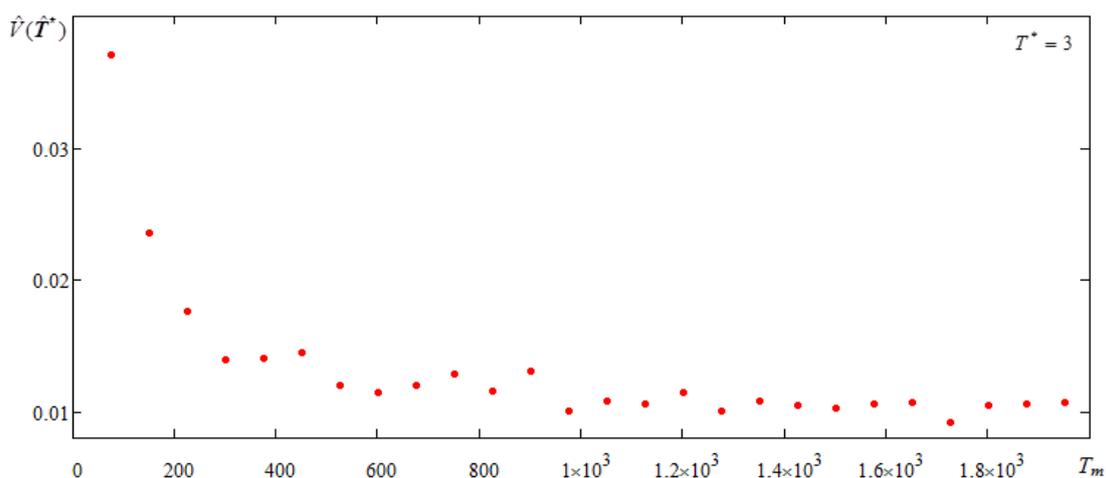


Рис.9. График зависимости $\hat{V}(\hat{T}^*)$ от T_m при $T^* = 3$ в особом случае

Fig. 9. Plot of $\hat{V}(\hat{T}^*)$ versus T_m with $T^* = 3$ in special case

Аналогично для особого случая рассматриваемого потока следует, что:

- 1) стационарный режим функционирования потока устанавливается при $T_m \geq 1300$ ед. времени;
- 2) оценка \hat{T}^* является смещенной оценкой.

Второй статистический эксперимент (исследование влияния параметра T^* на качество оценок). Второй статистический эксперимент поставлен при фиксированном времени моделирования, равном времени установления стационарного режима, полученному из первого статистического эксперимента. Варьирует параметр $T^* = 1, 2, 3, 4, 5$ остальные параметры принимаются такими же, как и в первом эксперименте.

Результаты второго статистического эксперимента для общего случая рассматриваемого потока приведены в табл. 5 ($T_m = 1\ 600$ ед. времени).

Таблица 5

Численные результаты второго статистического эксперимента, общий случай

T^*	1	2	3	4	5
$\hat{M}(T^*)$	0,9085	1,9106	2,91	3,9152	4,9104
$\hat{V}(T^*)$	0,0106	0,0106	0,016	0,0247	0,0413

На рис. 10, 11 приведены графики зависимости выборочного среднего $\hat{M}(T^*)$ и выборочной вариации $\hat{V}(T^*)$ от параметра $T^* = 1, 2, 3, 4, 5$, построенные по данным табл. 5.

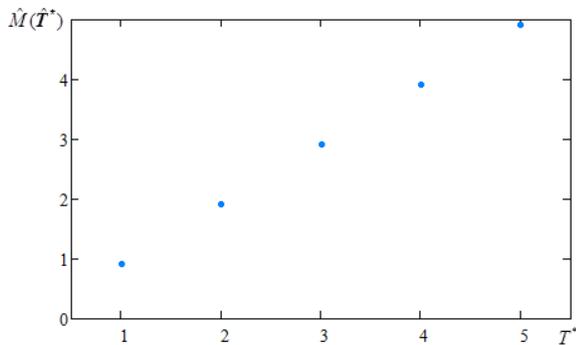


Рис. 10. График зависимости $\hat{M}(T^*)$ от T^* в общем случае
Fig. 10. Plot of $\hat{M}(T^*)$ versus T^* in general case

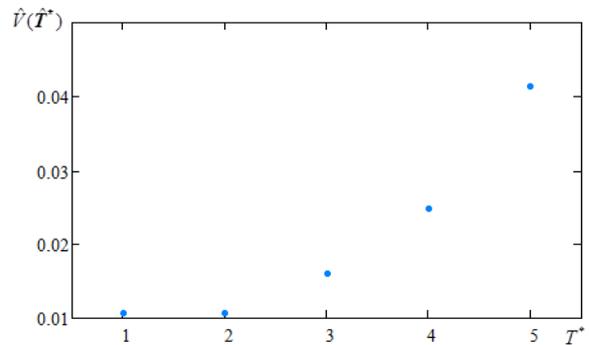


Рис. 11. График зависимости $\hat{V}(T^*)$ от T^* в общем случае
Fig. 11. Plot of $\hat{V}(T^*)$ versus T^* in general case

Результаты второго статистического эксперимента для особого случая приведены в табл. 6 ($T_m = 1\ 300$ ед. времени), а также на рис. 12, 13.

Таблица 6

Численные результаты второго статистического эксперимента, особый случай

T^*	1	2	3	4	5
$\hat{M}(T^*)$	0,8761	1,8795	2,886	3,8805	4,8837
$\hat{V}(T^*)$	0,0185	0,0187	0,0265	0,0350	0,0512

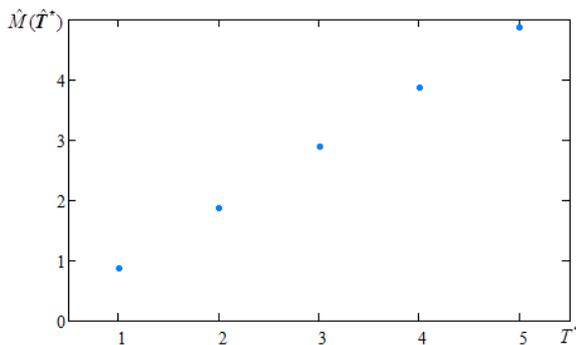


Рис. 12. График зависимости $\hat{M}(T^*)$ от T^* в общем случае
Fig. 12. Plot of $\hat{M}(T^*)$ versus T^* in general case

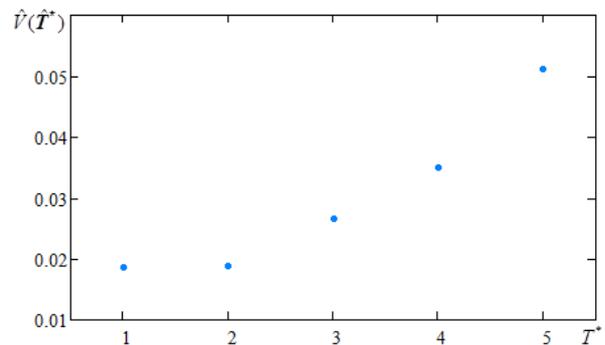


Рис. 13. График зависимости $\hat{V}(T^*)$ от T^* в общем случае
Fig. 13. Plot of $\hat{V}(T^*)$ versus T^* in general case

Данный эксперимент демонстрирует тот факт, что при увеличении значений параметра равномерного распределения мертвого времени T^* увеличивается выборочная вариация оценки. Это объясняется тем, что при больших значениях T^* увеличивается число потерянных событий исходного потока и, как следствие, ухудшается качество оценивания.

Заключение

В данной работе рассмотрен рекуррентный полусинхронный дважды стохастический поток событий в общем и особом случаях соотношения его параметров с непродлевающимся случайным мертвым временем, распределенным по равномерному закону.

Аналитически получены формулы (4), (5), определяющие плотность вероятности значений длительности интервала между соседними событиями в наблюдаемом потоке при случайном мертвом времени для общего случая $\lambda_1 - \lambda_2 - \alpha_2 \neq 0$ и аналогичные формулы (6), (7) для особого случая $\lambda_1 - \lambda_2 - \alpha_2 = 0$; выведены формулы (8), (9) для математического ожидания длительности интервала между соседними событиями для общего и особого случая соответственно.

Методом моментов найдены состоятельные ММ-оценки параметра T^* равномерного распределения длительности случайного мертвого времени; полученные оценки экспериментально исследованы на качество. Приведенные результаты численных расчетов указывают на приемлемое качество оценивания.

Список источников

1. 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.
2. Вишневецкий В.М., Дудин А.Н., Клименок В.Н. Стохастические системы с коррелированными потоками. Теория и применение в телекоммуникационных сетях. М. : Техносфера, 2018. 564 с.
3. Cox D.R. The analysis of non-Markovian stochastic processes by the inclusion of supplementary variables // Proceedings of the Cambridge Philosophical Society. 1955. V. 51 (3). P. 433–441.
4. Kingman Y.F.C. On doubly stochastic Poisson process // Proceedings of the Cambridge Philosophical Society. 1964. V. 60 (4). P. 923–930.
5. Neuts M.F. A versatile Markovian point process // Journal of Applied Probability. 1979. V. 16. P. 764–779.
6. Lucantoni D.M. New results on the single server queue with a batch markovian arrival process // Communication in Statistics Stochastic Models. 1991. V. 7. P. 1–46.
7. Горцев А.М., Нежелская Л.А. Оценивание длительности мертвого времени и параметров синхронного альтернирующего потока событий // Вестник Томского государственного университета. 2003. № S6. С. 232–239.
8. Gortsev A.M., Nezhel'skaya L.A. Estimation of parameters of synchronously alternating Poisson stream of events by the moment method // Telecommunications and Radio Engineering. 1996. V. 50 (1). P. 56–63.
9. Горцев А.М., Нежелская Л.А. Оценивание параметров синхронного дважды стохастического потока событий методом моментов // Вестник Томского государственного университета. 2002. № S1-1. С. 24–29.
10. Nezhel'skaya L.A. Probability density function for modulated MAP event flows with unextendible dead time // Communication in Computer and Information Science. 2015. V. 564. P. 141–151.
11. Горцев А.М., Нежелская Л.А. Оценивание параметров асинхронного потока с иницированием лишних событий методом моментов // Вестник Томского государственного университета. 2006. № S18. С. 267–273.
12. Леонова М.А., Нежелская Л.А. Вероятность ошибки при оценивании состояний обобщенного асинхронного потока событий // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2012. № 2 (19). С. 88–101.
13. Горцев А.М., Зуев В.Л. Оптимальная оценка состояний асинхронного дважды стохастического потока событий с произвольным числом состояний // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2010. № 2 (11). С. 44–65.
14. Горцев А.М., Леонова М.А., Нежелская Л.А. Сравнение МП- и ММ-оценок длительности мертвого времени в обобщенном асинхронном потоке событий // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2013. № 4 (25). С. 32–42.
15. Калягин А.А., Нежелская Л.А. Сравнение МП- и ММ-оценок длительности мертвого времени в обобщенном полусинхронном потоке событий // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2015. № 3 (32). С. 23–32.

16. Горцев А.М., Калягин А.А., Нежелская Л.А. Оценка максимального правдоподобия длительности мертвого времени в обобщенном полусинхронном потоке // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2015. № 1 (30). С. 27–37.
17. Нежелская Л.А. Оптимальное оценивание состояний полусинхронного потока событий в условиях его частичной наблюдаемости // Вестник Томского государственного университета. 2000. № 269. С. 95–98.
18. Горцев А.М., Калягин А.А., Нежелская Л.А. Совместная плотность вероятностей длительности интервалов обобщенного полусинхронного потока событий при непродолеваемом мертвом времени // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2014. № 2 (27). С. 19–29.
19. Апанасович В.В., Коляда А.А., Чернявский А.Ф. Статистический анализ случайных потоков в физическом эксперименте. Минск : Университетское, 1988. 256 с.
20. Нежелская Л.А. Оценка состояний и параметров дважды стохастических потоков событий : дис. ... д-ра физ.-мат. наук. Томск, 2016. 341 с.
21. Малинковский Ю.В. Теория вероятностей и математическая статистика. Гомель : ГГУ им. Ф. Скорины, 2004. Ч. 2: Математическая статистика. 146 с.

References

1. 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.
2. Vishnevsky, V.M., Dudin, A.N. & Klimenok, V.N. (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.
3. Cox, D.R. (1955) The analysis of non-Markovian stochastic processes by the inclusion of supplementary variables. *Proceedings of the Cambridge Philosophical Society*. 51(3). pp. 433–441. DOI: 10.1017/S0305004100030437
4. Kingman, Y.F.C. (1964) On doubly stochastic Poisson process. *Proceedings of the Cambridge Philosophical Society*. 60(4). pp. 923–930.
5. Neuts, M.F. (1979) A versatile Markovian point process. *Journal of Applied Probability*. 16(4). pp. 764–779. DOI: 10.2307/3213143.
6. Lucantoni, D.M. (1991) New results on the single server queue with a batch Markovian arrival process. *Communication in Statistics Stochastic Models*. 7. pp. 1–46. DOI: 10.1080/15326349108807174
7. Gortsev, A.M. & Nezhelskaya, L.A. (2003) Otsenivaniye dlitel'nosti mertvogo vremeni i parametrov sinkhronnogo al'terniruyushchego potoka sobytiy [Dead time period and parameter estimation of synchronous alternating flow of events]. *Vestnik Tomskogo gosudarstvennogo universiteta – Tomsk State University Journal*. 6. pp. 232–239.
8. Gortsev, A.M. & Nezhelskaya, L.A. (1996) Estimation of parameters of synchronously alternating Poisson stream of events by the moment method. *Telecommunications and Radio Engineering*. 50(1). pp. 56–63.
9. Gortsev, A.M. & Nezhelskaya, L.A. (2002) Estimation of the parameters of a synchronous doubly stochastic event flow by the method of moments. *Vestnik Tomskogo gosudarstvennogo universiteta – Tomsk State University Journal*. S1-1. pp. 24–29.
10. Nezhelskaya, L.A. (2015) Probability density function for modulated MAP event flows with unextendible dead time. *Communication in Computer and Information Science*. 564. pp. 141–151.
11. Gortsev, A.M. & Nezhelskaya, L.A. (2006) Otsenivaniye parametrov asinhronnogo potoka s iniciirovaniem lishnih sobytiy metodom momentov. *Vestnik Tomskogo gosudarstvennogo universiteta – Tomsk State University Journal*. S18. pp. 267–273.
12. Leonova, M.A. & Nezhelskaya, L.A. (2012) The probability of wrong decisions in the estimation of states of 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(19). pp. 88–101.
13. Gortsev, A.M. & Zuevich, V.L. (2010) Optimal estimation of states of the asynchronous doubly stochastic flow of events with arbitrary number of the states. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 2(11). pp. 44–65.
14. Gortsev, A.M., Leonova, M.A. & Nezhelskaya, L.A. (2013) The comparison of maximum likelihood estimation and method of moments estimation of dead time value in 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*. 4(25). pp. 32–42.
15. Kalyagin, A.A. & Nezhelskaya, L.A. (2015) The comparison of maximum likelihood estimation and method of moments estimation of dead time value in a generalized semisynchronous flow of events. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 3(32). pp. 23–32.
16. Gortsev, A.M., Kalyagin, A.A. & Nezhelskaya, L.A. (2015) Maximum likelihood estimation of dead time at a generalized semisynchronous flow of events. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 1(30). pp. 27–37.

17. Nezhelskaya, L. (2000) Optimal'noe otsenivanie sostoyaniy polusinkhronnogo potoka sobytii v usloviyakh ego chastichnoy nablyudaemosti [Optimal estimation of the states of semi-synchronous event flow under conditions of its partial observability]. *Vestnik Tomskogo gosudarstvennogo universiteta – Tomsk State University Journal*. 269. pp. 95–98.
18. Gortsev, A.M., Kalyagin, A.A. & Nezhelskaya, L.A. (2014) The joint probability density of duration of the intervals in a generalized semisynchronous flow of events with unprolonging dead time. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 2(27). pp. 19–29.
19. Apanasovich, V.V., Kolyada, A.A. & Chernyavsky, A.F. (1988) *Statisticheskiy analiz sluchaynykh potokov v fizicheskom eksperimente* [The statistical analysis of series of random events in physical experiment]. Minsk: Universitetskoe.
20. Nezhelskaya, L.A. (2016) *Otsenka sostoyaniy i parametrov dvazhdy stokhasticheskikh potokov sobytii* [Estimation of States and Parameters of Doubly Stochastic Event Flows]. Physics and Mathematics Dr. Diss. Tomsk.
21. Malinkovsky, Yu.V. (2004) *Teoriya veroyatnostey i matematicheskaya statistika* [Probability Theory and Mathematical Statistics]. Gomel: Francisk Skorina Gomel State University.

Информация об авторах:

Горцев Александр Михайлович – профессор, доктор технических наук, заведующий кафедрой прикладной математики Института прикладной математики и компьютерных наук Национального исследовательского Томского государственного университета (Томск, Россия). E-mail: a-gortsev@mail.ru

Веткина Анна Васильевна – аспирант кафедры прикладной математики Института прикладной математики и компьютерных наук Национального исследовательского Томского государственного университета (Томск, Россия). E-mail: anyavetkina@gmail.com

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

Information about the authors:

Gortsev Alexander M. (Doctor of Technical Sciences, Professor, National Research Tomsk State University, Tomsk, Russian Federation). E-mail: a-gortsev@mail.ru

Vetkina Anna V. (Post-graduate Student, National Research Tomsk State University, Tomsk, Russian Federation). E-mail: anyavetkina@gmail.com

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

Поступила в редакцию 23.08.2022; принята к публикации 29.11.2022

Received 23.08.2022; accepted for publication 29.11.2022

Научная статья

УДК 519.24

doi: 10.17223/19988605/61/6

Распределения второго порядка: построение, операции, приложения

Борис Станиславович Добронет¹, Ольга Аркадьевна Попова²

^{1, 2} Сибирский федеральный университет, Красноярск, Россия

¹ BDobronets@yandex.ru

² OlgaArc@yandex.ru

Аннотация. Рассмотрены методы исследования неопределенностей вероятностных оценок эмпирических функций распределения. Для этого предлагается новый подход, основанный на понятии распределения второго порядка. Для численных реализаций операций над распределениями второго порядка применяются вероятностные расширения их параметризованных представлений. Приведены примеры использования данного подхода для оценки надежности технических объектов в условиях малых выборок.

Ключевые слова: вычислительный вероятностный анализ; распределения второго порядка; неточные вероятности; вероятностные расширения

Для цитирования: Добронет Б.С., Попова О.А. Распределения второго порядка: построение, операции, приложения // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2022. № 61. С. 61–68. doi: 10.17223/19988605/61/6

Original article

doi: 10.17223/19988605/61/6

Second-order distributions: construction, operations, applications

Boris S. Dobronets¹, Olga A. Popova²

^{1, 2} Siberian Federal University, Krasnoyarsk, Russia

¹ BDobronets@yandex.ru

² OlgaArc@yandex.ru

Abstract. The article considers methods for studying the uncertainties of probabilistic estimates of empirical distribution functions. To this end, a new approach based on the concept of a second-order distribution is proposed. For numerical implementations of operations on second-order distributions, probabilistic extensions of their parametrized representations are used. Examples of using this approach to assess the reliability of technical objects under conditions of small samples are given.

Keywords: computational probabilistic analysis; second-order distributions; imprecise probabilities; probabilistic extensions

For citation: Dobronets, B.S. Popova, O.A. (2022) Second-order distributions: construction, operations, applications. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naja tehnik i informatika – Tomsk State University Journal of Control and Computer Science*. 61. pp. 61–68. doi: 10.17223/19988605/61/6

Построение эмпирических функций распределения практически не бывает точным; оно сопровождается ошибками, что приводит к неопределенности вероятностных оценок. Неточные вероятности – один из источников вероятностей второго порядка. Предположим, что у нас есть оценка вероятности P некоего события. Когда мы можем оценить границы $P_1 \leq P \leq P_2$, можно говорить об интервальной оценке вероятности P . Такие интервальные оценки для гистограмм приводят к интервальным гисто-

граммам. В тех случаях, когда для P существует вероятностная оценка, можно говорить о вероятности второго порядка. Таким образом, вероятности второго порядка – это вероятностные оценки самих вероятностей.

Для работы с неопределенными вероятностями существует несколько подходов [1]. Один из распространенных методов работы с неопределенными вероятностями – использование P-бокс [2]. P-бокс представляет собой интервальную функцию распределения ($F_1(x) \leq F(x) \leq F_2(x)$).

В работе [3] показано использование неопределенных вероятностей в инженерных расчетах. При этом предполагается, что при стохастическом моделировании для входных данных известны семейства распределений. Входные данные для моделирования выбираются случайным образом из этих семейств, далее методом Монте-Карло производится стохастическое моделирование. Важно отметить, что при работе с неопределенными вероятностями существенным препятствием является отсутствие эффективных арифметических операций.

Одним из подходов, моделирующих неопределенность вероятностных оценок, стало использование вероятностей второго порядка. В работе [4] для представления вероятностей второго порядка предложено новое понятие гистограммы второго порядка. Применение вычислительного вероятностного анализа позволило развить идею далее на основе применения кусочно-полиномиальных моделей [5, 6]. Для численных реализаций операций над распределениями второго порядка в статье рассматривается новый подход, основанный на использовании вероятностных расширений их параметризованных представлений.

Следует отметить давний и устойчивый интерес к вероятностям второго порядка в различных областях, включая принятие решений [7, 8]. Среди русскоязычных монографий, посвященных методам работы с неопределенностями, можно выделить работы О.И. Ужга-Реброва [9].

1. Вероятностные расширения

Основа вычислительного вероятностного анализа – численные арифметические операции над функциями плотностей вероятностей и вероятностные расширения (построение законов распределений функций случайных аргументов). Функции плотности вероятности случайных величин x, y, z будем обозначать полужирным фонтном $\mathbf{x}, \mathbf{y}, \mathbf{z}$. Тогда функция плотности вероятности \mathbf{z} случайной величины $z = x * y$, $*$ $\in \{+, -, \cdot, /\}$, записывается в виде $\mathbf{z} = \mathbf{x} * \mathbf{y}$.

Рассмотрим задачу определения закона распределения функции нескольких случайных аргументов.

Пусть (x_1, x_2, \dots, x_n) – система случайных непрерывных переменных с совместной функцией плотности вероятности $p(x_1, x_2, \dots, x_n)$. Случайная переменная z

$$z = f(x_1, x_2, \dots, x_n),$$

где функция $f : R^n \rightarrow R$.

Определение 1. Будем говорить, что случайная функция $\mathbf{f} : R^n \rightarrow R$ является вероятностным продолжением вещественной функции $f : R^n \rightarrow R$ на множестве $D \subset R^n$, если $\mathbf{f}(x) = f(x)$ для всех точечных аргументов $x \in D$.

Определение 2. Случайная функция $\mathbf{f} : R^n \rightarrow R$ называется вероятностным расширением вещественной функции $f : R^n \rightarrow R$ на множестве $D \subset R^n$, если она

(i) является вероятностным продолжением f на D ,

(ii) функция плотности вероятности \mathbf{f} совпадает с функцией плотности вероятности \mathbf{z} случайной величины z :

$$z = f(x_1, x_2, \dots, x_n).$$

Таким образом, мы можем записать

$$\mathbf{z} = \mathbf{f}(x_1, \dots, x_n).$$

В тех случаях, когда надо указать непосредственно значение плотности вероятности f в некоторой точке ξ , будем использовать обозначение

$$z(\xi) = f(\mathbf{x}_1, \dots, \mathbf{x}_n)(\xi).$$

Теорема 1 [10]. Пусть (x_1, \dots, x_n) – случайные величины и $f(\mathbf{x}_1, \dots, \mathbf{x}_n)$ – вероятностное расширение $f(x_1, x_2, \dots, x_n)$, и для всех вещественных t функция $f(t, \mathbf{x}_2, \dots, \mathbf{x}_n)$ – вероятностное расширение $f(t, x_2, \dots, x_n)$. Тогда

$$f(\mathbf{x}_1, \dots, \mathbf{x}_n)(\xi) = \int_{x_1}^{\bar{x}_1} x_1(t) f(t, \mathbf{x}_2, \dots, \mathbf{x}_n)(\xi) dt.$$

2. Распределения второго порядка

В вероятностном пространстве $(\Omega, \mathcal{F}, \mathbb{P})$ случайный процесс представляет собой набор случайных величин

$$\{a(x, \omega), x \in D, \omega \in \Omega\}.$$

Термин «случайное поле» обычно относится к случайному процессу, принимающему значения в евклидовом пространстве R^d , $d = 1, 2, 3$. Случайное поле можно посмотреть двумя способами:

- для фиксированного $x \in D$, $a(x, \cdot)$ является случайной величиной в Ω ;
- для фиксированного $\omega \in \Omega$, $a(\cdot, \omega)$ является реализацией случайного поля в D .

Определение 3. Распределение второго порядка $f^{(2)}$ – случайное поле $f(x, \omega)$, $x \in D, \omega \in \Omega$, заданное на $D \subset R$, где $(\Omega, \mathcal{F}, \mathbb{P})$ – вероятностное пространство. Обладает следующими свойствами: для фиксированного $\omega \in \Omega$, $f(\cdot, \omega)$ является функцией распределения f_ω .

В тех случаях, когда для распределений второго порядка f_i , $f_i(\cdot, \omega)$ – функции плотности вероятности, для распределений второго порядка можно определить арифметические операции.

Определение 4. Операции над распределениями второго порядка. Пусть $f^{(2)}, g^{(2)}$ – распределения второго порядка, $(\Omega_f, \mathcal{F}_f, \mathbb{P}_f), (\Omega_g, \mathcal{F}_g, \mathbb{P}_g)$ – соответственно их вероятностные пространства. Тогда результат операции $f^{(2)} * g^{(2)}$, $*$ $\in \{+, -, \cdot, /$, – распределение второго порядка $F^{(2)}$:

$$F(\cdot, \omega_*) = \{f(\cdot, \omega_f) * g(\cdot, \omega_g) \mid (\omega_f, \omega_g) \in \Omega_*\},$$

где $(\Omega_*, \mathcal{F}_*, \mathbb{P}_*)$ – вероятностное пространство, $\Omega_* = \Omega_f \times \Omega_g$.

Рассмотрим пример построения кусочно-полиномиального представления для распределения второго порядка. Пусть (ξ_1, \dots, ξ_n) , $n = 9$, – выборка случайной величины X с функцией распределения $F(t), t \in [0, 2]$. Далее $z_i = F(\xi_i), i = 1, \dots, n$. Заметим, что $z_i, i = 1, \dots, n$ – равномерно распределенные случайные величины на отрезке $[0, 1]$. Если $z_1 \leq z_2 \leq \dots \leq z_n$, тогда z_k – k -я порядковая статистика и математическое ожидание $[Ez_k] = k/(n+1)$.

На рис. 1 представлена функция распределения второго порядка, аппроксимирующая распределение кусочно-линейных интерполяций распределений Ирвина–Холла третьей степени, построенная на выборке случайной величины размерности 9. Голубые линии – плотности вероятности случайной кусочно-линейной функции. Красная линия – точная функция распределения. Зеленые линии – границы 95%-ной доверительной области.

Функция плотности вероятности есть производная от функции распределения

$$f(x) = F'(x),$$

следовательно, производная от функции распределения второго порядка будет функцией плотности вероятности второго порядка.

Производная от кусочно-линейной функции – кусочно-постоянная функция. На рис. 2 приведен пример кусочно-постоянной функции – функции плотности вероятности второго порядка (надежной

оценки ф.п.в.). Красная линия – точная функция плотности вероятности. Синие линии – функции плотности вероятности надежной оценки, построенной на выборке случайной величины размерности 9.

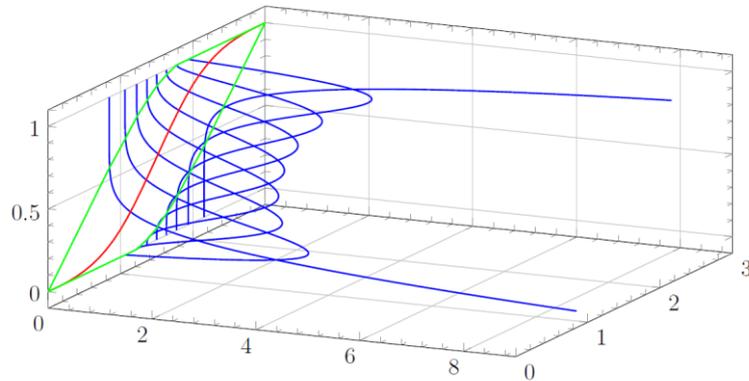


Рис. 1. Распределение второго порядка, аппроксимирующее распределение Ирвина–Холла третьей степени
Fig.1. Second order distribution, approximating the Irwin-Hall distribution of the third degree

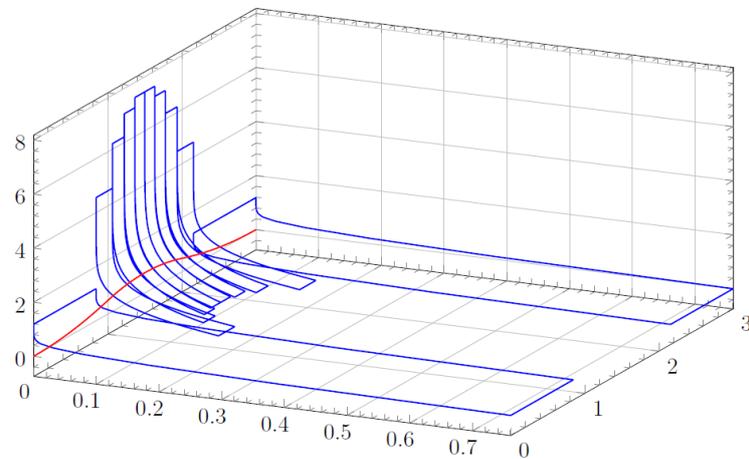


Рис. 2. Функция плотности вероятности второго порядка
Fig. 2. Second order probability density function

Далее будем использовать эрмитовы кубические сплайны S и точки $(x_i, i / (n + 1))$ для построения аппроксимации функции распределения F :

$$F(x) = \begin{cases} x^2 / 2, & x \in [0, 1], \\ x(2 - x / 2) - 1, & x \in [1, 2], \end{cases}$$

$$S(0) = 0, \quad S'(0) = 0, \quad S(1) = 1, \quad S'(1) = 0,$$

$$S(x) = fv(x - 1) + mw(x - 1) + v(x - 2).$$

$$v(x) = (|x - 1|)^2 (2|x - 1| + 1), \quad w(x) = x(|x - 1|)^2,$$

$$\sum_{i=1}^n (s(\xi_i) - z_i)^2 + \alpha \|s''\|_2^2 \rightarrow \min,$$

$$Ay = b,$$

где вектор $y = (f, m)^T$. Значение параметра α выбиралось таким образом, чтобы константы f, m определяли монотонный сплайн; легко видеть, что точка (f, m) должна лежать внутри треугольной области с вершиной в точке $(0, 5; 1, 5)$ и основанием $[0; 1]$. На рис. 3 приведены функции плотности вероятности констант m и f . Распределение второго порядка S можно представить в параметризованном виде

$$S(x, f, m) = fv(x - 1) + mw(x - 1) + v(x - 2).$$

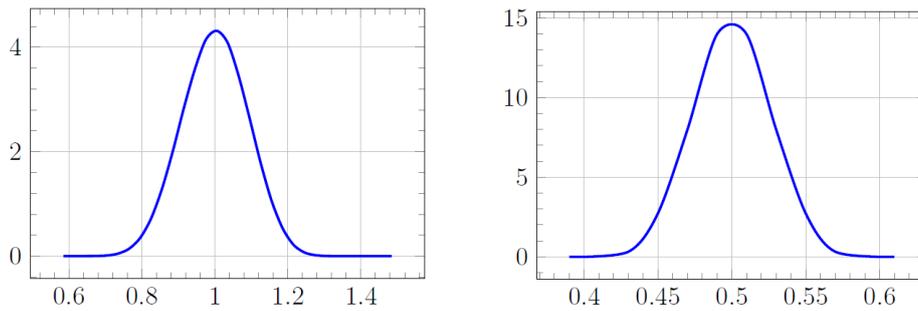


Рис. 3. Функции плотности вероятности констант m и f
 Fig. 3. Probability density functions of constants m and f

Для нахождения функции плотности вероятности распределения второго порядка продифференцируем S :

$$S'(x, f, m) = fv'(x-1) + mw'(x-1) + v'(x-2).$$

На рис. 4 представлена функция плотности вероятности второго порядка. Красная линия – математическое ожидание ES , зеленые линии – верхние и нижние границы доверительного интервала, синие линии – функции плотности вероятности при фиксированных значениях аргумента.

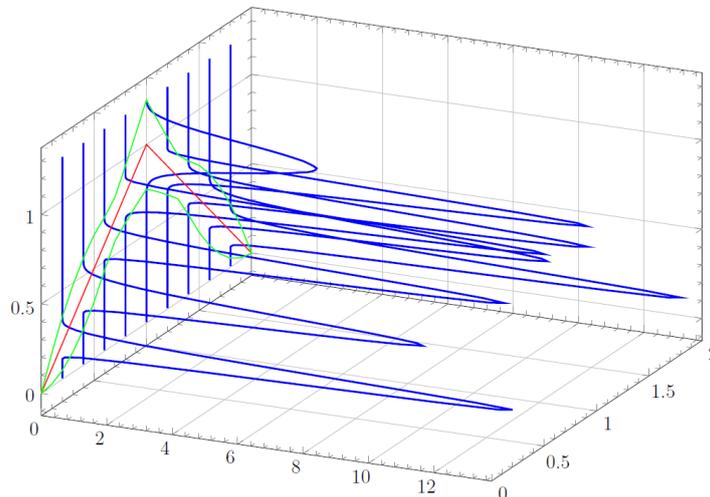


Рис. 4. Функция плотности вероятности второго порядка
 Fig. 4. Second order probability density function

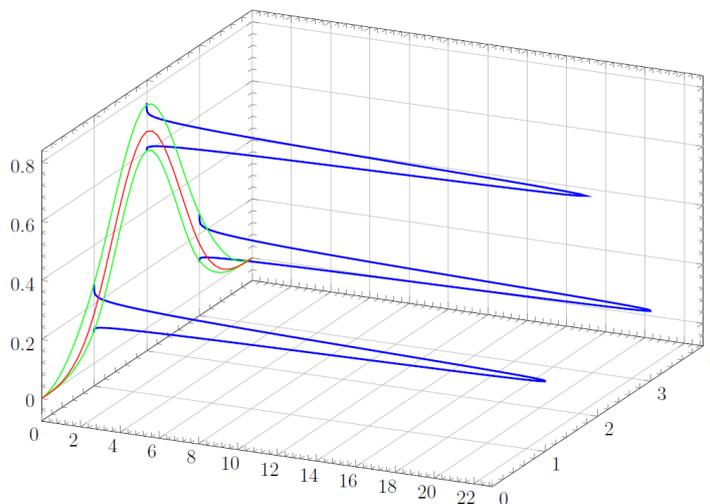


Рис. 5. Сумма распределений второго порядка
 Fig. 5. The sum of the second order distributions

В силу того, что распределения второго порядка мы представляем в параметризованном виде, для операций над распределениями второго порядка воспользуемся техникой вероятностных расширений. Таким образом, арифметические операции можно представить как вероятностные расширения соответствующих функций от параметров. На рис. 5 показан результат сложения двух распределений второго порядка, представленных случайными эрмитовыми сплайнами [6]. Красная линия – математическое ожидание распределения второго порядка, зеленые линии – верхние и нижние 95%-ные доверительные границы, синие линии – функции плотности вероятности при фиксированных значениях аргумента.

3. Достоверные оценки показателей надежности оборудования

В качестве примера использования распределений второго порядка рассмотрим построение достоверных оценок показателей надежности оборудования в условиях малых выборок статистических данных об отказах. Подход основан на использовании порядковых статистик и случайных интерполяционных многочленов для построения достоверных оценок функций распределения.

Вероятность безотказной работы $P(t)$ – это вероятность того, что в течение указанного времени работы не произойдет отказа. Время работы – это продолжительность или объем работы. Частота отказов – мера отказов за единицу времени. Частота отказов зависит от распределения отказов, которое представляет собой совокупную функцию распределения, описывающую вероятность отказов до момента времени t :

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq \xi < t + \Delta t | t \leq \xi)}{\Delta t} = \frac{f(t)}{P(t)} = \frac{f(t)}{1 - F(t)}.$$

Заметим, что $f(t) = P'(t)$ и

$$\lambda(t) = -\frac{P'(t)}{P(t)},$$

где $P(t)$ – вероятность безотказной работы.

Пусть $(\xi_1, \xi_2, \dots, \xi_n)$ – статистика отказов, полученная опытным путем. Тогда

$$-\ln(z_i) = \int_0^{\xi_i} \lambda(\xi) d\xi,$$

где $z_i = P(\xi_i)$.

Для нахождения $\lambda(t)$ будем использовать метод наименьших квадратов. Представим аппроксимацию $\lambda(t)$ в виде случайной кусочно-линейной функции $l(t)$

$$\lambda(t) \approx l(t) = \sum_{i=1}^m a_i \psi_i(t), t \in [0, T],$$

где $\{\psi_i, i = 1, \dots, m\}$ – базис в пространстве кусочно-линейных функций. Следуя работе [5], получаем приближение

$$\lambda(t) \approx \sum_{l=1}^n \left(\sum_{j=1}^m \psi_j(t) \right) \gamma_l \ln(z_l).$$

Поскольку известна совместная функция плотности вероятности $p(z_1, z_2, \dots, z_n)$, можем построить вероятностное расширение $l(t)$:

$$l(t) = \sum_{l=1}^n \left(\sum_{j=1}^m \psi_j(t) \right) \gamma_l \ln(z_l).$$

Пример. Имеем $(\xi_1, \xi_2, \dots, \xi_n)$, $n = 29$. Предположим, $\lambda(t)$ имеет вид:

$$\lambda(t) = \begin{cases} 0,2, & t \in [0; 0,7], \\ 0,2 + 12(t - 0,7) & t > 0,7. \end{cases}$$

На рис. 6 приведена достоверная оценка интенсивности отказов модельной задачи. Оттенками серого показаны плотности вероятности распределения второго порядка. На рис. 6 показана оценка функций плотности вероятности $P(t)$ в момент времени $t = 1$.

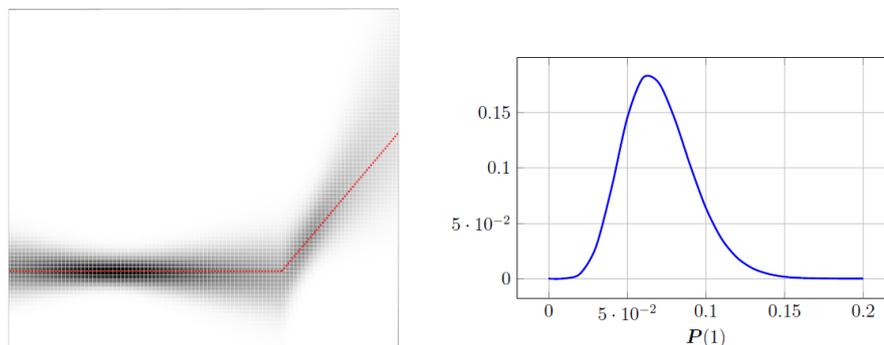


Рис. 6. Достоверная оценка интенсивности отказов и вероятности безотказной работы
Fig. 6. Reliable assessment of failure rate and probability of failure-free operation

Используя вероятностное расширение λ , можно вычислить значения оценки функции плотности вероятности $P(t)$ в любое время. Используя оценки функции плотности вероятности $P(t)$ в виде сплайнов, мы можем, например, оценить риск того, что вероятность $P(1) > 0,2$ или $P(1) < 0,05$.

Заключение

Применение кусочно-полиномиальных моделей распределений второго порядка позволяет строить надежные оценки эмпирических функций распределений. Рассмотренные примеры численных операций над распределениями второго порядка в задачах построения оценок надежности оборудования в условиях малых выборок подтверждают этот вывод. Дальнейшее использование распределений второго порядка может быть направлено на оценки рисков, принятие решений, стохастическое моделирование в условиях эпистемической неопределенности.

Список источников

1. Augustin T., Coolen F., Cooman G., Troffaes M. Introduction to Imprecise Probabilities. John Wiley & Sons, 2014. 448 p.
2. Ferson S., Kreinovich V., Ginzburg L., Myers D.S., Sents K. Constructing Probability Boxes and Dempster–Shafer Structures. Sandia National Laboratories, SAND2002-4015, 2003.
3. Swiler L.P., Giunta A.A. Aleatory and Epistemic Uncertainty Quantification for Engineering Applications. Sandia Technical Report, SAND2007-2670C, 2007.
4. Добронец Б.С., Попова О.А. Численный вероятностный анализ для исследования систем в условиях неопределенности // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2012. № 4 (21). С. 39–46.
5. Добронец Б.С., Попова О.А. Вычислительный вероятностный анализ: модели и методы. Красноярск : Сиб. федер. ун-т, Ин-т космических и информ. технологий, 2020. 236 с.
6. Попова О.А. Применение численного вероятностного анализа в задачах интерполяции // Вычислительные технологии. 2017. Т. 22, № 2. С. 99–114.
7. Sahlin N., Goldsmith R. The role of second-order probabilities in decision making // Advances in Psychology. 1983. V. 14. P. 455–467.
8. Utkin L. Augustin T. Decision Making with Imprecise Second-Order Probabilities // Proc. of the Third International Symposium on Imprecise Probabilities and Their Applications, Lugano, Switzerland, July 14–17, 2003. P. 547–561.
9. Ужга-Ребров О.И. Управление неопределенностями. Rēzekne, 2014. Ч. 4: Комбинирование неопределенностей. 408 с.
10. Dobronets B.S., Popova O.A. Computational Aspects of Probabilistic Extensions // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2019. № 47. С. 41–48.

References

1. Augustin, T., Coolen, F., Cooman, G. & Troffaes, M. (2014) *Introduction to Imprecise Probabilities*. John Wiley & Sons.
2. Ferson, S., Kreinovich, V., Ginzburg, L., Myers, D.S. & Sents, K. (2003) *Constructing Probability Boxes and Dempster–Shafer Structures*. Sandia National Laboratories, SAND2002-4015.

3. Swiler, L.P. & Giunta, A.A. (2007) *Aleatory and Epistemic Uncertainty Quantification for Engineering Applications*. Sandia Technical Report, SAND2007-2670C
4. Dobronets, B.S. & Popova, O.A. (2012) Numerical probabilistic analysis for the study of systems with uncertainty. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 4(21). pp. 39–46.
5. Dobronets, B.S. & Popova, O.A. (2020) *Vychislitel'nyy veroyatnostnyy analiz: modeli i metody* [Computational Probabilistic Analysis: Models and Methods]. Krasnoyarsk: Siberian Federal University, Institute of Space and Information Technologies.
6. Popova, O.A. (2017) Application of numerical probabilistic analysis in interpolation problems. *Vychislitel'nye tekhnologii – Computational Technologies*. 22(2). pp. 99–114.
7. Sahlin, N. & Goldsmith, R. (1983) The role of second-order probabilities in decision making. *Advances in Psychology*. 14. pp 455–467. DOI: 10.1016/S0166-4115(08)62250-5
8. Utkin, L. & Augustin, T. (2003) Decision Making with Imprecise Second-Order Probabilities. *Proceedings of the Third International Symposium on Imprecise Probabilities and Their Applications*. Lugano, Switzerland, July 14–17. pp. 547–561.
9. Uzhga-Rebrov, O. (2014) *Upravlenie neopredelennostyami* [Uncertainty Management]. Vol. 4. Rezekne: [s.n.].
10. Dobronets, B.S. & Popova, O.A. (2019) Computational aspects of probabilistic extensions. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 47. pp. 41–48. DOI: 10.17223/19988605/47/5

Информация об авторах:

Добронет Борис Станиславович – профессор, доктор физико-математических наук, профессор кафедры систем искусственного интеллекта Института космических и информационных технологий Сибирского федерального университета (Красноярск, Россия). E-mail: BDobronets@yandex.ru

Попова Ольга Аркадьевна – доцент, кандидат технических наук, доцент кафедры систем искусственного интеллекта Института космических и информационных технологий Сибирского федерального университета (Красноярск, Россия). E-mail: OlgaArc@yandex.ru

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

Information about the authors:

Dobronets Boris S. (Doctor of Physics and Mathematics, Professor, Siberian Federal University, Institute of Space and Information Technologies, Krasnoyarsk, Russian Federation). E-mail: BDobronets@yandex.ru

Popova Olga A. (Candidate of Technical Sciences, Associate Professor, Siberian Federal University, Institute of Space and Information Technologies, Krasnoyarsk, Russian Federation). E-mail: OlgaArc@yandex.ru

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

Поступила в редакцию 07.07.2021; принята к публикации 29.11.2022

Received 07.07.2021; accepted for publication 29.11.2022

Научная статья

УДК 519.21

doi: 10.17223/19988605/61/7

Оценка длительности мертвого времени в обобщенном МАР-потоке 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 МАР-потоке событий (Markovian Arrival Process) с двумя состояниями методом моментов в условиях непродлевающего мертвого времени фиксированной длительности. Приводятся явные виды плотностей вероятности значений длительности интервала между моментами наступления соседних событий потока в случае коррелированного и рекуррентного задания потоков. Формулируется алгоритм вычисления оценки периода ненаблюдаемости. Приводятся численные результаты расчетов оценок, полученных с использованием имитационной модели потока, и их анализ.

Ключевые слова: обобщенный МАР-поток событий; плотность вероятности длительности интервала между моментами наступления соседних событий; метод моментов; оценка непродлевающего мертвого времени фиксированной длительности; условия рекуррентности потока

Для цитирования: 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. doi: 10.17223/19988605/61/7

Original article

doi: 10.17223/19988605/61/7

Estimation of dead time duration in generalized MAP with two states

Anastasiya V. Keba¹, Ludmila A. Nezhel'skaya²

^{1, 2} Tomsk State University, Tomsk, Russian Federation

¹ mir.na.mig7@mail.ru

² ludne@mail.ru

Abstract. The paper deals with a generalized MAP with two states under conditions of partial observability (unextendable dead time of a fixed duration T). The accompanying random process $\lambda(t)$ of the flow is an unobservable piecewise constant process with n states: S_1, \dots, S_n . The process $\lambda(t)$ is unobservable in principle, the moments of occurrence of events t_1, t_2, \dots are observable.

We denote $\tau_k = t_{k+1} - t_k$, $\tau_k \geq 0$, is the value of the k th interval duration between events of the observed flow t_k and t_{k+1} , $k = 1, 2, \dots$. Due to the fact that the flow operates in a stationary mode, then for the probability density of values τ_k it is true $p_T(\tau_k) = p_T(\tau)$, $\tau_k \geq 0$, for any $k \geq 1$. This allows, without loss of generality, to set the moment of occurrence of the event t_k equal to zero or, which is the same, the moment the event occurs is $\tau = 0$.

It is necessary to use the method of moments to estimate the parameter T in generalized MAP with two states (we assume $n = 2$) at an arbitrary time instant τ , provided that the realization of the time intervals between events $\tau_1, \dots, \tau_{m-1}$ is known of the observed flow on the interval $(0, t)$, where $t = 0$ is the moment of occurrence of the event, τ is the end of observation, m is the number of observed events over the time interval $(0, t)$. It is assumed that $\lambda(t)$ is stationary. The estimation of the parameter T is carried out both in the correlated case of setting the flow under study, and in the recurrent one, the recurrence condition is determined from the analytical expression of the joint probability density $p_T(\tau_1, \tau_2)$ of the duration of two adjacent intervals between the moments of the occurrence of events.

In the paper we present an explicit form of the probability density parameters of the interval duration between events in generalized MAP with two states in the case of correlated and recurrent flows specification. The method of moments is used to find estimate of the duration of the unobservable period T . The algorithm were implemented by C# programming language in the Visual Studio 2015. Statistical experiments were conducted on the simulation model of the flow in order to establish the quality of the results obtained by evaluating the duration of the dead time T . The numerical results of these experiments are given in the paper and illustrate an acceptable estimate that does not contradict the physical interpretation.

Keywords: generalized MAP; probability density of the values of the interval duration between the moments of the events occurrence; method of moments; estimation of unextendible dead time of a fixed duration; flow recurrence conditions

For citation: Keba, A.V., Nezhel'skaya, L.A. (2022) Estimation of dead time duration 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*. 61. pp. 69–80. doi: 10.17223/19988605/61/7

Рассматривается обобщенный MAP-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; сопровождающий процесс потока есть кусочно-постоянный случайный процесс с конечным числом состояний [4–7].

В реальных ситуациях параметры, определяющие поток 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 следует учитывать возможные искажающие факторы, существенно влияющие на качество оценивания. Одним из таких факторов является мертвое время регистрирующих приборов [16–18], порождаемое каждым зарегистрированным событием: последующие события исходного потока, наступившие в течение периода мертвого времени, недоступны для наблюдения. Такой эффект характерен для большинства реальных систем. Предполагается, что период ненаблюдаемости потока имеет фиксированную длительность T [19].

В данной работе осуществляется исследование обобщенного MAP-потока событий с двумя состояниями в условиях частичной наблюдаемости. В общем случае дважды стохастические потоки событий, в том числе и рассматриваемый обобщенный MAP-поток, являются коррелированными потоками; в работе на основании явного вида совместной плотности вероятности двух смежных интервалов между моментами наступления событий выписываются условия рекуррентности рассматриваемого потока. Предлагается явный вид плотности вероятности значений длительности интервала между моментами наступления событий потока в случае как коррелированного, так и рекуррентного задания потоков. Приводятся численные результаты оценивания периода ненаблюдаемости соседних событий исследуемого потока. Данная статья является непосредственным развитием работ [20–22].

1. Постановка задачи

Исследуется обобщенный MAP-поток событий с произвольным числом состояний (далее – поток), функционирующий в установившемся (стационарном) режиме. Сопровождающий случайный процесс $\lambda(t)$ изучаемого потока представляет собой кусочно-постоянный принципиально ненаблюдаемый процесс с n состояниями: S_1, \dots, S_n . Полагается, что при $\lambda(t) = \lambda_i$ имеет место i -е состояние (S_i), $i = \overline{1, n}$, процесса $\lambda(t)$. При этом $\lambda_1 > \lambda_2 > \dots > \lambda_n > 0$. 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: $F_i(t) = 1 - e^{-\lambda_i t}$, $t \geq 0$, $i = \overline{1, n}$. 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, $i, j = \overline{1, n}$. Отметим, что для введенных вероятностей справедливо $\sum_{j=1}^n P_0(\lambda_j|\lambda_i) + \sum_{j=1}^n P_1(\lambda_j|\lambda_i) = 1, i = \overline{1, n}$.

Блочная матрица инфинитезимальных характеристик [19] процесса $\lambda(t)$ имеет вид $\mathbf{D} = \|\|\mathbf{D}_0|\mathbf{D}_1\|\|$:

$$\mathbf{D}_0 = \left\| \begin{array}{cccc} -\lambda_1(1 - P_0(\lambda_1|\lambda_1)) & \lambda_1 P_0(\lambda_2|\lambda_1) & \dots & \lambda_1 P_0(\lambda_n|\lambda_1) \\ \lambda_2 P_0(\lambda_1|\lambda_2) & -\lambda_2(1 - P_0(\lambda_2|\lambda_2)) & \dots & \lambda_2 P_0(\lambda_n|\lambda_2) \\ \dots & \dots & \dots & \dots \\ \lambda_n P_0(\lambda_1|\lambda_n) & \lambda_n P_0(\lambda_2|\lambda_n) & \dots & -\lambda_n(1 - P_0(\lambda_n|\lambda_n)) \end{array} \right\|_{n \times n},$$

$$\mathbf{D}_1 = \left\| \begin{array}{cccc} \lambda_1 P_1(\lambda_1|\lambda_1) & \lambda_1 P_1(\lambda_2|\lambda_1) & \dots & \lambda_1 P_1(\lambda_n|\lambda_1) \\ \lambda_2 P_1(\lambda_1|\lambda_2) & \lambda_2 P_1(\lambda_2|\lambda_2) & \dots & \lambda_2 P_1(\lambda_n|\lambda_2) \\ \dots & \dots & \dots & \dots \\ \lambda_n P_1(\lambda_1|\lambda_n) & \lambda_n P_1(\lambda_2|\lambda_n) & \dots & \lambda_n P_1(\lambda_n|\lambda_n) \end{array} \right\|_{n \times n}.$$

Каждое зарегистрированное событие потока создает период ненаблюдаемости фиксированной длительности (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 $S_i, i = \overline{1, n}$; 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.

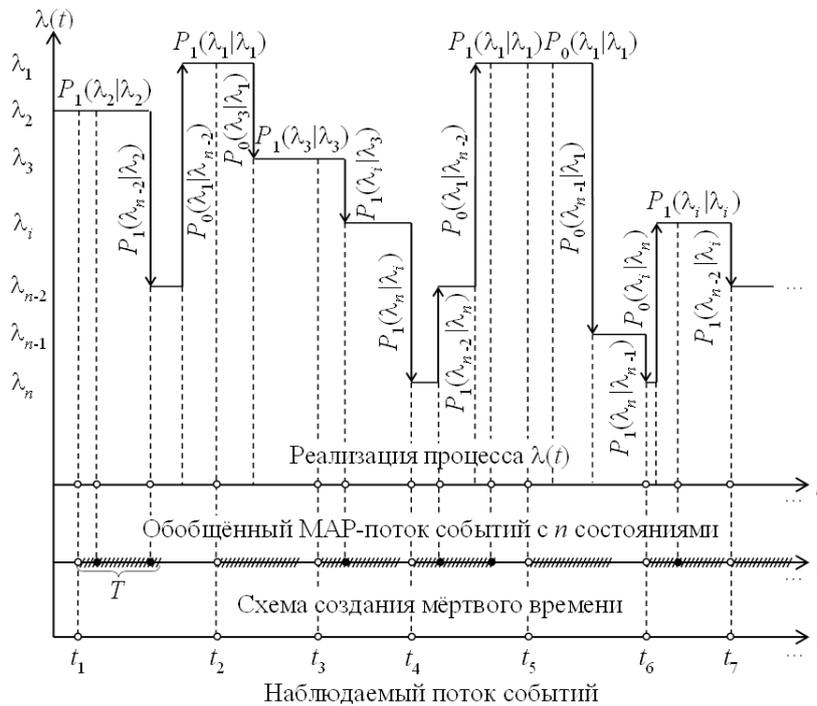


Рис. 1. Реализация обобщенного МАР-потока событий с произвольным числом состояний в условиях непродлевающегося мертвого времени длительности T

Fig. 1. Realization of Generalized MAP with an arbitrary number of states under conditions of unextendible dead time of duration T

Замечание 1. Введение вероятности $P_0(\lambda_i|\lambda_i) \neq 0$, $i = \overline{1, n}$, перехода процесса $\lambda(t)$ из состояния S_i в состояние S_i без наступления события приводит к обобщению классического МАР-потока с произвольным числом состояний.

Утверждение 1. Для обобщенного МАР-потока событий с n состояниями процесс $\lambda(t)$ является скрытым марковским процессом.

Цель работы заключается в решении задачи оценивания периода ненаблюдаемости для случая как коррелированного, так и для рекуррентного задания обобщенного МАР-потока событий с двумя состояниями и в анализе полученных оценок. Оценка находится методом моментов 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 коррелированном 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 $\tau = 0$.

Утверждение 2. Последовательность t_1, t_2, \dots порождает вложенную по моментам времени наступления событий цепь Маркова $\{\lambda(t_k)\}$.

Теорема 1. В обобщенном МАР-потоке, функционирующем в условиях непродлевающегося мертвого времени, плотность вероятности $p_T(\tau)$ длительности интервала между соседними событиями имеет вид:

$$p_T(\tau) = \begin{cases} 0, & 0 \leq \tau < T, \\ \gamma(T)z_1 e^{-z_1(\tau-T)} + (1-\gamma(T))z_2 e^{-z_2(\tau-T)}, & \tau \geq T, \end{cases}$$

$$\gamma(T) = \frac{z_2 - \lambda_1 \pi_1(T) [P_1(\lambda_1|\lambda_1) + P_1(\lambda_2|\lambda_1)] - \lambda_2 \pi_2(T) [P_1(\lambda_1|\lambda_2) + P_1(\lambda_2|\lambda_2)]}{z_2 - z_1},$$

$$\pi_i(T) = \pi_i - [\pi_i - \pi_i(0|T)] e^{-aT}, \quad i = 1, 2, \tag{1}$$

$$z_{1,2} = \frac{\lambda_1 (1 - P_0(\lambda_1|\lambda_1)) + \lambda_2 (1 - P_0(\lambda_2|\lambda_2)) \mp \sqrt{D}}{2},$$

$$D = [\lambda_1 (1 - P_0(\lambda_1|\lambda_1)) - \lambda_2 (1 - P_0(\lambda_2|\lambda_2))]^2 + 4\lambda_1 \lambda_2 P_0(\lambda_2|\lambda_1) P_0(\lambda_1|\lambda_2),$$

$$0 < z_1 < z_2,$$

где условные стационарные вероятности $\pi_i(0|T)$ того, что процесс $\lambda(\tau)$ в момент времени $\tau = 0$ находится в состоянии S_i , $i = 1, 2$, при условии, что в данный момент времени наступило событие наблюдаемого потока, породив период мертвого времени длительности T , определяются выражениями

$$\pi_1(0|T) = \frac{p_{21} + \pi_1(p_{11} - p_{21})(1 - e^{-aT})}{p_{12} + p_{21} + (p_{22} - p_{12})(1 - e^{-aT})}, \quad \pi_2(0|T) = \frac{p_{12} + \pi_2(p_{22} - p_{12})(1 - e^{-aT})}{p_{12} + p_{21} + (p_{22} - p_{12})(1 - e^{-aT})},$$

вероятности перехода p_{ij} , $i, j = 1, 2$, представляются в виде

$$p_{ii} = \frac{P_1(\lambda_i|\lambda_1)(1 - P_0(\lambda_2|\lambda_2)) + P_0(\lambda_2|\lambda_1)P_1(\lambda_i|\lambda_2)}{(1 - P_0(\lambda_1|\lambda_1))(1 - P_0(\lambda_2|\lambda_2)) - P_0(\lambda_1|\lambda_2)P_0(\lambda_2|\lambda_1)}, \tag{2}$$

$$P_{2i} = \frac{P_1(\lambda_i|\lambda_2)(1 - P_0(\lambda_1|\lambda_1)) + P_0(\lambda_1|\lambda_2)P_1(\lambda_i|\lambda_1)}{(1 - P_0(\lambda_1|\lambda_1))(1 - P_0(\lambda_2|\lambda_2)) - P_0(\lambda_1|\lambda_2)P_0(\lambda_2|\lambda_1)},$$

$$\sum_{j=1}^2 p_{ij} = 1, \quad i=1,2;$$

априорные финальные вероятности π_i состояний S_i , $i=1, 2$, и величина a равны соответственно

$$\pi_1 = \frac{\lambda_2(P_0(\lambda_1|\lambda_2) + P_1(\lambda_1|\lambda_2))}{a}, \quad \pi_2 = \frac{\lambda_1(P_0(\lambda_2|\lambda_1) + P_1(\lambda_2|\lambda_1))}{a}, \quad \sum_{i=1}^2 \pi_i = 1, \quad (3)$$

$$a = \lambda_1(P_0(\lambda_2|\lambda_1) + P_1(\lambda_2|\lambda_1)) + \lambda_2(P_0(\lambda_1|\lambda_2) + P_1(\lambda_1|\lambda_2)).$$

Доказательство основано на применении свойств вложенной цепи Маркова и здесь не приводится ввиду громоздкости выкладок.

3. Вид плотности вероятности длительности интервала между соседними событиями в рекуррентном обобщенном МАР-потоке при неполной наблюдаемости

Пусть (t_k, t_{k+1}) , (t_{k+1}, t_{k+2}) – два смежных интервала, длительности которых есть $\tau_k = t_{k+1} - t_k$ и $\tau_{k+1} = t_{k+2} - t_{k+1}$ соответственно, $\tau_k, \tau_{k+1} \geq 0$; их расположение на временной оси в силу стационарности потока произвольно. Тогда, полагая $k=1$, будем рассматривать два соседних интервала (t_1, t_2) , (t_2, t_3) с соответствующими значениями длительностей $\tau_1 = t_2 - t_1$ и $\tau_2 = t_3 - t_2$, $\tau_1, \tau_2 \geq 0$. При этом $\tau_1 = 0$ соответствует моменту t_1 наступления события наблюдаемого потока, а $\tau_2 = 0$ соответствует моменту t_2 наступления следующего события потока. С учетом функционирования потока в условиях непродлевающегося мертвого времени длительности двух смежных интервалов есть $\tau_1 = T + t^{(1)}$, $\tau_2 = T + t^{(2)}$, где $t^{(l)}$ – значение длительности интервала между соответствующими моментом окончания мертвого времени и моментом наступления следующего события наблюдаемого потока, $t^{(l)} > 0$, $l=1,2$. Совместная плотность вероятности при этом есть $p_T(\tau_1, \tau_2)$, $\tau_1, \tau_2 \geq 0$.

Теорема 2. В обобщенном МАР-потоке в условиях неполной наблюдаемости (при наличии непродлевающегося мертвого времени фиксированной длительности) совместная плотность вероятности $p_T(\tau_1, \tau_2)$ двух соседних интервалов имеет вид:

$$p_T(\tau_1, \tau_2) = \begin{cases} 0, & 0 \leq \tau_1 < T, \tau_2 \geq 0, \\ 0, & \tau_1 \geq 0, 0 \leq \tau_2 < T, \\ p_T(\tau_1)p_T(\tau_2) + \frac{\lambda_1\lambda_2}{z_1z_2}\gamma(T)(1 - \gamma(T))e^{-aT} \times \\ \times [P_1(\lambda_1|\lambda_1)P_1(\lambda_2|\lambda_2) - P_1(\lambda_1|\lambda_2)P_1(\lambda_2|\lambda_1)] \times & \tau_1, \tau_2 \geq T, \\ \times \left(z_1 e^{-z_1(t^{(1)}-T)} - z_2 e^{-z_2(t^{(1)}-T)} \right) \left(z_1 e^{-z_1(t^{(2)}-T)} - z_2 e^{-z_2(t^{(2)}-T)} \right), \end{cases} \quad (4)$$

где $\gamma(T)$, z_1, z_2 , $p_T(\tau_k)$, $k=1,2$, (при $\tau = \tau_k$) определены в (1); величина a определена в (3).

Доказательство здесь не приводится вследствие громоздкости выкладок.

На основании полученной аналитической формулы (4) совместной плотности вероятности $p_T(\tau_1, \tau_2)$ выпишем 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 (1) преобразуется к виду:

$$p_T(\tau) = \begin{cases} 0, & 0 \leq \tau < T, \\ \gamma(T)z_1 e^{-z_1(\tau-T)} + (1-\gamma(T))z_2 e^{-z_2(\tau-T)}, & \tau \geq T, \end{cases}$$

$$\gamma(T) = \frac{z_2 - \lambda_1 \pi_1(T) [P_1(\lambda_1 | \lambda_1) + P_1(\lambda_2 | \lambda_1)] - \lambda_2 \pi_2(T) [P_1(\lambda_1 | \lambda_2) + P_1(\lambda_2 | \lambda_2)]}{z_2 - z_1}, \quad (5)$$

$$\pi_i(T) = \pi_i - [\pi_i - p_{ji}] e^{-aT}, \quad i, j = 1, 2, \quad i \neq j,$$

где z_1, z_2 определены в (1), вероятности перехода $p_{ij}, i, j = 1, 2, -$ в (2), величины $\pi_i, i = 1, 2, a -$ в (3).

Два другие условия рекуррентности, вытекающие из выражения $\gamma(T)(1-\gamma(T)) = 0$, далее не рассматриваются, так как при их выполнении обобщенный МАР-поток событий становится простейшим.

4. Уравнения моментов и результаты численных расчетов

Для оценки параметра потока – длительности мертвого времени – используется метод моментов. Введем в рассмотрение выборку $\tau_1, \dots, \tau_{m-1}$ из распределения $p_T(\tau) = \gamma(T)z_1 e^{-z_1(\tau-T)} + (1-\gamma(T))z_2 e^{-z_2(\tau-T)}$, $\tau \geq 0$, 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 обобщенном МАР-потоке событий. Таким образом, для оценки параметра T необходимо иметь уравнение моментов вида:

$$M_T \tau = C_1. \quad (6)$$

Выполняя интегрирование левой части (6), находим

$$M_T \tau = \int_T^\infty \tau (\gamma(T)z_1 e^{-z_1(\tau-T)} + (1-\gamma(T))z_2 e^{-z_2(\tau-T)}) d\tau = T + \frac{\gamma(T)}{z_1} + \frac{1-\gamma(T)}{z_2} = C_1,$$

где $\gamma(T)$ определена в (1) для коррелированного обобщенного МАР-потока, в (5) – для рекуррентного обобщенного МАР-потока.

Подчеркнем, что решение уравнений моментов возможно только численно; в данной работе решения уравнений моментов находятся с применением метода простой итерации.

Отметим, что при выбранных параметрах коррелированного либо рекуррентного потоков, как показывают результаты численных экспериментов, приведенные ниже, математическое ожидание $M_T \tau$ есть возрастающая функция переменной (параметра) $T, T > 0$. Вследствие этого каждое из уравнений моментов имеет единственное решение.

С целью получения численных результатов разработан алгоритм вычисления оценки \hat{T} (решение уравнения моментов (6)) и построена имитационная модель обобщенного МАР-потока событий с двумя состояниями в условиях непродлевающегося мертвого времени. Основой имитационной модели являются датчики псевдослучайных чисел и метод обратных функций [23]. На первом этапе расчета осуществляется имитационное моделирование обобщенного МАР-потока события с двумя состояниями с тем, чтобы получить статистику C_1 на основе выборки моментов времени наступления событий t_1, \dots, t_m . На втором этапе расчета вычисляется оценка \hat{T} .

Для каждой реализации найдены оценки $\hat{T}^{(v)}, v = \overline{1, N}$, на основании которых вычислены выборочное среднее $\hat{M}(\hat{T}) = \frac{1}{N} \sum_{v=1}^N \hat{T}^{(v)}$ и выборочная вариация $\hat{V}(\hat{T}) = \frac{1}{N} \sum_{v=1}^N (\hat{T}^{(v)} - \hat{T})^2$.

В первом и втором экспериментах будем рассматривать случай коррелированного задания потока, а в эксперименте 3 исследуем влияние времени моделирования на оценку периода ненаблюдаемости в случае рекуррентного задания потока; условие рекуррентности при этом определено как $P_1(\lambda_1|\lambda_1)P_1(\lambda_2|\lambda_2) - P_1(\lambda_1|\lambda_2)P_1(\lambda_2|\lambda_1) = 0$.

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.

Целью данного эксперимента является нахождение временного интервала установления стационарного режима функционирования обобщенного МАР-потока события с двумя состояниями в условиях непродлевающегося мертвого времени фиксированной длительности в случае задания параметров для коррелированного потока или, иными словами, установление зависимостей $\hat{M}(\hat{T})$, $\hat{V}(\hat{T})$ от длительности времени моделирования T_m , 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 коррелированного обобщенного МАР-потока событий с двумя состояниями в рамках 1-го и 2-го статистических экспериментов

$P_0(\lambda_1 \lambda_1) = 0,05$	$P_0(\lambda_1 \lambda_2) = 0,05$
$P_0(\lambda_2 \lambda_1) = 0,05$	$P_0(\lambda_2 \lambda_2) = 0,05$
$P_1(\lambda_1 \lambda_1) = 0,15$	$P_1(\lambda_1 \lambda_2) = 0,75$
$P_1(\lambda_2 \lambda_1) = 0,75$	$P_1(\lambda_2 \lambda_2) = 0,15$

Таблица 2

Результаты 1-го статистического эксперимента

T_m	100	200	300	400	500
$\hat{M}(\hat{T})$	0,9445	0,9497	0,9515	0,9523	0,9529
$\hat{V}(\hat{T}) \times 10^4$	29,8556	24,3601	22,6321	21,8365	21,3135
T_m	600	700	800	900	1 000
$\hat{M}(\hat{T})$	0,9531	0,9534	0,9536	0,9538	0,9539
$\hat{V}(\hat{T}) \times 10^4$	21,0447	20,7491	20,6081	20,4755	20,3541
T_m	1 100	1 200	1 300	1 400	1 500
$\hat{M}(\hat{T})$	0,9540	0,9541	0,9541	0,9542	0,9542
$\hat{V}(\hat{T}) \times 10^4$	20,2591	20,1868	20,1356	20,0653	20,0202

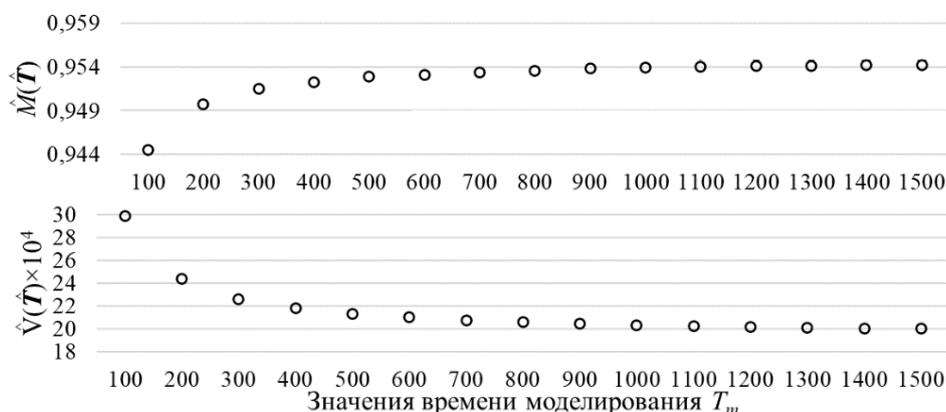


Рис. 2. График зависимости $\hat{M}(\hat{T})$, $\hat{V}(\hat{T}) \times 10^4$ от значения T_m

Fig. 2. $\hat{M}(\hat{T})$, $\hat{V}(\hat{T}) \times 10^4$ as functions of T_m

Из анализа численных результатов, приведенных на рис. 2, следует, что имеет место смещение оценок относительно исходных значений оцениваемых параметров, однако вариация стремится к близкому к нулю числу, т.е. методика оценивания качественна, и полученную оценку можно принимать за истинную с достаточно малой погрешностью. Анализ представленных в табл. 2 результатов эксперимента приводит к утверждению о зависимости получаемых оценок от времени моделирования, а именно: с увеличением значения T_m выборочные средние и выборочная вариация стабилизируются при $T_m \geq 500$, что объясняется концепцией метода моментов.

Эксперимент 2. Фиксируются число состояний $n = 2$, количество опытов $N = 1\,000$, значение времени моделирования $T_m = 1\,000$, значения $\lambda_1 = 25$, $\lambda_2 = 10$ процесса $\lambda(t)$ и переходные вероятности, представленные в табл. 1.

В рамках эксперимента исследуются зависимости $\hat{M}(T)$, $\hat{V}(T)$ от изменения длительности мертвого времени T , где $T \in \{0,5; 0,6; \dots; 1,9\}$.

Результаты второго статистического эксперимента представлены в табл. 3 и на рис. 3.

Результаты табл. 3 указывают на то, что увеличение длительности мертвого времени T влечет увеличение выборочной вариации оценки. Это объясняется тем, что при увеличении значения T увеличивается период ненаблюдаемости потока, что приводит к увеличению числа потерянных событий и, как следствие, к ухудшению качества оценивания. Оценки также получаются смещенными, о чем свидетельствуют полученные результаты для $\hat{M}(T)$.

Таблица 3

Результаты 2-го статистического эксперимента

T	0,5	0,6	0,7	0,8	0,9
$\hat{M}(T)$	0,4551	0,5547	0,6546	0,7542	0,8538
$\hat{V}(T) \times 10^4$	19,7215	20,0124	19,9936	20,2736	20,5286
T	1,0	1,1	1,2	1,3	1,4
$\hat{M}(T)$	0,9539	1,0533	1,1533	1,2527	1,3523
$\hat{V}(T) \times 10^4$	20,3396	20,7526	20,7149	21,1622	21,4898
T	1,5	1,6	1,7	1,8	1,9
$\hat{M}(T)$	1,4521	1,5515	1,6507	1,7508	1,8505
$\hat{V}(T) \times 10^4$	21,6761	21,9654	22,6457	22,4818	22,7099

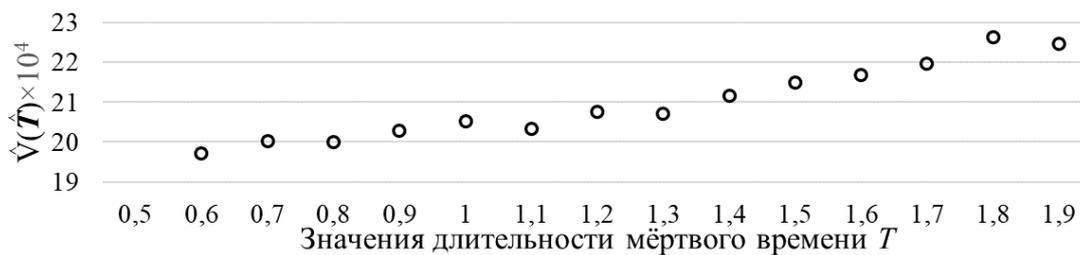


Рис. 3. График зависимости $\hat{V}(T) \times 10^4$ от значения T

Fig. 3. $\hat{V}(T) \times 10^4$ as functions of T

Замечание 2. В случае коррелированного задания обобщенного МАР-потока событий говорить о состоятельности получаемых методом моментов оценок не представляется возможным [24, 25].

Эксперимент 3. Фиксируются число состояний $n = 2$, количество опытов $N = 1\,000$, длительность мертвого времени $T = 1$, значения $\lambda_1 = 25$, $\lambda_2 = 10$ процесса $\lambda(t)$ и переходные вероятности, представленные в табл. 4. Значения длительности моделирования T_m изменяются в диапазоне $T_m \in \{100; 200; \dots; 1\,500\}$.

Результаты третьего статистического эксперимента представлены в табл. 5 и на рис. 4.

Таблица 4

Значения переходных вероятностей рекуррентного обобщенного МАР-потока событий с двумя состояниями в рамках 3-го статистического эксперимента

$P_0(\lambda_1 \lambda_1) = 0,05$	$P_0(\lambda_1 \lambda_2) = 0,05$
$P_0(\lambda_2 \lambda_1) = 0,05$	$P_0(\lambda_2 \lambda_2) = 0,05$
$P_1(\lambda_1 \lambda_1) = 0,45$	$P_1(\lambda_1 \lambda_2) = 0,45$
$P_1(\lambda_2 \lambda_1) = 0,45$	$P_1(\lambda_2 \lambda_2) = 0,45$

Таблица 5

Результаты 3-го статистического эксперимента

T_m	100	200	300	400	500
$\hat{M}(\hat{T})$	0,9451	0,9504	0,9521	0,9529	0,9535
$\hat{V}(\hat{T}) \times 10^4$	29,1964	23,7264	22,0135	21,2357	20,7416
T_m	600	700	800	900	1 000
$\hat{M}(\hat{T})$	0,9538	0,9541	0,9543	0,9544	0,9545
$\hat{V}(\hat{T}) \times 10^4$	20,3976	20,1648	19,9875	19,8844	19,7622
T_m	1100	1200	1300	1400	1500
$\hat{M}(\hat{T})$	0,9546	0,9547	0,9548	0,9548	0,9549
$\hat{V}(\hat{T}) \times 10^4$	19,6881	19,5981	19,5555	19,4938	19,4513

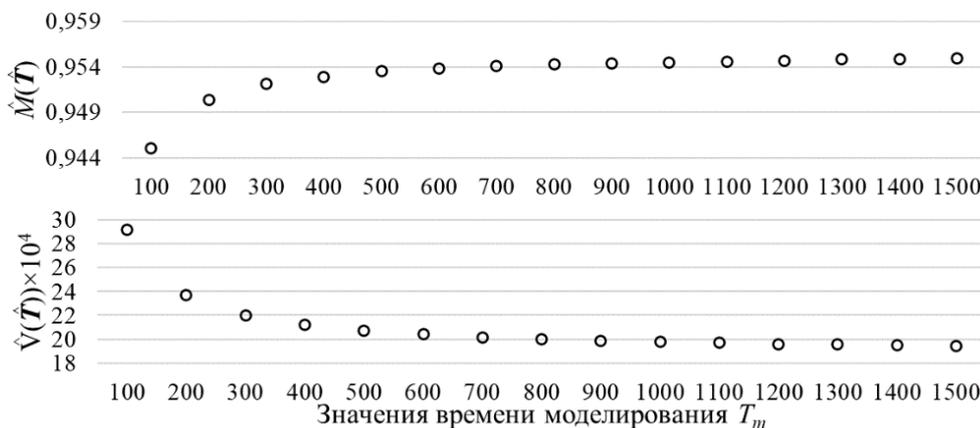


Рис. 4. График зависимости $\hat{M}(\hat{T})$, $\hat{V}(\hat{T}) \times 10^4$ от значения T_m

Fig. 4. $\hat{M}(\hat{T})$, $\hat{V}(\hat{T}) \times 10^4$ as functions of T_m

Из графика, приведенного на рис. 4, следует, что стационарный режим функционирования наблюдаемого потока устанавливается при $T_m \geq 500$. Полученные оценки параметра T являются смещенными.

Замечание 3. В случае рекуррентного задания обобщенного МАР-потока событий оценки 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 (6) имеет единственное решение [25].

Заключение

В данной статье рассмотрен коррелированный и рекуррентный обобщенный МАР-поток событий с двумя состояниями в условиях непродлевающегося мертвого времени фиксированной длительности T . Приведены явные виды плотностей вероятности значений длительности интервала между

соседними событиями обобщенного MAP-потока событий для коррелированного (1) и рекуррентного (5) случаев задания потоков. Методом моментов найдена оценка параметра \hat{T} . Алгоритм вычисления оценки реализован на языке программирования C# в среде Visual Studio 2015. Проведены статистические эксперименты, указывающие на работоспособность алгоритма оценки параметра T и приемлемое качество получаемых оценок.

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. Вишневецкий 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. Леонова М.А., Нежелская Л.А. Вероятность ошибки при оценивании состояний обобщенного асинхронного потока событий // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2012. № 2 (19). С. 88–101.
9. Нежелская Л.А. Оптимальное оценивание состояний полусинхронного потока событий в условиях его частичной наблюдаемости // Вестник Томского государственного университета. 2000. № 269. С. 95–98.
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. Gortsev A.M., Klimov I.S. Estimation of the parameters of an alternating Poisson stream of events // Telecommunications and Radio Engineering. 1993. V. 48 (10). P. 40–45.
15. Gortsev A.M., Klimov I.S. Estimation of intensity of Poisson stream of events for conditions under which it is partially unobservable // Telecommunications and Radio Engineering. 1992. V. 47 (1). P. 33–38.
16. Normey-Rico J.E. Control of dead-time processes // Advanced textbooks in control and signal processing. London : Springer-Verlag, 2007. 462 p.
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. Горцев А.М., Ниссенбаум О.В. Оценивание длительности мертвого времени и параметров асинхронного альтернирующего потока событий с инициированием лишнего события // Вестник Томского государственного университета. 2004. № 284. С. 137–145.
20. Кеба А.В., Нежелская Л.А. Статистические эксперименты на имитационной модели обобщенного MAP-потока событий с произвольным числом состояний в условиях непродлевающегося мертвого времени // Новые информационные технологии в исследовании сложных структур : материалы XIII Междунар. конф., 7 – 9 сентября 2020. Томск : Изд. Дом Том. гос. ун-та, 2020. С. 85–86.
21. Кеба А.В., Нежелская Л.А. Оптимальное оценивание состояний обобщенного MAP-потока событий с произвольным числом состояний в условиях непродлевающегося мертвого времени // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2021. № 56. С. 68–80.
22. Кеба А.В., Нежелская Л.А. Совместная плотность вероятности длительностей интервалов обобщенного MAP-потока событий при его неполной наблюдаемости и условия рекуррентности потока // Труды Томского государственного университета. Сер. физико-математическая. 2022. Т. 307. С. 27–33.
23. Соболев И.М. Численные методы Монте-Карло. М. : Наука, 1973. 312 с.
24. Шуленин В.П. Математическая статистика. Томск : Изд-во НТЛ, 2012. Ч. 1. 540 с.
25. 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. [On the equivalent substitutions method for computing fragments of communication networks]. *Izvestiya AN SSSR. Tekhnicheskaya kibernetika*. 6. pp. 92–99.
5. Basharin, G.P., Kokotushkin, V.A. & Naumov, V.A. (1980) O metode ekvivalentnykh zamen rascheta fragmentov setey svyazi. [On the equivalent substitutions method for computing fragments of communication networks]. *Izvestiya AN SSSR. Tekhnicheskaya kibernetika*. 1. pp. 55–61.
6. Neuts, M.F. (1979) A versatile Markov point process. *Journal of Applied Probability*. 16. pp. 764–779. 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. Leonova, M.A. & Nezhelskaya, L.A. (2012) The probability of wrong decisions in the estimation of states of 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(19). pp. 88–101.
9. Nezhelskaya, L.A. (2000) Optimal state estimation of the semi-synchronous event flow under conditions of its partial observability. *Vestnik Tomskogo gosudarstvennogo universiteta – Tomsk State University Journal*. 269. pp. 95–98.
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. Gortsev, A.M. & Klimov, I.S. (1993) Estimation of the parameters of an alternating Poisson stream of events. *Telecommunications and Radio Engineering*. 48(10). pp. 40–45.
15. Gortsev, A.M. & Klimov, I.S. (1992) Estimation of intensity of Poisson stream of events for conditions under which it is partially unobservable. *Telecommunications and Radio Engineering*. 47(1). pp. 33–38.
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. Translated from English. Moscow: Mir.
19. Gortsev, A.M. & Nissenbaum, O.V. (2004) Estimating dead time and parameters of an asynchronous elternating event flow with triggering an extra event. *Vestnik Tomskogo gosudarstvennogo universiteta – Tomsk State University Journal*. 284. pp. 137–145.
20. Keба, A.V. & Nezhelskaya, L.A. (2020) Statisticheskie eksperimenty na imitatsionnoy modeli obobshchennogo MAP-potoka sobytii s proizvol'nym chislom sostoyaniy v usloviyakh neprodlevayushchegosya mertvogo vremeni [Statistical experiments on a simulation model of a generalized MAP with an arbitrary number of states under conditions of unextendable dead time]. *Novye informatsionnye tekhnologii v issledovanii slozhnykh struktur* [New information technologies in the study of complex structures]. Proc. of the 13th International Conference. Tomsk: Tomsk State University. pp. 85–86.
21. Keба, A.V. & Nezhelskaya, L.A. (2021) Optimal state estimation of generalized MAP with an arbitrary number of states under conditions of unextendable dead time. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 56. pp. 68–80. DOI: 10.17223/19988605/56/8
22. Keба, 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]. In: Smyrin, I.S. (ed.) *Mathematical and software informational, technical and economic systems*. Vol. 307. Tomsk: Tomsk state University. pp. 27–33.
23. Sobol, I.M. (1973) *Chislennyye metody Monte-Karlo* [Numerical Methods of Monte Carlo]. Moscow: Nauka.
24. Shulenin, V.P. (2012) *Matematicheskaya statistika* [Mathematical Statistics]. Vol. 1. Tomsk: NTL.
25. Malinkovsky, Yu.V. (2004) *Teoriya veroyatnostey i matematicheskaya statistika* [Probability Theory and Mathematical Statistics]. Gomel: Francisk Skorina Gomel State University.

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, профессор кафедры 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

Nezhe'skaya Lyudmila A. (Doctor of Physical and Mathematical Sciences, Professor 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 05.07.2022; принята к публикации 29.11.2022

Received 05.07.2022; accepted for publication 29.11.2022

Original article

doi: 10.17223/19988605/61/8

The sensitivity coefficients for dynamic systems described by nonlinear difference generalized equations with the distributed memory and characterized by generalized functionals

Anatoly I. Rouban

Siberian Federal University, Krasnoyarsk, Russian Federation, ai-rouban@mail.ru

Abstract. The variational method of calculation of sensitivity coefficients connecting first variation of quality functional with variations of variable and constant parameters for multivariate non-linear dynamic systems described by generalized difference equations with the distributed memory on phase coordinates and variable parameters is developed. The nonlinear quality functional also has a generalized form. Sensitivity coefficients are components of sensitivity functional and they are before variations of variable and constant parameters. The base of sensitivity coefficients calculation are the decision of generalized equations of the object model in the forward direction of discrete time and corresponding difference conjugate equations for Lagrange's multipliers in the opposite direction of discrete time.

Keywords: variational method; sensitivity coefficient; generalized difference equation; distributed memory; conjugate equation; Lagrange's multiplier

For citation: Rouban, A.I. (2022) The sensitivity coefficients for dynamic systems described by nonlinear difference generalized equations with the distributed memory and characterized by generalized functionals. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naja tehnik i informatika – Tomsk State University Journal of Control and Computer Science*. 61. pp. 81–87. doi: 10.17223/19988605/61/8

Научная статья

УДК 62–50

doi: 10.17223/19988605/61/8

Коэффициенты чувствительности для динамических систем, описываемых нелинейными разностными обобщенными уравнениями с распределенной памятью и характеризующихся обобщенными функционалами

Анатолий Иванович Рубан

Сибирский Федеральный университет, Красноярск, Россия, ai-rouban@mail.ru

Аннотация. Вариационный метод применен для расчета коэффициентов чувствительности, которые связывают первую вариацию функционала качества работы системы с вариациями переменных и постоянных параметров, для многомерных нелинейных динамических систем, описываемых нелинейными разностными обобщенными уравнениями с распределенной памятью по фазовым координатам и переменным параметрам. Показатели качества работы систем также являются обобщенными нелинейными функционалами.

Ключевые слова: вариационный метод; коэффициент чувствительности; разностное уравнение; распределенная память; функционал качества работы системы; сопряженное уравнение; множитель Лагранжа

Для цитирования: Рубан А.И. Коэффициенты чувствительности для динамических систем, описываемых нелинейными разностными обобщенными уравнениями с распределенной памятью и характеризующихся обобщенными функционалами // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2022. № 61. С. 81–87. doi: 10.17223/19988605/61/8

Digital approaches to creation of new methods and algorithms of identification and control of dynamic systems lead to necessity of use discrete in time dynamic equations [1]. For dynamic systems the main problem at the analysis and syntheses methods of identification, control, optimization, stability [2–10] is the calculation of sensitivity functions and sensitivity coefficients (SC). The first-order sensitivity characteristics mostly are used. Later on we shall examine only SC of the first-order.

The SC are components of vector gradient from quality functional according to parameters and the sensitivity functional connects the first variation of quality functional with variations of variable and constant parameters.

Consider a vector output $y(t)$ of dynamic object model under discrete time $t \in [0, 1, \dots, N + 1]$ implicitly depending on vector constant α parameters and generalized $I(\alpha)$ functional constructed on a basis of $y(t)$ under $t \in [0, 1, \dots, N + 1]$ and on α parameters:

$$I(\alpha) = f_0(y(N + 1), y(N), \dots, y(1), y(0), \alpha) \equiv f_0(\cdot, \alpha).$$

SC with respect to constant α parameters are called a gradient from $I(\alpha)$ on α : $(dI(\alpha)/d\alpha)^T \equiv \nabla_\alpha I(\alpha)$. SC are a coefficients of single-line relationship between the first variation of functional $\delta_\alpha I(\alpha)$ and the variations $d\alpha$ of constant parameters α :

$$\delta_\alpha I(\alpha) = (\nabla_\alpha I(\alpha))^T d\alpha = \frac{dI(\alpha)}{d\alpha} d\alpha \equiv \sum_{j=1}^m \frac{\partial I(\alpha)}{\partial \alpha_j} d\alpha_j.$$

The direct method of SC calculation inevitably requires a solution of cumbersome sensitivity equations to sensitivity functions $W(t)$. For instance, for functional $I(\alpha)$ we have following SC vector (row vector):

$$\frac{dI(\alpha)}{d\alpha} = \sum_{t=0}^{N+1} \frac{\partial f_0(\cdot, \alpha)}{\partial y(t)} W(t) + \frac{\partial f_0(\cdot, \alpha)}{\partial \alpha}.$$

Here $W(t)$ is the matrix of single-line relationship of the first variation of dynamic model output with parameter variations: $\delta y(t) = W(t)d\alpha$. For obtaining the matrix $W(t)$ it is necessary to decide bulky system equations – sensitivity equations. The j -th column of matrix $W(t)$ is made of the sensitivity functions $\partial y(t)/\partial \alpha_j$ with respect to component α_j of vector α . They satisfy a vector equation (if y is a vector) resulting from dynamic model (for y) by derivation on a parameter α_j .

For variable parameters such method essentially becomes complicated and practically is not applicable.

Application of the variation approach allows fundamentally to simplify process of calculation of SC concerning variable and constant parameters. To this problem it is devoted given paper for the dynamic systems described by generalized difference equations with distributed memory on phase coordinates and variable parameters. The nonlinear quality functional also has a generalized form.

Variational method [6] makes possible to simplify the process of determination of conjugate equations and formulas of account of SC. On the basis of this method it is an extension of quality functional by means of inclusion into it dynamic equations of object with use of Lagrange's multipliers and obtaining the first variation of extended functional on phase coordinates of object and on interesting parameters. Dynamic equations for Lagrange's multipliers are obtained due to set equal to a zero (in the first variation of extended functional) the functions before the variations of phase coordinates. Given simplification first variation of extended functional brings at presence in the right part only parameter variations, i.e. it is got the sensitivity functional according to concerning parameters.

It is proved that both methods to calculation of SC (either with use of Lagrange's functions or with use of sensitivity functions) yield the same result, but the first method it is essential more simple in the computing relation.

1. Problem definition

We suppose that the dynamic object is described by system of generalized non-linear difference equations [1, 5]

$$\begin{aligned} y(t+1) &= f(y(t), \dots, y(1), y(0); \tilde{\alpha}(t), \dots, \tilde{\alpha}(1), \tilde{\alpha}(0); \alpha, t), \quad t=0, 1, \dots, N, \\ y(0) &= y_0(\alpha). \end{aligned} \quad (1)$$

Here: y is a vector-column of phase coordinates; $\tilde{\alpha}(t), \alpha$ are a vector-columns of interesting variable and constant parameters; $f(\cdot), y_0(\alpha)$ are known continuously differentiated limited nonlinear vector-functions.

The quality of functioning of system it is characterised of generalized nonlinear functional

$$I(\tilde{\alpha}, \alpha) = f_0(y(N+1), \dots, y(1), y(0); \tilde{\alpha}(N+1), \dots, \tilde{\alpha}(1), \tilde{\alpha}(0); \alpha), \quad (2)$$

depending from values of all phase coordinates and variable parameters: $y(t), \tilde{\alpha}(t), t = N+1, \dots, 1, 0$, and from α . The conditions for function $f_0(\cdot)$ are the same as for functions $f(\cdot), y_0(\alpha)$.

With the purpose of simplification of appropriate deductions with preservation of a generality in all transformations (1), (2) there are two vectors of parameters $\tilde{\alpha}(t), \alpha$. If in the equations (1), (2) parameters are different then it is possible formally to unit them in two vectors $\tilde{\alpha}(t), \alpha$, to use obtained outcomes and then to make appropriate simplifications, taking into account a structure of a vectors $\tilde{\alpha}(t), \alpha$.

By obtaining of results the obvious designations:

$$\begin{aligned} f(t) &\equiv f(y(t), \dots, y(1), y(0); \tilde{\alpha}(t), \dots, \tilde{\alpha}(1), \tilde{\alpha}(0); \alpha, t), \quad t=0, 1, \dots, N, \\ f_0(\cdot) &\equiv f_0(y(N+1), \dots, y(1), y(0); \tilde{\alpha}(N+1), \dots, \tilde{\alpha}(1), \tilde{\alpha}(0); \alpha), \end{aligned} \quad (3)$$

are used.

In difference from other papers devoted to calculation of SC in given paper the generalized difference model (1) and generalized nonlinear functional (2) are used. In the previous papers of functions $f(t)$ and $f_0(\cdot)$ also were nonlinear, but they contained the linear operator of summation, for example

$$\begin{aligned} f(t) &\equiv \sum_{s=0}^t K(t, y(s), \tilde{\alpha}(s), \alpha, s), \quad t=0, 1, 2, \dots, N, \quad y(0) = y_0(\alpha), \\ f_0(\cdot) &\equiv \sum_{t=0}^{N+1} f_0(y(t), \tilde{\alpha}(t), \alpha, t). \end{aligned}$$

The index t in functions $f(t)$ also reflects not only obvious dependence on step number, but also that the kind of functions from a step to a step can change.

Is shown also that the variation method allows to receive the SC in relation to variable and constant parameters:

$$\begin{aligned} \delta I(\tilde{\alpha}, \alpha) &= \sum_{t=0}^{N+1} \frac{\partial I(\tilde{\alpha}, \alpha)}{\partial \tilde{\alpha}(t)} d\tilde{\alpha}(t) + \frac{\partial I(\tilde{\alpha}, \alpha)}{\partial \alpha} d\alpha. \\ \nabla_{\tilde{\alpha}(t)} I(\tilde{\alpha}, \alpha) &= \left(\frac{\partial I(\tilde{\alpha}, \alpha)}{\partial \tilde{\alpha}_1(t)} \quad \dots \quad \frac{\partial I(\tilde{\alpha}, \alpha)}{\partial \tilde{\alpha}_{m_1}(t)} \right)^T, \quad t=0, 1, 2, \dots, N, N+1, \\ \nabla_{\alpha} I(\tilde{\alpha}, \alpha) &= \left(\frac{\partial I(\tilde{\alpha}, \alpha)}{\partial \alpha_1} \quad \dots \quad \frac{\partial I(\tilde{\alpha}, \alpha)}{\partial \alpha_{m_2}} \right)^T. \end{aligned} \quad (4)$$

Let's receive the conjugate equations for calculation of Lagrange's multipliers and on the basis of them formulas for calculation of the SC.

2. Principal results

Complement a quality functional (2) by restrictions-equalities (1) by means of Lagrange's multipliers $\lambda(t), t = 0, 1, 2, \dots, N+1$ (column vectors) and get the extended functional

$$I = I(\tilde{\alpha}, \alpha) + \sum_{t=0}^N \lambda^T(t+1) [-y(t+1) + f(t)] + \lambda^T(0) [-y(0) + y_0(\alpha)]. \quad (5)$$

Functional (5) complies with $I(\tilde{\alpha}, \alpha)$ when (1) is fulfilled.

We calculate the first variation of extended functional, caused by a variation of phase coordinates and also a variation of variables and constant parameters:

$$\delta I = \delta_y I + \delta_{\tilde{\alpha}} I + \delta_{\alpha} I \equiv \sum_{t=0}^{N+1} \frac{\partial I}{\partial y(t)} dy(t) + \sum_{t=0}^{N+1} \frac{\partial I}{\partial \tilde{\alpha}(t)} d\tilde{\alpha}(t) + \frac{\partial I}{\partial \alpha} d\alpha. \quad (6)$$

Here:

$$\begin{aligned} \delta_y I = & \left[\frac{\partial f_0(\cdot)}{\partial y(N+1)} - \lambda^T(N+1) \right] dy(N+1) + \sum_{t=0}^N \left[\frac{\partial f_0(\cdot)}{\partial y(t)} - \lambda^T(t) \right] dy(t) + \\ & + \sum_{t=0}^N \sum_{s=0}^t \lambda^T(t+1) \frac{\partial f(t)}{\partial y(s)} dy(s). \end{aligned} \quad (7)$$

We apply the equation

$$\sum_{t=0}^N \sum_{s=0}^t A(t,s) = \sum_{s=0}^N \sum_{t=s}^N A(s,t), \quad (8)$$

which has been received in [10] by calculation of SC for dynamic systems, described by more simple difference equations also with the distributed memory on phase coordinates and variable parameters. The proof of correctness of (8) equality is realized by an mathematical induction method.

For summand $\sum_{t=0}^N \sum_{s=0}^t \lambda^T(t+1) \frac{\partial f(t)}{\partial y(s)} dy(s)$ in variation $\delta_y I$ the equality (8) has the following form:

$$\sum_{t=0}^N \sum_{s=0}^t \lambda^T(t+1) \frac{\partial f(t)}{\partial y(s)} dy(s) = \sum_{s=0}^N \sum_{t=s}^N \lambda^T(s+1) \frac{\partial f(s)}{\partial y(t)} dy(t). \quad (9)$$

Then

$$\begin{aligned} \delta_y I = & \left[\frac{\partial f_0(\cdot)}{\partial y(N+1)} - \lambda^T(N+1) \right] dy(N+1) + \\ & + \sum_{t=0}^N \left[\frac{\partial f_0(\cdot)}{\partial y(t)} - \lambda^T(t) + \sum_{s=t}^N \lambda^T(s+1) \frac{\partial f(s)}{\partial y(t)} \right] dy(t). \end{aligned} \quad (10)$$

Derivatives standing in the formula (10) before variations of phase coordinates look like:

$$\begin{aligned} \frac{\partial I}{\partial y(N+1)} &= -\lambda^T(N+1) + \frac{\partial f_0(\cdot)}{\partial y(N+1)}, \\ \frac{\partial I}{\partial y(t)} &= -\lambda^T(t) + \sum_{s=t}^N \lambda^T(s+1) \frac{\partial f(s)}{\partial y(t)} + \frac{\partial f_0(\cdot)}{\partial y(t)}, \quad t = N, N-1, \dots, 1, 0. \end{aligned} \quad (11)$$

From equality to zero of these derivatives we receive the equations for Lagrange's multipliers:

$$\begin{aligned} \lambda^T(N+1) &= \frac{\partial f_0(\cdot)}{\partial y(N+1)}, \\ \lambda^T(t) &= \sum_{s=t}^N \lambda^T(s+1) \frac{\partial f(s)}{\partial y(t)} + \frac{\partial f_0(\cdot)}{\partial y(t)}, \quad t = N, N-1, \dots, 1, 0. \end{aligned} \quad (12)$$

These equations are decided in the opposite direction changes of an independent integer t variable.

Let's receive now formulas of SC calculation.

In the equations (in (6)) for $\delta_{\tilde{\alpha}} I$, $\delta_{\alpha} I$ the SC concerning variables and constant parameters look like:

$$\begin{aligned} \frac{\partial I}{\partial \tilde{\alpha}(N+1)} &= \frac{\partial f_0(\cdot)}{\partial \tilde{\alpha}(N+1)}, \\ \frac{\partial I}{\partial \tilde{\alpha}(t)} &= \frac{\partial f_0(\cdot)}{\partial \tilde{\alpha}(t)} + \sum_{s=t}^N \lambda^T(s+1) \frac{\partial f(s)}{\partial \tilde{\alpha}(t)}, \quad t = N, N-1, \dots, 1, 0, \\ \frac{\partial I}{\partial \alpha} &= \frac{\partial f_0(\cdot)}{\partial \alpha} + \sum_{t=0}^N \lambda^T(t+1) \frac{\partial f(t)}{\partial \alpha} + \lambda^T(0) \frac{dy_0(\alpha)}{d\alpha}. \end{aligned} \quad (13)$$

This result is more common in relation to appropriate results of monograph [5] and paper [10].

At reception SC (13) it was used the variations (6) of extended functional (5).

We prove equivalence of sensitivity coefficients for initial (2) and extended (5) functionals.

We take extended functional, presented in the formula (5). Before $\lambda^T(\cdot)$ in brackets there are the dynamic equations of the object which has been written down in the form of the equations of equality type. Hence, values of functions in brackets are always equal to zero.

Let's calculate from both parts of the previous equation derivatives in the beginning on a vector of constant α parameters:

$$\frac{\partial I}{\partial \alpha} = \frac{\partial I(\tilde{\alpha}, \alpha)}{\partial \alpha} + \sum_{t=0}^N \lambda^T(t+1) \left[-W_{\alpha}(t+1) + \sum_{s=0}^t \frac{\partial f(t)}{\partial y(s)} W_{\alpha}(s) + \frac{\partial f(t)}{\partial \alpha} \right] + \lambda^T(0) \left[-W_{\alpha}(0) + \frac{dy_0(\alpha)}{d\alpha} \right].$$

Before $\lambda^T(\cdot)$ now there are sensitivity equations for a matrix of sensitivity functions. These equations are written down as in the form of restrictions of equality type. Values of functions in brackets also are always equal to zero.

Hence, SC rather both for initial functional and for its extended variant have identical values.

For reception of the sensitivity equations it is necessary in equations (1) to impose a condition of differentiability for $f(t)$ on phase coordinates and on considered parameters. On α parameters should be differentiated initial functions $y_0(\alpha)$.

It is possible to receive the same result and for SC on relation to variable parameters. The sensitivity equations for each fixed value of argument of variable parameters $\tilde{\alpha}(j)$, $j=0,1,\dots,N+1$ have more complex form. They demand special consideration. Important that such sensitivity equations are objectively exist.

Если при формировании функционала качества работы системы используется дополнительный преобразователь фазовых координат, то необходимо учесть модель этого измерительного устройства

At additional use of model of the measuring device it is necessary to make changes to problem statement:

$$I(\tilde{\alpha}, \alpha) = f_0(\eta(N+1), \dots, \eta(1), \eta(0); \tilde{\alpha}(N+1), \dots, \tilde{\alpha}(1), \tilde{\alpha}(0); \alpha) \equiv f_0(\cdot);$$

$$\eta(t) \equiv \eta(y(t), \tilde{\alpha}(t), \alpha, t), \quad t = 0, 1, 2, \dots, N+1.$$

In the received results it is necessary to execute small replacements:

$$\frac{\partial f_0(\cdot)}{\partial y(t)} \text{ to replace on } \frac{\partial f_0(\cdot)}{\partial \eta(t)} \frac{\partial \eta(t)}{\partial y(t)};$$

$$\frac{\partial f_0(\cdot)}{\partial \tilde{\alpha}(t)} \text{ to replace on } \frac{\partial f_0(\cdot)}{\partial \tilde{\alpha}(t)} + \frac{\partial f_0(\cdot)}{\partial \eta(t)} \frac{\partial \eta(t)}{\partial \tilde{\alpha}(t)};$$

$$\frac{\partial f_0(\cdot)}{\partial \alpha} \text{ to replace on } \frac{\partial f_0(\cdot)}{\partial \alpha} + \sum_{t=0}^{N+1} \frac{\partial f_0(\cdot)}{\partial \eta(t)} \frac{\partial \eta(t)}{\partial \alpha}.$$

3. Examples

Example 1. We consider that in dynamic system there is a memory on phase coordinates, but there is no memory on variable parameters, i.e.:

$$f(t) \equiv f(y(t), \dots, y(1), y(0); \tilde{\alpha}(t), \alpha, t), \quad t = 0, 1, \dots, N. \quad (14)$$

Indicator of quality of system work is the same.

Structures of the conjugate equations (12) and SC (13) for constant parameters remain, but structure SC (13) for variable parameters becomes simpler:

$$\frac{\partial I}{\partial \tilde{\alpha}(t)} = \frac{\partial f_0(\cdot)}{\partial \tilde{\alpha}(t)} + \lambda^T(t+1) \frac{\partial f(t)}{\partial \tilde{\alpha}(t)}, \quad t = N, N-1, \dots, 1, 0.$$

Example 2. In mathematical model of dynamic system there is no additional memory on phase coordinates, but there is a memory on variable parameters. At this variant

$$f(t) \equiv f(y(t); \tilde{\alpha}(t), \dots, \tilde{\alpha}(1), \tilde{\alpha}(0); \alpha, t), \quad t=0, 1, \dots, N. \quad (15)$$

The indicator of a quality functional of system has the same appearance (2).

The structure of the conjugate equations (12) becomes simpler:

$$\lambda^T(N+1) = \frac{\partial f_0(\cdot)}{\partial y(N+1)}, \quad \lambda^T(t) = \lambda^T(t+1) \frac{\partial f(t)}{\partial y(t)} + \frac{\partial f_0(\cdot)}{\partial y(t)}, \quad t = N, N-1, \dots, 1, 0.$$

The SC structures (13) for variables and constant parameters remain.

Example 3. In dynamic system there is no additional memory on phase coordinates and on variable parameters. Then

$$f(t) \equiv f(y(t); \tilde{\alpha}(t); \alpha, t), \quad t=0, 1, \dots, N, \quad (16)$$

but a quality functional of system has the same appearance (2).

In this variant the conjugate equations and SC to variable parameters become more simple:

$$\lambda^T(N+1) = \frac{\partial f_0(\cdot)}{\partial y(N+1)}, \quad \lambda^T(t) = \lambda^T(t+1) \frac{\partial f(t)}{\partial y(t)} + \frac{\partial f_0(\cdot)}{\partial y(t)}, \quad t = N, N-1, \dots, 1, 0;$$

$$\frac{\partial I}{\partial \tilde{\alpha}(t)} = \frac{\partial f_0(\cdot)}{\partial \tilde{\alpha}(t)} + \lambda^T(t+1) \frac{\partial f(t)}{\partial \tilde{\alpha}(t)}, \quad t = N, N-1, \dots, 1, 0.$$

The SC to constant parameters remain former (13).

Conclusion

On the basis of a variation method the problem of effective calculation of SC for multivariate nonlinear dynamic systems described by the generalized difference equations with the distributed memory on phase coordinates and variable parameters is solved. The quality of functioning of systems it is characterized of generalized nonlinear functional (2) too. Variables and constant parameters are present at object model, at model of the measuring device and at generalized quality functional for system.

At the heart of SC calculation the decision of the difference equations of object model in a forward direction of time and obtained difference equations for Lagrange's multipliers in the opposite direction of time lies.

It is proved that both methods to calculation of SC (with use of Lagrange's functions or with use of sensitivity functions) yield the same result, but the first method it is essential more simple in the computing relation.

Results of present paper are applicable at design of high-precision systems and devices.

It is possible generalization of the received results on the dynamic systems described by. interconnected difference the ordinary equations and generalized equations with the distributed memory in time on phase coordinates and on variable parameters.

References

1. Katkovnik, V.Ya. & Poluektov, R.A. (1966) *Multidimensional Discrete Control Systems*. Moscow: Nauka.
2. Kozlov, Yu.M. & Yusupov, R.M. (1969) *Without Search Self-Adjusted Systems*. Moscow: Nauka.
3. Tomovich, R. & Vukobratovich M. (1972) *The General Sensitivity Theory*. Moscow: Sovetskoe radio.
4. Fomin, B.N., Fradkov A.L. & Yakubovich V.A. (1981) *Adaptive Control of Dynamic Objects*. Moscow: Nauka.
5. Ruban, A.I. (1982) *Identification and Sensitivity of Complex Systems*. Tomsk: Tomsk State University.
6. Voronov, A.A. (1985) *Introduction in Dynamics of Complex Control Systems*. Moscow: Nauka.
7. Rosenwasser, E. & Yusupov, R. (2000) *Sensitivity of Automatic Control Systems*. Boca Raton, London, New York, Washington, D.C.: CRC press Cop.
8. Wilkins, A.K., Tidor, B., White, J. & Barton, P.I. (2009) Sensitivity analysis for oscillating dynamical systems. *SIAM Journal on Scientific Computing*. 31(4). pp. 2706–2732.
9. Kudryashov, V.S. Ryazantsev, S.V. & Sviridov, D.A. (2014) Designing of multiconnected independent digital control system of low sensitivity. *Instruments and systems. Control. Inspection. Diagnostics*. 5. pp. 1–11.

10. Rouban A.I. (2021) The sensitivity coefficients for dynamic systems described by difference equations with the distributed memory on phase coordinates and variable parameters. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie vychislitel'naya tehnika i informatika – Tomsk State University Journal of Control and Computer Science*. 57. pp. 95–100. DOI: 10.17223/19988605/57/10

Information about the author:

Rouban Anatoly Ivanovich (Doktor of Technical Sciences, Professor of Computer Science Department of Institute of Space and Information Technologies, Siberian Federal University, Krasnoyarsk, Russian Federation). E-mail: ai-rouban@mail.ru

The author declares no conflicts of interests.

Информация об авторе:

Рубан Анатолий Иванович – профессор, доктор технических наук, профессор кафедры информатики Института космических и информационных технологий Сибирского Федерального университета (Красноярск, Россия). E-mail: ai-rouban@mail.ru

Автор заявляет об отсутствии конфликта интересов.

Поступила в редакцию 28.02.2022; принята к публикации 29.11.2022

Received 28.02.2022; accepted for publication 29.11.2022

Научная статья

УДК 519-7

doi: 10.17223/19988605/61/9

Использование ансамбля сверточных нейронных сетей для распознавания дорожных знаков

Игорь Константинович Харченко¹, Игорь Георгиевич Боровской²,
Елена Александровна Шельмина³

Томский государственный университет систем управления и радиоэлектроники, Томск, Россия

¹ igor.k.kharchenko@tusur.ru

² igor.g.borovskoi@tusur.ru

³ elena.a.shelmina@tusur.ru

Аннотация. Предлагается для распознавания дорожных знаков использовать ансамбль сверточных нейронных сетей, который является модификацией робастного метода распознавания на основе нейронных сетей глубокого обучения. Данный ансамбль повышает скорость работы робастного метода распознавания, а также позволяет увеличить быстродействие с сохранением высокой точности распознавания за счет удаления из набора данных значений, которые не представляют полезной нагрузки.

Ключевые слова: распознавание дорожных знаков; сверточные нейронные сети; нейросетевой ансамбль

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: Харченко И.К., Боровской И.Г., Шельмина Е.А. Использование ансамбля сверточных нейронных сетей для распознавания дорожных знаков // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2022. № 61. С. 88–96. doi: 10.17223/19988605/61/9

Original article

doi: 10.17223/19988605/61/9

Usage of convolutional neural network ensemble for traffic sign recognition

Igor I. Kharchenko¹, Igor G. Borovskoy², Elena A. Shelmina³

Tomsk State University of Control Systems and Radioelectronics, Tomsk, Russian Federation

¹ igor.k.kharchenko@tusur.ru

² igor.g.borovskoi@tusur.ru

³ elena.a.shelmina@tusur.ru

Abstract. The article suggests using an ensemble of convolutional neural networks for the recognition of road signs, which is a modification of a robust recognition method based on deep learning neural networks.

This ensemble improves the speed of the robust recognition method, and also allows you to increase the speed while maintaining high recognition accuracy by removing values from the data set that do not represent a payload.

Keywords: recognition of road signs; convolutional neural networks; neural network ensemble

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

For citation: Kharchenko, I.K., Borovskoy, I.G., Shelmina, E.A. (2022) Usage of convolutional neural network ensemble for traffic sign recognition. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naja tehnika i informatika – Tomsk State University Journal of Control and Computer Science*. 61. pp. 88–96. doi: 10.17223/19988605/61/9

В современном мире идентификация различных дорожных знаков является частью системы помощи водителю (СПБ, ADAS – Advanced Driver-Assistance System) [1], а также систем беспилотного управления автомобилем. Многие передовые компании, такие как Google, Uber, Tesla Motors, Volkswagen, Hyundai, NVIDIA, Baidu, ведут активную деятельность по реализации и тестированию подобных систем. Например, в Германии, Китае, США беспилотные автомобили уже используются в качестве такси и для грузоперевозок на дальние расстояния. Поэтому можно говорить об актуальности данной задачи. Для использования подобных систем в реальных дорожных условиях они должны быть устойчивыми к влиянию на качество распознавания различных факторов среды, таких как дождь, снег, туман, размытие и загрязнение линзы устройства видеозахвата.

Проблематике распознавания и локализации дорожных знаков посвящено немало количество научных работ. Для этих целей используются различные алгоритмы, детальный обзор которых представлен в [1, 2]. Рассмотрим классические алгоритмы, которые основываются на цвете и форме изображения. Качество распознавания у методов цветовой сегментации может зависеть от выбранного цветового пространства: к примеру, в работе [3] обнаружено, что использование цветового пространства Hue-Saturation-Intensity (HSI; оттенок–насыщенность–интенсивность) улучшает качество распознавания по сравнению с результатами, полученными на цветовом пространстве RGB в работе [4]. Авторы исследований [5, 6] установили, что применение цветовых пространств YCbCr и YUV также дает приемлемое качество распознавания. Однако такие методы очень чувствительны к изменению освещения, что является важным фактором при их применении в реальных условиях: распознавание должно быть корректным для разного времени суток. Если обратиться к методам сегментации по форме, то достаточно распространенным методом является детектор границ Кенни [7–9], а также каскад Хаара [10], быстрые трансформации Фурье [11], гистограмма ориентированных градиентов [5]. Применение методов сегментации по форме изображения конкретно к распознаванию дорожных знаков достаточно затруднительно, поскольку размеры и масштабы знака зависят от его удаленности от устройства захвата видео. Ситуацию усложняет еще и то, что знак может быть перекрыт другим объектом. Следовательно, можно прийти к выводу, что не существует какого-то конкретного способа качественного определения наилучшего классификатора для задачи распознавания дорожных знаков.

Недавние исследования в области машинного обучения привели к созданию инновационных подходов к решению данной задачи. Авторы работы [12] представили архитектуру сети, позволяющей одновременно вычислять и локализовать участок, на котором расположен знак. В работе [13] авторы пошли дальше и разделили локализацию и распознавание знака на две отдельные независимые нейросети: в качестве локализатора выступает гибридная архитектура SegU-Net, созданная на основе двух архитектур – SegNet [14] и U-Net [15], а для распознавания классов дорожных знаков используется архитектура VGG16 [16]. Далее авторы в рамках работы [17] решили модифицировать свой подход, добавив улучшение изображения в качестве отдельных этапов: классификатор на VGG16 обучен на распознавание погодных условий, если погодная ситуация требует нормализации (снег, дождь, туман), то изображение проходит через генеративно-сопоставительную сеть, которая приводит изображение к виду без плохих погодных условий, и после этого сеть SegU-Net локализует области дорожных знаков и распознает их.

Для локализации и распознавания объектов на изображениях используются два вида архитектур: одноэтапные и двухэтапные. В двухэтапных архитектурах применяются две сети: одна для локализации области расположения объекта, а вторая для распознавания его класса. В одноэтапных архитектурах оба действия производятся одной сетью. В работе [18] приводится сравнение нескольких архитектур нейросетевых детекторов: Faster R-CNN, R-FCN, Mask R-CNN, YOLO, SSD; из сравнения можно сделать вывод, что архитектура Faster R-CNN обладает высокой точностью, но низкой скоро-

стью работы, при этом алгоритмы семейства YOLO обладают самой высокой скоростью, при этом точность работы чуть ниже альтернатив.

Одним из существенных недостатков в данных работах является необходимость обучения сети на большом объеме обучающей выборки. К тому же скорость распознавания у YOLO гораздо выше предложенной авторами архитектуры SegU-Net.

Цель же данной работы – реализация нейросетевого ансамбля, подобного реализованному в [17], с заменой локализатора с SegU-Net на YOLO v4 [19]. Также для уменьшения скорости обучения предложено удалять области изображения, не содержащие никаких знаков. Получены следующие результаты: точность работы YOLO v4 оказалась в среднем на 5% ниже по каждому типу погодных условий, что обосновано неполной обучающей выборкой в данном исследовании в отличие от исследования [17]. При этом скорость увеличилась до 13 кадров в секунду по сравнению с 3 кадрами в секунду в работе [17]. Также за счет отсека областей без дорожных знаков на изображении удалось добиться уменьшения объема обучающей выборки в два раза по сравнению с исследованием [13].

1. Нейросетевой ансамбль

1.1. Описание подхода

Предложенный подход распознавания дорожных знаков состоит из трех архитектур нейронных сетей, каждая из которых выполняет отдельный набор действий. Первый модуль – классификатор погодных условий, отвечающий за распознавание типа погодных условий на изображении. Второй модуль отвечает за удаление плохих погодных условий на изображении, т.е. за приведение его к виду без плохих условий. Наконец, третий модуль – локализация и распознавание дорожных знаков на изображении.

Алгоритм данного нейросетевого ансамбля:

1. Исходное изображение передается классификатору погодных условий.
2. Если классификатором определяется одно из ухудшающих погодных условий, то изображение пропускается через модуль удаления плохих погодных условий.
3. Результирующее изображение подается на вход локализатора дорожных знаков, и в случае их наличия локализатор возвращает их ограничивающие прямоугольники (bounding box) с точностью распознавания для каждого такого прямоугольника.

Для проверки качества обучения нейронных сетей в области распознавания дорожных знаков существуют специальные наборы данных, такие как GTSDDB [20], BelgiumTS [21]. Однако в данных наборах нет достаточного количества погодных условий, на которых можно было бы обучить нейросеть без применения методов аугментации обучающей выборки [22] (которые в свою очередь тоже не гарантируют полноту данных), для обучения. Чтобы преодолеть это ограничение, в работе [23] авторы реализовали набор данных, состоящий из видеорядов с добавленными погодными условиями, влияющими на качество распознавания изображения. Набор состоит из двух поднаборов из 49 видеорядов в каждом: один набор – это реальная съемка из автомобиля, второй – съемка в виртуальной смоделированной среде. В рамках данной работы используется только первая часть данного набора, снятая из реального автомобиля.

1.2. Модуль классификации плохих погодных условий

Классификатором погодных условий выступает сеть VGG19 [16], предобученная на наборе данных ImageNet и обученная на наборе данных CURE-TSD путем метода «передачи обучения» (transfer learning). Идея «передачи обучения» заключается в использовании свойств сверточных сетей: такие сети на каждом последующем слое распознают все более сложные фрагменты, например на неглубоких слоях сеть распознает углы, края, простые формы, на более глубоких слоях – простые фигуры и т.д. Таким образом, вся информация о признаках изображений хранится в 16 сверточных

слоях, а обобщения – в трех полносвязных выходных слоях. Следовательно, обучению подлежат только последние три слоя; остальные обучению не подлежат.

Экспериментальным путем установлено, что разрешения входного изображения в 224×224 пикселя достаточно для точности распознавания погодных условий более 99%. Поэтому на вход сети подаются изображения, уменьшенные до размера 224×224 пикселя; каждому изображению соответствует метка с погодными условиями: 0 – хорошие условия, от 1 до 6 – ухудшающие условия. Модуль обучается на 15 эпохах обучения. Если ошибка обучения не уменьшается в течение трех эпох, шаг оптимизатора уменьшается наполовину (начальный шаг оптимизатора равен 0,0001).

1.3. Модуль нормализации плохих погодных условий

Для целей нормализации могут подойти такие архитектуры, как сверточный автоэнкодер и генеративно-состязательная сеть. Однако сверточный автоэнкодер неприменим для данной работы, поскольку его основное назначение – не улучшение изображения, а извлечение признаков из последнего. Поэтому для нормализации изображения применяется генеративно-состязательная сеть.

Архитектура состоит из двух сетей: генератора и дискриминатора. Генератор – это сеть, генерирующая изображения с шумом, которые подаются на вход дискриминатора, который, в свою очередь, учится «отделять» зашумленные изображения от незашумленных и тем самым удалять шум из изображения. Топология сети генератора и дискриминатора соответствует используемой в работе [17].

При обучении данной архитектуры используется результат суммы двух функций ошибки, приведенных авторами в работе [17].

Первая – это функция потерь с использованием средней абсолютной ошибки

$$L_{Wasserstein} = \frac{1}{H * W * C} \sum_{k=0}^C \sum_{j=0}^W \sum_{i=0}^H |R_{ijk} - T_{ijk}|, \quad (1)$$

где H и W – высота и ширина изображения, C – количество каналов изображения, R и T – реконструированное изображение и оригинальное изображение соответственно. Смысл использования именно такой функции (1) вместо наиболее часто используемой функции среднеквадратической ошибки в том, что, как обнаружено в работе [24], последняя накладывает эффект размытия на изображение, из-за чего нейросеть не может определить на нем мелкие детали. Применение функции средней абсолютной ошибки [25] дает лучшие результаты при реконструкции мелких деталей изображения.

Вторая – это так называемая перцептивная функция потерь

$$L_{perceptive} = \frac{1}{H * W * C} \sum_{k=0}^C \sum_{j=0}^W \sum_{i=0}^H |\varphi(R)_{ijk} - \varphi(T)_{ijk}|, \quad (2)$$

где $\varphi(R)$ и $\varphi(T)$ – признаки изображения, извлеченные из седьмого слоя сети VGG16, предобученной на наборе данных ImageNet, для реконструированного и оригинального изображения. Основная идея такого подхода с использованием функции (2) заключается в том, что необходимо максимально сохранить информацию о мелких частях дорожных знаков, а внутренние слои предобученной сети VGG16 уже содержат необходимую нам информацию о таких признаках, как углы, края и т.п.

При обучении модуля выбираются два равных фрагмента размером 1024×1024 пикселя с двух верхних углов каждого изображения. Обоснование такого подхода приведено в разд. 2. Обучение производится на 30 эпохах; в случае если в течение трех эпох нет качественного улучшения ошибки сети, то шаг обучения оптимизатора сокращается вдвое (начальный шаг равен 0,0001).

1.4. Модуль локализации и распознавания дорожных знаков

Принцип обучения модуля локализации и распознавания дорожных знаков на архитектуре YOLO v4 [19] схож с принципом обучения модуля нормализации: на вход поступают два равных фрагмента 1024×1024 пикселя по двум верхним углам изображения. Обучение производится на

4 000 эпохах: такое большое количество обосновано самой архитектурой сети, поскольку каждое изображение проходит через несколько этапов обработки.

2. Анализ набора данных, используемого для обучения модулей

В качестве набора данных выступает набор данных CURE-TSD, представляющий собой набор из 49 видеорядов по 300 кадров в каждом. Для каждого видеоряда есть по 70 фрагментов с нанесенными эффектами, чтобы смоделировать плохие погодные условия: 14 погодных условий с пятью уровнями ухудшения каждого условия. В рамках данного исследования отобраны только такие, которые встречаются в реальных условиях: дождь, снег, туман, загрязнение и размытие линзы, и снятые в реальном автомобиле; видеоряды, снятые в смоделированной среде, в данной работе не рассматриваются. Каждое из погодных условий также подразделяется на пять уровней: на первом хорошо различимы почти все знаки (погодные условия почти не влияют на качество распознавания), на последнем же знаки, находящиеся вдали, практически полностью невидны. Таким образом, для распознавания необходимо использовать по 26 видеорядов для каждого фрагмента снятого видео, в общей сложности 1 274 фрагмента.

Необходимо также принимать во внимание угол наклона камеры относительно дороги: от этого зависит, какая часть изображения действительно несет в себе полезную нагрузку, а сколько пикселей никогда не будет задействовано при распознавании. Соответственно, большая часть знаков расположена на правой части, по оси OY области пикселей знаков имеют нормальное распределение и расположены в основном в области от нуля до 646 пикселей. Следовательно, при таком угле наклона камеры можно отсечь почти 52% от всей области изображения без потери точности распознавания.

Также выявлено, что на одном кадре может располагаться от одного до пяти знаков с расстоянием между ними, превышающим 1 024 пикселя. Это говорит о том, что метод, который используют авторы в статье [17] (выборка фрагмента изображения размером $1\,024 \times 1\,024$ пикселя случайным образом), может быть применен только в ситуации, когда на изображении есть не более одного-двух знаков и они гарантированно будут находиться в одном фрагменте. Следовательно, это вполне могло оказать негативное влияние на точность распознавания знаков. С другой стороны, авторы статьи [13] для обучения сети SegU-Net выбирают четыре фрагмента размером $1\,024 \times 1\,024$ пикселя путем обрезки изображения по четырем углам (верхним и нижним).

Минимальная ширина знака, на котором обучается локализатор дорожных знаков, равна восьми, максимальная же составляет 246. Следовательно, минимальный размер фрагмента, на котором можно обучить второй и третий модули, может быть 512 или 1 024 (256 не берется из-за возможной потери информации о краях знака). Размер фрагмента должен быть кратен двум, поскольку в ином случае возможна потеря информации о контексте изображения при операции макс-пулинга во внутренних слоях сверточной сети [13]. Уменьшение размера фрагмента до 512 пикселей сократит размер сети, но не сократит время обучения, поэтому размер фрагмента обучения в рамках данной работы выбран равным $1\,024 \times 1\,024$ пикселя.

3. Результаты тестирования

Обучение производилось с использованием следующих вычислительных ресурсов: процессор Intel® Xeon® Gold 6226R, видеокарта Nvidia RTX 3090, объем оперативной памяти 64 ГБ.

По техническим ограничениям в рамках данной работы обучающая выборка ограничена только первыми 20 видеорядами. Для предотвращения переобучения обучающая, тестовая и валидационная выборки состоят из 16, 2 и 2 фрагментов соответственно.

Точность обучения классификатора погодных условий на валидационной выборке составила 99,54%. Ошибки распознавания возникают при распознавании дождя и размытия линзы на первом уровне, что в целом является приемлемым результатом. Авторы архитектуры [17] получили результат в 99,89%; такое весомое расхождение можно обосновать тем, что авторы использовали всю обучающую выборку, состоящую из 49 видеофрагментов, для обучения ансамбля.

Результаты работы модуля нормализации и локализации для каждого типа погодных условий на валидационной выборке (в столбце «Полученная») по сравнению с полученными в работе [17] (в столбце «Эталонная») приведены в таблице.

Точность модуля нормализации и локализации дорожных знаков

Условие	Уровень	Полученная	Эталонная
Без плохих условий	–	95,91	99,02
Дождь	1	97,28	98,15
	2	97,41	97,72
	3	97,41	97,38
	4	96,17	96,88
	5	96,05	96,03
Снег	1	96,67	98,95
	2	96,42	98,24
	3	96,54	98,57
	4	96,42	96,04
	5	93,29	94,94
Загрязнение линзы	1	95,99	98,55
	2	94,81	98,56
	3	94,19	98,33
	4	92,58	97,94
	5	92,11	97,33
Размытие линзы	1	96,30	98,62
	2	95,19	98,51
	3	94,45	98,22
	4	93,99	97,57
	5	93,12	97,31
Туман	1	95,68	98,31
	2	94,08	98,76
	3	93,71	98,85
	4	93,02	98,38
	5	92,83	97,74

Из результатов видно, что точность обучения на погодных условиях, которые перекрывают часть знака (дождь, снег), сопоставима с эталонной, однако на остальных условиях точность меньше в среднем на 5%. Это можно объяснить тем, что YOLO v4 немного хуже справляется с более мелкими элементами на изображении и не всегда может с высокой точностью распознать мелкие детали на большом удалении от знака.

В работе [17] скорость работы всего ансамбля в среднем равна 2,79 кадров в секунду. При использовании модифицированного ансамбля скорость работы увеличена до 12,76 кадров в секунду.

Заключение

В данной статье представлен нейросетевой ансамбль, являющийся модификацией работы [17], улучшающий скорость работы робастного метода распознавания, а также сокращающий скорость обучения с сохранением высокой точности распознавания за счет удаления из набора данных части изображений, которые не представляют полезной нагрузки. Классификатор погодных условий определяет, нужно ли улучшать изображение: в случае необходимости в работу подключается нейросеть для улучшения изображения. Нейросеть YOLO v4 локализует область знака и распознает его класс: использование данной нейросети вместо связки SegU-Net и VGG16 дает выигрыш в скорости распознавания классификатора при потере точности распознавания в 5% при таких условиях, как туман, размытие и загрязнение линзы. Обученные нейросети могут использоваться и на российских дорогах, поскольку видеоряды в обучающей выборке были сняты в Бельгии, где движение правостороннее, а знаки схожи с российскими.

При этом данный подход к распознаванию применим и к смежным задачам: например, к распознаванию госномеров автомобилей или определению на проезжей части людей, пересекающих проезжую часть в неполюженном месте, для предотвращения аварийных ситуаций.

Список источников

1. Mogelmoose A., Trivedi M.M., Moeslund T.B. Vision-based traffic sign detection and analysis for intelligent driver assistance systems: Perspectives and survey // *IEEE Trans. Intell. Transp. Syst.* 2012. V. 13 (4). P. 1484–1497.
2. Gudigar A., Shreesha C., Raghavendra U. A review on automatic detection and recognition of traffic sign // *Multimedia Tools and Applications*. 2011. V. 75 (1). P. 333–364.
3. Chen Y., Xie Y., Wang Y. Detection and recognition of traffic signs based on HSV vision model and shape features // *Journal of Computers*. 2013. V. 8 (5). P. 1366–1370.
4. Zaklouta F., Stanciulescu B. Real-time traffic sign recognition in three stages // *Robotics and Autonomous Systems*. 2014. V. 62 (1). P. 16–24.
5. Creusen I.M., Wijnhoven R.G.J., Herbschleb E., de With P.H.N. Color exploitation in HOG-based traffic sign detection. URL: <https://ieeexplore.ieee.org/document/5651637> (accessed: 11.08.2022).
6. Prieto M.S., Allen A.R. Using self-organising maps in the detection and recognition of road signs // *Image and Vision Computing*. 2009. V. 27 (6). P. 673–683.
7. Garcia-Garrido M.A., Ocaña M., Llorca D.F., Sotelo M.A., Arroyo E., Llamazares A. Robust traffic signs detection by means of vision and v2i communications. URL: <https://ieeexplore.ieee.org/document/6082844> (accessed: 11.08.2022).
8. Gonzalez A., Garrido M.A.G., Fernandez-Llorca D., Gavilan M. Automatic traffic signs and panels inspection system using computer vision // *IEEE Transactions on Intelligent Transportation Systems*. 2011. V. 12 (2). P. 485–499.
9. Greenhalgh J., Mirmehdi M. Recognizing text-based traffic signs. // *IEEE Transactions on Intelligent Transportation Systems*. 2015. V. 16 (3). P. 1360–1369.
10. Baro X., Escalera S., Vitria J., Pujol O., Radeva P. Traffic sign recognition using evolutionary AdaBoost detection and Forest-ECOC classification // *IEEE Transactions on Intelligent Transportation Systems*. 2009. V. 10 (1). P. 113–126.
11. Jimenez P.G., Bascon S.M., Moreno H.G., Arroyo S.L., Ferreras F.L. Traffic sign shape classification and localization based on the normalized FFT of the signature of blobs and 2D homographies // *Signal Processing*. 2008. V. 88 (12). P. 2943–2955.
12. Lee H.S., Kim K. Simultaneous traffic sign detection and boundary estimation using convolutional neural network // *IEEE Transactions on Intelligent Transportation Systems*. 2018. V. 19 (5). P. 1652–1663.
13. Kamal U., Tonmoy T.I., Das S., Hasan M.K. Automatic traffic sign detection and recognition using SegU-Net and a modified Tversky loss function with L1-constraint // *IEEE Transactions on Intelligent Transportation Systems*. 2020. V. 21 (4). P. 1467–1479.
14. Badrinarayanan V., Kendall A., Cipolla R. SegNet: A deep convolutional encoder-decoder architecture for image segmentation // *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2017. V. 39 (12). P. 2481–2495.
15. Ronneberger O., Fischer P., Brox T. U-net: Convolutional networks for biomedical image segmentation // *Medical Image Computing and Computer Assisted Intervention : Proc. of the 21st International Conference*. 2015. V. 9351. P. 234–241.
16. Simonyan K., Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. URL: <https://arxiv.org/abs/1409.1556v6> (accessed: 11.08.2022).
17. Ahmed S., Kamal U., Hasan M.K. DFR-TSD: A Deep Learning Based Framework for Robust Traffic Sign Detection Under Challenging Weather Conditions // *IEEE Transactions on Intelligent Transportation Systems*. 2021. V. 23 (6). P. 5150–5162.
18. Sanchez S.A., Romero H.J., Morales A.D. A review: Comparison of performance metrics of pretrained models for object detection using the TensorFlow framework. URL: https://www.researchgate.net/publication/342570032_A_review_Comparison_of_performance_metrics_of_pretrained_models_for_object_detection_using_the_TensorFlow_framework (accessed: 11.08.2022).
19. Bochkovskiy A., Wang C.-Ya., Mark L.H.-Y. YOLOv4: Optimal Speed and Accuracy of Object Detection. URL: <https://arxiv.org/abs/2004.10934> (accessed: 11.08.2022).
20. Houben S., Stallkamp J., Salmen J., Schlipsing M., Igel C. Detection of traffic signs in real-world images: The German traffic sign detection benchmark. URL: <https://ieeexplore.ieee.org/document/6706807> (accessed: 11.08.2022).
21. Timofte R., Zimmermann K., Van Gool L. Multi-view traffic sign detection, recognition, and 3D localization // *Machine Vision and Applications*. 2014. V. 25 (3). P. 633–647.
22. 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.
23. Temel D., Alshawi T., Chen M.-H., AlRegib G. CURE-TSD: Challenging unreal and real environments for traffic sign detection // *IEEE Dataport*. URL: <https://iee-dataport.org/open-access/cure-tds-challenging-unreal-and-real-environment-traffic-sign-detection> (accessed: 11.08.2022).
24. Mathieu M., Couprie C., LeCun Y. Deep multi-scale video prediction beyond mean square error. URL: <https://arxiv.org/abs/1511.05440> (accessed: 11.08.2022).
25. Zhao H., Gallo O., Frosio I., Kautz J. Loss functions for neural networks for image processing. URL: <http://arxiv.org/abs/1511.08861> (accessed: 11.08.2022).

References

1. Mogelmoose, A., Trivedi, M.M. & Moeslund, T.B. (2012) Vision-based traffic sign detection and analysis for intelligent driver assistance systems: Perspectives and survey. *IEEE Transactions on Intelligent Transportation Systems*. 13(4). pp. 1484–1497. DOI: 10.1109/TITS.2012.2209421
2. Gudigar, A., Shreesh, C. & Raghavendra, U. (2016) A review on automatic detection and recognition of traffic sign. *Multimedia Tools and Applications*. 75(1). pp. 333–364. DOI: 10.1007/s11042-014-2293-7
3. Chen, Y., Xie, Y. & Wang, Y. (2013) Detection and recognition of traffic signs based on HSV vision model and shape features. *Journal of Computers*. 8(5). pp. 1366–1370. DOI: 10.4304/jcp.8.5.1366-1370
4. Zaklouta, F. & Stanculescu, B. (2014) Real-time traffic sign recognition in three stages. *Robotics and Autonomous Systems*. 62(1). pp. 16–24. DOI: 10.1016/j.robot.2012.07.019
5. Creusen, I.M., Wijnhoven, R.G.J., Herbschleb, E. & de With, P.H.N. (2010) *Color exploitation in HOG-based traffic sign detection*. [Online] Available from: <https://ieeexplore.ieee.org/document/5651637>. (Accessed: 11th August 2022).
6. Prieto, M.S. & Allen, A.R. (2009) Using self-organising maps in the detection and recognition of road signs. *Image and Vision Computing*. 27(6). pp. 673–683. DOI: 10.1016/j.imavis.2008.07.006
7. Garcia-Garrido, M.A., Ocaña, M., Llorca, D.F., Sotelo, M.A., Arroyo, E. & Llamazares, A. (2011) *Robust traffic signs detection by means of vision and v2i communications*. [Online] Available from: <https://ieeexplore.ieee.org/document/6082844>. (Accessed: 11th August 2022).
8. Gonzalez, A., Garrido, M.A.G., Fernandez-Llorca, D. & Gavilan, M. (2011) Automatic traffic signs and panels inspection system using computer vision. *IEEE Transactions on Intelligent Transportation Systems*. 12(2). pp. 485–499. DOI: 10.1109/TITS.2010.2098029
9. Greenhalgh, J. & Mirmehdi, M. (2015) Recognizing text-based traffic signs. *IEEE Transactions on Intelligent Transportation Systems*. 16(3). pp. 1360–1369. DOI: 10.1109/TITS.2014.2363167
10. Baro, X., Escalera, S., Vitria, J., Pujol, O. & Radeva, P. (2009) Traffic sign recognition using evolutionary AdaBoost detection and Forest-ECOC classification. *IEEE Transactions on Intelligent Transportation Systems*. 10(1). pp. 113–126. DOI: 10.1109/TITS.2008.2011702
11. Jimenez, P.G., Bascon, S.M., Moreno, H.G., Arroyo, S.L. & Ferreras, F.L. (2008) Traffic sign shape classification and localization based on the normalized FFT of the signature of blobs and 2D homographies. *Signal Processing*. 88(12). pp. 2943–2955. DOI: 10.1016/j.sigpro.2008.06.019
12. Lee, H.S. & Kim, K. (2018) Simultaneous traffic sign detection and boundary estimation using convolutional neural network. *IEEE Transactions on Intelligent Transportation Systems*. 19(5). pp. 1652–1663. DOI: 10.48550/arXiv.1802.10019
13. Kamal, U., Tonmoy, T.I., Das, S. & Hasan, M.K. (2020) Automatic traffic sign detection and recognition using SegU-Net and a modified Tversky loss function with L1-constraint. *IEEE Transactions on Intelligent Transportation Systems*. 21(4). pp. 1467–1479.
14. Badrinarayanan, V., Kendall, A. & Cipolla, R. (2017) SegNet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 39(12). pp. 2481–2495. DOI: 10.48550/arXiv.1511.00561
15. Ronneberger, O., Fischer, P. & Brox, T. (2015) U-net: Convolutional networks for biomedical image segmentation. *Proc. of the 21st Medical Image Computing and Computer Assisted Intervention International Conference*. 9351. pp. 234–241.
16. Simonyan, K. & Zisserman, A. (n.d.) *Very Deep Convolutional Networks for Large-Scale Image Recognition*. [Online] Available from: <https://arxiv.org/abs/1409.1556v6>. (Accessed: 11th August 2022).
17. Ahmed, S., Kamal, U. & Hasan, M.K. (2021) DFR-TSD: A Deep Learning Based Framework for Robust Traffic Sign Detection Under Challenging Weather Conditions. *IEEE Transactions on Intelligent Transportation Systems*. 23(6). pp. 5150–5162. DOI: 10.13140/RG.2.2.18341.86249
18. Sanchez, S.A., Romero, H.J. & Morales, A.D. (2020) A review: Comparison of performance metrics of pretrained models for object detection using the TensorFlow framework. *IOP Conference Series Materials Science and Engineering*. 844:012024. DOI: 10.1088/1757-899X/844/1/012024
19. Bochkovskiy, A., Wang, C.-Ya. & Mark, L.H.-Y. (2020) *YOLOv4: Optimal Speed and Accuracy of Object Detection*. [Online] Available from: <https://arxiv.org/abs/2004.10934>. (Accessed: 11th August 2022).
20. Houben, S., Stallkamp, J., Salmen, J., Schlipsing, M. & Igel, C. *Detection of traffic signs in real-world images: The German traffic sign detection benchmark*. [Online] Available from: <https://ieeexplore.ieee.org/document/6706807> (Accessed: 11th August 2022).
21. Timofte, R., Zimmermann, K. & Van Gool, L. (2014) Multi-view traffic sign detection, recognition, and 3D localization. *Machine Vision and Applications*. 25(3). pp. 633–647. DOI: 10.1007/s00138-011-0391-3
22. 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 sistemy – Sensory Systems*. 32(3). pp. 236–245. DOI: 10.1134/S0235009218030058
23. Temel, D., Alshawi, T., Chen, M.-H. & AlRegib, G. (2021) *CURE-TSD: Challenging unreal and real environments for traffic sign detection*. *IEEE Dataport*. [Online] Available from: <https://iee-dataport.org/open-access/cure-tds-challenging-unreal-and-real-environment-traffic-sign-detection>. (Accessed: 11th August 2022).
24. Mathieu, M., Couprie, C. & LeCun, Y. (2016) *Deep multi-scale video prediction beyond mean square error*. [Online] Available from: <https://arxiv.org/abs/1511.05440>. (Accessed: 11th August 2022).
25. Zhao, H., Gallo, O., Frosio, I. & Kautz, J. (n.d.) *Loss functions for neural networks for image processing*. [Online] Available from: <http://arxiv.org/abs/1511.08861>. (Accessed: 11th August 2022).

Информация об авторах:

Харченко Игорь Константинович – аспирант кафедры автоматизированных систем управления Томского государственного университета систем управления и радиоэлектроники (Томск, Россия). E-mail: igor.k.kharchenko@tusur.ru

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

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

Information about the authors:

Kharchenko Igor Konstantinovich (Post-graduate student, Tomsk State University of Control Systems and Radioelectronics, Tomsk, Russian Federation). E-mail: igor.k.kharchenko@tusur.ru

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

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

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

Received 12.08.2022; accepted for publication 29.11.2022

Поступила в редакцию 12.08.2022; принята к публикации 29.11.2022

ИНФОРМАТИКА И ПРОГРАММИРОВАНИЕ

INFORMATICS AND PROGRAMMING

Научная статья

УДК 519.233.22

doi: 10.17223/19988605/61/10

Перечисление симметричных ожерелий

Юрий Дмитриевич Григорьев

Государственный электротехнический университет «ЛЭТИ», Санкт-Петербург, Россия, yuri_grigoriev@mail.ru

Аннотация. Рассматривается задача перечисления симметричных ожерелий. С помощью разработанного алгоритма множество ожерелий с заданными распределениями бусин и заданной длины расщепляется на множества симметричных и несимметричных ожерелий. Рассматриваются зеркальная симметрия и ее модификации – ахиральная и хиральная. Вводится классификация ожерелий по типам и классам симметрии. Уточняются результаты Яковенко для подсчета мощности классов симметрии. Рассматриваются приложения полученной классификации к анализу восьмистиший – ожерелий длины 8, среди которых выделяются твердые формы – триолет, сицилиана, октава и другие, а также иные симметричные восьмистишные строфы, встречающиеся в практике стихосложения.

Ключевые слова: ожерелье; перечисление симметричных ожерелий; группа симметрий; строфа

Для цитирования: Григорьев Ю.Д. Перечисление симметричных ожерелий // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2022. № 61. С. 97–107. doi: 10.17223/19988605/61/10

Original article

doi: 10.17223/19988605/61/10

Enumeration of symmetrical necklaces

Yury D. Grigoriev

Electrotechnical University “LETI”, St. Petersburg, Russian Federation, yuri_grigoriev@mail.ru

Abstract. The problem of the symmetrical stanzas enumeration, whose models are necklaces or simple cycles of a given length, is considered. The automorphisms of the necklace symmetry group form a dihedral group, whose action on the set of necklaces allows us to distribute them over orbits. With help of a specially developed algorithm, the orbits are divided into sets of symmetric and asymmetric necklaces, while the symmetrical necklaces are classified into types and classes of symmetry. Mirror or axial symmetry, as well as its variants, achiral and chiral symmetries, are considered. Analytical results are given for calculating the cardinality of symmetry classes (of representative graphs) belonging to the same type of symmetry.

Applications of the resulting classification to the analysis of octets, for which various rhyming schemes are used as a attribute of symmetry, are considered. The analysis of the applicability of symmetrical stanzas in the practice of versification has been carried out. It is shown that within the framework of the proposed approach, all known hard forms of versification for octets, such as the triolet, siciliane, octave, and others, are symmetrical; examples of rarely used symmetrical stanzas are listed, indicating the possible reasons for this phenomenon. The possibility of using the proposed approach in the problem of choosing the rhythmic profile of a stanza, which generates the symmetry of the rhythm, is shown.

Keywords: necklace; symmetry of necklaces; groups of necklaces symmetries; enumeration of necklaces; stanza

For citation: Grigoriev Yu.D. (2022) Enumeration of symmetrical necklaces. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naja tehnika i informatika – Tomsk State University Journal of Control and Computer Science*. 61. pp. 97–107. doi: 10.17223/19988605/61/10

Задача об ожерелье является одной из хорошо известных в дискретной математике [1. С. 246; 2. С. 212]. На ее примере наглядно демонстрируется один из таких конструктивных методов теории групп, как действие некоторой группы симметрий $G = \langle X, * \rangle$ на заданное множество ожерелий [3. С. 301]. Основной вопрос, который встает при исследовании симметрий ожерелий, заключается не столько в их подсчете в некотором множестве, сколько в их реальном перечислении – построении алгоритма получения ожерелий в явном виде [4]. В рассматриваемом случае мы сталкиваемся с феноменом так называемой хиральности – разновидностью симметрии, не совпадающей с осевой симметрией. Хиральность как атрибут многих сложных систем является объектом изучения в общей теории симметрии [5–7], органической химии [8, 9], кристаллографии [10] и многих других научных дисциплинах.

Цель статьи – рассмотрение алгоритмического подхода к решению задачи перечисления симметричных ожерелий и перенесение ее на теорию поэтических структур – строф, моделями которых служат ожерелья. Такое моделирование позволяет не только выявить симметрии поэтических строф, но и расширить их классификацию за счет внесения в нее различных признаков симметрии типа правила альтернанса или правила скрещения рифмических цепей [11. С. 195]. Эта задача является тем более актуальной, что проблема аналитического перечисления симметричных ожерелий в явном виде на сегодняшний день остается в полном объеме нерешенной.

1. Постановка задачи

В [12. С. 14, 18] была предложена модель строфических структур в виде кольцевых диаграмм, или *ожерелий* [1. С. 246]. Такое представление позволяет эффективно исследовать свойства поэтической строфы с позиций симметрии при условии, что в качестве признаков симметрии строфы рассматриваются только такие, которые допускают строгую формализацию, – системы рифмовок, слоговый объем стихов, ритмический профиль и т.д.

Рассмотрим стихотворение А. Блока «На зов метелей» [13. С. 165]. Система рифмовки его строфы IV соответствует симметричному ожерелью *abcb adcd* (см. ниже: рис. 3, ожерелье 2 в табл. 2), где одинаковыми буквами обозначены звенья соответствующих рифмических цепей:

На зов метелей

IV'

И мгла заломила руки,
Заломила руки ввысь.
Ты опустила очи,
И мы понеслись.
И навстречу вставали новые звуки:
Летели снега,
Налетающей ночи
Звенели рога.

Эта восьмистишная строфа симметрична, и все бы хорошо, но дело в том, что здесь я осуществляю подлог – на самом деле у Блока два последних стиха переставлены местами. В результате этого симметрия нарушается, и все наши красивые рассуждения рушатся. Возникает вопрос: может, Блок на подсознательном уровне стремился к симметрии и просто ошибся в чередовании двух последних стихов? О приведении строф к симметрии говорится в [14], где, в частности, анализируется

отклонение от симметрии в стихотворении Д. Мережковского «Март» и высказывается предположение, что в этом стихотворении вообще пропущен один стих, нарушающий симметрию.

В статье решается вопрос, сколько симметричных (m_1, m_2, \dots, m_r) -ожерелий длины n , $\sum_{i=1}^r m_i = n$, существует при заданном распределении его бусин (m_i – число бусин i -го цвета). Цвета бусин применительно к стихосложению определяют систему рифмовки и рассматриваются как признак симметрии строфы. Так, варианту Блока соответствует несимметричное $(2, 2, 2, 2)$ -ожерелье, а после его реконструкции (второй вариант) – симметричное, хотя в обоих случаях распределение бусин одно и то же. Задача алгоритмического подсчета, перечисления и анализа симметричных ожерелий длины n иллюстрируется в статье на примере 8-стишных строф.

2. Симметричные ожерелья

Пусть T – множество ожерелий $x \in T$, в котором определено отношение эквивалентности \sim , G – множество операторов $g \in G$, $g: T \rightarrow T$, т.е. $gx \in T$, $G \times G \rightarrow G$ – композиция операторов, относительно которой G является группой, в общем случае неабелевой.

2.1. Симметрия

Пусть G действует на $x \in T$, при этом $g_1(g_2x) = (g_1g_2)x$, $ex = x$. Ожерелье $x \in T$ называется симметричным, если существует оператор $g \neq e$ такой, что $gx \sim x$. Метризуем множество T . Пусть $d(x, y)$ – число несовпадающих бусин ожерелий $x, y \in T$, взятых в порядке следования. Это известное расстояние Хемминга [15. С. 52], используемое для строк одинаковой длины любых r -ичных алфавитов. Очевидно, ожерелье $x \in T$ симметрично тогда и только тогда, когда существует элемент $g \neq e$ такой, что $d(gx, x) = 0$. В связи с этим такие операторы g называются *изометриями*. Это определение симметричности эквивалентно предыдущему.

Вопрос о количестве $N(m_1, \dots, m_r)$ орбит, индуцируемых действием на множество T какой-либо группы G , решается с помощью леммы Бернсайда. Однако на вопрос о том, сколько существует симметричных орбит (орбит, содержащих симметричные ожерелья), лемма Бернсайда ответа не дает.

2.2. Представительные графы

Расстояние $\rho(a, b)$ между двумя бусинами a и b ожерелья x определим как цепь наименьшей длины, связывающую a и b . Кроме того, ясно что для $\forall a, b: \rho(a, b) \in \{1, 2, \dots, [n/2]\}$. Например, для соседних узлов a и b имеем $\rho(a, b) = 1$. Легко проверить что $\rho(a, b)$ – метрика [16. С. 42].

Сопоставим (m_1, m_2, \dots, m_r) -ожерелью $x \in T$ длины n следующую $(n, n - q)$ -диаграмму H с n вершинами и $n - q$ ребрами – отрезками прямых линий: (а) вершины диаграммы H определим как вершины правильного n -угольника; (б) ребра H – как звенья r компонент связности длины m_i , образующие цепи наименьшей длины (если $m_i = 2$, то полагаем, что цепь состоит из одного звена); (в) q – количество координат m_i , равных 2. Так, если рассматривается $(3, 3, 2)$ -ожерелье, то $n = 8$, $r = 3$, $q = 1$. Следовательно, ожерелью x сопоставляется $(8, 7)$ -диаграмма H , содержащая 8 вершин и 7 ребер, образующих 3 компоненты связности. Ребра диаграммы H при их изображении на плоскости могут пересекаться, но точки их пересечения вершинами H не являются. Описанная $(n, n - q)$ -диаграмма однозначно определяет ожерелье $x \in T$ и отличается от обычного определения $(n, n - q)$ -графа лишь дополнительным требованием (а). Такая диаграмма названа в [12. С. 21] *представительным графом* симметричного ожерелья x . Этот граф будем считать непомеченным.

Каждый граф H имеет группу автоморфизмов $\Gamma = \Gamma(H)$, состоящую из изоморфизмов H на себя [16. С. 311]. Среди $n!$ перестановок n узлов H только $2n$ перестановок являются изометриями в смыс-

ле метрики $\rho(\cdot, \cdot)$. Данные изометрии сохраняют инвариантными $n(n-1)/2$ расстояний между узлами графа H . Очевидно, это вращения и зеркальные отражения, которые в совокупности образуют группу диэдра

$$D_n = \{s, t : s^n = t^2 = (st)^2 = e\},$$

где s – вращение на угол $2\pi/n$, t – отражение. Поскольку далее предметом нашего интереса будут восьмистишия, кратко охарактеризуем свойства симметрий группы D_8 [7]. Для этого выделим в ней подгруппу G^+ и ее дополнение $G^- = G \setminus G^+$:

$$G^+ = \{s^{8-2i} : i = 0, \dots, 3\}, \quad G^- = \{s^{9-2i}, ts^{8-j} : i = 0, \dots, 3, j = 0, \dots, 7\}.$$

Группа $G^+ = G'$ – коммутант D_8 , при этом $G/G' \cong D_2$. Изометрии $g \in G^+$ – это вращения. Дополнение $G^- = G \setminus G^+$ группой не является, но, как и G^+ , состоит из объединения классов сопряженности группы D_8 , из них два множества – $S_1, S_2 \in G^-$ – представляют особый интерес:

$$S_1 = \{t, ts^2, ts^4, ts^6\}, \quad S_2 = \{ts, ts^3, ts^5, ts^7\}.$$

Множество S_1 содержит отражения относительно диагоналей правильного 8-угольника. Их можно представить в виде композиций отражений относительно осей симметрии, проходящих через середины сторон 8-угольника, и вращений. Множество S_2 – это отражения относительно осей симметрии, проходящих через середины противоположных сторон 8-угольника. Эти отражения нельзя получить с помощью композиций отражений, принадлежащих S_1 , и последующих вращений. Оба множества при любом $n = 2m$ содержат одинаковое число элементов. Существование множеств S_1 и S_2 – источник ахиральной и хиральной симметрии ожерелий.

2.3. Хиральность

С понятием симметрии тесно связан феномен хиральности. Согласно лорду Кельвину, классическое определение хиральности звучит так: «Я называю любую геометрическую фигуру или группу точек *хиральной* и говорю, что она обладает *хиральностью*, если ее отражение в идеально плоском зеркале не может быть перенесено так, чтобы оно совпало с нею самой» [7]. Применительно к нашим обозначениям это определение формулируется следующим образом: ожерелье $x \in T$ *хирально*, если оно неэквивалентно (относительно вращений) своему зеркальному отражению $y = \varphi(x)$. В противном случае ожерелье x *ахирально*, т.е. $x \sim \varphi(x)$. Исходя из этого определения ахиральность трактуется как искаженная зеркальная симметрия, а хиральность означает отсутствие зеркальной симметрии. Таким образом, ахиральность и хиральность являются характеристическими свойствами множеств $S_1, S_2 \in D_8$.

Комбинаторная проблема перечисления хиральных объектов в R^2 связана с проблемой перечисления автоморфизмов 2-го порядка группы симметрий G [5. С. 278]. Поскольку группой автоморфизмов ожерелья как простого (n, n) -цикла [2. С. 26] является группа диэдра D_n , содержащая отражения, т.е. автоморфизмы порядка 2, то в множестве T каждое ожерелье a сопровождает двойственный ему элемент b , образующий с ним *энантиомерную пару* (a, b) , при этом энантиомеры a и b принадлежат разным орбитам. В приложениях этот факт хорошо известен [8. С. 63]. Очевидно также, что если ожерелье a ахирально, т.е. $a \sim \varphi(a)$, то энантиомеры a и $b = \varphi(a)$ принадлежат одной орбите.

Перечисленные виды симметрии могут иметь разный порядок, т.е. представительные графы H могут иметь несколько осей симметрии того или иного порядка. В связи с этим возможны следующие *типы* симметрий $[k, l]$:

$$[k, 0], [k, k], [0, l], [0, 0], \quad k, l \geq 1,$$

где $k > 0$ – порядок зеркальной или ахиральной симметрии, $l > 0$ – порядок хиральной симметрии. Ахиральная симметрия может встречаться только в сочетании с осевой, при этом всегда $k = l > 0$. Если $k = l = 0$, то ожерелье по определению несимметрично. К этому случаю из чисто алгоритмиче-

ских соображений будем относить и центральную симметрию – вращение вокруг точки, которое можно представить как композицию двух зеркальных симметрий [16. С. 52–56].

Одному и тому же типу симметрии может соответствовать несколько классов симметрии. Под ожерельями разных классов, принадлежащими одному типу симметрии, понимаются ожерелья, графы H которых изоморфны [2. С. 24], но не могут быть переведены друг в друга никаким движением (сдвигом, вращением, отражением или их композициями), т.е. каким-либо изометрическим преобразованием [18. С. 278]. Ожерелья, которые могут быть переведены одно в другое некоторым движением, назовем *конгруэнтными*. Такие ожерелья принадлежат одному классу симметрии.

Пример 1. На рис. 1, 2 представлена энантиомерная пара хиральных ожерелий, представляющая строфы I и II стихотворения А. Пушкина «Осень». Строфы отличаются тем, что в них А. Пушкин меняет местами мужские и женские рифмы. Аналогично построены и последующие пары строф. Сами же строфы представляют знаменитую твердую форму – октаву [11. С. 172]. Ей соответствует симметрия $[0, 1]$.

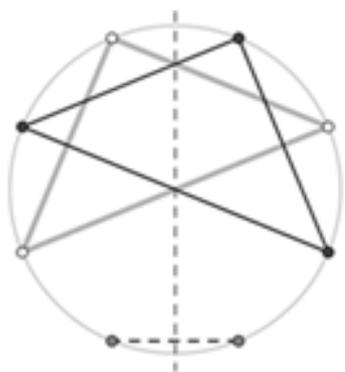


Рис. 1. А. Пушкин. Осень («Октябрь уж наступил...»).

Строфа I: abab abcc (ожерелье 11 в табл. 1)

Fig. 1. A. Pushkin. Autumn («October has already come...»).

The stanza I: abcb adcd (necklace 11 in the table 1)



Рис. 2. А. Пушкин. Осень («Теперь моя пора...»).

Строфа II: baba baccc (ожерелье 11 в табл. 1)

Fig. 2. A. Pushkin. Autumn («Now is my time...»).

The stanza II: abcd cabd (necklace 11 in the table 1)

3. Перечисление ожерелий

Аналитический подсчет числа представительных графов является нерешенной в общем случае задачей. Некоторое продвижение в ее решении можно получить в частных случаях, используя лемму Бернсайда.

Теорема 1. Количество $N_C(2, n-2)$ представительных графов $(2, n-2)$ -ожерелий равно

$$N_C(2, n-2) = [n/2]. \quad (1)$$

Доказательство. Пусть $\varphi(x)$ – функция Эйлера [19. С. 28], $P(k, l) = (k+l)!/k!l!$ – биномиальный коэффициент, НОД (k, l) – наибольший общий делитель целых чисел k и l . Согласно лемме Бернсайда количество орбит $(2, n-2)$ -ожерелий равно

$$N_O(2, n-2) = n^{-1} \sum_{d|\text{НОД}(2, n-2)} \varphi(d) P(2/d, (n-2)/d), \quad (2)$$

где $d = 1, 2$ или $d = 1$ в случаях четного и нечетного n соответственно. Рассматривая эти случаи отдельно, а затем объединяя полученный результат в один, получаем $N_O(2, n-2) = [n/2]$. С другой стороны, очевидно, что конгруэнтными являются только те ожерелья, расстояния $\rho(a, b) = 1, 2, \dots, [n/2]$ между бусинами a и b которых одного цвета совпадают. Из геометрических соображений следует, что для любых этих значений $\rho(a, b)$ данные ожерелья симметричны. Поскольку все $(2, n-2)$ -ожерелья независимо от их длины являются зеркально симметричными, это означает, что $N_C(2, n-2) = N_O(2, n-2)$, т.е. справедливо равенство (1).

Задача перечисления представительных графов как задача о (m_1, m_2, \dots, m_r) -ожерельях в общем виде рассматривалась Д.И. Яковенко [20]. В этой работе с опорой на лемму Бернсайда сформулировано несколько результатов для случаев четного и нечетного n . Однако проверка показала, что не все они являются верными. После уточнения некоторых ее утверждений может быть сформулирована следующая

Теорема 2. Пусть $x - (m_1, m_2, \dots, m_r)$ -ожерелье длины n , $P(m_1, m_2, \dots, m_r)$ – полиномиальный коэффициент, $N_C(m_1, m_2, \dots, m_r)$ – количество представительных графов ожерелья x . Имеют место следующие утверждения:

P1. Пусть n нечетно. Тогда если среди чисел m_1, m_2, \dots, m_r только одно нечетно, а все остальные попарно различны и четны, то

$$N_C(m_1, m_2, \dots, m_r) = P([m_1/2], [m_2/2], \dots, [m_r/2]) . \quad (3)$$

P2. Пусть n четно. Тогда если все числа m_1, m_2, \dots, m_r четны и попарно различны, то имеет место (3).

Доказательство теоремы 2 по аналогии с [20].

Задачу перечисления ожерелий в общем виде теорема 2 не решает. В связи с этим в [12] разработан универсальный алгоритм перечисления представительных графов, на основе которого составлены обширные таблицы симметричных ожерелий [12. Табл. 2.6–2.15]. К сожалению, описание алгоритма перечисления в [12. С. 23–25] недостаточно конструктивно для возможности его программной реализации, что фактически означает утрату этого алгоритма.

Разработанный нами в среде MAPLE алгоритм перечисления относится к переборному типу. Вследствие этого он требует больших затрат машинного времени и на данный момент не дает возможности решать задачи большой размерности. Отметим, что результаты, полученные Л.Г.Портером и с помощью нашего алгоритма до $n = 8$ включительно, полностью совпадают.

4. Перечисление симметрий восьмистиший

Восьмистишие – одна из самых распространенных и семантически нейтральных строф [11. С. 172]. Разнообразие таких строф определяется широкой возможностью разбиения числа 8 на слагаемые, нетривиальные из которых имеют вид:

$$\begin{aligned} 8 &= 6 + 2 \\ &= 5 + 3 = 4 + 4 \\ &= 3 + 3 + 2 = 4 + 2 + 2 = 2 + 2 + 2 + 2 \end{aligned} \quad (4)$$

Из (4) следует, что существует 6 видов 8-стишных строф со схемами рифмовки на две, три и четыре рифмы. В табл. 1 представлены сводные данные о симметричных ожерельях длины 8 на две и три рифмы.

Таблица 1

Перечень симметрий (k, l) - и (k, l, m) -ожерелий

	Симметрии (k, l) -ожерелий			Симметрии (k, l, m) -ожерелий					
	(6, 2)		(5, 3)	(4, 4)		(3, 3, 2)		(4, 2, 2)	
1	[1, 0]	aaaa aabb	aaab bbbb	[1, 0]	aaab babb	[1, 0]	aaab cbc b	[1, 0]	aaaa bccb
2	[1, 0]	aaaa abab	abaa abab	[1, 1]	aaaa bbbb	[1, 0]	abab acbc	[1, 0]	aaab cacb
3	[1, 0]	aaaa abba	aaba abab	[1, 1]	aaba bbab	[1, 1]	aaac bbbc	[1, 0]	aaba acbc
4	[1, 0]	aaab aaab		[2, 2]	aabb aabb	[1, 1]	abac babc	[1, 0]	acca baab
5				[4, 4]	abab abab	[0, 1]	aaab bbcc	[1, 1]	aabb aacc
6				[0, 1]	aaab abbb	[0, 1]	aaba bbcc	[1, 1]	aabc aacb
7				[0, 1]	abba abab	[0, 1]	aabb accb	[1, 1]	abab acac
8						[0, 1]	aabb cabc	[2, 2]	abac abac
9						[0, 1]	aabb cbac	[0, 1]	aaaa bbcc
10						[0, 1]	aabc abbc	[0, 1]	aaaa bc bc
11						[0, 1]	abab abcc	[0, 1]	aaab bacc
12						[0, 1]	abab cabc	[0, 1]	aaab cabc

	Симметрии (k, l)-ожерелий			Симметрии (k, l, m)-ожерелий	
	(6, 2)	(5, 3)	(4, 4)	(3, 3, 2)	(4, 2, 2)
13				[0, 1] abab cbac	[0, 1] abab bcbc
14					[0, 1] aaba cbac
15					[0, 2] aabc aabc
N_c	4	3	7	13	15
N_o	4	7	10	70	54
P	28	56	70	560	420

Примечание. N_c – количество симметричных ожерелий, N_o – количество орбит, $P = P(k, l)$ – всего ожерелий.

Поясним последнюю строку в табл. 1, например число 420. Орбиты в зависимости от симметрии ожерелий содержат разное число ожерелий. В данном случае три (4, 2, 2)-орбиты содержат по 4 ожерелья, а остальные – по 8, что в итоге дает $(54 - 3) \times 8 + 3 \times 4 = 420$ ожерелий. Симметрии (2, 2, 2, 2)-ожерелий приведены табл. 2 и табл. 3 соответственно.

Таблица 2

Перечень симметрий (2, 2, 2, 2)-ожерелий

	Тип	Класс			Тип	Класс	
		Ожерелье	Рифмовка			Ожерелье	Рифмовка
1	[1, 1]	abba cdde	CC	9	[0, 1]	aabc cdbd	
2	[1, 1]	abac dbdc		10	[0, 1]	abac dbcd	
3	[0, 1]	aabb cdde	AC BC – ноэль	11	[0, 2]	aabc ddbc	BB
4	[0, 1]	abab cdde		12	[0, 2]	abcd acbd	
5	[0, 1]	abcd bdac		13	[0, 2]	abab cdcd	
6	[0, 1]	aabc bdc d		14	[0, 4]	aabb ccdd	
7	[0, 1]	aabc dbcd	15	[0, 4]	abca dcbd		
8	[0, 1]	aabb cdcd	AB	16	[0, 8]	abcd abcd	

Примечание. Рифмовки в катрене: A – смежная aabb, B – перекрестная abab, C – охватная abba.

Поясним табл. 2. Число (2, 2, 2, 2)-ожерелий и соответствующее им число орбит согласно лемме Бернсайда соответственно равны:

$$P(2, 2, 2, 2) = 2\,520, \quad N_o(2, 2, 2, 2) = 318.$$

Среди 318 орбит 312 состоят из 8 ожерелий, а 6 – из 4, т. е. $(318 - 6) \times 8 + 6 \times 4 = 2\,520$ ожерелий. В рассматриваемом случае 48 ожерелий образуют класс [0, 0] несимметричных ожерелий, а оставшиеся в 16 классах $318 - 48 = 270$ симметричных ожерелий распределяются по пяти типам симметрии согласно табл. 3.

Таблица 3

Распределение симметрий (2, 2, 2, 2)-ожерелий по типам и классам

	Тип симметрии						Всего
	[1, 1]	[0, 1]	[0, 2]	[0, 4]	[0, 8]	[0, 0]	
Кол-во классов	2	8	3	2	1		318
Мощность класса	12	24	12	6	6	48	
Кол-во ожерелий	2×12	8×24	3×12	2×6	1×6	48	

5. Ожерелья, описывающие метрический профиль строфы

Будем рассматривать ритм как систему расположения ударений на сильных слоговых позициях метра [21. С. 109]. Пусть заданный размер строфы – 4-стопный ямба. Подсчитаем, сколько раз стоят ударения на сильных слогах каждого стиха, и какой процент это составляет от общего числа n стихов в строфе. Этот процентный ряд образует *ритмический профиль* данной строфы [11. С. 90]. Так как на последней стопе это всегда 100%, то 4-ю стопу далее вообще исключаем из рассмотрения и задаем ритмический профиль вектором $p \in R^3$. В соответствии с этим возможны шесть различных *ритмических форм* 4-стопного ямба:

$$2 = 010, 3 = 011, 4 = 100, 5 = 101, 6 = 110, 7 = 111 \quad (5)$$

(форма 1 = 001 никогда не встречается), где единицы – это стопы, несущие ударение.

Нас будет интересовать зависимость между p и распределением числа ритмических форм $x = (x_2, x_3, \dots, x_7) \in R^6$, представляющих строфу. От взаимного расположения в строфе ритмических форм ритмический профиль p не зависит. Следовательно, один и тот же профиль может достигаться как на симметричных, так и на несимметричных ожерельях. Для однозначного определения распределения (x_2, x_3, \dots, x_7) шести ритмических форм, необходимо задать шесть независимых уравнений. Положим

$$A = F_1^2, B = F_2^2, C = F_3^2, D = F_1F_2, E = F_1F_3, F = F_2F_3, \quad (6)$$

где $F_iF_j = \langle F_i, F_j \rangle$ – скалярные произведения в R^n , F_i – вектор-столбец иктов на i -й стопе.

Теорема 3. Распределение $x = (x_2, x_3, \dots, x_7)$ ритмических форм 4-стопного ямба имеет вид:

$$\begin{aligned} x_2 &= B - C - D + E, & x_3 &= C - E, & x_4 &= A - C - D + F, \\ x_5 &= C - F, & x_6 &= C + D - E - F, & x_7 &= -C + E + F. \end{aligned} \quad (7)$$

Доказательство. Обозначим $y = (n, A, B, C, D, E, F) \in R^7$. Из (5), (6) и тождества $x_2 + x_3 + \dots + x_7 = n$ приходим к равенству $Qx = y$, $Q \in R^{7 \times 6}$. Поскольку $A + B - D = n$, то, исключая координату n из y , можем записать $x = S\tilde{y}$, где $S \in R^{6 \times 6}$, $\tilde{y} = (A, B, C, D, E, F) \in R^6$, а x определяется согласно (7), при этом матрицы Q и S имеют вид:

$$Q = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 \end{pmatrix} \quad S = \begin{pmatrix} 0 & 1 & -1 & -1 & 1 & 0 \\ 0 & 0 & 1 & 0 & -1 & 0 \\ 1 & 0 & -1 & -1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & -1 \\ 0 & 0 & 1 & 1 & -1 & -1 \\ 0 & 0 & -1 & 0 & 1 & 1 \end{pmatrix}$$

Пример. 2. Рассмотрим два первых катрена стихотворения Блока «Я был смущенный и веселый...» [13. С. 188]:

<p>I Я был смущенный и веселый. Меня дразнил твой темный шелк, Когда твой занавес тяжелый Раздвинулся — театр умолк.</p>	<p>II Живым огнем разъединило Нас рампы светлое кольцо, И музыка преобразила И обожгла твое лицо.</p>
--	---

В совокупности будем рассматривать два этих четверостишия как 8-стишную строфу abcd caeb. Ритмическое поле $F_1F_2F_3$, распределение $x = (x_2, x_3, \dots, x_7) \in R^6$ ритмических форм и ритмический профиль (6) имеют вид:

Стихи	Ожерелье	$F_1F_2F_3$	Форма	Профиль
1	a	110	$x_2 = 2$	$A = 4$
2	b	011	$x_3 = 2$	$B = 6$
3	c	010	$x_4 = 1$	$C = 3$
4	d	101	$x_5 = 1$	$D = 2$
5	c	010	$x_6 = 2$	$E = 1$
6	a	110	$x_7 = 0$	$F = 2$
7	e	100		
8	b	011		

Равенства (7) легко проверяются. (2, 2, 2, 1, 1)-ожерелье abcd caeb, описывающее данную строфу, несимметрично. Если, однако, выполнить несложную реконструкцию, а именно заменить ритмиче-

ские формы стихов 7 и 8 на $3 = 011$ и $5 = 101$ соответственно, то профиль $p = (A, B, C)$ изменится с $(4, 6, 3)$, на $(4, 6, 4)$, а ожерелье x преобразуется к $(2, 2, 2, 2)$ -виду $abcd\ cabd$ и станет симметричным (рис. 4; ожерелье 10 в табл. 2). Таким образом, видим, как при изменении ритмического профиля изменяется симметрия ожерелья ритма.



Рис. 3. А. Блок. На зов метелей (реконструкция).
Строфа $abcb\ adcd$ (ожерелье 2 в табл. 2)

Fig. 3. A. Block. To the call of the blizzards (reconstruction).

The stanza $abcb\ adcd$ (necklace 2 in the table 2)

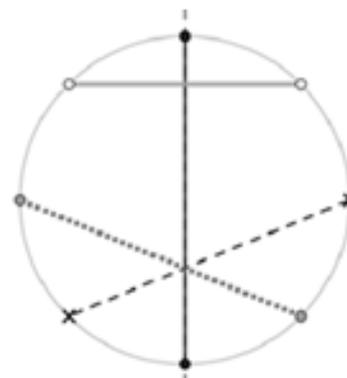


Рис. 4. А. Блок. «Я был смущенный и веселый...» (реконструкция).
Строфа $abcd\ cabd$ (ожерелье 10 в табл. 2)

Fig. 4. A. Block. «I was confused and cheerful...» (reconstruction).

The stanza $abcd\ cabd$ (necklace 10 in the table 2)

Несовпадение кодировок ожерелий в табл. 2 с указанными в тексте и на рис. 3, 4 не должно смущать. С точностью до обозначений и начальной точки отсчета на круговых диаграммах эти кодировки эквивалентны. Последнее связано с тем, что каждое ожерелье на самом деле представляет класс строф с одной и той же системой рифмовки, но с n разными точками отсчета. С математической точки зрения мы имеем здесь n эквивалентных ожерелий, а с точки зрения поэзии – n различных вариантов строфы, которые могут сильно отличаться в плане эстетического восприятия.

Заключение

В статье приведены некоторые результаты перечисления ожерелий с помощью разработанного алгоритма. На основе этих результатов предложены подходы к формальному анализу стихотворных текстов на основе стержневого принципа стихосложения – симметрии. Для принятого в качестве модели строфы понятия ожерелья предложена классификация симметричных ожерелий по типам и классам, которая увязана со свойствами группы автоморфизмов простого цикла. Получена более точная формулировка теоремы Яковенко для подсчета мощности классов симметрии. Перечислены симметричные ожерелья длины 8 с рифменными цепями длины 2, 3 и 4. На языке ожерелий найдена связь распределения ритмических форм 4-стопного ямба с ритмическим профилем строфы.

Список источников

1. Сачков В.Н. Введение в комбинаторные методы дискретной математики. М. : Наука, 1982. 384 с.
2. Харари Ф. Теория графов. М. : Мир, 1973. 302 с.
3. Кострикин А.И. Введение в алгебру. М. : Наука, 1977. 496 с.
4. Харари Ф., Палмер Э. Перечисление графов. М. : Мир, 1977. 324 с.
5. Petitjean M. Chirality and Symmetry Measures. A Transdisciplinary Review // Entropy. 2003. V. 5 (3), P. 217–312.
6. Petitjean M. Definition on Symmetry // Symmetry. Culture and Science. 2007. V. 18 (2-3). P. 99–119.
7. Petitjean M. Chirality in Metrix Spaces // Optim. Lett. 2020. V. 14. P. 329–338.
8. Зоркий П.М. Симметрия молекул и кристаллических структур. М. : Изд-во Моск. ун-та, 1986. 232 с.
9. Марков В.М. Новые критерии степени симметрии и хиральности молекул : автореф. дис. ... канд. хим. наук : 92.00.04. Челябинск, 2001. 23 с.
10. Шубников А.В., Копчик В.А. Симметрия в науке и искусстве. М. : Наука, 1972. 340 с.
11. Гаспаров М.Л. Русский стих начала XX века в комментариях. М. : Фортуна Лимитед, 2001. 288 с.

12. Портер Л.Г. Симметрия – владычица стихов : очерк начал общей теории поэтических структур. М. : Языки славянской культуры, 2003. 256 с.
13. Блок А.А. Покой нам только снится... : [сборник]. М. : АСТ, 2021. 480 с.
14. Владимирова О.В., Григорьев Ю.Д. Поэтические реконструкции, или приведение к симметрии (на примере «Персидских мотивов» С. Есенина) // Вестник Новосибирского государственного университета. Сер. Лингвистика и межкультурная коммуникация. 2020. Т. 18, № 4. С. 16–31.
15. Питерсон У., Уэлдон Э. Коды, исправляющие ошибки. М. : Мир, 1976. 600 с.
16. Кокстер Г.С.М. Введение в геометрию. М. : Наука, 1966. 648 с.
17. Оре О. Теория графов. М. : Наука, 1968. 352 с.
18. Александров П.С. Лекции по аналитической геометрии. М. : Наука, 1968. 912 с.
19. Виноградов И.М. Основы теории чисел. М. : Наука, 1965. 172 с.
20. Яковенко Д.И. Задача об ожерельях // Вестник Омского университета. 1998. № 2. С. 21–24.
21. Баевский В.С. Лингвистические, математические, семиотические и компьютерные модели в истории и теории литературы. М. : Языки славянской культуры, 2001. 336 с. (Studia philologica).

References

1. Sachkov, V.N. (1982) *Vvedenie v kombinatornye metody diskretnoy matematiki* [Introduction to Combinatorial Methods of Discrete Mathematics]. Moscow: Nauka.
2. Harary, F. (1969) *Teoriya grafov* [Graph Theory]. Translated from English by V.P. Kozyreva. Reading, Massachusetts: Addison-Wesley Publishing Company.
3. Kostrikin, A.I. (1977) *Vvedenie v algebru* [Introduction to Algebra]. Moscow: Nauka.
4. Harary, F. & Palmer, E.M. (1973) *Perechislenie grafov* [Graphical Enumeration]. Translated from English. Moscow: Mir.
5. Petitjean, M. (2003) Chirality and Symmetry Measures. A Transdisciplinary Review. *Entropy*. 5(3). pp. 217–312. DOI: 10.3390/e5030271
6. Petitjean, M. (2007) Definition on Symmetry. *Symmetry. Culture and Science*. 18(2-3). pp. 99–119.
7. Petitjean, M. (2020) Chirality in Metrix Spaces. *Optim. Lett.* 14. pp. 329–338.
8. Zorkiy, P.M. (1986) *Simmetriya molekul i kristallicheskih struktur* [Symmetry of Molecules and Crystal Structures]. Moscow: Moscow State University.
9. Markov, V.M. (2001) *Novye kriterii stepeni simmetrii i khiral'nosti molekul* [New criteria for the degree of symmetry and chirality of molecules]. Abstract of Chemistry Cand. Diss. South Ural State University. Chelyabinsk.
10. Shubnikov, A.V. & Koptsik, V.A. (1972) *Simmetriya v nauke i iskusstve* [Symmetry in Science and Art]. Moscow: Nauka.
11. Gasparov, M.L. (2001) *Russkiy stikh nachala XX veka v kommentariyakh* [Russian verse from the early 20th century in the comments]. Moscow: Fortuna Limited.
12. Porter, L.G. (2003) *Simmetriya – vladychitsa stikhov: ocherk nachal obshchey teorii poeticheskikh struktur* [Symmetry is the sovereign of poetry: An essay on the beginnings of a general theory of poetic structures]. Moscow: Yazyki slavyanskoy kul'tury.
13. Block, A.A. (2021) *Pokoy nam tol'ko snitsya* [We only dream of peace]. Moscow: AST.
14. Vladimirova, O.V. & Grigoriev, Yu.D. (2020) Poetic Reconstructions, or Reduction to Symmetry (Based on S. Esenin's Cycle "Persian Motives"). *Vestnik Novosibirskogo gosudarstvennogo universiteta. Ser. Lingvistika i mezhkul'turnaya kommunikatsiya – NSU Vestnik. Series: Linguistics and Intercultural Communication*. 18(4). pp. 16–31. DOI: 10.25205/1818-7935-2020-18-4-16-31
15. Peterson, W.W. & Weldon, E.J. (1972) *Kody, ispravlyayushchie oshibki* [Error-Correcting Codes]. Translated from English. Cambridge, Massachusetts; London, England: The MIT Press.
16. Coxeter, H.S.M. (1961) *Vvedenie v geometriyu* [Introduction to Geometry]. Translated from English. New York; London: John Wiley & Sons, Inc.
17. Ore, O. (1968) *Teoriya grafov* [Theory of Graphs]. Translated from English. Moscow: Nauka.
18. Aleksandrov, P.S. (1968) *Lektsii po analiticheskoy geometrii* [Lectures on Analytic Geometry]. Moscow: Nauka.
19. Vinogradov, I.M. (1965) *Osnovy teorii chisel* [Principles of Number Theory]. Moscow: Nauka.
20. Yakovenko D.I. (1998) Zadacha ob ozherel'yakh [The necklaces problem]. *Vestnik Omskogo universiteta – Herald of Omsk University*. 2. pp. 21–24.
21. Baevsky, V.S. (2001) *Lingvisticheskie, matematicheskie, semioticheskie i komp'yuternye modeli v istorii i teorii literatury* [Linguistic, Mathematical, Semiotic, and Computer Models in the History and Theory of Literature]. Moscow: Yazyki slavyanskoy kul'tury.

Информация об авторе:

Григорьев Юрий Дмитриевич – профессор, доктор технических наук, Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» (Санкт-Петербург, Россия). E-mail: a-gortsev@mail.ru

Автор заявляет об отсутствии конфликта интересов.

Information about the author:

Grigoriev Yury D. (Doctor of Technical Sciences, Professor, Electrotechnical University “LETI”, St. Petersburg, Russian Federation). E-mail: yuri_grigoriev@mail.ru

The author declares no conflicts of interests.

Поступила в редакцию 19.05.2022; принята к публикации 29.11.2022

Received 19.05.2022; accepted for publication 29.11.2022

Original article

doi: 10.17223/19988605/61/11

Improving reputation and trust-based approach with reliability indicators for autonomous vehicles intergroup communication

Timofey Y. Melnikov¹, Sergey S. Chuprov², Eduard A. Lazarev³,
Ruslan I. Gataullin⁴, Ilia I. Viksnin⁵

¹ ITMO University, St. Petersburg, Russian Federation

^{2, 3, 4, 5} Electrotechnical University "LETI", St. Petersburg, Russian Federation

¹ tim_melnikov@mail.ru

² drmyscull@gmail.com

³ lazarev.eduard00@gmail.com

⁴ rusfiner@mail.ru

⁵ wixnin@mail.ru

Abstract. Reputation and Trust models have successfully been implemented to ensure security and maintain nodes' trustworthiness in decentralized networks, including autonomous vehicles (AVs). However, the implementations of these models usually lack of objective parameters to calculate the initial Reputation values. In this study, we employ reliability-based approach to evaluate the initial Reputation value.

Keywords: Reputation; trust; reliability; autonomous vehicles; communication

Acknowledgments: The work was supported by the Ministry of Science and Higher Education of the Russian Federation "Goszadanie" №075-01024-21-02 from 29.09.2021 (project FSEE-2021-0014).

For citation: Melnikov, T.Y., Chuprov, S.S., Lazarev, E.A., Gataullin, R.I., Viksnin, I.I. (2022) Improving reputation and trust-based approach with reliability indicators for autonomous vehicles intergroup communication. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naja tehnika i informatika – Tomsk State University Journal of Control and Computer Science*. 61. pp. 108–116. doi: 10.17223/19988605/61/11

Научная статья

УДК 519-7

doi: 10.17223/19988605/61/11

Улучшение модели репутации и доверия путем внедрения индикаторов надежности для внутригрупповой коммуникации между автономными транспортными средствами

Тимофей Юрьевич Мельников¹, Сергей Сергеевич Чупров², Эдуард Артемович Лазарев³,
Руслан Ильнурович Гатауллин⁴, Илья Игоревич Вискнин⁵

¹ Университет ИТМО, Санкт-Петербург, Россия

^{2, 3, 4, 5} Государственный электротехнический университет «ЛЭТИ», Санкт-Петербург, Россия

¹ tim_melnikov@mail.ru

² drmyscull@gmail.com

³ lazarev.eduard00@gmail.com

⁴ rusfiner@mail.ru

⁵ wixnin@mail.ru

Аннотация. Модели репутации и доверия успешно применяются для обеспечения безопасности и поддержания надежности узлов в децентрализованных сетях, включая автономные транспортные средства. Однако в реализациях этих моделей обычно отсутствуют объективные параметры для расчета начальных значе-

ний репутации. В данном исследовании предлагается подход, основанный на теории надежности, для оценки исходного значения репутации участников группы автономных транспортных средств.

Ключевые слова: репутация; доверие; надежность; автономные транспортные средства; коммуникация

Благодарности: Работа выполнена при поддержке Министерства науки и высшего образования Российской Федерации «Госзадание» № 075-01024-21-02 от 29.09.2021 (проект FSEE-2021-0014).

Для цитирования: Мельников Т.Ю., Чупров С.С., Лазарев Е.А., Гатауллин Р.И., Викснин И.И. Улучшение подхода, основанного на доверии и репутации, с помощью индикаторов надежности для межгрупповой коммуникации автономными транспортными средствами // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2022. № 61. С. 108–116. doi: 10.17223/19988605/61/11

The evolution of Information and Communication technology over the past few decades has resulted in the emergence of Internet of Things (IoT) and Smart Everything concepts. These concepts are generally based on the integration of physical and information components, and on the interconnections between them. Such an integration became known as Cyber-Physical Systems (CPSs) [1]. An example of a CPS is an autonomous monitoring system that has a sensor component (e.g., camera) and a video processing component (e.g., Machine Learning classifier) that performs video classification and pattern recognition operations.

Reputation and Trust models usually find their application in the systems whose elements communicate directly with each other and are essentially decentralized. An example of such systems is peer-to-peer (P2P) computer networks, where there no central server and the nodes communicate directly with each other [2]. For the nodes of P2P networks, it can be extremely important to evaluate the information received from others in order to choose their further actions within the network (e.g., packets forwarding towards a final destination). In VANETs, the communication between AVs and infrastructure roadside units are also decentralized, and can be considered as a P2P network [3]. There are various models for man aging such networks and practical examples of their application [4].

Reputation and Trust indicators can be implemented and calculated in different ways in the context of computer systems. Generally, most of them define some quantitative indicators to evaluate how trustworthy is the agent by the other agents. Mui, Mohtashemi, and Halberstadt define Reputation as an indicator that an agent creates through their previous actions [5]. The Trust indicator, in this case, is a subjective expectation one agent has about another agent's future actions based on the history of their interactions.

The application of Reputation and Trust-based approach to AVs intergroup communication allows to improve its security and safety [6, 7]. However, Reputation and Trust approach has a substantial drawback. Let us consider two cases: AV's group initialization moment, when the agents do not have any information on the previous actions of each other; and when a new agent joins to an already existing AVs group, and the group participants also do not have any information about the new agent's previous actions. In these two cases, the initial Reputation indicator value cannot be calculated based on the previous actions or "correctness" of the transmitted data by the agents. In these scenarios, the Reputation initial value is initialized as a constant (e.g., 0.5 in [6]). This initial constant initialization allows the potential adversarial or failed agent to transmit "incorrect" before their Reputation will drop below the acceptable threshold.

To address this drawback, an approach based on Data Quality (DQ) concept was proposed [8]. This solution allows to use such data source's characteristics as accuracy and security to calculate the initial Reputation value. This solution allows to rely on not only subjective evaluation of the agent's Reputation by the other agents, but also on the objective parameters based on the agent's hardware components. Agents DQ value depends on how "well" their sensors and devices are working at the evaluation moments: if there any failures or deviations from normal, the DQ value will be decreased [8]. However, the methods on deriving these data source's parameters to calculate DQ value are highly scenario-specific and under-researched [9]. An approach partially related to evaluating of the robotic device's hardware characteristics was presented in [10]. The authors proposed to consider various probability-dependent parameters in the robotic device's control strategy, for example, a failure rate of the robotic device's subsystem. Such an approach has motivated us to investigate and develop probability-dependent parameters based on AV's device characteristics to calculate

DQ values. We base our metrics on the reliability theory [11], which considers the probability of the system's components operation failures. We assume that the "correctness" of the data transmitted by the agents is related to this probability. The less reliable the agent's components, the lower its Reputation value should be initialized, and otherwise. This approach allows us to obtain objective basis for the initial Reputation calculation, which previous models lack [6, 12].

We incorporate the terms of failure, time between failure, and probability of failure-free operation into the initial Reputation value calculus. We incorporate these terms into the previously developed Reputation and Trust approach to calculate the initial Reputation value based on AV's hardware reliability characteristics. This solution allows to rely real physical AVs characteristics with the probability of "incorrect" data production and transmission. To test the proposed approach, we verify it on a well-known real-world hardware IoT components. The contribution of this paper is twofold. First, we provide a theoretical reliability based approach to calculate the AV's initial Reputation value. The approach allows to rely the hardware characteristics and the data produced by an AV model based on these hardware. In addition, we verify the proposed approach implementation on the intersection management problem with a software simulation. The initial Reputation values for the software simulation are based on the real-world device collected hardware characteristics.

1. Initial Reputation Calculus

1.1. System Description

Let us define a group of AVs represented by the agents of decentralized network. The agents are able to communicate directly with each other according to the P2P network communication model [2]. The data agents exchange consist of their final destination and agent's current condition (e.g., current location, energy resource, etc.). The main objective of the agents' group is to safely traverse the intersection [6, 13].

We assume, that each agent's architecture includes several sub-systems, the failure of which affect the reliability agent's reliability. Such sub-systems depend on the agent's functionality and architecture. Below we provide some examples of such sub-systems.

- Communication sub-system. If an agent is disconnected from the communication channel or its communication sub-system fails, it becomes unable to update or transmit "correct" information to other agents.
- The agent's computing sub-system is responsible for the agent's decision-making. If it fails, the information transmitted by the agent cannot be considered as reliable.
- Sensing sub-system for obtaining data from the environment (e.g., camera, radar, or other sensors). Failure of this sub-system leads to the obtaining of "incorrect" data from the environment and its transmission to other agents.

Based on the probabilities of these sub-systems' failure, we propose to evaluate the agent's initial Reputation value. To formalize these probabilities, we employ the exponential distribution law [11. P. 86]. In our case, we assume that the sub-systems' failures are constantly distributed, can occurred in a random manner, and are not caused by the deterioration or aging of the agent's parts. If the agent's components are deteriorated, it will be necessary to use a different distribution, but the basic calculation principle will not change.

1.2. Initial Reputation Calculus

Here we define failure probabilities and other terms according to [11]. Probability of failure during the sub-system operation time t can be defined as (1).

$$F(t) = 1 - e^{-\lambda t}, \quad (1)$$

where λ is a constant failure rate, and e is an exponential distribution.

From (1) we can derive the equation for failure-free operation during time t (2)

$$P(t) = 1 - F(t) = e^{-\lambda t}. \quad (2)$$

The parameter λ can be calculated according to (3)

$$\lambda = \frac{1}{T_0}, \quad (3)$$

where T_0 is the mean time between failures. It is calculated as (4).

$$T_0 = \frac{\sum_{i=1}^N t_i}{\sum_{i=1}^N m_i}, \quad (4)$$

where:

- N is the total number of sub-system's components for which the indicator is calculated;
- t_i is the operating time of the i -th sub-system's element;
- m_i is the number of failures of the i -th sub-system's element for the entire operation period.

For the i -th agent, as the initial Reputation value R_{0ei} , we consider the probability of its non-failure operation P_{ei} . Let each i -th agent is composed of a set of sub-systems: $S_{ei} = \{s_1, \dots, s_m\}$, and $P_{ij}(t)$ is the probability of non-failure operation of the j -th subsystem of the i -th agent during the time t , which can be calculated as (5).

$$P_{ij}(t) = e^{-t \cdot \lambda_{ij}}, \quad (5)$$

where λ_{ij} is the corresponding parameter of the i -th agent's j -th sub-system. According to (3) failure rate for (5) can be calculated according to (6).

$$\lambda_{ij} = \frac{1}{T_{0ij}}, \quad (6)$$

where T_{0ij} is the mean time between failures of the i -th agent's j -th subsystem. Further, according to the [11. P. 120–124], we need to calculate the probability of failure-free operation for the combination of these sub-systems, which is an entire agent. To do this, we need to consider how the agent sub-systems are interrelated in terms of failures. If the sub-systems are connected in such a way, that the failure of any of them leads to the failure of the entire system, then it can be referred to a serial connection. In another case, the sub-systems are connected in a parallel manner. For such a connection, the probability of failure-free operation during the required time t_0 considering (6) can be calculated as (7).

$$P_{ei} = \prod_{j=1}^m P_{ij}(t_0), \quad (7)$$

where m is the total number of the agent's sub-systems; and t_0 is the required operating time.

After that, the value calculated as (7) can be incorporated into the Reputation calculus instead of R_0 value, described in [6]. The modified initial Reputation value can be calculated as (8):

$$R_t = \begin{cases} \prod_{j=1}^m P_{ij}(t_0) + \sum_{k=1}^t I_k, & I_k > 0,5 \\ \prod_{j=1}^m P_{ij}(t_0) + \sum_{k=1}^t I_k - (R_{t-1} - e^{-(1-I_t) \cdot t}), & I_k < 0,5, \end{cases} \quad (8)$$

- $\sum_{k=1}^t I_k$, is the sum of Truth values for the previous iterations [6];
- R_{t-1} is the Reputation value on previous iteration;
- I_t is the Truth value on current iteration.

From the developed calculus, the following points can be derived.

1) Every Reputation model has its penalty system, when the agents with low reputation are punished in some way. Usually they are excluded from further group communication [6]. If any of the agent's sub-system is initially considered faulty (the probability of its failure-free operation is close or equal to zero), then the agent's Reputation will also be close to 0. Therefore, this agent will be excluded from the group communication after the first Reputation evaluation round. Such a scenario is true only if the sub-systems are connected serially. This means that if the key agent's sub-systems fail, the agent is excluded from the group communication.

2) The proposed initial Reputation value calculation approach sets a specified quality bar for all agent's sub-systems and their components. So, for example, if we have three serially connected sub-systems,

then to obtain an initial Reputation value of 0,5, the reliability indicators of the sub-systems should be about 0,8. With the increase in the sub-systems number, the requirements for these indicators will increase too.

3) The values obtained with (8) are always in the $[0, 1]$ interval, as the probability of a random event is also lies in this interval.

One can see that the initial Reputation value is based on the sub-system's uptime, as we are considering the probability of uptime for a certain time t . This is justified by the fact that when the system is required to work longer, we will obviously impose higher requirements on the agents' reliability. Let us suppose that we have a group of agents, each of which with a high probability will not fail for an hour of real time. Then, we can employ these agents in a system that should run for half an hour. However, these agents are not enough reliable for a system that needs to run for 6 hours, for example. This fact allows us to clarify the value of the initial agent's Reputation of agents, depending on the user and/or application requirements.

2. Empirical Study

2.1. Experimental setup

To verify the proposed reliability-based Reputation calculation approach, we conduct an empirical study. First, we employ real hardware components of the AV's models developed for our previous study [6] to calculate their reliability values. Then, we integrate the developed reliability-based calculus in our intersection management simulation environment, and compare the effectiveness with the traditional constant Reputation value initialization.

Each agent has the following subsystems:

- Computing sub-system, which is a small on-board computer used by the agent for decision-making and evaluation of the received information.
- The controller sub-system, which is used to provide connectivity between the computing and other sub-systems.
- The communication sub-system, used by an agent to transmit and receive information from other group participants.
- The ultrasonic rangefinder (URF) sub-system, which is used to avoid obstacles while driving.
- The vision sub-system (on-board camera), that the agent needs for localization and mapping.

Each of these sub-systems has its reliability indicator, which depends on the probability of failure-free operation for a given time. Let us assume that the agent should work without failures within 30 minutes. Such a minor time interval is chosen in the consideration of the widely available and cheap hardware components, employed in this study. We assume, that each of the defined sub-system's depends on a particular hardware component that corresponding to this sub-system. The employed AV model's hardware components are listed below.

- 1) Raspberry Pi 3 single-board computer, which serves as the computing sub system. It directly receives the data from an on-board camera, and from the communication sub-system via Arduino Nano;
- 2) Arduino Nano is connected to the URF and communication sub-system, and serves as a controller sub-system;
- 3) NRF24L01 serves as a communication sub-system. It is connected to Arduino Nano via SPI;
- 4) URF sub-system is connected to the Arduino Nano digital outputs;
- 5) RaspiCam is chosen as the AV's on-board camera, and it serves as the vision sub-system. It is connected with the computing sub-system, which processes the obtained image via OpenCV.

In Fig. 1 connection between hardware components is presented.

For this empirical study, we use two previously developed AV's models with two single sets of components, listed above. We ran these two models to drive through the intersections on our testbed [6] for 30 minutes, and monitored their sub-system's failures. Usually, it is possible to determine quite accurately that one of the agent's sub-systems has failed, as, depending on the subsystem, the agent can start to transmit meaningless data or stops updating information about their actions. According to the obtained results, we

calculated the reliability indicators for both sets, which are presented in Table 1. Based on the 7 and 8, and assuming that all sub-systems are connected in series, we calculated the new initial Reputation value for two sets of sub-systems. These values are also presented in Table 1.

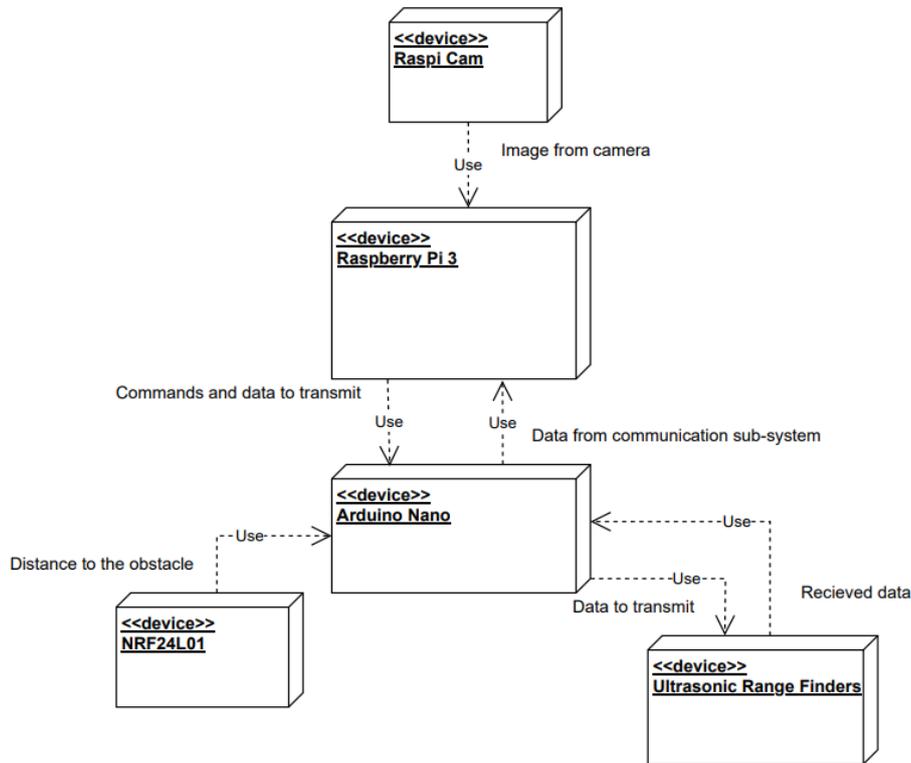


Fig. 1. Hardware connection UML-diagram

Table 1

Reliability values for the studied real hardware components

Hardware component	First set	Second set
Raspberry Pi 3	0,98	0,95
NRF24L01	0,94	0,90
Arduino Nano	0,82	0,70
Camera	0,97	0,89
Rangefinders	0,94	0,85
Reliability value	0,68	0,45

2.2. Software Simulation Assumptions

Since we have a limited number of AVs models, we leveraged the software simulator for the AVs intersection management system [14]. We used the Reputation values obtained with the hardware components to initialize agents' Reputation indicators. Let us assume that about half of the agents in the system have sub-systems' reliability indicators higher than 0,5, and the other half have lower than 0,5.

The simulation parameters can be represented as follows.

1) Two groups of experiments were conducted: with the constant Reputation initialization (0,5); and based on the calculated reliability characteristics.

2) In both experiment groups, there were two sub-groups: for each values calculated based on the studied hardware set.

3) A group of 20 agents operates in the system. The main objective of the agent is to safely and optimally traverse the intersection. Detailed simulation environment description is provided in [14, 15].

4) 1000 iterations of the simulation are conducted for each sub-group.

5) Initially agents in the group have different reliability indicators for their sub systems. To avoid setting the same reliability values for different agents, we took values from Table 1 and randomly distribute it over the agents. As we studied two hardware sets, we divided our agents into two sub-groups of 10 participants: with initially more reliable components, and with less reliable ones. The reliability values are initialized in the range of 0,65 – 0,72 for the first sub-group, and in the range of 0,42 – 0,45 for the second set.

6) Agents’ sub-systems can fail in the operation process with a probability corresponding to the exponential distribution law. The parameter λ for a given distribution law is obtained from the reliability indicators of agent sub-systems. After sub-system fails agent starts transmitting “incorrect” data. Data is considered “incorrect” when it does not represent the real condition of the agent or surrounding environment.

7) Agents are able to determine when one of their subsystem fails.

8) The assumed uptime for all agent’s subsystems is set to 30 minutes, and every 30 seconds agents communicate with each other and perform one round of the Reputation values calculation.

9) When agent’s Reputation drops below the 0,3 threshold, they are excluded from the group communication.

2.3. Results

As an evaluation criteria, we considered the following ones: average time T_{avg} and its standard deviation σ_T of the agents’ operation during the experiment; and the average time required to exclude failed agent from the group communication T_{err_avg} and its standard deviation σ_{T_err} . Moreover, to investigate the effect of reliability-based Reputation indicators affect the operation time of the agents. To address this question, we calculate the linear correlation coefficient between the time spent by agents in the system and their initial reputation value r_{rep} . The empirical study results are represented in Table 2.

Table 2

Empirical study results

Initial Reputation calculus	Components’ set	T_{avg} , sec	σ_T , sec	T_{err_avg} , sec	σ_{T_err} , sec
Reliability-based	First set	1453	387	827	701
	Second set	1030	395	602	554
Constant-based (0,5)	First set	1374	417	787	744
	Second set	1081	431	643	600

In Fig. 2 The relationship between the agent's initial Reputation value and the time of agent's operation in the system is presented.

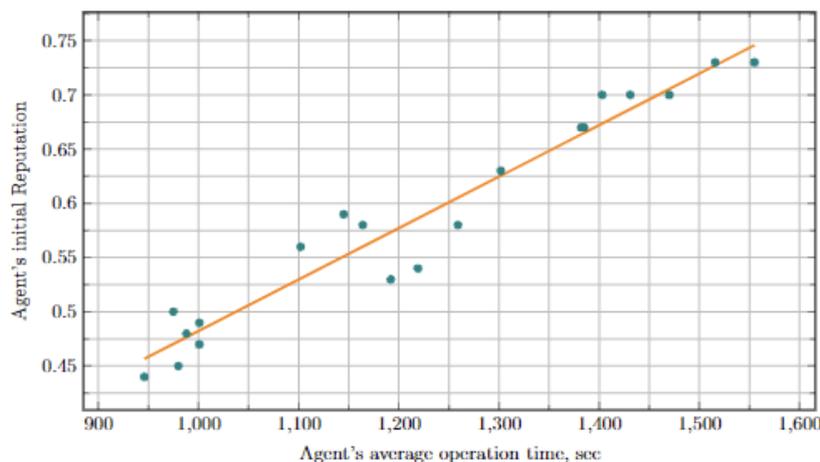


Fig. 2. The relationship between the agent's initial Reputation value and the time of agent's operation in the system (dots – initial Reputation and time values, orange line – linear regression)

As one can see from the obtained results, the employment of reliability indicators allows to slightly decrease average time to exclude agents that provide “incorrect” data due to sub-systems failures. The linear

correlation coefficient between the overall average agent's operation time (calculated based on both sub-groups), and initial reliability indicator is $r_{\text{rep}} = 0,97$, which can be interpreted as a strong relationship between the initial reliability indicators and the agent's operation time. Moreover, it can be seen that in all cases with reliability indicators, the standard deviation decreases. This can be interpreted by the fact that all agents are initially divided into two groups according to the initialized Reputation values. Thus, the spread of values both for the time spent in the system and for the agent's exclusion from the group communication decreased.

The results on correlation allows us to say that with reliability indicators it becomes possible to predict the average time of how long the agent with a specified reliability value will operate in the system. With the traditional Reputation initialization approach, such prediction is impossible, as there is no objective indicators can be used. In addition, reliability-based approach allows to reduce the time to exclude a failed agent from the group communication.

Conclusion

In this paper, we proposed to integrate reliability indicators to improve Reputation and Trust-based approach for AVs intergroup communication. Reputation and Trust models have found its implementation to enhance security in VANETs. However, these models has a substantial security gap – they rely on retrospective actions of the agent, and the initial Reputation value are usually set as a constant. In our preliminary research, we developed various approaches to address this issue [8, 12], but the direct relationships between the “correctness” of the data produced by the AVs and its hardware components were not considered in those studies. In this study, we employ reliability indicators to assess the AVs hardware parameters and to use these parameters to define initial Reputation value. We lever aged terms and models from the reliability theory and integrated them into the initial Reputation value calculus. Furthermore, to verify the reliability-based approach, we conducted an empirical study, We employed two real-world AVs models, developed in our preliminary intersection management research, and used its hardware components to empirically obtain reliability indicators. Furthermore, we conducted a set of software simulations to investigate the effect of the developed calculus against tradition Reputation initialization. The results showed that in the reliability-based approach the initial reliability indicator and the agents operation time are strongly correlated with each other (0,94). In addition, the approach allowed to decrease the time for exclusion failed agents' from the intergroup communication.

References

1. Wolf, W. (2009) Cyber-physical systems. *Computer*. 42(03). pp. 88–89.
2. Schollmeier, R. (2001) A definition of peer-to-peer networking for the classification of peer-to-peer architectures and applications. *Proceedings First International Conference on Peer-to-Peer Computing*. IEEE. pp. 101–102.
3. Verma, A., Raghavendra, P., Arun, P. & Rajeev, T. (2018) Information retrieval in two-tier VANET/P2P using RSU as a super-peer. *Wireless Communication Technology*. 2(1). pp. 1–9.
4. Kalyaev, I.A., Gaiduk, A.R. & Kapustyan, S.G. (2009) *Models and Algorithms of Collective Control in Groups of Robots*. Moscow: Fizmatlit.
5. Lik, M., Mohtashemi, M. & Halberstadt, A. (2002) A computational model of trust and reputation. *Proceedings of the 35th Annual Hawaii International Conference on System Sciences*. pp. 2431–2439. DOI:10.1109/HICSS.2002.994181
6. Chuprov, S.S., Viksnin, I.I., Kim I.V., Marinenkov, E.D., Usova, M., Lazarev, E.A., Melnikov, T.Y. & Zakoldaev, D. (2019) Reputation and trust approach for security and safety assurance in intersection management system. *Energies*. 12(23). 4527. pp. 1–19.
7. Viksnin, I.I., Iureva, R.A., Komarov, I.I. & Drannik, A.L. (2016) Assessment of stability of algorithms based on trust and reputation model. *2016 18th Conference of Open Innovations Association and Seminar on Information Security and Protection of Information Technology (FRUCT-ISPIT)*. IEEE. pp. 364–369.
8. Chuprov, S., Viksnin, I., Kim, I., Resnikand, L. & Khokhlov, I. (2020) Reputation and trust models with data quality metrics for improving autonomous vehicles traffic security and safety. *2020 IEEE systems security symposium (SSS)*. pp. 1–8. DOI: 10.1109/SSS47320.2020.9174269
9. Pipino, L.L., Lee, Y.W. & Wang, R.Y. (2002) Data Quality Assessment. *Communications of the ACM*. 45(4). pp. 211–218. DOI: <https://doi.org/10.1145/505248.506010>
10. Thrun, S., Burgard, W. & Fox, D. (2005) *Probabilistic Robotics*. 492 p. [Online] Available from: <https://docs.ufpr.br/~danielsantos/ProbabilisticRobotics.pdf>

11. Gnedenko, B.V., Belyaev, Yu.K. & Solovyev, A.D. (2014) *Mathematical Methods of Reliability Theory*. Academic Press.
12. Marinencov, E., Chuprov, S., Viksnin, I. & Kim, I. (2020) Empirical study on trust, reputation, and game theory approach to secure communication in a group of unmanned vehicles. *CEUR Workshop Proceedings*. 2590. Paper 28. pp. 1–12.
13. Dresner, K. & Stone, P. (2008) A multiagent approach to autonomous intersection management. *Journal of Artificial Intelligence Research*. 31. pp. 591–656.
14. Chuprov, S., Viksnin, I., Kim, I. & Nedosekin, G. (2019) Optimization of autonomous vehicles movement in urban intersection management system. *Proceedings of the 24th Conference of Fruct Association*. St. Petersburg, Russia. pp. 60–66.
15. Chuprov, S., Viksnin, I., Kim, I., Tursukov, N. & Nedosekin, G. (2020) Empirical study on discrete modeling of urban intersection management system. *International Journal of Embedded and Real-Time Communication Systems (IJERTCS)*. 11(2). pp. 16–38. DOI: 10.4018/IJERTCS.2020040102

Information about the authors:

Melnikov Timofei Yu. (Student of the Faculty of Secure Information Technologies of the National Research University ITMO, St. Petersburg, Russian Federation). E-mail: tim_melnikov@mail.ru

Chuprov Sergey S. (Software Developer in MIS Laboratory, The Institute of A.S. Popov, Electrotechnical University “LETI”, St. Petersburg, Russian Federation). E-mail: drmyscull@gmail.com

Lazarev Eduard A. (Technician in MIS Laboratory, The Institute of A.S. Popov, Electrotechnical University “LETI”, St. Petersburg, Russian Federation). E-mail: lazarev.eduard00@gmail.com

Gataullin Ruslan I. (Software Developer in MIS Laboratory, The Institute of A.S. Popov, Electrotechnical University “LETI”, St. Petersburg, Russian Federation). E-mail: rusfiner@mail.ru

Viksnin Ilya I. (Candidate of Technical Science, Researcher in MIS Laboratory, The Institute of A.S. Popov, Electrotechnical University “LETI”, St. Petersburg, Russian Federation). E-mail: wixnin@mail.ru

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

Информация об авторах:

Мельников Тимофей Юрьевич – студент факультета безопасности информационных технологий Национального исследовательского Университета ИТМО (Санкт-Петербург, Россия). E-mail: suncheture@gmail.com

Чупров Сергей Сергеевич – программист лаборатории МИС Института им. А.С. Попова Санкт-Петербургского государственного электро-технического университета «ЛЭТИ» (Санкт-Петербург, Россия). E-mail: drmyscull@gmail.com

Лазарев Эдуард Артемович – техник лаборатории МИС Института им. А.С. Попова Санкт-Петербургского государственного электро-технического университета «ЛЭТИ» (Санкт-Петербург, Россия). E-mail: lazarev.eduard00@gmail.com

Гатауллин Руслан Ильнурович – программист лаборатории МИС Института им. А.С. Попова Санкт-Петербургского государственного электро-технического университета «ЛЭТИ» (Санкт-Петербург, Россия). E-mail: rusfiner@mail.ru

Виксин Илья Игоревич – кандидат технических наук, научный сотрудник лаборатории МИС Института им. А.С. Попова Санкт-Петербургского государственного электро-технического университета «ЛЭТИ» (Санкт-Петербург, Россия). E-mail: wixnin@mail.ru

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

Received 24.05.2022; accepted for publication 29.11.2022

Поступила в редакцию 24.05.2022; принята к публикации 29.11.2022

Научная статья

УДК 004.032.26, 004.272

doi: 10.17223/19988605/61/12

Реализация сигмоидной функции активации с помощью концепции перестраиваемых вычислительных сред

Дмитрий Вадимович Шашев¹, Владислав Владимирович Шатравин²

^{1, 2} Томский государственный университет, Томск, Россия

¹ dshashev@mail.ru

² shatravin@stud.tsu.ru

Аннотация. Рассматривается вариант реализации сигмоидной функции активации для ускорителей нейронных сетей, целиком реализованных на перестраиваемых вычислительных средах (ПВС). Показаны преимущества применения подобных ускорителей в маломощных интеллектуальных системах. Предложены модели ускорителей на основе распределенной кусочно-линейной аппроксимации сигмоиды. Приведены результаты временной симуляции Verilog-модулей разработанных моделей.

Ключевые слова: сигмоида; функция активации; нейронные сети; аппроксимация; перестраиваемая вычислительная среда

Благодарности: Исследование выполнено за счет гранта Российского научного фонда № 21-71-00012, <https://rscf.ru/project/21-71-00012/>.

Для цитирования: Шашев Д.В., Шатравин В.В. Реализация сигмоидной функции активации с помощью концепции перестраиваемых вычислительных сред // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2022. № 61. С. 117–127. doi: 10.17223/19988605/61/12

Original article

doi: 10.17223/19988605/61/12

Implementation of the sigmoid activation function using the reconfigurable computing environments

Dmitriy V. Shashev¹, Vladislav V. Shatravin²

^{1, 2} Tomsk State University, Tomsk, Russian Federation

¹ dshashev@mail.ru

² shatravin@stud.tsu.ru

Abstract. In this paper, we consider options for implementation of the sigmoid activation function for hardware accelerators of neural networks implemented entirely on reconfigurable computing environments (RCE). The advantages of using such accelerators in low-power intelligent systems are shown. Accelerator models based on distributed piecewise linear approximation of the sigmoid are proposed. The time simulation results of the developed Verilog-models are presented.

Keywords: sigmoid; activation function; neural networks; approximation; reconfigurable computing environment

Acknowledgments: The research was supported by the Russian Science Foundation, grant No. 21-71-00012, <https://rscf.ru/project/21-71-00012/>.

For citation: Shashev, D.V., Shatravin, V.V. (2022) Implementation of the sigmoid activation function using the reconfigurable computing environments. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naja tehnika i informatika – Tomsk State University Journal of Control and Computer Science*. 61. pp. 117–127. doi: 10.17223/19988605/61/12

Нейронные сети находят активное применение в разнообразных интеллектуальных системах для задач, решение которых классическими методами является трудоемким или даже невозможным. Среди таких задач – обработка естественного языка, классификация и распознавание образов, предиктивный анализ и многие другие. Одним из ключевых недостатков нейронных сетей является высокая вычислительная сложность. Особенно это характерно для активно применяемых сверточных нейронных сетей, размер которых сейчас достигает сотен миллиардов параметров [1].

Для эффективного применения нейронных сетей в маломощных автономных системах (например, сенсорные сети, мобильные сети) требуются разработка специализированных аппаратных ускорителей и тонкая адаптация применяемых алгоритмов.

Существует множество работ на тему оптимизации аппаратного и алгоритмического обеспечения систем машинного обучения для маломощных устройств. В качестве развития аппаратного обеспечения обычно предлагается замена мощных, но энергозатратных графических процессоров на более эффективные FPGA и ASIC [2–5]. При этом осуществляется адаптация алгоритмов под особенности конкретной архитектуры для эффективного использования вычислительных ресурсов. Зачастую также производится компромиссное ухудшение некоторых характеристик исходной модели для уменьшения вычислительной сложности. К примеру, в [6, 7] предлагается переход к числам меньшей точности, а в [8, 9] – полная бинаризация сигналов. В работе [10] рассматривается уменьшение сложности сети через устранение части внутренних связей – разряжение сети.

Данная работа посвящена аспектам построения перестраиваемых ускорителей для реализации нейронных сетей на основе концепции перестраиваемых вычислительных сред. Данный подход открывает новые возможности для реализации нейросетевых алгоритмов, в том числе с точки зрения повышения качественных характеристик. Одним из ключевых вопросов проектирования ускорителя на перестраиваемых средах является реализация функций активации нейронов. В данной работе предложен вариант реализации на перестраиваемых вычислительных средах одной из наиболее популярных функций активации – сигмоидной.

1. Динамически перестраиваемые ускорители нейронных сетей

Искусственные нейронные сети представляют собой математическую модель вычислительной системы, имитирующей структуру и базовые принципы функционирования биологических нейронных сетей. В классической архитектуре искусственной сети нейроны распределены в несколько слоев (рис. 1). Каждый нейрон выполняет простую функцию: суммирует значения на всех своих входах, добавляет к ним собственное значение смещения и применяет к результату заданную функцию активации (рис. 2). Полученный результат передается нейронам следующего слоя, умножаясь на вес связи [11].

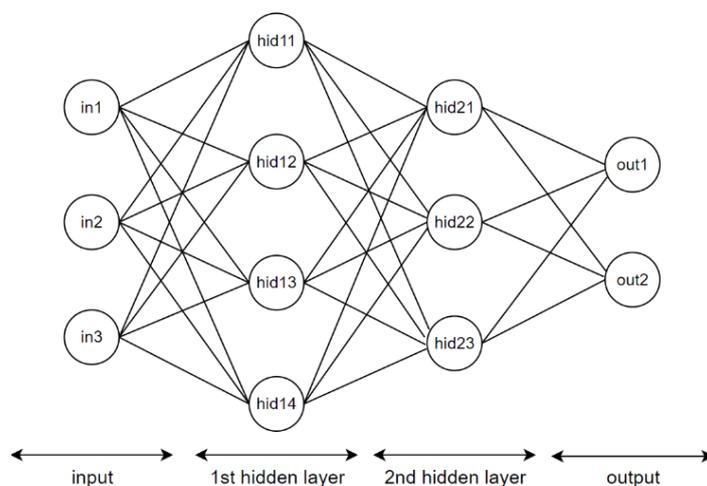


Рис. 1. Полносвязная сеть прямого распространения с двумя скрытыми слоями
Fig. 1. A fully connected direct distribution network with two hidden layers

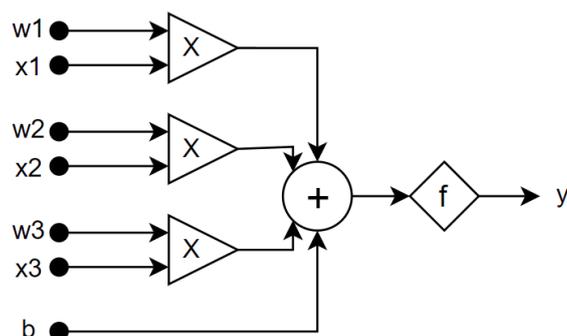


Рис. 2. Структура искусственного нейрона с тремя входами
 Fig. 2. The structure of an artificial neuron with three inputs

Функция активации нейрона – нелинейная функция одного аргумента, которая приводит взвешенную сумму входов нейрона к некоторому распределению. На практике применяются различные функции активации в зависимости от назначения слоя и решаемой сетью задачи. В задачах классификации в нейронах скрытых слоев зачастую используется функция активации ReLU, а на выходном слое – сигмоидная функция. Это позволяет получить на выходе сети результат в виде вероятности принадлежности объекта целевому классу. Использование в сигмоидной функции операций деления и возведения в произвольную степень усложняет ее реализацию, в связи с чем применяются различные способы адаптации и упрощения. В данной работе сигмоидная функция будет заменена ее кусочно-линейной аппроксимацией.

Внедрение глубоких нейронных сетей в реальные системы приводит к необходимости применения специальных вычислительных устройств. Большое количество однотипных вычислений даже на современных центральных процессорах требует значительных затрат времени и энергии. В последние годы набирает популярность использование гибридных систем, где применяются одновременно центральный процессор и специализированный под задачи машинного обучения сопроцессор [12]. Благодаря распределению задач между центральным процессором и сопроцессором такие системы достаточно эффективно справляются с широким классом задач. Но при этом сохраняется проблема узкой направленности сопроцессоров. Одним из способов устранения такого ограничения является применение динамически перестраиваемых аппаратных ускорителей.

Отличительной особенностью динамически перестраиваемых ускорителей является способность менять внутреннюю структуру в процессе функционирования для изменения реализуемой модели сети. Иными словами, одно и то же вычислительное устройство в разные моменты времени может реализовывать нейронные сети принципиально разных архитектур. Это дает следующие преимущества:

- способность использовать разные архитектуры нейронных сетей без изменения аппаратной базы;
- поддержку различных моделей сети для различных режимов функционирования (к примеру, использование в режиме ожидания менее точной, но экономной модели);
- поддержку удаленной настройки и модификации реализуемых алгоритмов, что может быть очень полезно, когда отсутствует априорная информация, необходимая для подготовки целевой модели сети, или когда условия функционирования изменчивы;
- возможность восстановления работоспособности устройства посредством изменения структуры вычислителя и перераспределения вычислений по неповрежденным участкам.

На данный момент разрабатываются различные подходы к построению перестраиваемых ускорителей. В [13] предлагается иерархическая «многогранульная» архитектура на основе разделяемой нейронной сети (Bisection Neural Network), допускающая несколько уровней настройки. Работа [14] предлагает реализацию перестраиваемости посредством выделения некоторого количества реконфигурируемых нейронов, которые могут свободно перераспределяться между скрытыми слоями сети. В данной работе авторами предлагается построение динамически перестраиваемых ускорителей на основе концепции перестраиваемых вычислительных сред.

2. Перестраиваемые вычислительные среды

Перестраиваемые вычислительные среды (ПВС) – математическая модель широкого класса вычислительных устройств, основанных на идее организованного совместного функционирования большого количества однотипных, относительно простых вычислительных элементов (ВЭ), расположенных в виде регулярной решетки и попарно соединенных с соседними вычислительными элементами [15, 16]. Каждый ВЭ может быть независимо настроен при помощи внешнего управляющего сигнала на одну операцию из некоторого predetermined базиса. Организованное функционирование большого количества вычислителей позволяет реализовать на среде сложные алгоритмы обработки. Таким образом, вычислительные возможности перестраиваемой среды ограничиваются количеством элементов и заданным базисом операций.

В общем случае ПВС может иметь ВЭ произвольной формы и быть одно-, двух- и трехмерной. В данной работе будут рассматриваться двумерные среды с квадратными ВЭ. Таким образом, каждый неграничный ВЭ будет соединен с четырьмя соседними элементами (рис. 3).

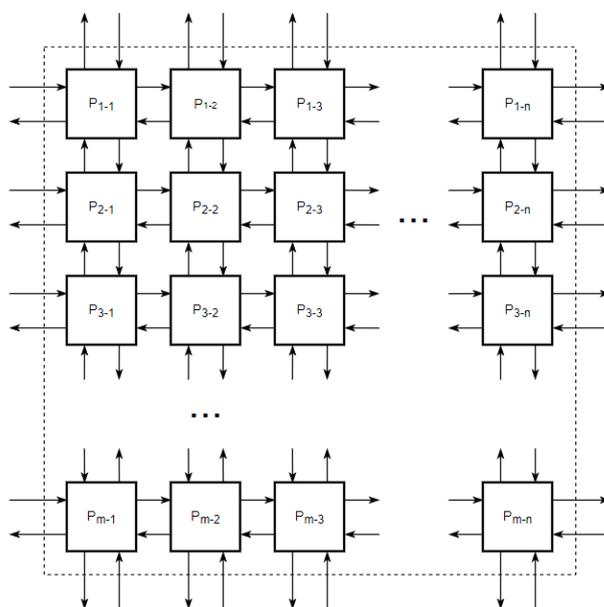


Рис. 3. Перестраиваемая вычислительная среда
Fig. 3. Reconfigurable computing environment

Одной из важных задач при проектировании ускорителей является уменьшение движения данных внутри системы в связи с сильным влиянием на быстродействие и энергопотребление. В целях снижения потока данных их стараются эффективно переиспользовать. Выделяют четыре основные стратегии переиспользования данных: фиксация весов, входов, выходов и строк [17]. При фиксации весов каждый ВЭ хранит собственное значение веса и использует его для каждого поступающего входного сигнала. Фиксация входов подразумевает, что каждый ВЭ хранит фрагмент входного сигнала, а значение весов поступает на вход элемента. Для фиксации выходов характерно пошаговое накопление в каждом ВЭ его промежуточных результатов (полусумм) до получения конечного результата. В ускорителях с фиксацией строк каждый ВЭ хранит строку данных и значение веса и последовательно рассчитывает несколько промежуточных результатов посредством умножения векторов в плавающем окне [18]. Предлагаемая нами модель ускорителя использует гибридную стратегию переиспользования (табл. 1). При обработке входного сигнала большого размера используется фиксация весов, а в остальных сценариях переиспользование не используется. Отсутствие переиспользования компенсируется замыканием всех вычислений внутри среды и конвейеризацией. Благодаря тому, что все вычисления осуществляются внутри среды, отсутствует необходимость ввода / вывода промежуточных результатов, устраняются накладные расходы на обращение к внешней памяти и другим модулям.

Сравнение архитектур ускорителей на основе однородных сред

Характеристика	TPU systolic array [19]	Eyeriss [18]	Предложенная модель
Вычислительный элемент	8-битовый MAC	16-битовый MAC с модулями памяти и накоплением промежуточных результатов	16-битовый элемент с поддержкой 7 атомарных операций
Память ВЭ	Нет	448KB SRAM + 72KB Reg	21 бит
Размер ВЭ	Очень малый	Крупный	Средний
Количество вычислительных элементов	Очень много: $256 \times 256 = 65\,536$	Мало: $12 \times 14 = 168$	Много, зависит от задачи
Перестраиваемость ВЭ	Нет	Нет	Да
Задача однородной среды	Матричное умножение	Матричное умножение	Полный цикл вычислений
Стратегия переиспользования данных	Фиксация весов	Фиксация строк	Гибридная (фиксация весов либо отсутствие переиспользования)
Место хранения промежуточных результатов	Буфер	Буфер	ВЭ
Постобработка (активация, субдискретизация)	Отдельный блок	Отдельный блок	Внутри среды

В отличие от некоторых альтернативных решений [13, 18], предлагаемая нами архитектура ускорителя основана на принципе предельной атомарности вычислительных элементов. Это означает, что ВЭ имеют простое строение, а базис операций включает минимально необходимое число операций. Это позволяет добиться высокой степени гибкости среды и допускает тонкую настройку параметров реализуемой модели. Анализ классических сетей прямого распространения позволил выделить следующий базис операций: «источник сигнала», «передача сигнала», «умножение с накоплением», «ReLU», «сигмоида» [20]. С их помощью каждый нейрон сети может быть представлен в виде цепочки из нескольких ВЭ (рис. 4) [21]. Длина цепочки определяет количество входов, что позволяет создавать нейрон любой конфигурации. Разработанная модель ВЭ оперирует 16-битными числами с фиксированной запятой.

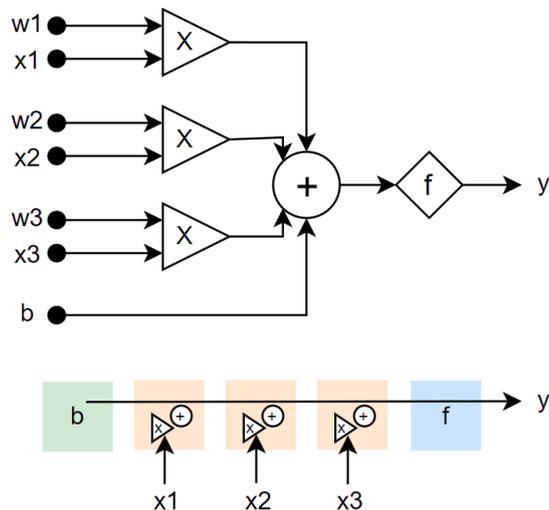


Рис. 4. Реализация нейрона на вычислительных элементах среды
Fig. 4. Implementation of a neuron on computational elements of the environment

Для поддержания реализации глубоких нейронных сетей и обеспечения конвейерной обработки в разрабатываемую вычислительную среду заложен механизм многотактной обработки сигнала посредством внутреннего заикливания [22]. Он заключается в разбиении среды на несколько сегментов, каждый из которых реализует один из слоев сети. Сигнал, перемещаясь по сегменту, не только проходит необходимую обработку, но и разворачивается в направлении следующего сегмента (рис. 5). По мере движения сигнала сегменты динамически перестраиваются на реализацию последующих

слоев сети. Такая архитектура позволяет реализовать на среде сеть произвольной глубины. Связанные с многотактной обработкой потери производительности могут быть частично устранены при помощи конвейеризации. Для четырехугольной среды с четырьмя сегментами можно на каждом такте перестраивать один сегмент, а на трех других производить обработку трех смещенных друг относительно друга на один такт входных сигналов. Реализация многотактной обработки требует добавления в базис операции «фиксация» для хранения промежуточного результата между сегментами среды. Функция поворота сигнала на 90° заложена в операцию «умножение с накоплением».

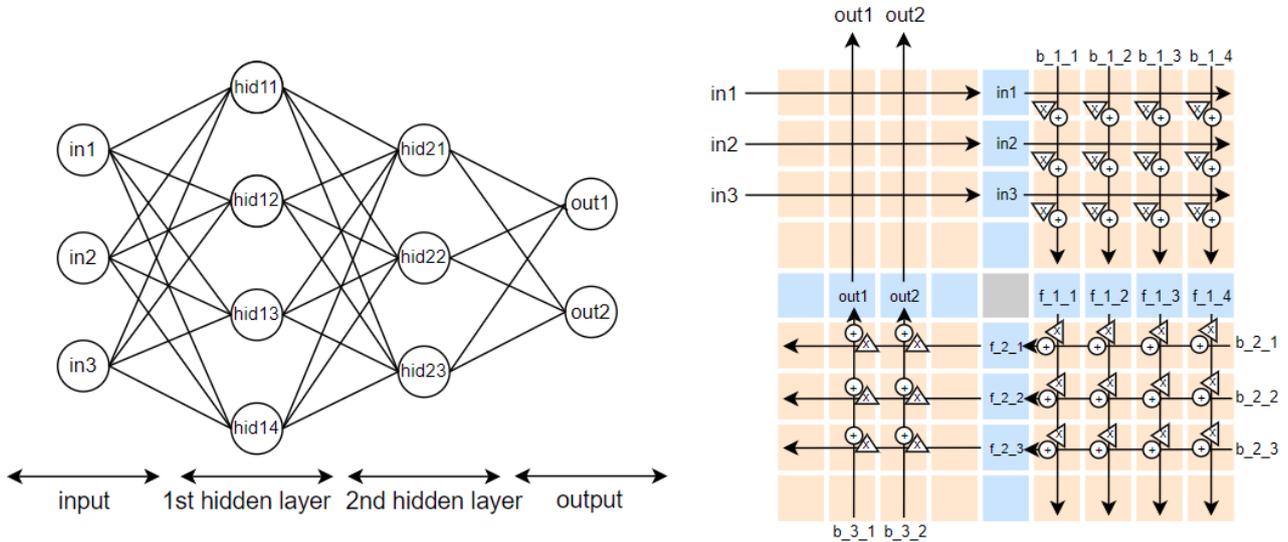


Рис. 5. Реализация многотактной обработки сигнала на сегментированной среде
 Fig. 5. Implementation of multi-step signal processing in a segmented environment

Другим недостатком многотактной реализации является ограничение максимально поддерживаемого размера слоя в связи с сегментированием среды. Для решения этой проблемы может быть использован цельный режим функционирования, когда для реализации слоя сети используется вся площадь среды. Данный режим требует применения дополнительной внешней (относительно среды) памяти и контроллера маршрутизации входного сигнала (рис. 6).

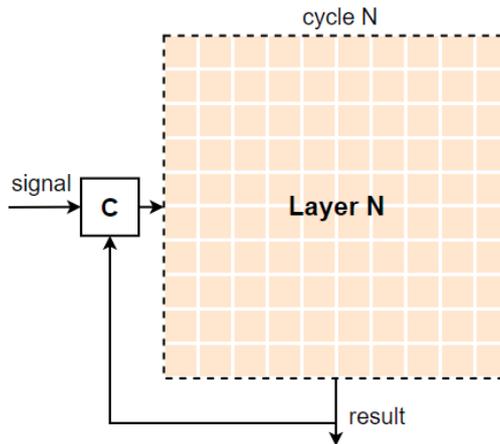


Рис. 6. Структура ускорителя в режиме без сегментирования
 Fig. 6. Accelerator structure in non-segmentation mode

Временная симуляция нейронов, выполненных согласно обозначенным принципам, показала приемлемые результаты временной симуляции. Однако одна из базовых операций – сигмоидная функция активации – в изначальном виде является значительно более трудоемкой, что приводит к большой площади каждого вычислительного элемента и их низкому быстродействию.

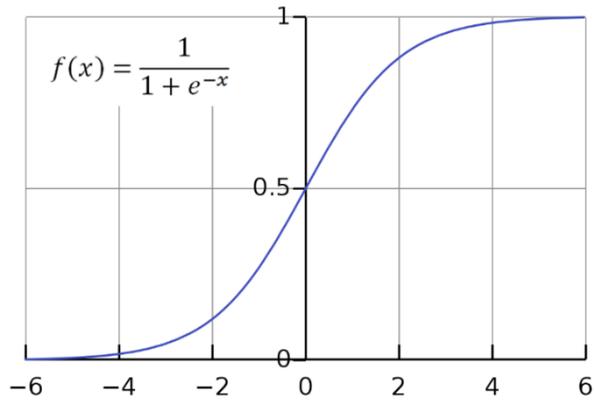


Рис. 7. Сигмоидная функция активации
Fig. 7. Sigmoid activation function

График и математическое выражение сигмоиды (ее также называют логистической функцией) приведены на рис. 7. Применение предложенной в [23] кусочно-линейной аппроксимации отчасти решает проблему, но операция по-прежнему остается самой трудоемкой во всей модели ВЭ. В связи с этим в данной работе рассматривается распределенная реализация сигмоиды на однородных средах.

3. Распределенная реализация сигмоидной функции активации

Распределенная реализация опирается на возможность независимого расчета значений на разных интервалах аппроксимации. Это позволяет распределить вычисления интервалов между разными вычислительными элементами среды, соответствующий распределенной реализации сигмоиды, представлен на рис. 8.

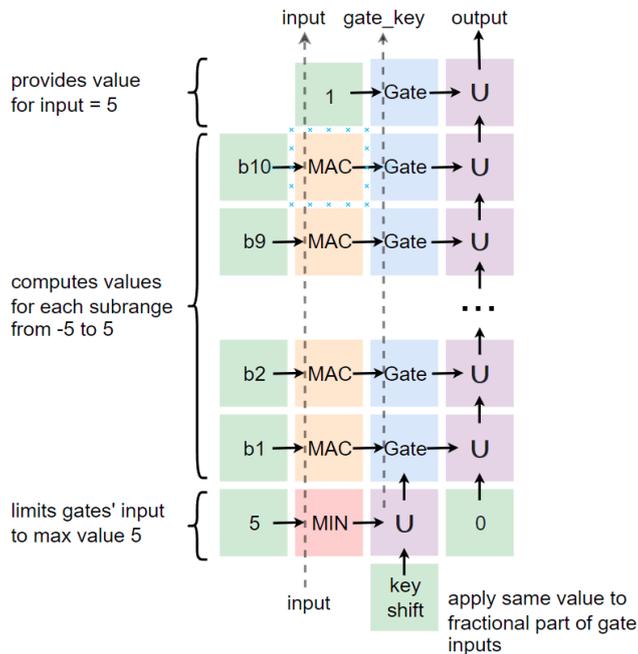


Рис. 8. Распределённая реализация сигмоиды
Fig. 8. Distributed implementation of sigmoid

Для наглядности используется цветовая дифференциация операций: зеленый – «источник сигнала», красный – «минимум двух чисел», желтый – «умножение с накоплением» (первый множитель поступает на нижний вход элемента, второй хранится во внутренней памяти элемента, аккумулирующее значение поступает на левый вход), синий – «затвор», фиолетовый – «объединение». Как видно,

для реализации распределенной сигмоиды было необходимо ввести в базис несколько новых операций – «минимум», «затвор» и «объединение». На данный момент эти операции используются только для сигмоиды, но они могут быть переиспользованы для реализации других функций, поэтому их введение в базис операций целесообразно.

Операция «затвор» используется для ограничения распространения сигнала в зависимости от поступающего на вход-ключ значения. На представленной схеме вход-ключ расположен на нижней грани элемента, а управляемый сигнал поступает на левый вход. Затвор сравнивает поступающее на вход значение ключа и, если он совпадает с заложенным во внутреннюю память элемента значением, пропускает входной сигнал через себя (слева направо). В противном случае на выходе элемента (справа) будет 16-битный ноль. Значение ключа передается вверх другим элементам-затворам независимо от срабатывания текущего затвора. Таким образом, у всех затворов, чей внутренний параметр не совпадает со поступающим значением ключа, выход будет равен нулю.

Как было показано ранее, при сравнении входного сигнала с узлами аппроксимации нас интересует только целая часть числа. В то же время операция «затвор» осуществляет сравнение ключа целиком. Это необходимо для переиспользования операции в других задачах, где сравнения только целой части числа может быть недостаточно. Для разрешения этого противоречия используется нижняя пара вычислительных элементов «key shift» и «объединение». Они осуществляют дополнительное смещение значения ключа – все биты дробной части (младшие 8 бит ключа) устанавливаются равными единице.

Операция объединения заключается в применении побитового ИЛИ к входным сигналам. Она позволяет объединить промежуточные результаты, полученные на разных интервалах аппроксимации (выходные значения элементов-затворов), в конечный результат. Также операция используется для осуществления сдвига ключа.

Минимум используется для ограничения максимального значения ключа. В принятом алгоритме аппроксимации всем значениям аргумента свыше 5 соответствует значение функции 1. Поэтому множество проверок узлов от 5 до 127 можно заменить ограничением ключа значением 5 и одним сравнением с этим значением. Для значений от -128 до -5 дополнительной логики не предусмотрено, так как на этом диапазоне значение функции равно 0, что соответствует нахождению всех затворов в закрытом состоянии.

Таким образом, на реализацию распределенной сигмоиды требуется 48 вычислительных элементов.

4. Симуляция разработанных моделей

Для сравнения предложенной реализации сигмоидной функции была разработана модель вычислительного элемента на языке описания аппаратуры Verilog. Разработанный Verilog-модуль включает весь описанный базис операций ВЭ. Оценивались две характеристики предложенных модулей – размер на среде (в логических элементах, ЛЭ) и быстродействие. Полученные результаты приведены в табл. 2.

Таблица 2

Результаты экспериментов

Реализация	Полный размер, ЛЭ	Размер ячейки, ЛЭ	Наибольшая задержка, нс	Абс. ошибка средняя	Абс ошибка макс.
Распределенная	13 175	296	18,5	$4e^{-3}$	$1e^{-2}$

Для оценки размера был осуществлен синтез модуля в среде Quartus Prime (версия 20.1.0, сборка 711 от 06.05.2020 SJ Edition) для устройства Cyclone V 5CGXFC9E7F35C8. В настройках синтеза отключено использование DSP блоков. Для защиты всех внутренних подмодулей от удаления в процессе оптимизации применялась Verilog-директива synthesis keep.

Для анализа быстродействия использовался встроенный в среду Quartus инструмент Timing Analyzer соответствующей версии. Симуляция происходила в режиме Fast 1000mV 0C посредством

определения наибольшей задержки между входом и выходом исследуемого модуля. Оба модуля были предварительно сконфигурированы, т.е. время перенастройки ячеек не включено в результаты измерения.

Поскольку в структуре однородных сред существуют двусторонние связи между соседними элементами, при временной симуляции распределенной реализации Timing Analyzer обнаруживает петли комбинационной логики. Для устранения этой проблемы на момент измерения быстродействия распределенной реализации были отключены неиспользуемые соединения между отдельными ВЭ.

Измерение ошибки аппроксимации осуществлялось при помощи тестового приложения. Приложение берет случайное число в диапазоне аппроксимации $[-5, 5]$, округляет его до точности 16-битного числа, рассчитывает значение аппроксимации в этой точке и сравнивает со значением натуральной сигмоиды для исходного неокругленного числа. Эксперимент повторяется $1e^6$ раз, в результате чего определяются наибольшая и средняя абсолютные ошибки.

Результаты аналогичных исследований приведены в [24]. Как видно из табл. 2, предложенная реализация сигмоидной функции имеет приемлемую среднюю ошибку и высокое быстродействие. Однако распределенная реализация требует значительно большего количества логических элементов. Это связано с распределением вычислений по большому количеству вычислительных элементов, каждый из которых содержит логику всех базисных операций среды. Также существенную роль играет комбинационная реализация арифметических операций, что увеличивает быстродействие за счет большей площади модулей.

Заключение

В данной работе предложен способ реализации кусочно-линейной аппроксимации сигмоиды на однородной среде. Распределенная реализация имеет меньшую производительность и требует большего количества ВЭ, но строится из простых базовых операций, которые могут быть переиспользованы для других задач и обеспечивают меньший размер отдельного ВЭ. Эксперименты показали высокое быстродействие (18,5 нс) и приемлемую точность (наибольшая и средняя абсолютные ошибки $1e^{-2}$ и $4e^{-3}$ соответственно). Сравнение с аналогичными исследованиями показывает перспективность дальнейшего изучения и применения разработанных моделей.

Список источников

1. Chen J., Li J., Majumder R. Make every feature binary: A 135B parameter sparse neural network for massively improved search relevance. 2021. URL: <https://www.microsoft.com/en-us/research/blog/make-every-feature-binary-a-135b-parameter-sparse-neural-network-for-massively-improved-search-relevance/> (accessed: 18.07.2022).
2. Nabavinejad S.M., Reda S., Ebrahimi M. Coordinated Batching and DVFS for DNN Inference on GPU Accelerators // IEEE Transactions on Parallel and Distributed Systems. 2022. 12 p. URL: <https://people.kth.se/~mebr/assets/files/TPDS2022-%20Coordinated%20Batching%20and%20DVFS%20for%20DNN%20Inference%20on%20GPU%20Accelerators.pdf> (accessed: 18.07.2022).
3. Guo J. et al. AccUDNN: A GPU Memory Efficient Accelerator for Training Ultra-Deep Neural Networks // IEEE 37th International Conference on Computer Design (ICCD). 2019. P. 65–72. doi: 10.1109/ICCD46524.2019.00017
4. Chang K.C., Fan C.P. Cost-Efficient Adaboost-based Face Detection with FPGA Hardware Accelerator // 2019 IEEE International Conference on Consumer Electronics – Taiwan (ICCE-TW). 2019. P. 1–2.
5. Lee J., He J., Wang K. Neural Networks and FPGA Hardware Accelerators for Millimeter-Wave Radio-over-Fiber Systems // 2020 22nd International Conference on Transparent Optical Networks (ICTON). 2020. P. 1–4.
6. Sakai Y. Quantization for Deep Neural Network Training with 8-bit Dynamic Fixed Point // 2020 7th International Conference on Soft Computing & Machine Intelligence (ISCMI). 2020. P. 126–130. doi: 10.1109/ISCMI51676.2020.9311563
7. Trusov A., Limonova E., Slugin D., Nikolaev D., Arlazarov V.V. Fast Implementation of 4-bit Convolutional Neural Networks for Mobile Devices // 2020 25th International Conference on Pattern Recognition (ICPR). 2021. P. 9897–9903. doi: 10.1109/ICPR48806.2021.9412841
8. Liu Z., Zhang H., Su Z., Zhu X. Adaptive Binarization Method for Binary Neural Network // 2021 40th Chinese Control Conference (CCC). 2021. P. 8123–8127. doi: 10.23919/CCC52363.2021.9549344
9. Zhu B., Al-Ars Z., Hofstee H.P. NASB: Neural Architecture Search for Binary Convolutional Neural Networks // 2020 International Joint Conference on Neural Networks (IJCNN). 2020. P. 1–8. doi: 10.1109/IJCNN48605.2020.9207674
10. Tang Z. et al. Automatic Sparse Connectivity Learning for Neural Networks // IEEE Transactions on Neural Networks and Learning Systems. 2022. P. 1–15. doi: 10.1109/TNNLS.2022.3141665

11. Haykin S. *Neural Network: A Comprehensive foundation*. Prentice Hall International, 1999. 842 p.
12. Chajan E., Schulte-Tiggas J., Reke M., Ferrein A., Matheis D., Walter T. GPU based model-predictive path control for self-driving vehicles // 2021 IEEE Intelligent Vehicles Symposium (IV). 2021. P. 1243–1248. doi: 10.1109/IV48863.2021.9575619
13. Kan Y., Wu M., Zhang R., Nakashima Y. A multi-grained reconfigurable accelerator for approximate computing // IEEE Computer Society Annual Symposium on VLSI (ISVLSI). 2020. P. 90–95.
14. Khalil K., Eldash O., Dey D., Kumar A., Bayoumi M. A Novel Reconfigurable Hardware Architecture of Neural Networ // IEEE 62nd International Midwest Symposium on Circuits and Systems (MWSCAS). 2019. P. 618–621.
15. Shashev D.V., Shidlovskiy S.V. Morphological processing of binary images using reconfigurable computing environments // Optoelectronics, Instrumentation and Data Processing. 2015. № 51. P. 227–233. doi: 10.3103/S8756699015030036
16. Евреинов Э.В. Однородные вычислительные системы, структуры и среды. М. : Радио и связь, 1981. 208 с.
17. Ghimire D., Kil D., Kim S-h. A Survey on Efficient Convolutional Neural Networks and Hardware Acceleration // MDPI J. Electronics. 2022. V. 945. P. 1–23.
18. Chen Y., Krishna T., Emer J.S., Sze V. Eyeriss: An Energy-Efficient Reconfigurable Accelerator for Deep Convolutional Neural Networks // IEEE J. Solid-State Circuits. 2017. V. 52. P. 127–138.
19. An in-depth look at Google’s first Tensor Processing Unit (TPU). 2017. URL: <https://cloud.google.com/blog/products/ai-machine-learning/an-in-depth-look-at-googles-first-tensor-processing-unit-tpu> (accessed: 18.07.2022).
20. Shatravin V., Shashev D.V. Designing high performance, power-efficient, reconfigurable compute structures for specialized applications // Journal of Physics: Conference Series. 2020. V. 1611. P. 1–6.
21. Shatravin V., Shashev D.V., Shidlovskiy S.V. Applying the Reconfigurable Computing Environment Concept to the Deep Neural Network Accelerators Development // International Conference on Information Technology (ICIT). 2021. P. 842–845. doi: 10.1109/ICIT52682.2021.9491771
22. Shatravin V., Shashev D.V., Shidlovskiy S.V. Developing of models of dynamically reconfigurable neural network accelerators based on homogeneous computing environments // Proc. of the XXIV international scientific conference Distributed Computer and Communication Networks: Control, Computation, Communications (DCCN). 2021. P. 102–107.
23. Faiedh H., Gafsi Z., Besbes K. Digital Hardware Implementation of Sigmoid Function and its Derivative for Artificial Neural Networks // The 13 International Conference on Microelectronics. 2001. P. 189–192.
24. Pan Z., Gu Z., Jiang X., Zhu G., Ma D. A Modular Approximation Methodology for Efficient Fixed-Point Hardware Implementation of the Sigmoid Function // IEEE Transactions on Industrial Electronics. 2022. P. 10694–10703. doi: 10.1109/TIE.2022.3146573

References

1. Chen, J., Li, J. & Majumder, R. (2021) *Make every feature binary: A 135B parameter sparse neural network for massively improved search relevance*. [Online] Available from: <https://www.microsoft.com/en-us/research/blog/make-every-feature-binary-a-135b-parameter-sparse-neural-network-for-massively-improved-search-relevance/> (Accessed: 17th July 2022).
2. Nabavinejad, S.M., Reda, S. & Ebrahimi, M. (2022) Coordinated Batching and DVFS for DNN Inference on GPU Accelerators. *IEEE Transactions on Parallel and Distributed Systems*. pp. 1–12.
3. Guo, J. et al. (2019) AccUDNN: A GPU Memory Efficient Accelerator for Training Ultra-Deep Neural Networks. *IEEE 37th International Conference on Computer Design (ICCD)*. pp. 65–72. DOI: 10.1109/ICCD46524.2019.00017
4. Chang, K.C. & Fan, C.P. (2019) Cost-Efficient Adaboost-based Face Detection with FPGA Hardware Accelerator. *2019 IEEE International Conference on Consumer Electronics – Taiwan (ICCE-TW)*. pp. 1–2.
5. Lee, J., He, J. & Wang, K. (2020) Neural Networks and FPGA Hardware Accelerators for Millimeter-Wave Radio-over-Fiber Systems. *22nd International Conference on Transparent Optical Networks (ICTON)*. pp. 1–4.
6. Sakai, Y. (2020) Quantization for Deep Neural Network Training with 8-bit Dynamic Fixed Point. *7th International Conference on Soft Computing & Machine Intelligence (ISCMI)*. pp. 126–130. DOI: 10.1109/ISCMI51676.2020.9311563
7. Trusov, A., Limonova, E., Slugin, D., Nikolaev, D. & Arlazarov, V.V. (2021) Fast Implementation of 4-bit Convolutional Neural Networks for Mobile Devices. *25th International Conference on Pattern Recognition (ICPR)*. pp. 9897–9903. DOI: 10.1109/ICPR48806.2021.9412841
8. Liu, Z., Zhang, H., Su, Z. & Zhu, X. (2021) Adaptive Binarization Method for Binary Neural Network. *40th Chinese Control Conference (CCC)*. pp. 8123–8127. DOI: 10.23919/CCC52363.2021.9549344
9. Zhu, B., Al-Ars, Z. & Hofstee, H.P. (2020) NASB: Neural Architecture Search for Binary Convolutional Neural Networks. *International Joint Conference on Neural Networks (IJCNN)*. pp. 1–8. DOI: 10.1109/IJCNN48605.2020.9207674
10. Tang, Z. et al. (2022) Automatic Sparse Connectivity Learning for Neural Networks. *IEEE Transactions on Neural Networks and Learning Systems*. pp. 1–15. DOI: 10.1109/TNNLS.2022.3141665
11. Haykin, S. (1999) *Neural Network: A Comprehensive Foundation*. Prentice Hall International.
12. Chajan, E., Schulte-Tiggas, J., Reke, M., Ferrein, A., Matheis, D. & Walter, T. (2021) GPU based model-predictive path control for self-driving vehicles. *IEEE Intelligent Vehicles Symposium (IV)*. pp. 1243–1248. DOI: 10.1109/IV48863.2021.9575619
13. Kan, Y., Wu, M., Zhang, R. & Nakashima, Y. (2020) A multi-grained reconfigurable accelerator for approximate computing. *IEEE Computer Society Annual Symposium on VLSI (ISVLSI)*. pp. 90–95.
14. Khalil, K., Eldash, O., Dey, B., Kumar, A. & Bayoumi, M. (2019) A Novel Reconfigurable Hardware Architecture of Neural Network. *IEEE 62nd International Midwest Symposium on Circuits and Systems (MWSCAS)*. pp. 618–621.

15. Shashev, D.V. & Shidlovskiy, S.V. (2015) Morphological processing of binary images using reconfigurable computing environments. *Optoelectronics, Instrumentation and Data Processing*. 51. pp. 227–233. DOI: 10.3103/S8756699015030036.
16. Evreinov, E.V. (1981) *Odnorodnye vychislitel'nye sistemy, struktury i sredy* [Homogeneous computing systems, structures and environments]. Moscow: Radio i svyaz'.
17. Ghimire, D., Kil, D. & Kim, S-h. (2022) A Survey on Efficient Convolutional Neural Networks and Hardware Acceleration. *MDPI J. Electronics*. 945. pp. 1–23.
18. Chen, Y., Krishna, T., Emer, J.S. & Sze, V. (2017) Eyeriss: An Energy-Efficient Reconfigurable Accelerator for Deep Convolutional Neural Networks. *IEEE J. Solid-State Circuits*. 52. pp. 127–138.
19. TPU. (2017) *An in-depth look at Google's first Tensor Processing Unit (TPU)*. [Online] Available from: <https://cloud.google.com/blog/products/ai-machine-learning/an-in-depth-look-at-googles-first-tensor-processing-unit-tpu> (Accessed: 17th July 2022).
20. Shatravin, V. & Shashev, D.V. (2020) Designing high performance, power-efficient, reconfigurable compute structures for specialized applications. *Journal of Physics: Conference Series*. 1611. pp. 1–6.
21. Shatravin, V., Shashev, D.V. & Shidlovskiy, S.V. (2021) Applying the Reconfigurable Computing Environment Concept to the Deep Neural Network Accelerators Development. *International Conference on Information Technology (ICIT)*. pp. 842–845. DOI: 10.1109/ICIT52682.2021.9491771.
22. Shatravin, V., Shashev, D.V. & Shidlovskiy, S.V. (2021) Developing of models of dynamically reconfigurable neural network accelerators based on homogeneous computing environments. *Distributed Computer and Communication Networks: Control, Computation, Communications (DCCN)*. Proceedings of the 24th International Conference. pp. 102–107.
23. Faiedh, H., Gafsi, Z. & Besbes, K. (2001) Digital Hardware Implementation of Sigmoid Function and its Derivative for Artificial Neural Networks. *The 13 International Conference on Microelectronics*. pp. 189–192.
24. Pan, Z., Gu, Z., Jiang, X., Zhu G. & Ma, D. (2022) A Modular Approximation Methodology for Efficient Fixed-Point Hardware Implementation of the Sigmoid Function. *IEEE Transactions on Industrial Electronics*. pp. 10694–10703. DOI: 10.1109/TIE.2022.3146573

Информация об авторах:

Шашев Дмитрий Вадимович – кандидат технических наук, доцент кафедры управления качеством факультета инновационных технологий Национального исследовательского Томского государственного университета (Томск, Россия). E-mail: dshashev@mail.ru

Шатравин Владислав Владимирович – аспирант кафедры управления качеством факультета инновационных технологий Национального исследовательского Томского государственного университета (Томск, Россия). E-mail: shatravin@stud.tsu.ru

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

Information about the authors:

Shashev Dmitriy Vadimovich (Candidate of Technical Sciences, Associate Professor, National Research Tomsk State University, Tomsk, Russian Federation). E-mail: dshashev@mail.ru

Shatravin Vladislav (Post-graduate Student, National Research Tomsk State University, Tomsk, Russian Federation). E-mail: shatravin@stud.tsu.ru

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

Received 06.04.2022; accepted for publication 29.11.2022

Поступила в редакцию 06.04.2022; принята к публикации 29.11.2022

ДИСКРЕТНЫЕ ФУНКЦИИ И АВТОМАТЫ

DISCRETE FUNCTION AND AUTOMATONS

Научная статья

УДК 519.7

doi: 10.17223/19988605/61/13

Графовые представления множеств всех достижимых
реакций комбинационной схемыВиктор Алексеевич Провкин¹, Анжела Юрьевна Матросова²^{1,2}Томский государственный университет, Томск, Россия¹*prowkan@mail.ru*²*mau11@yandex.ru*

Аннотация. Рассматривается задача получения множества всех достижимых реакций комбинационной логической схемы. Предлагается алгоритм построения ROBDD-графа, представляющего все достижимые реакции схемы. Получаемый граф содержит внутренние вершины, которые помечены только выходными переменными схемы. Алгоритм может быть использован в тех случаях, когда ROBDD-граф, зависящий от входных и выходных переменных, не может быть построен из-за экспоненциального роста количества вершин.

Ключевые слова: комбинационные схемы; ROBDD-графы; булевы функции

Для цитирования: Провкин В.А., Матросова А.Ю. Графовые представления множеств всех достижимых реакций комбинационной схемы // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2022. № 61. С. 128–138. doi: 10.17223/19988605/61/13

Original article

doi: 10.17223/19988605/61/13

Graph representations of the sets of all reachable
reactions of the combinational circuitViktor A. Provkina¹, Anzhela Yu. Matrosova²^{1,2}Tomsk State University, Tomsk, Russian Federation¹*prowkan@mail.ru*²*mau11@yandex.ru*

Abstract. The problem of obtaining the set of all achievable reactions of a combinational logic circuit is considered. An algorithm for constructing a ROBDD graph representing all the achievable reactions of the circuit is proposed. The resulting graph contains internal vertices, which are labeled only with schema output variables. The algorithm can be used in cases where an ROBDD graph that depends on input and output variables cannot be obtained because of the exponential growth of the number of vertices.

Keywords: combinational circuits; reduced ordered binary decision diagrams; Boolean functions

For citation: Provkina V.A., Matrosova A.Yu. (2022) Graph representations of the sets of all reachable reactions of the combinational circuit. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naja tehnika i informatika – Tomsk State University Journal of Control and Computer Science*. 61. pp. 128–138. doi: 10.17223/19988605/61/13

При решении ряда задач диагностики и тестирования логических схем полезными оказываются особые функции для внутренних полюсов, которые зависят от некоторого подмножества предшествующих полюсов схемы. Эти функции являются частично определенными. Область их определения состоит из таких наборов предшествующих полюсов, что смена значения сигнала на соответствующем набору внутреннем полюсе вызывает смену значений сигналов хотя бы на одном из выходов схемы. Такие функции могут использоваться для маскирования неисправностей и вредоносных подсхем [1, 2], верификации частичных реализаций схем [3, 4], структурного кодирования («обфускации») [5, 6], защиты схем со структурной избыточностью (троирование) [7, 8].

Такие функции могут вычисляться либо с помощью частично определенных функций внутренних полюсов, зависящих от входов схемы (область определения таких функций – множество тестовых наборов для одиночной константной неисправности в соответствующем полюсе) [9, 10], либо без использования таких функций. Второй способ более предпочтителен, так как построение частично определенной функции внутренних полюсов, зависящей от входов схемы, возможно далеко не всегда, поскольку в некоторых случаях размеры ROBDD-графов, требуемых для построения функции, растут экспоненциально. Алгоритм построения, реализующий этот способ, приведен в работе [11]. На первом шаге алгоритма вычисляется множество наборов, которые могут появиться на некотором подмножестве внутренних полюсов схемы. Эта задача эквивалентна задаче получения множества всех возможных выходных реакций некоторой известной комбинационной схемы.

В работе [11] данная задача решается путем перебора всех возможных наборов (выходных реакций на подмножестве полюсов) с проверкой их на достижимость. При маскировании неисправностей число переменных, от которых зависит частично определенная функция, обычно достигает одного-двух десятков, и поэтому полный перебор выполняется за приемлемое время. Однако при решении других задач, например при верификации частичных реализаций логических схем, число переменных может достигать нескольких десятков, а число возможных реакций, как известно, увеличивается экспоненциально с ростом числа переменных. Поэтому желательно иметь алгоритмы, которые определяют множество достижимых реакций схемы без полного перебора и компактно их представляют.

В данной работе предлагается алгоритм построения ROBDD-графа, представляющего множество всех достижимых реакций комбинационной логической схемы. Получаемый граф содержит внутренние вершины, которые помечены только выходными переменными схемы. Алгоритм предлагается использовать в тех случаях, когда ROBDD-граф, зависящий от входных и выходных переменных, не может быть построен из-за экспоненциального роста количества его внутренних вершин.

1. Основные определения и постановка задачи

Комбинационная логическая схема – модель устройства без памяти. У таких устройств состояния выходов однозначно определяются набором входных сигналов (булевым вектором на множестве входных переменных). Состояние выходов комбинационной схемы (булев вектор на множестве выходных переменных схемы) при подаче на входы определенного набора называется реакцией схемы на этот входной набор.

Например, рассмотрим комбинационную схему (рис. 1), реализующую систему булевых функций:

$$\begin{aligned} y_1 &= x_1 \overline{x_2} \vee x_3, \\ y_2 &= \overline{x_1 x_2 x_3}, \\ y_3 &= \overline{x_3} \vee x_4 \vee x_5, \\ y_4 &= \overline{x_3} (x_4 \vee x_5). \end{aligned}$$

Если на входы схемы поступает набор 10101, то на выходах схемы достигается набор 1100, т.е. набор 1100 является реакцией схемы на входной набор 10101.

Булев вектор в пространстве выходных переменных называется достижимой реакцией, если существует набор, при подаче на входы которого заданный булев вектор появляется в качестве реакции схемы.

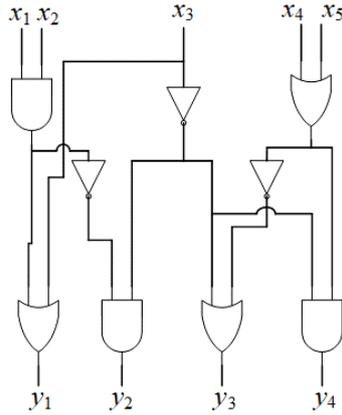


Рис. 1. Пример комбинационной схемы
Fig. 1. Example of combinational circuit

Формально задача может быть поставлена следующим образом. Имеется многовыходная комбинационная схема C , реализующая систему булевых функций $f_i(x_1, x_2, \dots, x_n)$, $i = 1, \dots, m$. Требуется найти такое множество $N = \{(\beta_1, \beta_2, \dots, \beta_m)\} \subseteq B_2^m$, что для каждого $(\beta_1, \beta_2, \dots, \beta_m) \in N$ существует непустое множество, такое что для каждого $(\alpha_1, \alpha_2, \dots, \alpha_n) \in M$ $f_i(\alpha_1, \alpha_2, \dots, \alpha_n) = \beta_i$, $i = 1, \dots, m$. Элементы множества $N = \{(\beta_1, \beta_2, \dots, \beta_m)\} \subseteq B_2^m$ будем называть достижимыми реакциями схемы. Множество $M = \{(\alpha_1, \alpha_2, \dots, \alpha_n)\} \subseteq B_2^n$ будем называть полным прообразом реакции $(\beta_1, \beta_2, \dots, \beta_m) \in N$. Отметим, что для различных наборов $(\beta_1^1, \beta_2^1, \dots, \beta_m^1) \in N$ и $(\beta_1^2, \beta_2^2, \dots, \beta_m^2) \in N$ соответствующие множества M^1 и M^2 не пересекаются (их пересечение означало бы, что подача одного набора на входы схемы приводит к двум разным реакциям одновременно, что невозможно).

2. Определение функции, содержащей информацию о достижимых реакциях комбинационной схемы, и ее свойства

Пусть задана многовыходная комбинационная схема C , реализующая систему булевых функций $y_1 = f_1(x_1, x_2, \dots, x_n)$, $y_2 = f_2(x_1, x_2, \dots, x_n)$, ..., $y_m = f_m(x_1, x_2, \dots, x_n)$. Рассмотрим следующую функцию, которая зависит от входных и выходных переменных комбинационной схемы:

$$f(x_1, x_2, \dots, x_n, y_1, y_2, \dots, y_m) = \bigwedge_{i=1}^m [y_i \sim f_i(x_1, x_2, \dots, x_n)],$$

где \sim – логическая операция «эквивалентность». Эта функция обладает следующими свойствами:

1. Если вместо входных переменных подставить константы из $\{0, 1\}$, то получим функцию, которая принимает единичное значение на единственном наборе. Этот набор является реакцией схемы на данный входной набор. На остальных наборах полученная функция принимает нулевое значение. Пусть $(\alpha_1, \alpha_2, \dots, \alpha_n)$ – некоторый входной набор, тогда существует один и только один выходной набор $(\beta_1, \beta_2, \dots, \beta_m)$ такой, что $f(\alpha_1, \alpha_2, \dots, \alpha_n, \beta_1, \beta_2, \dots, \beta_m) = 1$, причем $\beta_1 = f_1(\alpha_1, \alpha_2, \dots, \alpha_n)$, $\beta_2 = f_2(\alpha_1, \alpha_2, \dots, \alpha_n)$, ..., $\beta_m = f_m(\alpha_1, \alpha_2, \dots, \alpha_n)$. Для всякого набора $(\gamma_1, \gamma_2, \dots, \gamma_m) \neq (\beta_1, \beta_2, \dots, \beta_m)$

$$f(\alpha_1, \alpha_2, \dots, \alpha_n, \gamma_1, \gamma_2, \dots, \gamma_m) = 0.$$

2. Если вместо выходных переменных подставить константы из $\{0, 1\}$, то получится функция, которая принимает единичное значение на всех наборах, которые обеспечивают проявление заданной реакции на выходах схемы. В частном случае, если заданная реакция недостижима, то полученная функция тождественно равна 0. Рассмотрим некоторый выходной набор $(\beta_1, \beta_2, \dots, \beta_m)$. Тогда

$f(\alpha_1, \alpha_2, \dots, \alpha_n, \beta_1, \beta_2, \dots, \beta_m) = 1$ для любого набора $(\alpha_1, \alpha_2, \dots, \alpha_n)$ такого, что $\beta_1 = f_1(\alpha_1, \alpha_2, \dots, \alpha_n)$, $\beta_2 = f_2(\alpha_1, \alpha_2, \dots, \alpha_n)$, ..., $\beta_m = f_m(\alpha_1, \alpha_2, \dots, \alpha_n)$. Если таких наборов не существует, то $f(\alpha_1, \alpha_2, \dots, \alpha_n, y_1 = \beta_1, y_2 = \beta_2, \dots, y_m = \beta_m) = 0$ для всех наборов $(\alpha_1, \alpha_2, \dots, \alpha_n)$.

Для получения всех достижимых реакций схемы можно построить ROBDD-граф функции $f(x_1, x_2, \dots, x_n, y_1, y_2, \dots, y_m)$, причем выбираем следующий порядок переменных при использовании формулы разложения Шеннона: $y_1, y_2, \dots, y_m, x_1, x_2, \dots, x_n$. Полученный граф будет обладать следующими свойствами.

Утверждение 1. Каждая простая цепь, заканчивающаяся в вершине со значением 0 и не содержащая вершин, помеченных входными переменными, соответствует недостижимому набору (возможно, некоторому множеству наборов).

Доказательство. Корневой вершине графа соответствует функция $f(x_1, x_2, \dots, x_n, y_1, y_2, \dots, y_m)$, и этой вершине приписана переменная y_1 . При переходе из корневой вершины по дуге, помеченной значением 1 (0), попадаем в вершину, которой соответствует функция $f(x_1, x_2, \dots, x_n, 1, y_2, \dots, y_m)$ ($f(x_1, x_2, \dots, x_n, 0, y_2, \dots, y_m)$). Эта вершина помечена переменной y_2 . Перейдя из нее по дуге, помеченной значением 1 (0), получаем функцию, которая построена из функции, соответствующей предыдущей вершине, подстановкой значения 1 (0) в переменную y_2 . Продолжим этот процесс до тех пор, пока не дойдем до 0-концевой вершины. Так как в цепи присутствуют только вершины, помеченные выходными переменными, то это значит, что подстановки значений выполнялись только в те аргументы функции, которые соответствуют выходным переменным схемы. Поскольку цепь заканчивается в 0-концевой вершине, то $f(\alpha_1, \alpha_2, \dots, \alpha_n, y_1 = \beta_1, y_2 = \beta_2, \dots, y_m = \beta_m) = 0$ для всех наборов $(\alpha_1, \alpha_2, \dots, \alpha_n)$ и, следовательно, набор $(\beta_1, \beta_2, \dots, \beta_m)$ недостижим.

Утверждение 2. Каждая простая цепь, содержащая вершины, помеченные входными переменными схемы, и заканчивающаяся в 1-терминальной вершине графа, соответствует некоторой достижимой реакции, представленной всеми выходными переменными схемы, за счет отрезка цепи, связывающего рассматриваемую простую цепь с корнем графа.

Доказательство. Если пройти по отрезку цепи из корневой вершины до первой вершины v простой цепи, помеченной входной переменной, то этой вершине сопоставляется функция $f(x_1, x_2, \dots, x_n, y_1 = \beta_1, y_2 = \beta_2, \dots, y_m = \beta_m)$, которая принимает единичное значение на таких наборах $(\alpha_1, \alpha_2, \dots, \alpha_n)$, для которых справедливы равенства $\beta_1 = f_1(\alpha_1, \alpha_2, \dots, \alpha_n)$, $\beta_2 = f_2(\alpha_1, \alpha_2, \dots, \alpha_n)$, ..., $\beta_m = f_m(\alpha_1, \alpha_2, \dots, \alpha_n)$. Иными словами, граф с корнем в вершине v представляет функцию, являющуюся полным прообразом реакции $(\beta_1, \beta_2, \dots, \beta_m)$.

Пусть задана комбинационная схема, поведение которой описывается следующей таблицей истинности (табл. 1).

Таблица 1

Пример таблицы истинности, описывающей поведение комбинационной схемы

x_1	x_2	x_3	x_4	y_1	y_2	y_3
0	0	0	0	0	1	1
0	0	0	1	1	1	0
0	0	1	0	1	0	1
0	0	1	1	0	1	1
0	1	0	0	1	1	0
0	1	0	1	1	0	1
0	1	1	0	0	1	1
0	1	1	1	1	1	0

x_1	x_2	x_3	x_4	y_1	y_2	y_3
1	0	0	0	1	0	1
1	0	0	1	0	1	1
1	0	1	0	1	1	0
1	0	1	1	1	0	1
1	1	0	0	0	1	1
1	1	0	1	1	1	0
1	1	1	0	1	0	1
1	1	1	1	0	1	1

ROBDD-граф функции $f(x_1, x_2, x_3, x_4, y_1, y_2, y_3)$ представляется следующим образом (рис. 2).

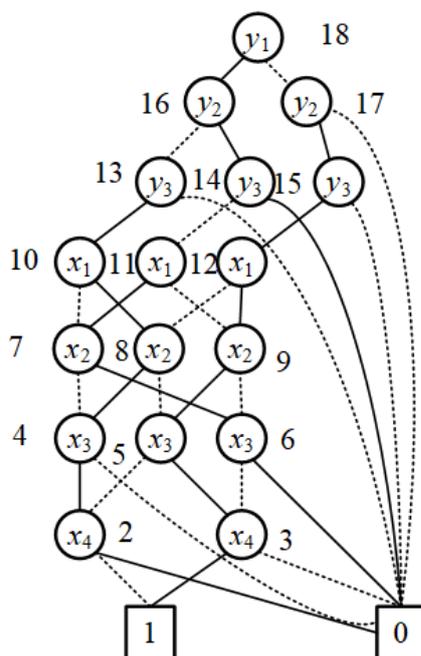


Рис. 2. ROBDD-граф функции $f(x_1, x_2, x_3, x_4, y_1, y_2, y_3)$

Fig. 2. ROBDD of function $f(x_1, x_2, x_3, x_4, y_1, y_2, y_3)$

Будем искать в ROBDD-графе пути, начинающиеся в корневой вершине, до первой вершины, помеченной входной переменной схемы. На рис. 2 видно, что таких путей три: 18–16–13–10, 18–16–15–11 и 18–17–15–12. Им сопоставляются двоичные векторы 101, 110 и 011 соответственно. Выделив подграфы с корневыми вершинами 10, 11 и 12, получаем входные наборы, поступление которых на входы схемы приводит к появлению соответствующей реакции на выходах. Так, для реакции 101 это наборы 0010, 0101, 1000, 1011 и 1110, для реакции 110 – наборы 0001, 0100, 0111, 1010 и 1101, а для реакции 011 – наборы 0000, 0011, 0110, 1001, 1100 и 1111.

3. Модификация общего алгоритма построения ROBDD-графа для построения графа реакций

С помощью ROBDD-графа, построенного по функции $f(x_1, x_2, \dots, x_n, y_1, y_2, \dots, y_m)$, можно получить все достижимые реакции схемы. Однако часто в задачах требуется найти только достижимые реакции схемы, и не требуется знать, какие входные наборы дают соответствующие реакции. Например, для предыдущего примера граф, который представляет только достижимые реакции схемы, выглядит следующим образом (рис. 3).

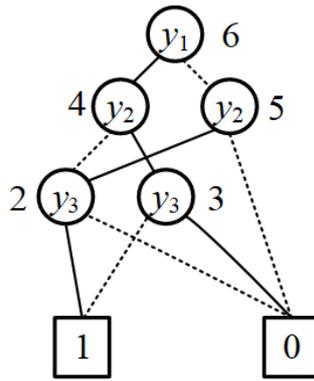


Рис. 3. ROBDD-граф достижимых реакций схемы
Fig. 3. ROBDD of reachable circuit patterns

Этот граф, назовем его в дальнейшем графом достижимых реакций, получается из исходного следующим образом: если 1(0)-дуга из вершины, помеченной выходной переменной схемы, ведет в вершину, помеченную входной переменной схемы, то из вершины, помеченной выходной переменной схемы, проводится 1(0)-дуга в 1-терминальную вершину графа. Затем из графа удаляются все вершины, помеченные входными переменными схемы. Далее в графе выполняются соответствующие для ROBDD операции упрощения.

Очевидно, что граф достижимых реакций содержит существенно меньше внутренних вершин, чем граф функции $f(x_1, x_2, \dots, x_n, y_1, y_2, \dots, y_m)$ (7 вершин вместо 19). Хотя граф достижимых реакций может быть получен из графа функции $f(x_1, x_2, \dots, x_n, y_1, y_2, \dots, y_m)$, желательно иметь алгоритм его построения без предварительного получения графа функции f . Такой алгоритм может быть использован в тех случаях, когда, например, построение графа функции f невозможно из-за экспоненциального роста количества его вершин.

Предлагаемый алгоритм построения ROBDD-графа, представляющего все достижимые реакции схемы, основан на алгоритме построения ROBDD-графа произвольной булевой функции. В графе сначала выполняется разложение по выходным переменным схемы, а затем по ее входным переменным. Модификация этого алгоритма основана на следующем правиле.

Пусть в графе дуга из некоторой вершины, помеченной выходной переменной схемы, ведет в вершину, помеченную входной переменной схемы. Тогда вершина, помеченная входной переменной схемы, представляет функцию $f(x_1, x_2, \dots, x_n) = f(x_1, x_2, \dots, x_n, y_1 = \beta_1, y_2 = \beta_2, \dots, y_m = \beta_m)$, причем эта функция не равна тождественно ни константе 0, ни константе 1 (так как в противном случае эта вершина была бы 1- или 0-терминальной вершиной графа). Эта функция принимает значение 1 на таких двоичных векторах, подача которых на входы схемы приводит к появлению на ее выходах реакции $(\beta_1, \beta_2, \dots, \beta_m)$. Если важна только достижимость реакции, но нет необходимости знать, на каких входных наборах она достигается, то значение функции

$$f(x_1, x_2, \dots, x_n) = f(x_1, x_2, \dots, x_n, y_1 = \beta_1, y_2 = \beta_2, \dots, y_m = \beta_m)$$

можно положить равным 1. Это значит, что в соответствующем ROBDD-графе достижимых реакций дуга из вершины, помеченной выходной переменной схемы, должна вести в 1-терминальную вершину. На основе вышесказанного предлагается поступать следующим образом.

Если в процессе построения графа на множестве входных и выходных переменных обнаружено, что функция f принимает значение 1 на некотором наборе, то соответствующая 0- или 1-дуга из последней вершины v , помеченной выходной переменной схемы, ведет в 1-терминальную вершину графа реакций, а граф с корнем в вершине v , состоящий из вершин, помеченных входными переменными схемы, удаляется.

Если для функции f такого набора не существует, т.е. она тождественно равна нулю, то соответствующая 0- или 1-дуга из вершины v ведет в 0-терминальную вершину строящегося графа.

Далее в полученном графе, состоящем из вершин, помеченных только выходными переменными, выполняются операции упрощения с целью получения ROBDD-графа достижимых реакций.

Утверждение 3. Если для двух схем при одном и том же порядке разложения выходных переменных графы достижимых реакций различны, то схемы реализуют различные системы булевых функций.

Доказательство следует из свойств ROBDD-графов, полученных при одном и том же порядке разложения переменных.

Покажем работу алгоритма на примере. Рассмотрим комбинационную схему с четырьмя входами и тремя выходами, поведение которой описывается системой булевых функций:

$$y_1 = f_1(x_1, x_2, x_3, x_4) = (\overline{x_1 x_2} \vee x_1 x_2)(\overline{x_3 x_4} \vee x_3 x_4) \vee \overline{x_1 x_2}(\overline{x_3} \vee x_4) \vee x_1 \overline{x_2}(x_3 \vee \overline{x_4}),$$

$$y_2 = f_2(x_1, x_2, x_3, x_4) = (\overline{x_1 x_2} \vee x_1 x_2)(\overline{x_3} \vee x_4) \vee \overline{x_1 x_2}(x_3 \vee \overline{x_4}) \vee x_1 \overline{x_2}(\overline{x_3 x_4} \vee x_3 x_4),$$

$$y_3 = f_3(x_1, x_2, x_3, x_4) = (\overline{x_1 x_2} \vee x_1 x_2)(x_3 \vee \overline{x_4}) \vee \overline{x_1 x_2}(\overline{x_3 x_4} \vee x_3 x_4) \vee x_1 \overline{x_2}(\overline{x_3} \vee x_4).$$

Тогда функция

$$\begin{aligned} f(x_1, x_2, x_3, x_4, y_1, y_2, y_3) &= (y_1 \sim f_1(x_1, x_2, x_3, x_4))(y_2 \sim f_2(x_1, x_2, x_3, x_4))(y_3 \sim f_3(x_1, x_2, x_3, x_4)) = \\ &= \left[y_1 \sim \left((\overline{x_1 x_2} \vee x_1 x_2)(\overline{x_3 x_4} \vee x_3 x_4) \vee \overline{x_1 x_2}(\overline{x_3} \vee x_4) \vee x_1 \overline{x_2}(x_3 \vee \overline{x_4}) \right) \right] \wedge \\ &\wedge \left[y_2 \sim \left((\overline{x_1 x_2} \vee x_1 x_2)(\overline{x_3} \vee x_4) \vee \overline{x_1 x_2}(x_3 \vee \overline{x_4}) \vee x_1 \overline{x_2}(\overline{x_3 x_4} \vee x_3 x_4) \right) \right] \wedge \\ &\wedge \left[y_3 \sim \left((\overline{x_1 x_2} \vee x_1 x_2)(x_3 \vee \overline{x_4}) \vee \overline{x_1 x_2}(\overline{x_3 x_4} \vee x_3 x_4) \vee x_1 \overline{x_2}(\overline{x_3} \vee x_4) \right) \right]. \end{aligned}$$

Пусть переменные y_1, y_2 и y_3 имеют номера 1, 2 и 3 соответственно, переменные x_1, x_2, x_3 и x_4 – номера 4, 5, 6 и 7. Число входных переменных $n = 4$, число выходных переменных $m = 3$.

Номер текущей переменной $i = 1$. Подставляем в переменную y_1 значение 0:

$$\begin{aligned} f(x_1, x_2, x_3, x_4, 0, y_2, y_3) &= \left[\left((\overline{x_1 x_2} \vee x_1 x_2)(\overline{x_3 x_4} \vee x_3 x_4) \vee \overline{x_1 x_2}(\overline{x_3} \vee x_4) \vee x_1 \overline{x_2}(x_3 \vee \overline{x_4}) \right) \right] \wedge \\ &\wedge \left[y_2 \sim \left((\overline{x_1 x_2} \vee x_1 x_2)(\overline{x_3} \vee x_4) \vee \overline{x_1 x_2}(x_3 \vee \overline{x_4}) \vee x_1 \overline{x_2}(\overline{x_3 x_4} \vee x_3 x_4) \right) \right] \wedge \\ &\wedge \left[y_3 \sim \left((\overline{x_1 x_2} \vee x_1 x_2)(x_3 \vee \overline{x_4}) \vee \overline{x_1 x_2}(\overline{x_3 x_4} \vee x_3 x_4) \vee x_1 \overline{x_2}(\overline{x_3} \vee x_4) \right) \right]. \end{aligned}$$

Переходим к переменной y_2 ($i = i + 1 = 2$) – подставляем в эту переменную значение 0:

$$\begin{aligned} f(x_1, x_2, x_3, x_4, 0, 0, y_3) &= \left[\left((\overline{x_1 x_2} \vee x_1 x_2)(\overline{x_3 x_4} \vee x_3 x_4) \vee \overline{x_1 x_2}(\overline{x_3} \vee x_4) \vee x_1 \overline{x_2}(x_3 \vee \overline{x_4}) \right) \right] \wedge \\ &\wedge \left[\left((\overline{x_1 x_2} \vee x_1 x_2)(\overline{x_3} \vee x_4) \vee \overline{x_1 x_2}(x_3 \vee \overline{x_4}) \vee x_1 \overline{x_2}(\overline{x_3 x_4} \vee x_3 x_4) \right) \right] \wedge \\ &\wedge \left[y_3 \sim \left((\overline{x_1 x_2} \vee x_1 x_2)(x_3 \vee \overline{x_4}) \vee \overline{x_1 x_2}(\overline{x_3 x_4} \vee x_3 x_4) \vee x_1 \overline{x_2}(\overline{x_3} \vee x_4) \right) \right]. \end{aligned}$$

Переходим к переменной y_3 ($i = i + 1 = 3$) – подставляем в эту переменную значение 0:

$$\begin{aligned} f(x_1, x_2, x_3, x_4, 0, 0, 0) &= \left[\left((\overline{x_1 x_2} \vee x_1 x_2)(\overline{x_3 x_4} \vee x_3 x_4) \vee \overline{x_1 x_2}(\overline{x_3} \vee x_4) \vee x_1 \overline{x_2}(x_3 \vee \overline{x_4}) \right) \right] \wedge \\ &\wedge \left[\left((\overline{x_1 x_2} \vee x_1 x_2)(\overline{x_3} \vee x_4) \vee \overline{x_1 x_2}(x_3 \vee \overline{x_4}) \vee x_1 \overline{x_2}(\overline{x_3 x_4} \vee x_3 x_4) \right) \right] \wedge \\ &\wedge \left[\left((\overline{x_1 x_2} \vee x_1 x_2)(x_3 \vee \overline{x_4}) \vee \overline{x_1 x_2}(\overline{x_3 x_4} \vee x_3 x_4) \vee x_1 \overline{x_2}(\overline{x_3} \vee x_4) \right) \right]. \end{aligned}$$

Переходим к переменной x_1 ($i = i + 1 = 4$) – подставляем в эту переменную значение 0:

$$\begin{aligned} f(0, x_2, x_3, x_4, 0, 0, 0) &= \left[\left(\overline{x_2}(\overline{x_3 x_4} \vee x_3 x_4) \vee x_2(\overline{x_3} \vee x_4) \right) \right] \wedge \\ &\wedge \left[\left(\overline{x_2}(\overline{x_3} \vee x_4) \vee x_2(x_3 \vee \overline{x_4}) \right) \right] \wedge \left[\left(\overline{x_2}(x_3 \vee \overline{x_4}) \vee x_2(\overline{x_3 x_4} \vee x_3 x_4) \right) \right]. \end{aligned}$$

Переходим к переменной x_2 ($i = i + 1 = 5$) – подставляем в эту переменную значение 0:

$$f(0, 0, x_3, x_4, 0, 0, 0) = \overline{(x_3 x_4 \vee x_3 \overline{x_4})} \overline{(x_3 \vee x_4)} (x_3 \vee \overline{x_4}).$$

Переходим к переменной x_3 ($i = i + 1 = 6$) – подставляем в эту переменную значение 0:

$$f(0, 0, 0, x_4, 0, 0, 0) = 0.$$

При подстановке в переменную x_4 и значения 0, и значения 1 значение функции равно 0. Поэтому функции $f(0, 0, 0, x_4, 0, 0, 0)$ соответствует 0-терминальная вершина.

Подставляем в переменную x_3 значение 1: $f(0, 0, 1, x_4, 0, 0, 0) = \overline{x_4} x_4 = 0$.

При подстановке в переменную x_4 и значения 0, и значения 1 значение функции равно 0. Поэтому функции $f(0, 0, 1, x_4, 0, 0, 0)$ соответствует 0-терминальная вершина.

Следовательно, и функции $f(0, 0, x_3, x_4, 0, 0, 0)$ соответствует 0-терминальная вершина.

Возвращаемся к переменной x_2 ($i = 5$) и подставляем в эту переменную значение 1:

$$f(0, 1, x_3, x_4, 0, 0, 0) = \overline{(x_3 \vee x_4)} (x_3 \vee \overline{x_4}) \overline{(x_3 x_4 \vee x_3 \overline{x_4})}.$$

Переходим к переменной x_3 и подставляем в нее значения 0 и 1:

$$f(0, 1, 0, x_4, 0, 0, 0) = 0, \quad f(0, 1, 1, x_4, 0, 0, 0) = 0.$$

Получаем, что функции $f(0, 1, x_3, x_4, 0, 0, 0)$ соответствует 0-терминальная вершина.

Возвращаемся к переменной x_2 : получаем, что $f(0, 0, x_3, x_4, 0, 0, 0) = f(0, 1, x_3, x_4, 0, 0, 0) = 0$. Поэтому функции $f(0, x_2, x_3, x_4, 0, 0, 0)$ также соответствует 0-терминальная вершина.

Возвращаемся к переменной x_1 ($i = 4$) и подставляем в эту переменную значение 1:

$$\begin{aligned} f(1, x_2, x_3, x_4, 0, 0, 0) &= \overline{\left[\overline{x_2} \overline{(x_3 x_4 \vee x_3 \overline{x_4})} \vee \overline{x_2} (x_3 \vee \overline{x_4}) \right]} \wedge \\ &\wedge \overline{\left[\overline{x_2} (x_3 \vee x_4) \vee \overline{x_2} \overline{(x_3 x_4 \vee x_3 \overline{x_4})} \right]} \overline{\left[\overline{x_2} (x_3 \vee x_4) \vee \overline{x_2} (x_3 \vee \overline{x_4}) \right]} \\ f(1, 0, x_3, x_4, 0, 0, 0) &= \overline{(x_3 \vee x_4)} \overline{(x_3 x_4 \vee x_3 \overline{x_4})} (x_3 \vee \overline{x_4}) = 0, \\ f(1, 1, x_3, x_4, 0, 0, 0) &= \overline{(x_3 x_4 \vee x_3 \overline{x_4})} \overline{(x_3 \vee x_4)} (x_3 \vee \overline{x_4}) = 0. \end{aligned}$$

В результате получили, что $f(x_1, x_2, x_3, x_4, 0, 0, 0) = 0$, т. е. функции $f(x_1, x_2, x_3, x_4, 0, 0, 0)$ соответствует 0-терминальная вершина.

Вернувшись к переменной y_3 ($i = 3$) и подставив в нее значение 1, получим, что

$$f(x_1, x_2, x_3, x_4, 0, 0, 1) = 0.$$

Таким образом, имеем

$$f(x_1, x_2, x_3, x_4, 0, 0, y_3) = 0.$$

Теперь вернемся к переменной y_2 ($i = 2$) и подставим в нее значение 1. Далее подставим значение 0 в переменную y_3 . Получим, что $f(x_1, x_2, x_3, x_4, 0, 1, 0) = 0$. Подставим значение 1 в переменную y_3 :

$$\begin{aligned} f(x_1, x_2, x_3, x_4, 0, 1, 1) &= \overline{\left[\overline{\left(\overline{x_1 x_2} \vee x_1 x_2 \right)} \overline{(x_3 x_4 \vee x_3 \overline{x_4})} \vee \overline{x_1 x_2} (x_3 \vee x_4) \vee x_1 \overline{x_2} (x_3 \vee \overline{x_4}) \right]} \wedge \\ &\wedge \overline{\left[\overline{\left(\overline{x_1 x_2} \vee x_1 x_2 \right)} (x_3 \vee x_4) \vee \overline{x_1 x_2} (x_3 \vee \overline{x_4}) \vee x_1 \overline{x_2} \overline{(x_3 x_4 \vee x_3 \overline{x_4})} \right]} \wedge \\ &\wedge \overline{\left[\overline{\left(\overline{x_1 x_2} \vee x_1 x_2 \right)} (x_3 \vee \overline{x_4}) \vee \overline{x_1 x_2} \overline{(x_3 x_4 \vee x_3 \overline{x_4})} \vee x_1 \overline{x_2} (x_3 \vee x_4) \right]}. \end{aligned}$$

Переходим к переменной x_1 – подставляем в эту переменную значение 0:

$$\begin{aligned} f(0, x_2, x_3, x_4, 0, 1, 1) &= \overline{\left[\overline{\left(\overline{x_2} \overline{(x_3 x_4 \vee x_3 \overline{x_4})} \vee x_2 (x_3 \vee x_4) \right)} \right]} \overline{\left[\overline{\left(\overline{x_2} (x_3 \vee x_4) \vee x_2 (x_3 \vee \overline{x_4}) \right)} \right]} \wedge \\ &\wedge \overline{\left[\overline{\left(\overline{x_2} (x_3 \vee \overline{x_4}) \vee x_2 (x_3 x_4 \vee x_3 \overline{x_4}) \right)} \right]}. \end{aligned}$$

Подставим значение 0 в переменную x_2 : $f(0,0,x_3,x_4,0,1,1) = (\overline{x_3x_4} \vee x_3\overline{x_4})(\overline{x_3} \vee x_4)(x_3 \vee \overline{x_4})$.

Подставим значение 0 в переменную x_3 : $f(0,0,0,x_4,0,1,1) = \overline{x_4}$.

Подставим значение 1 в переменную x_4 и получим, что $f(0,0,0,0,0,1,1) = 1$, т.е. функции $f(0,0,0,0,0,1,1)$ соответствует 1-терминальная вершина. Это значит, что функциям $f(0,0,0,x_4,0,1,1)$, $f(0,0,x_3,x_4,0,1,1)$, $f(0,x_2,x_3,x_4,0,1,1)$ и $f(x_1,x_2,x_3,x_4,0,1,1)$ также соответствует 1-терминальная вершина. Ранее мы получили, что функции $f(x_1,x_2,x_3,x_4,0,1,0)$ соответствует 0-терминальная вершина. Поэтому функции $f(x_1,x_2,x_3,x_4,0,1,y_3)$ соответствует внутренняя вершина, помеченная переменной y_3 , из которой 0-дуга ведёт в 0-терминальную вершину, а 1-дуга – в 1-терминальную вершину. Добавим в ROBDD-граф такую вершину (рис. 4).

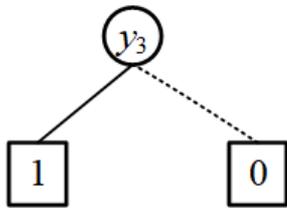


Рис. 4. ROBDD-граф функции $f(x_1, x_2, x_3, x_4, 0, 1, y_3)$

Fig. 4. ROBDD of function $f(x_1, x_2, x_3, x_4, 0, 1, y_3)$

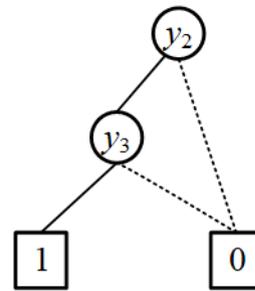


Рис. 5. ROBDD-граф функции $f(x_1, x_2, x_3, x_4, 0, y_2, y_3)$

Fig. 5. ROBDD of function $f(x_1, x_2, x_3, x_4, 0, y_2, y_3)$

Ранее мы получили, что $f(x_1, x_2, x_3, x_4, 0, 0, y_3) = 0$. Поэтому функции $f(x_1, x_2, x_3, x_4, 0, y_2, y_3)$ соответствует внутренняя вершина, помеченная переменной y_2 , из которой 0-дуга ведёт в 0-терминальную вершину, а 1-дуга – в ранее добавленную внутреннюю вершину. Теперь граф примет вид, представленный на рис. 5.

Далее подставим значение 1 в переменную y_1 и продолжим работу алгоритма. В результате получаем ROBDD-граф, изображенный на рис. 6.

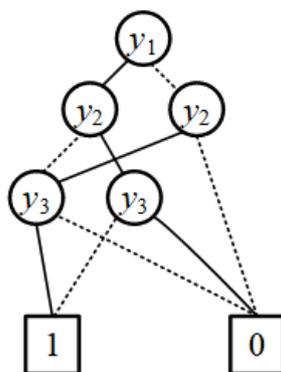


Рис. 6. ROBDD-граф, представляющий все достижимые реакции

Fig. 6. ROBDD representing all reachable patterns

4. Экспериментальное сравнение размеров полного и сокращённого графов

Разработана программа, реализующая алгоритм. Проведено сравнение количества вершин графа функции f и графа достижимых реакций схемы на контрольных примерах схем из набора LGSynth89. Результаты сравнения приведены в табл. 2.

Экспериментальные результаты: сравнение числа вершин в полном и сокращенном графах

Название схемы	Число входов	Число выходов	Число достижимых реакций	Число вершин в графе функции f	Число вершин в графе реакций
alu2	10	6	38	736	10
alu4	14	8	146	8 174	12
C432	36	7	128	2 608	1
x2	10	7	14	107	24

Из таблицы видно, что предлагаемый алгоритм позволяет строить ROBDD-графы достижимых реакций схемы, содержащие существенно меньше, иногда на порядки, внутренних вершин, чем графы соответствующих функций f , зависящих от входных и выходных переменных схемы. В частности, если все возможные двоичные наборы достижимы на выходах схемы, то получается граф реакций, состоящий из одной терминальной вершины 1 (схема C432).

Заключение

Предложен алгоритм построения ROBDD-графа, представляющего множество всех достижимых реакций комбинационной логической схемы, зависящего только от выходных переменных схемы. При построении графа реакций не требуется получения предложенного в работе более громоздкого графа функции f , представляющей входные наборы схемы вместе с порождаемыми ими реакциями. Проведено сравнение количества вершин в графе функции f и соответствующем графе реакций схемы. Эксперименты показали, что предлагаемый алгоритм позволяет строить графы достижимых реакций схемы, значительно более компактные, чем графы функций f .

Список источников

- Jiang J.-H.R., Kravets V.N., Lee N.-Z. Engineering Change Order for Combinational and Sequential Design Rectification // Design, Automation & Test in Europe Conference & Exhibition (DATE) : Proc. 2020. P. 726–731.
- Dhar T., Roy S.K., Giri C. Hardware Trojan Detection by Stimulating Transitions in Rare Nets // Proc. 2019 32nd International Conference on VLSI Design and 2019 18th International Conference on Embedded Systems (VLSID). 2019. P. 537–538.
- Wimmer R., Wimmer K., Scholl C., Becker B. Analysis of Incomplete Circuits Using Dependency Quantified Boolean Formulas // Advanced Logic Synthesis / A. Reis, R. Drechsler (eds). Cham : Springer, 2018. P. 151–168.
- Becker B., Scholl C., Wimmer R. Verification of Incomplete Designs. Formal System Verification / R. Drechsler (eds). Cham : Springer, 2018. P. 37–72.
- Золоторевич Л.А. Аппаратная защита цифровых устройств // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2020. № 50. С. 69–78.
- Yasin M., Rajendran J., Sinanoglu O., Karri R. On Improving the Security of Logic Locking // IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems. 2016. V. 35 (9). P. 1411–1424.
- Gunti N.B., Lingasubramanian K. Fault Sensitive Neutralization of Hardware Trojans Using Multi-level Triple Modular Redundancy Scheme // Proc. 2017 IEEE International Symposium on Nanoelectronic and Information Systems (iNIS). 2017. P. 105–110.
- Gunti N.B., Lingasubramanian K. Neutralization of the Effect of Hardware Trojan in SCADA System Using Selectively Placed TMR // Proc. 2017 IEEE International Symposium on Nanoelectronic and Information Systems (iNIS). 2017. P. 99–104.
- Matrosova A., Provkina V., Nikolaeva E. Masking Internal Node Faults and Trojan Circuits in Logical Circuits // Proc. of 2019 IEEE East-West Design & Test Symposium (EWDTS), 13–16 September 2019, Batumi. Kharkov : IEEE, 2019. P. 416–419.
- Matrosova A., Provkina V. Masking Internal Node Logical Faults and Trojan Circuits Injections with Using SAT Solvers // Proc. 2020 IEEE International Conference on Automation, Quality and Testing, Robotics (AQTR 2020), Cluj-Napoca, Romania 21–23 May 2020. New York : IEEE, 2020. P. 1–4.
- Matrosova A., Provkina V. Applying Incompletely Specified Boolean Functions for Patch Circuit Generation // Proc. 2021 IEEE East-West Design & Test Symposium (EWDTS). Batumi, Georgia, 10–13 September 2021. Red Hook : IEEE, 2021. P. 238–241.

References

- Jiang, J.-H.R., Kravets, V.N. & Lee, N.-Z. (2020) Engineering Change Order for Combinational and Sequential Design Rectification. *Design, Automation & Test in Europe Conference & Exhibition (DATE)*. Proceedings of the Conference. pp. 726–731.

2. Dhar, T., Roy, S.K. & Giri, C. (2019) Hardware Trojan Detection by Stimulating Transitions in Rare Nets. *Proc. of the 32nd International Conference on VLSI Design and 2019 18th International Conference on Embedded Systems (VLSID)*. pp. 537–538.
3. Wimmer, R., Wimmer, K., Scholl, C. & Becker, B. (2018) Analysis of Incomplete Circuits Using Dependency Quantified Boolean Formulas. In: Reis, A. & Drechsler, R. (eds) *Advanced Logic Synthesis*. Springer, Cham.
4. Becker, B., Scholl, C. & Wimmer, R. (2018) Verification of Incomplete Designs. In: Drechsler, R. (eds) *Formal System Verification*. Springer, Cham.
5. Zolotarevich, L.A. (2020) Hardware protection of digital device. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 50. pp. 69–78. DOI: 10.17223/19988605/50/9
6. Yasin, M., Rajendran, J., Sinanoglu, O. & Karri, R. (2016) On Improving the Security of Logic Locking. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*. 35(9). pp. 1411–1424. DOI: 10.1109/TCAD.2015.2511144
7. Gunti, N.B. & Lingasubramanian, K. (2017) Fault Sensitive Neutralization of Hardware Trojans Using Multi-level Triple Modular Redundancy Scheme. *IEEE International Symposium on Nanoelectronic and Information Systems (INIS)*. pp. 105–110.
8. Gunti, N.B. & Lingasubramanian K. (2017) Neutralization of the Effect of Hardware Trojan in SCADA System Using Selectively Placed TMR. *Proceedings 2017 IEEE International Symposium on Nanoelectronic and Information Systems (INIS)*. pp. 99–104.
9. Matrosova, A., Provkin, V. & Nikolaeva, E. (2019) Masking Internal Node Faults and Trojan Circuits in Logical Circuits. *Proceedings of 2019 IEEE East-West Design & Test Symposium (EWDTS)*. 13–16 September. Batumi, Kharkov: IEEE. pp. 416–419.
10. Matrosova, A. & Provkin, V. (2020) Masking Internal Node Logical Faults and Trojan Circuits Injections with Using SAT Solvers. *IEEE International Conference on Automation, Quality and Testing, Robotics (AQTR 2020)*. Cluj-Napoca, Romania, May 21–23, 2020. Proceedings. New York, USA: IEEE. pp. 1–4.
11. Matrosova, A. & Provkin, V. (2021) Applying Incompletely Specified Boolean Functions for Patch Circuit Generation. *Proceedings IEEE East-West Design & Test Symposium (EWDTS)*. Batumi, Georgia, September 10–13, 2021. [Red Hook]: IEEE. pp. 238–241.

Информация об авторах:

Провкин Виктор Алексеевич – аспирант кафедры компьютерной безопасности Института прикладной математики и компьютерных наук Национального исследовательского Томского государственного университета (Томск, Россия). E-mail: prowkan@mail.ru

Матросова Анжела Юрьевна – профессор, доктор технических наук, профессор кафедры компьютерной безопасности Института прикладной математики и компьютерных наук Национального исследовательского Томского государственного университета (Томск, Россия). E-mail: mau11@yandex.ru

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

Information about the authors:

Provkin Viktor A. (Post-graduate Student, National Research Tomsk State University, Tomsk, Russian Federation). E-mail: prowkan@mail.ru

Matrosova Anzhela Y. (Doctor of Technical Sciences, Professor, National Research Tomsk State University, Tomsk, Russian Federation). E-mail: mau11@yandex.ru

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

Received 14.07.2022; accepted for publication 29.11.2022

Поступила в редакцию 14.07.2022; принята к публикации 29.11.2022

Научный журнал

**ВЕСТНИК
ТОМСКОГО
ГОСУДАРСТВЕННОГО
УНИВЕРСИТЕТА**

**УПРАВЛЕНИЕ,
ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА
И ИНФОРМАТИКА**

**TOMSK STATE UNIVERSITY
JOURNAL OF CONTROL AND COMPUTER SCIENCE**

2022. № 61

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 27.12.2022 г. Формат 60x84^{1/8}.
Гарнитура Times. Усл. печ. л. 16,5.
Тираж 250 экз. Заказ № 5296. Цена свободная.

Дата выхода в свет 20.01.2023 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