ISSN 1816-0301 (Print)
ISSN 2617-6963 (Online)

ИНФОРМАТИКА

INFORMATICS

ТОМ 21
VOL.

1 | 2024

20 лет
ЮБИЛЕЙ ЖУРНАЛА

ОТ РЕДАКЦИИ

В журнале «Информатика» публикуются оригинальные и обзорные статьи, описывающие результаты фундаментальных и прикладных исследований специалистов академического и вузовского профиля в области информатики и информационных технологий.

Основной целью журнала является публикация наиболее значимых новых результатов в указанной области. Приветствуются статьи, описывающие заключительные результаты научных проектов и диссертационных исследований, открывающие новые направления исследований, которые находятся на стыке информатики и других наук.

Журнал рассчитан на широкий круг специалистов в области информатики и информационных технологий.

Основные разделы журнала:

- биоинформатика;
- математическое моделирование;
- защита информации и надежность систем;
- информационные технологии;
- логическое проектирование;
- обработка сигналов, изображений, речи, текста и распознавание образов;
- автоматизация проектирования;
- интеллектуальные системы.

Префикс DOI: 10.37661

Индексирование:

Высшей аттестационной комиссией Республики Беларусь журнал «Информатика» был включен в список научных изданий для опубликования результатов диссертационных исследований.

В декабре 2017 г. включен в базу данных Российского индекса научного цитирования (РИНЦ). С помощью инструментов и сервисов, доступных на платформе eLIBRARY (раздел «Личный кабинет»), можно самостоятельно корректировать список своих публикаций и цитирований в РИНЦ.

В июле 2017 г. включен в базу журналов открытого доступа Directory of Open Access Journals (DOAJ).

С помощью поисковых систем Google Scholar, WorldCat, Соционет можно получить свободный доступ к полному тексту научных публикаций журнала.

Адрес редакции:
ул. Сурганова, 6, к. 305, г. Минск, 220012, Беларусь
Тел. +375 (017) 351 26 22

Editorial address:
Surganova str., 6, of. 305, Minsk, 220012, Belarus
Phone +375 (017) 351 26 22

E-mail: rio@newman.bas-net.by
https://inf.grid.by/jour

THE EDITOR'S NOTE

The journal "Informatics" is a scientific publication in computer sciences and information technologies which reviews the results in basic and applied research of scientists from the universities and scientific centers.

The journal focuses on the most significant and modern papers of research projects results and PhD/DSc thesis in computer sciences.

The journal is edited for the specialists in IT and computer sciences research and application.

The main sections of the journal:

- bioinformatics;
- mathematical modeling;
- information protection and system reliability;
- information technology;
- logical design;
- signal, image, speech, text processing and pattern recognition;
- computer-aided design;
- artificial intelligence methods.

DOI Prefix: 10.37661

Indexation:

the journal "Informatics" is in the list of scientific publications recommended by the Higher Attestation Commission of the Republic of Belarus for scientists to publish the results of PhD/DSc research.

In December 2017 the journal was included in the database of the Russian Science Citation Index (RISC) and provides free access to reviewed electronic scientific paper, improving scientific information traffic and also raising quotation of works of the authors (please use https://elibrary.ru or section for authors https://elibrary.ru_author_tools).

In July 2017 included in the database of open access journals Directory of Open Access Journals (DOAJ).

Using the Google Scholar, WorldCat, Соционет search engine, you can get free access to full text of scientific publications of magazine.

ISSN 1816-0301 (Print)
ISSN 2617-6963 (Online)

ОБЪЕДИНЕННЫЙ ИНСТИТУТ ПРОБЛЕМ ИНФОРМАТИКИ
НАЦИОНАЛЬНОЙ АКАДЕМИИ НАУК БЕЛАРУСИ

ИНФОРМАТИКА

Informatika

Том 21, № 1, январь-март 2024

Ежеквартальный научный журнал

Издается с января 2004 г.

Учредитель и издатель – государственное научное учреждение «Объединенный институт проблем информатики Национальной академии наук Беларуси» (ОИПИ НАН Беларуси)

ИНФОРМАТИКА

Том 21, № 1, январь-март 2024

Ответственный за выпуск *Мойсейчик Светлана Сергеевна*
Редактор *Гончаренко Галина Борисовна*
Компьютерная верстка *Бутевич Ольга Борисовна*

Сдано в набор 26.02.2024. Подписано в печать 20.03.2024. Формат 60×84 1/8. Бумага офсетная. Гарнитура Таймс. Ризография. Усл. печ. л. 13,9. Уч.-изд. л. 13,7. Тираж 40 экз. Заказ 2.

Государственное научное учреждение «Объединенный институт проблем информатики Национальной академии наук Беларуси».
Свидетельство о государственной регистрации издателя, изготовителя, распространителя печатных изданий № 1/274 от 04.04.2014. ЛП № 02330/444 от 18.12.13. Ул. Сурганова, 6, 220012, Минск, Беларусь.

ISSN 1816-0301 (Print)
ISSN 2617-6963 (Online)

THE UNITED INSTITUTE OF INFORMATICS PROBLEMS
OF THE NATIONAL ACADEMY OF SCIENCES OF BELARUS

INFORMATICS

Vol. 21, no. 1, January-March 2024

Published quarterly

Issued since January 2004

Founder and publisher – State Scientific Institution "The United Institute of Informatics Problems of the National Academy of Sciences of Belarus" (UIIP NASB)

INFORMATICS

Vol. 21, no. 1, January-March 2024

Issue Head *Sviatlana S. Maiseichyk*
Editor *Halina B. Hancharenka*
Computer Imposition *Volha B. Butsevich*

Sent for press 26.02.2024. Output 20.03.2024. Format 60×84 1/8. Offset paper. Headset Times. Riesography. Printed sheets 13,9. Publisher's signatures 13,7. Circulation 40 copies. Order 2.

State Scientific Institution "The United Institute of Informatics Problems of the National Academy of Sciences of Belarus".
Certificate on the state registration of the publisher, manufacturer, distributor of printing editions no. 1/274 dated 04.04.2014. License for the press no. 02330/444 dated 18.12.13.
6, Surganov Str., 220012, Minsk, Belarus.

ISSN 1816-0301 (Print)
ISSN 2617-6963 (Online)

СОДЕРЖАНИЕ

ISSN 1816-0301 (Print)
ISSN 2617-6963 (Online)

CONTENTS

К 20-летию журнала «Информатика»

Уважаемые читатели, авторы, коллеги!

В марте 2004 г. в государственном научном учреждении «Объединенный институт проблем информатики» вышел в свет первый номер научного журнала «Информатика». Двадцать лет – знаковая дата, большой и трудный путь. За это время журнал преодолел период становления и перешел на уровень стабильного существования и развития.

Еще в советское время в Институте технической кибернетики АН БССР выпускался сборник трудов «Вычислительная техника в машиностроении». Это было регулярное издание с периодичностью четыре раза в год, пользующееся хорошей репутацией у ученых. В 1990-е гг. в институте расширилась издательская деятельность и стали выходить тематические сборники по таким направлениям, как «Логическое проектирование», «Автоматизация проектирования дискретных систем», «Комплексная защита информации», «Суперкомпьютерные системы и их применение», «Анализ цифровых изображений», «Моделирование и информационные технологии проектирования» и другие, труды конференций.

В 2001 г. директор Института технической кибернетики и НИО «Кибернетика» академик В. С. Танаев задумался о том, чтобы изменить название института, включив в него слово «информатика». Он предложил также подумать над созданием журнала «Информатика». В начале 2002 г. ученый совет проголосовал за новое название института – Объединенный институт проблем информатики НАН Беларуси и были поданы документы для официальной смены названия. К сожалению, внезапная смерть Вячеслава Сергеевича 19 июля 2002 г. не позволила ему увидеть результат, поскольку именно в июле 2022 г. завершилось преобразование института.

Исполняющим обязанности генерального директора ОИПИ НАН Беларуси был назначен профессор С. В. Абламейко, а в апреле 2003 г. он был утвержден в должности генерального директора. С. В. Абламейко предложил А. В. Тузикову стать заместителем генерального директора по научной работе, и летом 2003 г. они вместе вернулись к идее создания при институте научного журнала. Вместе с заведующей редакционно-издательским отделом Н. А. Рудой был разработан план подготовки и регистрации журнала, а 10 декабря 2003 г. ученый совет ОИПИ НАН Беларуси поддержал создание журнала «Информатика». С января 2004 г. журнал начал выходить.

В состав первой редакционной коллегии журнала (гл. редактор С. В. Абламейко, зам. гл. редактора А. В. Тузиков) вошли ведущие ученые Беларуси, представляющие основные организации, которые работают в области информатики и информационных технологий: В. В. Анищенко, А. И. Белоус, П. Н. Бибило, С. И. Верещагин, член-корреспондент А. Д. Закревский, С. П. Кундас, академик В. А. Лабунов, В. И. Махнач, М. М. Маханек, П. П. Матус, А. В. Прибыльский, А. А. Петровский, Ю. Н. Сотсков, Ю. С. Харин, академик А. Ф. Чернявский.

На страницах журнала освещаются многие принципиальные вопросы и проблемы информатики, новейшие результаты фундаментальных и прикладных исследований, информация о созданных и внедренных научно-технических и практических разработках и многое другое.

Тематика журнала охватывает широкий спектр задач по важнейшим научным направлениям в области информатики и информационных технологий. Публикуются оригинальные статьи по таким разделам, как:

автоматизация проектирования,

обработка сигналов, изображений, речи, текста и распознавание образов,

математическое моделирование,
информационные технологии,
защита информации и надежность систем,
космические информационные технологии и геоинформатика,
интеллектуальные системы,
биоинформатика,
логическое проектирование.

Список разделов меняется в зависимости от возникновения новых научных направлений и спроса читателей.

За все эти годы было выпущено 76 изданий, опубликовано более 1000 статей. Сложился уникальный авторский коллектив – болеющие за дело ученые и практики. Многие из них, опубликовав свои работы на страницах нашего журнала, защитили диссертации и успешно продолжают заниматься научными исследованиями. В целом тематика журнала всегда отражала приоритеты исследований и разработок в области информатики как в нашей стране, так и за рубежом, а журнал остается одним из наиболее важных научных изданий в Беларуси в данной предметной области.

Со временем журнал вошел в перечень научных изданий для опубликования диссертационных исследований Высшей аттестационной комиссии Республики Беларусь, поменялся состав его редколлегии.

С 2017 г. журнал включен в базу данных Российского индекса научного цитирования и базу журналов открытого доступа Directory of Open Access Journals. Он имеет свой сайт, обеспечивающий автоматизированную информационную поддержку процесса подачи, рецензирования и открытого доступа к публикуемым статьям, которые проходят обязательное рецензирование и редакторскую правку. Все выпуски размещены в открытом доступе на сайте журнала. Благодаря финансовой поддержке руководства института журнал выходит и в бумажном виде, ведется подписка.

В журнале публикуются статьи белорусских и зарубежных авторов на трех языках: русском, белорусском и английском. Опубликованные статьи отражаются в ряде информационных систем, включая eLIBRARY, Google Scholar, WorldCat, Research4life и др.

В настоящее время авторитетные члены редколлегии, в состав которой входят пять академиков, профессора, доктора и кандидаты наук, и редакционный коллектив проводят работу по продвижению журнала в реферативную базу данных рецензируемой научной литературы Scopus. Сформирован международный редакционный совет из известных зарубежных ученых, разработаны новые требования к оформлению статей, правила для авторов, рекомендации рецензентам. На будущее есть немало планов и замыслов.

Редколлегия журнала «Информатика» выражает искреннюю признательность авторам, подписчикам, коллегам и всем читателям за интерес к нашему журналу, за совместную плодотворную работу, за то, что долгие годы вы были вместе с нами!

Особую благодарность хочется выразить сотрудникам редакционно-издательского отдела, которые выполняют огромную работу по редактированию и формированию выпусков журнала.

Академик НАН Беларуси, профессор *С. В. Абламейко,*
главный редактор журнала «Информатика» в 2004–2009 гг.

Генеральный директор ОИПИ НАН Беларуси *С. В. Кругликов,*
доктор военных наук, доцент

Член-корреспондент НАН Беларуси, профессор *А. В. Тузиков,*
главный редактор журнала «Информатика» в 2009–2024 гг.

ЗАЩИТА ИНФОРМАЦИИ И НАДЕЖНОСТЬ СИСТЕМ
INFORMATION PROTECTION AND SYSTEM RELIABILITY

УДК 004.832.32
https://doi.org/10.37661/1816-0301-2024-21-1-9-27

Оригинальная статья
Original Paper

Симметричные физически неклонируемые функции типа арбитр

В. Н. Ярмолик✉, А. А. Иванюк

Белорусский государственный университет информатики и радиоэлектроники,
ул. П. Бровки, 6, Минск, 220013, Беларусь
✉*E-mail: yarmolik10ru@yahoo.com*

Аннотация

Цели. Решается задача построения нового класса физически неклонируемых функций типа арбитр (АФНФ), объединяющих достоинства как классических, так и сбалансированных АФНФ. Актуальность такого исследования связана с активным развитием физической криптографии. В работе преследуются следующие цели: исследование и анализ классических АФНФ, построение новой математической модели АФНФ и разработка нового базового элемента АФНФ.

Методы. Используются методы синтеза и анализа цифровых устройств, в том числе на программируемых логических интегральных схемах, основы булевой алгебры и схемотехники.

Результаты. Установлено, что в классических АФНФ применяется стандартный базовый элемент, выполняющий три функции, а именно функцию формирования двух случайных величин *Generate*, функцию выбора пары путей *Select* и функцию переключения путей *Switch*, которые задаются одним битом запроса. Показано, что совместное использование этих функций, с одной стороны, позволяет достичь высоких характеристик АФНФ, а с другой – приводит к формированию асимметричного поведения АФНФ. С целью анализа основных характеристик АФНФ и их идеального поведения была рассмотрена новая математическая модель АФНФ, аналогичная модели случайного подбрасывания монеты. Для реализации АФНФ, функционирующих согласно предложенной модели, был разработан новый базовый элемент. Показано, что применение предложенного базового элемента позволяет строить симметричные физически неклонируемые функции (С_АФНФ), отличающиеся от классических АФНФ тем, что функции *Generate*, *Select* и *Switch* базового элемента выполняются независимыми его компонентами и задаются разными битами запроса.

Заключение. Предложенный подход к построению симметричных физически неклонируемых функций, основанный на реализации функций *Generate*, *Select* и *Switch* различными компонентами базового элемента, показал свои работоспособность и перспективность. Экспериментально подтвержден эффект улучшения характеристик подобных С_АФНФ, и в первую очередь заметного улучшения их вероятностных свойств, выраженных в равной вероятности ответов. Перспективным представляется дальнейшее развитие идей построения С_АФНФ, экспериментальное исследование их характеристик, а также анализ устойчивости к различного рода атакам, в том числе и с использованием машинного обучения.

Ключевые слова: физическая криптография, физически неклонируемые функции, физические однонаправленные функции, физически неклонируемая функция типа арбитр

Для цитирования. Ярмолик, В. Н. Симметричные физически неклонируемые функции типа арбитр / В. Н. Ярмолик, А. А. Иванюк // Информатика. – 2024. − Т. 21, № 1. – С. 9–27.
https://doi.org/10.37661/1816-0301-2024-21-1-9-27

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Поступила в редакцию | Received 16.07.2023
Подписана в печать | Accepted 03.01.2024
Опубликована | Published 29.03.2024

Symmetric physically unclonable functions of the arbiter type

Vyacheslav N. Yarmolik✉, Alexander A. Ivaniuk

Belarusian State University of Informatics and Radioelectronics,
st. P. Brovki, 6, Minsk, 220013, Belarus
✉*E-mail: yarmolik10ru@yahoo.com*

Abstract
Objectives. The problem of constructing a new class of physically unclonable functions of the arbiter type (APUF) that combines the advantages of both classical and balanced APUF is solved. The relevance of such a study is associated with the active development of physical cryptography. The following goals are pursued in the work: research and analysis of classical APUF, construction of a new mathematical model of APUF and development of a new basic element of APUF.
Methods. The methods of synthesis and analysis of digital devices are used, including those based on programmable logic integrated circuits, the basics of Boolean algebra and circuitry.
Results. It has been established that classical APUF uses a standard basic element that performs three functions, namely, the function of generating two random variables *Generate*, the function of choosing a pair of paths *Select* and the function of switching paths *Switch*, which are specified by one bit of the challenge. It is shown that the joint use of these functions, on the one hand, makes it possible to achieve high characteristics of the APUF, and on the other hand, leads to the formation of an asymmetric behavior of the APUF. In order to analyze the main characteristics of APUF and their ideal behavior, a new mathematical model of APUF was considered, similar to the model of random coin toss. To implement APUF functioning according to the proposed model, a new basic element was developed. It is shown that the use of the proposed basic element allows to build symmetrical physically unclonable functions (C_APUF), which differ from the classical APUF in that the *Generate*, *Select* and *Switch* functions of the basic element are performed by their independent components and are specified by different bits of challenge.
Conclusion. The proposed approach to the construction of symmetrical physically unclonable functions, based on the implementation of the *Generate*, *Select* and *Switch* functions by various components of the base element, has shown its efficiency and promise. The effect of improving the characteristics of similar C_APUF has been experimentally confirmed, and, first of all, a noticeable improvement in their probabilistic properties expressed in equal probability of responses. It seems promising to further develop the ideas of building C_APUF, experimental study of their characteristics, as well as analysis of resistance to various types of attacks, including using machine learning.

Keywords: physical cryptography, physically unclonable functions, physical one-way functions, physically unclonable arbiter-type function

For citation. Yarmolik V. N., Ivaniuk A. A. *Symmetric physically unclonable functions of the arbiter type*. Informatika [*Informatics*], 2024, vol. 21, no. 1, pp. 9−27 (In Russ.).
https://doi.org/10.37661/1816-0301-2024-21-1-9-27

Conflict of interest. The authors declare of no conflict of interest.

Введение. В настоящее время многие цифровые устройства, связанные с нашей повседневной жизнью, подключены к вычислительным системам и сетям, что требует решения задач их идентификации и аутентификации. Физически неклонируемые функции (ФНФ) (Physical Unclonable Functions, *PUF*s) [1, 2] были предложены для решения этих задач [3]. В последние годы сфера применения PUF значительно расширилась за счет их активного использования в криптографии для генерирования криптографических ключей, а также реализации различных криптографических протоколов [4, 5].

Наиболее широко используемое на сегодняшний день определение ФНФ было предложено П. Туилсом (P. Tuyls) [3]. Согласно его формулировке физически неклонируемые функции – это физические системы, неотъемлемым свойством которых является неклонируемость, т. е. невозможность воспроизведения двух идентичных ФНФ. Подобные системы наследуют данное свойство неклонируемости из-за того, что состоят из множества компонентов, параметры которых в процессе создания подобных физических систем принимают случайные значения [3, 6–8]. Невозможность контролировать и управлять параметрами элементов ФНФ, принимающими случайные значения во время производства, делает их уникальными и физически неклонируемыми. ФНФ описываются значениями входных и выходных параметров (сигналов). Подобная пара, состоящая из входного физического параметра запроса (Challenge – *C*) и выходного параметра ответа (Response – *R*), называется парой запрос-ответ (Challenge-Response Pair, *CRP*). Таким образом, ФНФ можно рассматривать как функцию $R = F(C)$, которая преобразует запросы *C* в ответы *R* [3, 6–8].

Анализ большого числа исследований в области ФНФ [3, 6–12] показывает, что в общем случае из всех характеристик, описывающих поведение ФНФ, на первом месте стоит стабильность, характеризующаяся повторяемостью ответов на один и тот же запрос. Затем идет уникальность, которая напрямую связана с неклонируемостью ФНФ, далее отмечают простоту технической реализации и, наконец, непредсказуемость, описывающую случайность ФНФ. Наиболее полно всем приведенным характеристикам отвечают ФНФ, основанные на задержках распространения (delay based) электрических сигналов [6, 8–15].

Существует множество разнообразных реализаций ФНФ, использующих задержки распространения тестового сигнала, среди которых лидирующую позицию занимают так называемые АФНФ (*APUF*) [6, 16–19]. На основании запроса *C* в АФНФ задается конфигурация, как правило, двух функционально и топологически симметричных путей, по которым распространяются идентичные копии тестового сигнала. Ответом *R* АФНФ является результат сравнения временных задержек распространения сигнала по двум путям [6]. Симметричность путей, определяющая их идентичность, обеспечивает близкие значения величин задержек распространения сигналов, которые в силу технологических вариаций в процессе производства, имеющих случайный характер, будут иметь незначительные отличия. Пары симметричных путей АФНФ изготавливаются таким образом, чтобы подобных пар было большое множество, из которого по конкретному запросу *C* выбирается одна из них.

Классической схемой АФНФ является схема, изображенная на рис. 1 [6, 8]. Она строится с использованием *n* последовательно подключенных базовых элементов, состоящих из пар двухвходовых мультиплексоров MUX_1 и MUX_2.

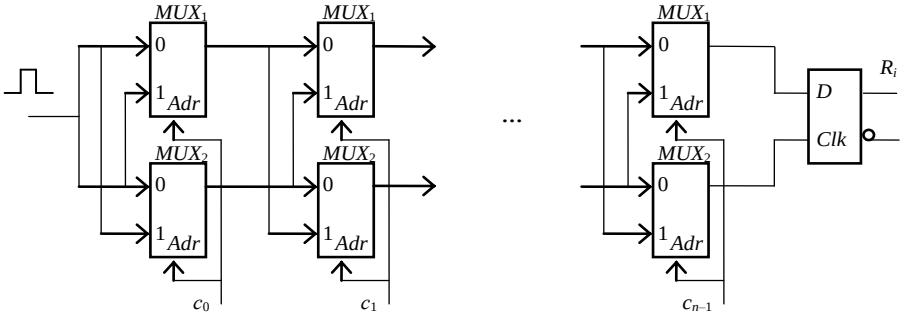


Рис. 1. ФНФ типа арбитр на базе двухвходовых мультиплексоров

Fig. 1. Arbiter-type PUF based on two-input multiplexers

Управляющие входы (*Adr*) мультиплексоров MUX_1 и MUX_2 каждой пары являются одним из входов для задания значения бита c_j запроса C_i, представляющего собой n-разрядный двоичный вектор $C_i = c_0 c_1 c_2 ... c_{n-1}$, где $c_j \in \{0, 1\}$, $j \in \{0, 1, 2, \ldots, n-1\}$. Запрос C_i в схеме АФНФ (см. рис. 1) формирует два пути таким образом, что если для j-й ступени АФНФ $c_j = 0$, то для построения первого пути используется мультиплексор MUX_1, для второго пути MUX_2, а если $c_j = 1$ – наоборот.

Основные проблемы при создании ФНФ заключаются в противоречии требования, которое характеризует стабильность их функционирования, с требованием о непредсказуемости, случайности таких функций. Попытка увеличить стабильность ФНФ увеличивает их уязвимость для различного рода атак, в особенности атак с применением современных методов машинного обучения [20, 21].

Случайность АФНФ оценивается метрикой единообразия (uniformity), которая определяет равновероятность появления ответов 0 и 1. Значения данной метрики меньше 1,0 свидетельствуют о наличии асимметрии в генерируемых парах путей, что особенно характерно для АФНФ, реализованных на программируемой логике (FPGA) [15, 16, 19, 22]. Одним из наиболее эффективных методов увеличения симметрии пар путей АФНФ является их балансировка [22–24]. Эта процедура, требующая дополнительных индивидуальных настроек АФНФ, технологически является сложной задачей, а в ряде случаев ASIC-реализаций и невыполнимой.

Как показано в работах [22, 24], основная причина асимметрии – это взаимозависимость трех функций, реализуемых базовым элементом одновременно. В общем случае задача балансировки пар путей АФНФ решается путем модификации классического базового элемента [22–24]. В статье [24] приведен оригинальный базовый элемент, позволяющий нивелировать асимметрию за счет реализации функции генерирования случайной величины добавленной задержки не на мультиплексорах, а на выделенных линиях задержки. Однако мультиплексоры базового элемента по-прежнему одновременно реализуют две остальные функции, связанные с выбором добавленной задержки и формированием пары путей. По сути, базовый элемент, рассмотренный в работе [24], определяет только знак добавленной задержки, который однозначно влияет на выбор одного из двух путей через базовый элемент. Вторым существенным недостатком базового элемента, приведенного в [24], является необходимость балансировки запросов $C_i = c_0 c_1 c_2 ... c_{n-1}$, $c_j \in \{0, 1\}$. Балансировка заключается в использовании только таких запросов, для которых выполняется баланс единичных и нулевых значений c_j. Соответственно, n должно быть четным, а количество и вид возможных значений запросов C_i ограниченными.

Таким образом, проблема построения эффективных АФНФ, характеризующихся высокой степенью симметрии, как наиболее распространенной разновидности ФНФ является практически нерешенной. В предложенной статье рассматривается задача построения симметричных АФНФ, которые характеризуются обеспечением высоких показателей их характеристик, таких как стабильность, уникальность и единообразие. Материал данной статьи является дальнейшим развитием идей балансировки пар путей АФНФ, изложенных в работе [24]. В сравнении с ранее полученными результатами по балансировке АФНФ [22–24] симметричные АФНФ не накладывают ограничения на количество и вид формируемых запросов и исключают процедуру обеспечения симметричности путей из процесса изготовления АФНФ.

1. Анализ АФНФ. При реализации АФНФ изготавливается множество функционально и топологически идентичных пар электрических путей, представляющих собой последовательно подключенные базовые элементы и их межсоединения. Для построения таких пар путей используются базовые элементы, которые реализуют три функции, а именно функцию генерирования (*Generate*) двух случайных величин добавленной временной задержки, функцию выбора (*Select*) одной из двух случайных величин и функцию переключения (*Switch*) путей [23, 24]. В классическом представлении j-й базовый элемент реализуется с использованием двух мультиплексоров MUX_{1j} и MUX_{2j} (см. рис. 1) [6, 8, 23, 24]. Соответственно, подобный базовый элемент выполняет три указанные ранее функции. Последовательно рассмотрим каждую из них.

Функция *Generate* реализует генерирование уникальных значений задержек как результат случайных факторов при производстве АФНФ. Каждый из двух мультиплексоров MUX_{1j}

и MUX_{2j} базового элемента представляет собой уникальный физический объект, описываемый характеристиками, имеющими фиксированные значения, но полученными как результат множества непредсказуемых и случайных факторов при его изготовлении. В первую очередь к таким характеристикам относятся величины задержек $\Delta(0)_{1,j}$, $\Delta(0)_{2,j}$, $\Delta(1)_{1,j}$ и $\Delta(1)_{2,j}$. Численное значение $\Delta(0)_{1,j}$ определяет временную задержку прохождения сигнала с нулевого входа, обозначенного символом 0, для первого мультиплексора (MUX_{1j}) j-й ступени АФНФ на его выход, а $\Delta(0)_{2,j}$ – задержку для второго мультиплексора (MUX_{2j}). Величины $\Delta(1)_{1,j}$ и $\Delta(1)_{2,j}$ представляют собой задержки сигналов по единичным входам соответствующих мультиплексоров. В процессе функционирования АФНФ эти величины в идеальном случае имеют отличающиеся, но неизменные значения и участвуют в определении величин добавленной разности задержек $\delta_{0,j}$ и $\delta_{1,j}$ согласно уравнениям

$$\delta_{0,j} = \Delta(0)_{1,j} - \Delta(0)_{2,j};\ \delta_{1,j} = \Delta(1)_{1,j} - \Delta(1)_{2,j}. \quad (1)$$

Разность задержки $\delta_{0,j}$ для j-й ступени АФНФ формируется при $c_j = 0$ как добавленная разность задержек $\Delta(0)_{1,j}$ и $\Delta(0)_{2,j}$ прохождения сигнала по двум путям через MUX_{1j} и MUX_{2j}, а разность задержки $\delta_{1,j}$ – как разность $\Delta(1)_{1,j}$ и $\Delta(1)_{2,j}$ при $c_j = 1$. Значения величин добавленной разности задержек $\delta_{0,j}$ и $\delta_{1,j}$ являются результатом функции *Generate* j-го базового элемента.

Функция *Select* в соответствии со значением бита $c_j \in \{0, 1\}$ запроса $C_i = c_0\, c_1 \ldots c_{n-1}$ выбирает одну из двух величин добавленной разности задержек: $\delta_{0,j}$ или $\delta_{1,j}$. Аргументом этой функции является значение бита $c_j \in \{0, 1\}$ запроса C_i, который определяет одну из двух пар путей через j-й базовый элемент. Разница задержки d_j после j-ступени вычисляется в соответствии со следующим рекуррентным уравнением [8, 23]:

$$d_j = \delta_{c_j,j} + d_{j-1} \cdot (-1)^{c_j};\ d_{-1} = 0;\ j = 0,1,2,\ldots,n-1. \quad (2)$$

Как видно из соотношения (2), функция *Select* j-го базового элемента определяет первое слагаемое, которое принимает одно из двух значений: $\delta_{0,j}$ при $c_j = 0$ или $\delta_{1,j}$ при $c_j = 1$.

Аргументом функции *Switch*, как и функции *Select*, является значение c_j, которое определяет знак накопленной на предыдущих базовых элементах по отношению к j-му базовому элементу добавленной задержки d_{j-1} (2). При $c_j = 1$ выполняется переключение одного пути на j-м базовом элементе АФНФ с MUX_{1j} на MUX_{2j}, а второго – с MUX_{2j} на MUX_{1j}, что эквивалентно изменению знака разницы задержек сигналов пары путей на предыдущих ступенях АФНФ.

Суммарное значение разности задержек d_{n-1} по выбранной запросом C_i паре путей, а именно его знак плюс либо минус, и определяет ответ R_i на запрос C_i:

$$d_{n-1} = \delta_{c_{n-1},n-1} + \sum_{j=0}^{n-2}(\delta_{c_j,j} \cdot \prod_{k=j+1}^{n-1}(-1)^{c_k}). \quad (3)$$

Аналогично, как и для случая сбалансированных АФНФ [24], соотношение (3) для вычисления d_{n-1} может быть представлено в следующем виде:

$$d_{n-1} = \delta_{c_{n-1},n-1} + \sum_{j=0}^{n-2}(\delta_{c_j,j} \cdot \prod_{k=j+1}^{n-1}(1-2\cdot c_k)) = \delta_{c_{n-1},n-1} + \sum_{j=0}^{n-2}(\delta_{c_j,j} \cdot (1-2\cdot \bigoplus_{k=j+1}^{n-1} c_k)). \quad (4)$$

Выражение (4) представляет собой две формулы для вычисления величины d_{n-1}, отличные от (3), в которых используются арифметические операции +, – и ·, а также логическая операция сложения по модулю два $\oplus$. Значение разности задержки $\delta_{c_j,j}$ каждой ступени АФНФ в приведенных формулах входит со знаком плюс либо минус в зависимости от запроса C_i. Соотноше-

ния (3) и (4) отличаются друг от друга формулами вычисления знака величины $\delta_{c_j,j}$, $j=\overline{0,n-2}$. При этом отметим, что знак добавленной задержки $\delta_{c_{n-1},n-1}$ всегда положителен.

Например, значение d_{n-1} для $n = 4$ и $C_i = c_0\,c_1\,c_2\,c_3 = 1\ 0\ 0\ 1$ вычисляется с применением формулы (3) следующим образом:

$$d_3 = \delta_{c_0,0}\cdot(-1)^{c_1}\cdot(-1)^{c_2}\cdot(-1)^{c_3} + \delta_{c_1,1}\cdot(-1)^{c_2}\cdot(-1)^{c_3} + \delta_{c_2,2}\cdot(-1)^{c_3} + \delta_{c_3,3} = -\delta_{1,0} - \delta_{0,1} - \delta_{0,2} + \delta_{1,3}.$$

Аналогичный результат может быть получен на основании формулы (4):

$$d_3 = \delta_{c_0,0}\cdot(1-2(c_1\oplus c_2\oplus c_3)) + \delta_{c_1,1}\cdot(1-2(c_2\oplus c_3)) + \delta_{c_2,2}\cdot(1-2(c_3)) + \delta_{c_3,3} = -\delta_{1,0} - \delta_{0,1} - \delta_{0,2} + \delta_{1,3}.$$

Если представить знаки величин $\delta_{c_j,j}$ в выражениях (3) и (4) в виде вектора $B_i = b_0\,b_1\,b_2\,...\,b_{n-2}+1$, где $b_j\in\{+1, -1\}$, $j\in\{0, 1, 2, \ldots, n-2\}$, а $b_{n-1} = +1$, то соотношение, определяющее зависимость B_i от C_i, имеет следующий вид:

$$b_j = (1-2\cdot\overset{n-1}{\underset{k=j+1}{\oplus}} c_k),\ j=\overline{0,n-2};\ b_{n-1}=+1. \tag{5}$$

Принимая во внимание $b_j\in\{+1, -1\}$, можно заключить, что $(1-b_j)/2$ равняется 0 для $b_j = +1$ и 1 для $b_j = -1$. Отсюда следует, что для обеспечения знака +, т. е. значения $b_j = +1$, необходимо, чтобы для значений элементов запроса C_i выполнялось условие $\overset{n-1}{\underset{k=j+1}{\oplus}} c_k = 0$, а для обеспечения $b_j = -1$ это условие представляется как $\overset{n-1}{\underset{k=j+1}{\oplus}} c_k = 1$.

Таким образом, зависимость знаков +1 или –1 добавленных задержек $\delta_{c_j,j}$, $j=\overline{0,n-2}$, от запроса C_i определяется системой из $n - 1$ уравнений, в которой используются арифметические операции умножения и вычитания, а также логическая операция сложения по модулю два:

$$\begin{aligned}
&b_0 = 1-2\cdot(c_1\oplus c_2\oplus c_3\oplus\ldots\oplus c_{n-2}\oplus c_{n-1});\\
&b_1 = 1-2\cdot(c_2\oplus c_3\oplus c_4\oplus\ldots\oplus c_{n-2}\oplus c_{n-1});\\
&\ldots\\
&b_{n-3} = 1-2\cdot(c_{n-2}\oplus c_{n-1});\\
&b_{n-2} = 1-2\cdot c_{n-1}.
\end{aligned} \tag{6}$$

Исходными данными для системы (6) являются элементы $c_1\,c_2\,c_3\,...\,c_{n-1}$ вектора запроса C_i, а результатом – вектор знаков $B_i = b_0\,b_1\,b_2\,...\,b_{n-2}+1$. В дальнейшем символы +1 и –1 знаков b_j заменяются логическими значениями 0 и 1. Соответственно, $b_j = +1$ заменяется на логический ноль, а $b_j = -1$ – на логическую единицу. Например, вектор, состоящий из четырех знаков $B_i = b_0\,b_1\,b_2\,b_3 = +1\ +1\ -1\ -1$, представляется в виде двоичного вектора $B_i = b_0\,b_1\,b_2\,b_3 = 0\ 0\ 1\ 1$.

Значения векторов запроса $C_i = c_0\,c_1\,c_2\,c_3$ и соответствующих им векторов знаков $B_i = b_0\,b_1\,b_2\,b_3$ для $n = 4$ приведены в табл. 1. В столбцах таблицы дается описание трех АФНФ, каждая из которых построена на четырех базовых элементах согласно рис. 1. Первая из них, а именно $АФНФ_0$, представляет собой классическую АФНФ, для которой значения добавленных задержек $\delta_{0,j}$ и $\delta_{1,j}$ для каждого из четырех базовых элементов являются уникальными случайными величинами. Для $АФНФ_0$ и остальных АФНФ приведены значения добавленной разности задержки d_3 на выходах пары путей, знак которой определяет ответ R_i на запрос C_i (см. рис. 1) для $n = 4$. $АФНФ_1$ представляет случай, когда задержки $\Delta(0)_{1,j}$ и $\Delta(1)_{1,j}$ мультиплексора MUX_{1j}, а также задержки $\Delta(0)_{2,j}$ и $\Delta(1)_{2,j}$ мультиплексора MUX_{2j} одинаковы, т. е. $\Delta(0)_{1,j} = \Delta(1)_{1,j}$ и $\Delta(0)_{2,j} = \Delta(1)_{2,j}$. Тогда для добавленных задержек $\delta_{0,j}$ и $\delta_{1,j}$ согласно (1) выполняется равен-

ство $\delta_{0,j} = \delta_{1,j} = \delta_j$. Подобное соотношение задержек является весьма вероятным, учитывая симметрию мультиплексоров, для которых задержки по единичному и нулевому входам должны быть одинаковыми либо, в худшем случае, близкими по величине.

Пример АФНФ$_2$ рассматривался ранее в работе [23] как результат изготовления АФНФ, когда из-за вариаций производственного процесса не только равны задержки $\Delta(0)_{1,j}$ и $\Delta(1)_{1,j}$ мультиплексоров MUX_{1j} и задержки $\Delta(0)_{2,j}$ и $\Delta(1)_{2,j}$ мультиплексоров MUX_{2j} всех $n = 4$ ступеней АФНФ$_2$, но и их разность для всех базовых элементов принимает одинаковое значение d, т. е. $\delta_{0,j} = \delta_{1,j} = \delta_j = d$. Отмечалась реальность такой ситуации в технологических процессах изготовления подобных функций, особенно при реализации АФНФ на программируемых структурах [15, 16, 22].

Таблица 1
Описание функционирования АФНФ$_0$, АФНФ$_1$ и АФНФ$_2$

Table 1
Description of $APUF_0$, $APUF_1$ and $APUF_2$ functioning

$C_i = c_0\,c_1\,c_2\,c_3$	$B_i = b_0\,b_1\,b_2\,b_3$	АФНФ$_0$ $APUF_0$	АФНФ$_1$ $APUF_1$	АФНФ$_2$ $APUF_2$	R_i
		d_3	d_3	d_3	
0 0 0 0	0 0 0 0	$+\delta_{0,0}+\delta_{0,1}+\delta_{0,2}+\delta_{0,3}$	$+\delta_0+\delta_1+\delta_2+\delta_3$	$+d+d+d+d = 4d$	0
0 0 0 1	1 1 1 0	$-\delta_{0,0}-\delta_{0,1}-\delta_{0,2}+\delta_{1,3}$	$-\delta_0-\delta_1-\delta_2+\delta_3$	$-d-d-d+d = -2d$	1
0 0 1 0	1 1 0 0	$-\delta_{0,0}-\delta_{0,1}+\delta_{1,2}+\delta_{0,3}$	$-\delta_0-\delta_1+\delta_2+\delta_3$	$-d-d+d+d = 0$	X
0 0 1 1	0 0 1 0	$+\delta_{0,0}+\delta_{0,1}-\delta_{1,2}+\delta_{1,3}$	$+\delta_0+\delta_1-\delta_2+\delta_3$	$+d+d-d+d = 2d$	0
0 1 0 0	1 0 0 0	$-\delta_{0,0}+\delta_{1,1}+\delta_{0,2}+\delta_{0,3}$	$-\delta_0+\delta_1+\delta_2+\delta_3$	$-d+d+d+d = 2d$	0
0 1 0 1	0 1 1 0	$+\delta_{0,0}-\delta_{1,1}-\delta_{0,2}+\delta_{1,3}$	$+\delta_0-\delta_1-\delta_2+\delta_3$	$+d-d-d+d = 0$	X
0 1 1 0	0 1 0 0	$+\delta_{0,0}-\delta_{1,1}+\delta_{1,2}+\delta_{0,3}$	$+\delta_0-\delta_1+\delta_2+\delta_3$	$+d-d+d+d = 2d$	0
0 1 1 1	1 0 1 0	$-\delta_{0,0}+\delta_{1,1}-\delta_{1,2}+\delta_{1,3}$	$-\delta_0+\delta_1-\delta_2+\delta_3$	$-d+d-d+d = 0$	X
1 0 0 0	0 0 0 0	$+\delta_{1,0}+\delta_{0,1}+\delta_{0,2}+\delta_{0,3}$	$+\delta_0+\delta_1+\delta_2+\delta_3$	$+d+d+d+d = 4d$	0
1 0 0 1	1 1 1 0	$-\delta_{1,0}-\delta_{0,1}-\delta_{0,2}+\delta_{1,3}$	$-\delta_0-\delta_1-\delta_2+\delta_3$	$-d-d-d+d = -2d$	1
1 0 1 0	1 1 0 0	$-\delta_{1,0}-\delta_{0,1}+\delta_{1,2}+\delta_{0,3}$	$-\delta_0-\delta_1+\delta_2+\delta_3$	$-d-d+d+d = 0$	X
1 0 1 1	0 0 1 0	$+\delta_{1,0}+\delta_{0,1}-\delta_{1,2}+\delta_{1,3}$	$+\delta_0+\delta_1-\delta_2+\delta_3$	$+d+d-d+d = 2d$	0
1 1 0 0	1 0 0 0	$-\delta_{1,0}+\delta_{1,1}+\delta_{0,2}+\delta_{0,3}$	$-\delta_0+\delta_1+\delta_2+\delta_3$	$-d+d+d+d = 2d$	0
1 1 0 1	0 1 1 0	$+\delta_{1,0}-\delta_{1,1}-\delta_{0,2}+\delta_{1,3}$	$+\delta_0-\delta_1-\delta_2+\delta_3$	$+d-d-d+d = 0$	X
1 1 1 0	0 1 0 0	$+\delta_{1,0}-\delta_{1,1}+\delta_{1,2}+\delta_{0,3}$	$+\delta_0-\delta_1+\delta_2+\delta_3$	$+d-d+d+d = 2d$	0
1 1 1 1	1 0 1 0	$-\delta_{1,0}+\delta_{1,1}-\delta_{1,2}+\delta_{1,3}$	$-\delta_0+\delta_1-\delta_2+\delta_3$	$-d+d-d+d = 0$	X

Ответы R_i на каждый из запросов C_i для АФНФ, приведенных в табл. 1, определяются знаком добавленной задержки d_3. Предположив, что для АФНФ$_2$ задержка d принимает положительное значение, получим ответ $R_i = 0$ для положительных d_3, $R_i = 1$ – для отрицательных d_3 и метастабильный ответ $R_i = X$ для $d_3 = 0$. Анализ приведенных в табл. 1 данных свидетельствует об асимметрии множества ответов R_i, что негативно сказывается на вероятностных характеристиках АФНФ. В случае АФНФ$_2$ вероятность формирования ответа $R_i = 0$ оказывается существенно больше вероятности появления ответа $R_i = 1$.

Приведенные аналитические соотношения, описывающие функционирование АФНФ, а также примеры АФНФ для $n = 4$ позволяют сформулировать следующее утверждение.

Утверждение 1. *Математическое ожидание* $\mu(d_{n-1})$ *добавленной задержки* d_{n-1} *классической АФНФ равняется* $(\delta_{0,n-1} + \delta_{1,n-1})/2$.

Доказательство. Запрос C_i для АФНФ, представляющей собой n последовательно подключенных базовых элементов, формируется случайным образом с вероятностью $p(C_i) = 1/2^n$. Каждому запросу C_i соответствует ответ R_i в виде добавленной задержки $d_{n-1}(C_i)$ (3). Тогда $\mu(d_{n-1})$ определяется согласно соотношению

$$\mu(d_{n-1}) = \sum_{i=0}^{2^n-1} d_{n\text{-}1}(C_i)\cdot p(C_i) = \frac{1}{2^n}\sum_{i=0}^{2^n-1} d_{n\text{-}1}(C_i). \qquad (7)$$

Величина $d_{n-1}(C_i)$ определяется алгебраической суммой добавленных задержек $b_j \cdot \delta_{c_j,j}$, $j=\overline{0,n-1}$ базовых элементов АФНФ согласно (3). Значение бита c_j, $j=\overline{0,n-2}$, запроса C_i определяет выбор одной из двух задержек $\delta_{0,j}$ или $\delta_{1,j}$, а биты $c_{j+1}\ c_{j+2}\ c_{j+3}\ \ldots\ c_{n-1}$ формируют ее знак b_j, равный +1 или –1 (5). Значения бит $c_j\ c_{j+1}\ c_{j+2}\ \ldots\ c_{n-1}$ запроса C_i принимают все возможные из 2^{n-j} комбинации для каждого двоичного кода $c_0\ c_1\ c_2\ \ldots\ c_{j-1}$ в младших его разрядах. Соответственно, ровно половина задержек $\delta_{c_j,j}$, входящих в выражение $d_{n-1}(C_i)$, примет значение $\delta_{0,j}$, а вторая – $\delta_{1,j}$. Это следует из того факта, что половина значений c_j для всех возможных комбинаций $c_0\ c_1\ c_2\ \ldots\ c_{j-1}$ равняется нулю, а вторая половина – единице. В силу того что в выражении (7) двоичное представление кода $c_{j+1}\ c_{j+2}\ \ldots\ c_{n-1}$ также принимает все возможные комбинации, то, соответственно, согласно (5) половина значений $\delta_{0,j}$ будет иметь знак +1 и столько же знак –1. Аналогично это справедливо и для $\delta_{1,j}$. Исключение составляют задержки $\delta_{0,n-1}$ и $\delta_{1,n-1}$ последнего базового элемента, которые имеют только знак +1. Таким образом, в выражении (7) суммируются все возможные значения $d_{n-1}(C_i)$, каждое из которых состоит из n слагаемых $b_j \cdot \delta_{c_j,j}$, $j=\overline{0,n-1}$. При этом половина $\delta_{0,j}$, $j=\overline{0,n-2}$, и половина $\delta_{1,j}$, $j=\overline{0,n-2}$, будут иметь знак $b_j = +1$, а вторые половины – знак $b_j = -1$, и только $\delta_{0,n-1}$ и $\delta_{1,n-1}$ входят в эту сумму со знаком +1. Все множество слагаемых $d_{n-1}(C_i)$ и их составляющих $b_j \cdot \delta_{c_j,j}$, $j=\overline{0,n-1}$, для $n = 4$ приведено в табл. 1 (см. $АФНФ_0$). Окончательно выражение для $\mu(d_{n-1})$ равняется $(\delta_{0,n-1} + \delta_{1,n-1})/2$, что и требовалось доказать.

Значение $\mu(d_{n-1})$ для $АФНФ_0$ принимает вид $(\delta_{0,3} + \delta_{1,3})/2$, что соответствует утверждению 1, для $АФНФ_1$ оно составляет δ_3, а для $АФНФ_2$ его величина равняется d (см. табл. 1). В идеальном случае для обеспечения равной вероятности получения для АФНФ нулевого и единичного ответов R_i необходимо, чтобы $\mu(d_{n-1}) = 0$.

Приведенный анализ АФНФ позволяет сделать следующие выводы. В первую очередь необходимо отметить удачный набор трех функций *Generate*, *Select* и *Switch*, реализуемых базовым элементом классической АФНФ [23]. Их сочетание обеспечивает формирование случайных значений добавленных задержек, выбор одной из них со знаком плюс либо минус осуществляется в соответствии с запросом C_i. Результирующая сумма (3) таких задержек на n последовательно соединенных базовых элементах определяет ответ R_i, который имеет достаточно хорошие характеристики, отмеченные в большинстве публикаций [8, 15, 23]. Именно сочетание указанных функций обеспечило вполне работоспособное состояние $АФНФ_2$, несмотря на исключительно аномальное сочетание добавленных задержек, которые зачастую не имеют ничего общего со случайными значениями, как это, например, показано для случая $АФНФ_2$.

Необходимо отметить и недостатки классической АФНФ, которые в первую очередь связаны с асимметрией пар путей, вызванной в том числе ненулевым значением $\mu(d_{n-1})$ (см. утверждение 1). Эффект асимметрии сказывается на том, что поведение классических АФНФ отличается от желаемого, особенно при их реализации на программируемой логике типа FPGA, что объясняется сложностью, а в большинстве случаев и невозможностью обеспечения физической идентичности элементов и симметричности их межсоединений [22]. Зачастую наблюдается абсолютная асимметрия, требующая дальнейшей балансировки путей, что относится к нежелательной, но вынужденной процедуре [22, 24]. Данная процедура для АФНФ, приведенных в табл. 1, будет состоять в нивелировании влияния добавленной задержки третьего базового элемента путем добавления по одному из его входов линии задержки. Величина значения временной задержки, используемой для балансировки $АФНФ_0$, равняется $(\delta_{0,3} + \delta_{1,3})/2$. Для $АФНФ_1$ она составляет δ_3, для $АФНФ_2$ – d. Во всех трех случаях после балансировки будет выполняться равенство $\mu(d_3) = 0$, что обеспечит симметричность выбираемых пар путей и, соответственно, улучшение основных характеристик АФНФ.

Эффект симметричности проявляется на уровне конкретных скорректированных ответов в виде добавленной задержки $d_3(C_i)$. Например, на запрос $C_i = 0\ 0\ 1\ 1$ сбалансированная $АФНФ_0$ генерирует задержку $d_3(0011) = +\delta_{0,0} +\delta_{0,1}–\delta_{1,2} +\delta_{1,3} – (\delta_{0,3} + \delta_{1,3})/2 = +\delta_{0,0} +\delta_{0,1}–\delta_{1,2}– \delta_{0,3}/2 + \delta_{1,3}/2$. В соответствии с утверждением 1 для запроса $C_i = 0\ 0\ 1\ 1$ существует запрос C_k, $i \neq k$,

для которого формируется значение $d_3(C_k)$, равное по абсолютной величине $d_3(0011)$, но имеющее противоположный знак. Действительно, для C_k = 0 0 1 0 получаем $d_3(0010) = -\delta_{0,0} -\delta_{0,1}+\delta_{1,2} +\delta_{0,3} - (\delta_{0,3} + \delta_{1,3})/2 = -\delta_{0,0} -\delta_{0,1}+\delta_{1,2} +\delta_{0,3}/2- \delta_{1,3}/2$. Видно, что $d_3(0011)$ и $d_3(0010)$ равны по абсолютной величине $|d_3(0011)| = |d_3(0010)|$, но имеют противоположные знаки. Это свидетельствует о симметричности относительно нулевого среднего $\mu(d_3) = 0$. Очевидно, что для общего случая сбалансированных АФНФ, в том числе и для всех трех сбалансированных АФНФ, приведенных в табл. 1, для запроса C_i всегда существует запрос C_k. Взаимосвязь C_i и C_k определяется равенством

$$d_{n-1}(C_i) = -d_{n-1}(C_k);\ i \neq k;\ i,k \in \{0,1,2,\ldots n-1\}. \quad (8)$$

Балансировка является одним из наиболее эффективных методов увеличения стабильности АФНФ [17, 22, 24]. Однако эта процедура, требующая дополнительных индивидуальных настроек АФНФ, технологически может быть сложной задачей, а в ряде случаев ASIC-технологий и невыполнимой. Балансировка путей означает отход от основополагающей концепции ФНФ, заключающейся в использовании при изготовлении ФНФ их единого схемотехнического описания для получения неповторяемого поведения ФНФ, описываемого уникальной функцией $R = F(C)$ [6, 8].

2. Математическая модель симметричных АФНФ. Анализ, проведенный в предыдущем разделе, показал, что в сбалансированной АФНФ наблюдается симметрия формируемых добавленных задержек d_{n-1} относительно их нулевого математического ожидания. Знак величины d_{n-1} плюс либо минус определяет ответ $R_i \in \{0, 1\}$, для которого и необходимо выполнение условия равной вероятности двух возможных ответов.

Приведенная формулировка поведения симметричных АФНФ напоминает классическую модель, описывающую подбрасывание монеты. Можно предположить, что структура АФНФ представляет собой монету, а подаваемые на нее запросы C_i имитируют ее подбрасывание, в результате которого формируются ответы $R_i = 0$ (орел) и $R_i = 1$ (решка). В классической постановке задачи результатом подбрасывания монеты являются независимые величины ответов (орел или решка), что нельзя утверждать в случае АФНФ. Более того, важным свойством АФНФ является повторяемость результатов эксперимента, заключающаяся в том, что для повторяемого запроса C_i ответ R_i должен быть таким же, т. е. повторяемым. В то же время уникальность, характеризующаяся различными значениями ответов R_i на одни и те же запросы C_i для физически различных АФНФ, определяется их структурой и внутренними параметрами (монетой). Что касается параметров структуры АФНФ, то их задание определяется множеством случайных факторов при изготовлении АФНФ, однако при проведении экспериментов эти параметры в идеальном случае должны быть неизменными.

Принимая во внимание приведенные замечания, а также анализ классической АФНФ, изложенный в разд. 1, можно заключить, что математическая модель функционирования симметричных АФНФ описывается моделью псевдоподбрасывания монеты. Единственным отличием модели псевдоподбрасывания монеты от классической модели подбрасывания является повторяемость результатов подбрасывания для идентичных подбрасываний, определяемых одним и тем же запросом C_i. Все остальные свойства псевдоподбрасывания должны быть такими же либо максимально близкими к свойствам процедуры подбрасывания монеты. Соотношение двух указанных математических моделей аналогично соотношению моделей генерирования псевдослучайных и случайных чисел [25].

Формулировка обобщенной математической модели, описывающей симметричные АФНФ, как аналога классической модели псевдоподбрасывания монеты принимает следующий вид:

1. Структура АФНФ (монета) однозначно задается множеством $\delta_0, \delta_1, \delta_2, \ldots, \delta_{n-1}$ произвольных, сгенерированных случайным образом величин добавленной разности задержек. Отметим, что значения $\delta_0, \delta_1, \delta_2, \ldots, \delta_{n-1}$ не изменяются в процессе функционирования АФНФ.

2. Запросы C_i, имитирующие подбрасывание монеты, формируются случайным образом в виде двоичного вектора $C_i = c_0\, c_1\, c_2 \ldots c_{n-1}$, где $c_j \in \{0, 1\}$, j = 0, 1, 2, …, n–1, а $p(c_j = 0) = p(c_j = 1) = 0{,}5$.

3. Процедура имитации подбрасывания монеты заключается в вычислении суммы d_{n-1}:

$$d_{n-1} = (1-2\cdot c_0)\cdot\delta_0 + (1-2\cdot c_1)\cdot\delta_1 + (1-2\cdot c_2)\cdot\delta_2 + \ldots + (1-2\cdot c_{n-1})\cdot\delta_{n-1}. \quad (9)$$

Выражение $(1-2c_j)$ представляет собой знак $b_j \in \{+1, -1\}$ слагаемого δ_j вектора знаков $B_i = b_0 b_1 b_2 \ldots b_{n-1}$ слагаемых суммы (9).

4. Значение знака суммы d_{n-1}, соответственно +1 или −1, определяет ответ $R_i = 0$ или $R_i = 1$ (орел или решка). Так как значения $\delta_0, \delta_1, \delta_2, \ldots, \delta_{n-1}$ не изменяются в процессе функционирования АФНФ, повторение запроса C_i для $d_{n-1} \neq 0$ приведет к повторению значения ответа R_i. В случае равенства нулю суммы d_{n-1} результатом будет метастабильное значение X ответа R_i. Оно характеризуется получением как нулевого, так и единичного результата при повторении одного и того же запроса C_i.

Представленная математическая модель характеризуется свойством симметричности получаемых значений суммы d_{n-1}, что является следствием следующего утверждения.

Утверждение 2. *Для математического ожидания* $\mu(d_{n-1})$ *суммы* d_{n-1}, *полученной согласно (9), справедливо равенство* $\mu(d_{n-1}) = 0$.

Доказательство. Запросы C_i для получения суммы (9) согласно описанной ранее обобщенной математической модели формируются случайным образом с вероятностью $p(C_i) = 1/2^n$. Согласно этой модели для запроса C_i формируется ответ $R_i(C_i)$ в виде знака +1 или −1 суммы d_{n-1}. В то же время для каждого запроса C_i существует инверсный запрос C_k, компоненты которого $\overline{c_0}\,\overline{c_1}\,\overline{c_2}\ldots\overline{c_{n-1}}$ принимают инверсные значения по отношению к компонентам $c_0 c_1 c_2 \ldots c_{n-1}$ запроса C_i. Использование запроса C_k приведет к инвертированию знаков слагаемых в выражении (9) и, соответственно, знака d_{n-1}. Таким образом, $d_{n-1}(C_i) = -d_{n-1}(C_k)$, что свидетельствует о выполнении условия симметричности (8), для которого сумма (7), определяющая значение математического ожидания $\mu(d_{n-1})$, равняется нулю. Что и требовалось доказать.

АФНФ, для которых выполняется свойство симметрии (8), обозначим как С_АФНФ. Примеры подобных функций для случая $n = 4$ и различных значений $\delta_0, \delta_1, \delta_2, \delta_3$ приведены в табл. 2.

Таблица 2
Описание функционирования симметричных С_АФНФ$_0$, С_АФНФ$_1$ и С_АФНФ$_2$

Table 2
Description of symmetric C_APUF$_0$, C_APUF$_1$ and C_APUF$_2$ functioning

$C_i = c_0\,c_1\,c_2\,c_3$	С_АФНФ$_0$ *C_APUF$_0$*		С_АФНФ$_1$ *C_APUF$_1$*		АФНФ$_1$ *APUF$_1$*		С_АФНФ$_2$ *C_APUF$_2$*	
	$\delta_0, \delta_1, \delta_2, \delta_3 =$ $= +1, -2, +3, -4$		$\delta_0, \delta_1, \delta_2, \delta_3 =$ $= +1, +2, +3, +4$		$\delta_0, \delta_1, \delta_2, \delta_3 =$ $= +1, +2, +3, +4$		$\delta_0, \delta_1, \delta_2, \delta_3 =$ $= +2, +2, +2, +2$	
	d_3	R_i	d_3	R_i	d_3	R_i	d_3	R_i
0 0 0 0	+1−2+3−4 =−2	1	+1+2+3+4 =+10	0	+1+2+3+4 =+10	0	+2+2+2+2 =+8	0
0 0 0 1	+1−2+3+4 =+6	0	+1+2+3−4 =+2	0	−1−2−3+4 =−2	1	+2+2+2−2 =+4	0
0 0 1 0	+1−2−3−4 =−8	1	+1+2−3+4 =+4	0	−1−2+3+4 =+6	0	+2+2−2+2 =+4	0
0 0 1 1	+1−2−3+4 = 0	X	+1+2−3−4 =−4	1	+1+2−3+4 =+4	0	+2+2−2−2 = 0	X
0 1 0 0	+1+2+3−4 =+2	0	+1−2+3+4 =+6	0	−1+2+3+4 =+8	0	+2−2+2+2 =+4	0
0 1 0 1	+1+2+3+4 =+10	0	+1−2+3−4 =−2	1	+1−2−3+4 = 0	X	+2−2+2−2 = 0	X
0 1 1 0	+1+2−3−4 =−4	1	+1−2−3+4 = 0	X	+1−2+3+4 = +6	0	+2−2−2+2 = 0	X
0 1 1 1	+1+2−3+4 =+4	0	+1−2−3−4 =−8	1	−1+2−3+4 =+2	0	+2−2−2−2 =−2	1
1 0 0 0	−1−2+3−4 =−4	1	−1+2+3+4 =+8	0	+1+2+3+4 =+10	0	−2+2+2+2 =+4	0
1 0 0 1	−1−2+3+4 =+4	0	−1+2+3−4 = 0	X	−1−2−3+4 = −2	1	−2+2+2−2 = 0	X
1 0 1 0	−1−2−3−4 =−10	1	−1+2−3+4 =+2	0	−1−2+3+4 =+4	0	−2+2−2+2 = 0	X
1 0 1 1	−1−2−3+4 =−2	1	−1+2−3−4 =−6	1	+1+2−3+4 =+4	0	−2+2−2−2 =−4	1
1 1 0 0	−1+2+3−4 = 0	X	−1−2+3+4 =+4	0	−1+2+3+4 =+8	0	−2−2+2+2 = 0	X
1 1 0 1	−1+2+3+4 =+8	0	−1−2+3−4 =−4	1	+1−2−3+4 = 0	X	−2−2+2−2 =−2	1
1 1 1 0	−1+2−3−4 =−6	1	−1−2−3+4 =−2	1	+1−2+3+4 = 6	0	−2−2−2+2 =−4	1
1 1 1 1	−1+2−3+4 =+2	0	−1−2−3−4 =−10	1	−1+2−3+4 =+2	0	−2−2−2−2 =−10	1

В табл. 2 также представлена реализация классической АФНФ$_1$ с такими же значениями δ_0, δ_1, δ_2, δ_3, = +1, +2, +3, +4, как и для С_АФНФ$_1$. Сравнение результатов поведения С_АФНФ$_1$ и АФНФ$_1$ показывает преимущества новой обобщенной модели симметричных АФНФ, обеспечивающих абсолютную симметрию ответов.

Приведенные конкретные реализации модели симметричных АФНФ позволяют сделать вывод об эффективности предложенной модели подобных АФНФ. Как видно из табл. 1, С_АФНФ$_0$, С_АФНФ$_1$ и С_АФНФ$_2$ характеризуются идеальной симметрией по сравнению с классической АФНФ$_1$.

3. Практическая реализация симметричных АФНФ. Множество разновидностей АФНФ представлено различными их модификациями для реализации – как схемой ASIC, так и схемой с использованием современных FPGA. В публикациях [8, 15, 17, 22, 24] указано, что идентичность элементов АФНФ, симметричность их геометрических параметров и межсоединений являлись и являются основополагающим требованием к подобным структурам. В то же время, понимая сложность, а в большинстве случаев и невозможность достижения топологической симметричности, в особенности межсоединений для FPGA, создаются и широко применяются АФНФ на программируемой логике. По мнению авторов, это объясняется тем, что значения величин добавленной разности задержек δ_0, δ_1, δ_2, …, δ_{n-1} в принципе могут быть произвольными, в том числе и детерминированными. Для любых значений этих величин, как следует из примеров, приведенных в табл. 1 и 2, достигается необходимая уникальность АФНФ. Однако закономерные зависимости между разностями задержек δ_0, δ_1, δ_2, …, δ_{n-1} приводят к уязвимости для взлома АФНФ путем применения методов машинного обучения [20, 21].

Требование формирования множества произвольных, сгенерированных случайным образом величин добавленной разности задержек δ_0, δ_1, δ_2, …, δ_{n-1} остается центральным для АФНФ. Попытки уменьшить зависимость между значениями задержек представлены в ряде работ [22, 24, 26, 27], которые направлены на совершенствование реализации функции *Generate* базового элемента АФНФ. Применение линий задержки с величинами задержки, существенно превышающими временные задержки распространения сигналов через элементы АФНФ и ее межсоединения, позволяет нивелировать топологическую асимметричность АФНФ [24]. Это, например, достигается включением по входам классического базового элемента АФНФ r последовательно подключенных логических элементов с целью увеличения значений задержек по каждому из двух входов базового элемента [24].

Реализация базового элемента АФНФ, использующего линии задержки *Del*1 и *Del*2, приведена на рис. 2, *а* [24]. Задержки $\Delta_{1,l}(j)$ и $\Delta_{2,l}(j)$, где $l \in \{0, 1, \ldots, r-1\}$, на каждом из r элементов линий задержки *Del*1 и *Del*2 по входам j-го базового элемента представляют собой значение случайной и независимой величины. Отметим, что для элементов, образующих линии задержки *Del*1 и *Del*2, например повторителей, в процессе их производства величина этих задержек случайным образом принимает конкретное значение. Этот факт и определяет уникальность и неповторяемость АФНФ. Задержка $\Delta_{v,l}(j)$, $v \in \{1, 2\}$, описывается математическим ожиданием (средним значением) $\mu(\Delta_{v,l}(j))$ и дисперсией (отклонением значений) $Var(\Delta_{v,l}(j))$. Величины добавленной разности δ_j задержек $\Omega1(j)$ и $\Omega2(j)$ последовательно соединенных логических элементов (например, повторителей) j-го базового элемента определяются как $\delta_j = \Omega1(j) - \Omega2(j) = [\Delta_{1,0}(j) + \Delta_{1,1}(j) + \ldots + \Delta_{1,r-1}(j)] - [\Delta_{2,0}(j) + \Delta_{2,1}(j) + \ldots + \Delta_{2,r-1}(j)]$. Мерой разброса значений добавленной разности задержек δ_j, которая является линейной комбинацией случайных величин $\Delta_{v,l}(j)$, будет значение дисперсии $Var(\delta_j) = 2r \times Var(\Delta_{v,l}(j))$. Среднеквадратическое (стандартное) отклонение $\sigma = \sqrt{2r \times Var(\Delta_{v,l}(j))}$ разности задержек δ_j растет с ростом величины r, которая определяет диапазон 3σ ее изменения относительно математического ожидания.

В рассмотренной структуре базового элемента (рис. 2, *а*) мультиплексоры выполняют только одну функцию, а именно *Switch*, так как величины задержек $\Omega1(j)$ и $\Omega2(j)$ на схемах *Del*1 и *Del*2 являются доминирующими, определяют значение δ_j и, соответственно, реализуют

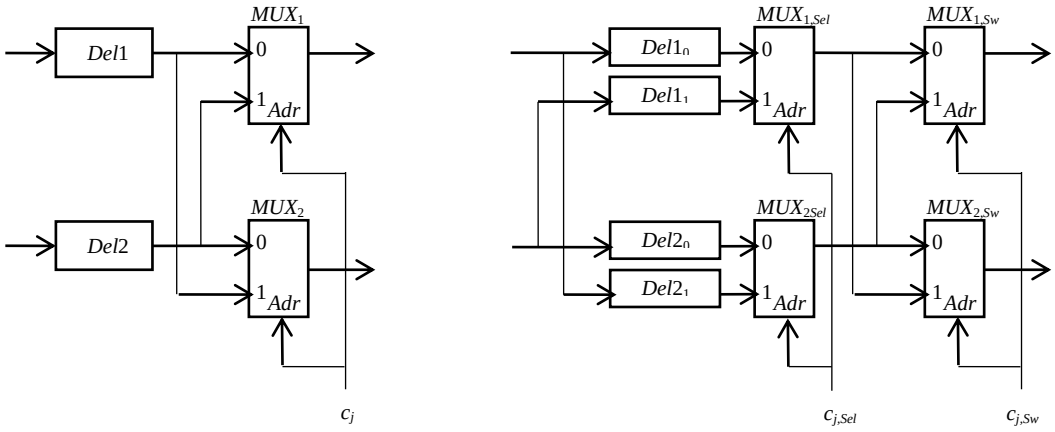
функцию *Generate*. Как показано в работе [24], описание функционирования *j*-го базового элемента АФНФ на линиях задержки отличается от описания (2) классической АФНФ и принимает вид

$$d_j = (d_{j-1} + \delta_j) \times (-1)^{c_j}. \tag{10}$$

Суммарное значение разности задержек d_{n-1} (11) по выбранной запросом C_i паре путей, аналогичное выражению (4), определяет ответ R_i на запрос C_i:

$$d_{n-1} = \sum_{j=0}^{n-1} \delta_j \times \prod_{i=j}^{n-1} (-1)^{c_i} = \sum_{j=0}^{n-1} \delta_j \times \prod_{i=j}^{n-1} (1 - 2 \times c_i) = \sum_{j=0}^{n-1} \delta_j \times (1 - 2 \times \bigoplus_{i=j}^{n-1} c_i). \tag{11}$$

Выражение (11) для вычисления d_{n-1} полностью соответствует реализации модели симметричных АФНФ, имитирующей псевдоподбрасывание монеты (9). В обоих случаях знак слагаемых, представляющих собой добавленные задержки δ_j, равновероятно принимает значения + и −.



a) *b)*

Рис. 2. Функциональные схемы базового элемента: *a)* сбалансированной АФНФ; *b)* симметричной АФНФ

Fig. 2. Functional diagrams of the base element: a) balanced PUF; b) symmetric PUF

Использование схем *Del*1 и *Del*2 в классическом базовом элементе позволяет нивелировать влияние задержек на межсоединениях и мультиплексорах [24]. Это достигается тем, что величины задержки сигнала $\Omega1(j)$ и $\Omega2(j)$ на схемах *Del*1 и *Del*2 задаются существенно бóльшими по сравнению с задержками на мультиплексорах $\Delta(0)_{1,j}$, $\Delta(0)_{2,j}$, $\Delta(1)_{1,j}$ и $\Delta(1)_{2,j}$ и межсоединениях между ними. Количество *r* элементов в линиях задержки *Del*1 и *Del*2 может быть произвольным и по необходимости увеличиваться для обеспечения требуемых соотношений задержек и отклонений тестового сигнала.

В структуре базового элемента сбалансированной АФНФ (рис. 2, *a*) [24] мультиплексоры по сравнению с классическим базовым элементом выполняют только одну функцию, а именно *Switch*, и не выполняют функцию *Select*. Соответственно, функция *Generate* выполняется на схемах *Del*1 и *Del*2. Ее результатом является одна случайная величина δ_j, определяемая разностью $\Omega1(j) - \Omega2(j)$ и имеющая знак + при $c_j = 0$ и знак − при $c_j = 1$. Высокая степень сбалансированности рассмотренных в работе [24] АФНФ позволяет достичь приемлемого уровня стабильности, единообразия и внутрикристальной уникальности АФНФ. Это в первую очередь достигается обеспечением симметричности гарантированной математической моделью псевдоподбрасывания монеты, описываемой соотношением (9). Однако отсутствие операции *Select* влияет на уникальность АФНФ, определяемую случайными факторами изготовления АФНФ.

Отметим, что в классической реализации базового элемента все три функции, а именно *Generate*, *Select* и *Switch*, выполняются на двух мультиплексорах, что приводит к асимметрии АФНФ (см. утверждение 1).

В качестве базового элемента симметричных АФНФ (С_АФНФ) используем структуру, объединяющую достоинства как классического базового элемента, так и базового элемента сбалансированных АФНФ. Функционально такой элемент выполняет все три функции классического элемента и обеспечивает условие симметричности пар путей, описываемое утверждением 2. Реализация базового элемента симметричных АФНФ показана на рис. 2, *b*.

Предлагаемый базовый элемент симметричных АФНФ выполняет все три функции классического базового элемента. В то же время в отличие от классических и сбалансированных АФНФ все эти функции выполняются на различных компонентах базового элемента.

Функция *Generate* реализуется на линиях задержки $Del1_0$, $Del1_1$, $Del2_0$ и $Del2_1$ с задержками $\Omega1_0(j)$, $\Omega1_1(j)$, $\Omega2_0(j)$ и $\Omega2_1(j)$, которые должны быть несравнимо больше задержек на мультиплексорах. Таким образом, временные задержки мультиплексоров практически не влияют на величины генерируемых добавленных разностей задержки. Аналогично, как и для классической АФНФ (1), эти величины определяются из соотношений

$$\delta_{0,j} = \Omega1_0(j) - \Omega2_0(j); \quad \delta_{1,j} = \Omega1_1(j) - \Omega2_1(j). \tag{12}$$

Функцию *Select* выполняют два мультиплексора $MUX_{1,Sel}$ и $MUX_{2,Sel}$, которые в зависимости от значения бита запроса $c_{j,Sel} \in \{0, 1\}$ выбирают одну из двух случайных величин добавленных разностей задержек $\delta_{0,j}$ или $\delta_{1,j}$. В рамках рассмотренной ранее математической модели симметричных АФНФ данная функция определяет структуру монеты (см. разд. 2), которая описывается множеством задержек $\delta_{c_{0,Sel},0}, \delta_{c_{1,Sel},1}, \delta_{c_{2,Sel},2}, ..., \delta_{c_{m-1,Sel},n-1}$. Таким образом, биты запроса $c_{0,Sel}\, c_{1,Sel}\, c_{2,Sel}$ и $c_{n-1,Sel}$, по сути, выбирают одну их 2^n возможных монет, для которых и реализуется процедура псевдоподбрасывания.

Два следующих мультиплексора $MUX_{1,Sw}$ и $MUX_{2,Sw}$ каждого базового элемента отвечают за реализацию функции *Switch*, выполняющей переключение пары путей (она же монета), выбранной битами запроса $c_{j,Sel}$. В терминах математической модели псевдоподбрасывания монеты функция *Switch* реализует подбрасывание монеты согласно (9), т. е. определяет знаки слагаемых $\delta_{c_{0,Sel},0}, \delta_{c_{1,Sel},1}, \delta_{c_{2,Sel},2}, ..., \delta_{c_{m-1,Sel},n-1}$, выбранных функцией *Select*. Аргументом этой функции j-го базового элемента (рис. 1, *b*) является бит запроса $c_{j,Sw} \in \{0, 1\}$. Так же, как и в случае классического базового элемента, на значение знаков слагаемых, представляющих собой добавленные разности задержки, оказывают влияние и биты запроса $c_{j,Sel}$ на знаки указанных величин, но только для предыдущих базовых элементов по отношению к текущему j-му. Бит запроса $c_{j,Sw}$ влияет на знак добавленной разности задержки j-го базового элемента непосредственно, как это видно из рекуррентного выражения

$$d_j = (d_{j-1} \cdot (-1)^{c_{j,Sel}} + \delta_{c_{j,Sel},j}) \cdot (-1)^{c_{j,Sw}} = d_{j-1} \cdot (-1)^{c_{j,Sel} \oplus c_{j,Sw}} + \delta_{c_{j,Sel},j} \cdot (-1)^{c_{j,Sw}};$$
$$d_{-1} = 0; \; j = 0,1, \ldots, n-1.$$

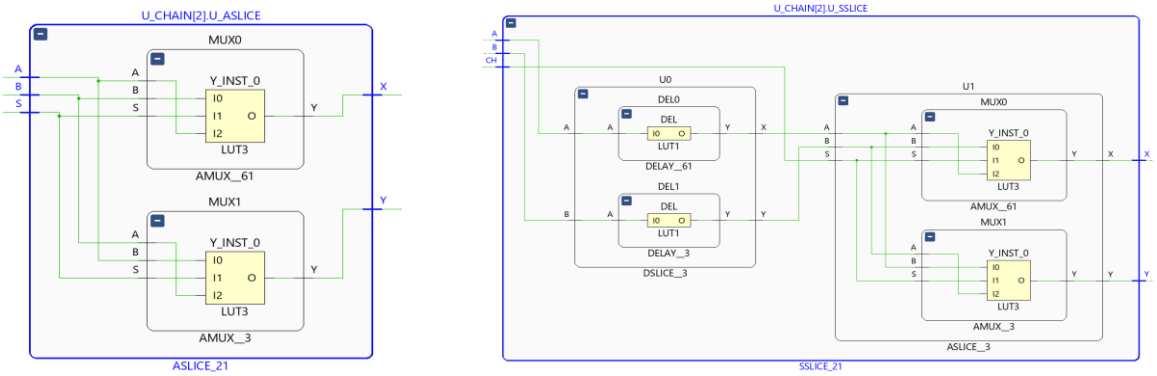
Приведенное выражение объединяет соотношения (2) и (10) в связи с тем, что предложенный новый базовый элемент включает основные свойства как классической АФНФ, так и сбалансированной. В части классической АФНФ новый базовый элемент реализует все три его функции, в том числе весьма значимую функцию *Switch*. В то же время идеи, заложенные в сбалансированных АФНФ, нашли свое отражение в обеспечении симметрии пар путей, выбираемых запросами в С_АФНФ.

Суммарное значение разности задержек d_{n-1} С_АФНФ по выбранной запросом C_i паре путей, а именно его знак + либо –, и определяет ответ R_i на запрос C_i:

$$d_{n-1} = \delta_{c_{(n-1),Sel},n-1} \cdot (-1)^{c_{(n-1),Sw}} + \sum_{j=0}^{n-2} (\delta_{c_{j,Sel},j} \cdot \prod_{k=j+1}^{n-1} (-1)^{c_{k,Sel} \oplus c_{k,Sw}}). \tag{13}$$

Выражение (13) для вычисления d_{n-1}, так же как и соотношение (11), соответствует реализации модели симметричных АФНФ, имитирующей псевдоподбрасывание монеты (9), так как знак слагаемых, представляющих собой выбранные функцией *Select* добавленные задержки $\delta_{c_{j,Sel},j}$, равновероятно принимает значение + и −.

4. Описание экспериментальных исследований. Для подтверждения эффективности предложенных в статье новых решений по построению симметричных АФНФ был проведен ряд экспериментов на программируемых логических интегральных схемах FPGA Xilinx Zynq7, входящих в состав плат быстрого прототипирования цифровых устройств Digilent Zybo Z7-10. Реализовывались четыре идентичных экземпляра классической схемы АФНФ и четыре экземпляра симметричной схемы АФНФ с применением базового элемента, представленного на рис. 2, *a*, для $n = 32$. На рис. 3 приведены примеры реализации базового элемента как для классической схемы АФНФ, так и для симметричной схемы в терминах технологических блоков FPGA. Элементы задержки для симметричной схемы АФНФ были реализованы на LUT-блоках в качестве логических повторителей сигналов (элементы LUT1 на рис. 3, *b*).



a) *b)*

Рис. 3. Базовый элемент для классической схемы АФНФ (*a*) и симметричной схемы АФНФ (*b*)

Fig. 3. Basic element for classical APUF scheme (a) and symmetrical APUF scheme (b)

В ходе экспериментов исследовались временные параметры множества пар путей, а именно значения d_{31} для различных запросов, выраженных как $\Delta_{C_i}^f$, где $f \in [0,3]$ есть индекс экземпляра одной из четырех исследуемых схем АФНФ, реализованных на одном кристалле FPGA. Отсортированные результаты указанных параметров приведены на рис. 4.

В отличие от проведенных ранее экспериментов [24] значения подаваемых запросов не подвергались балансировке, а полученные данные по основным характеристикам АФНФ коррелируются с приведенными ниже.

Значения математического ожидания $\mu(\Delta_{C_i}^f)$ (см. табл. 3) для 10^4 различных запросов, полученных с помощью 32-разрядого генератора М-последовательностей, показывают асимметрию во всех четырех реализациях классической АФНФ, в то время как для реализаций симметричной АФНФ свидетельствуют об их большей симметричности.

Таблица 3

Математическое ожидание $\mu(\Delta_{C_i}^f)$, нс

Table 3

Expected value $\mu(\Delta_{C_i}^f)$, *ns*

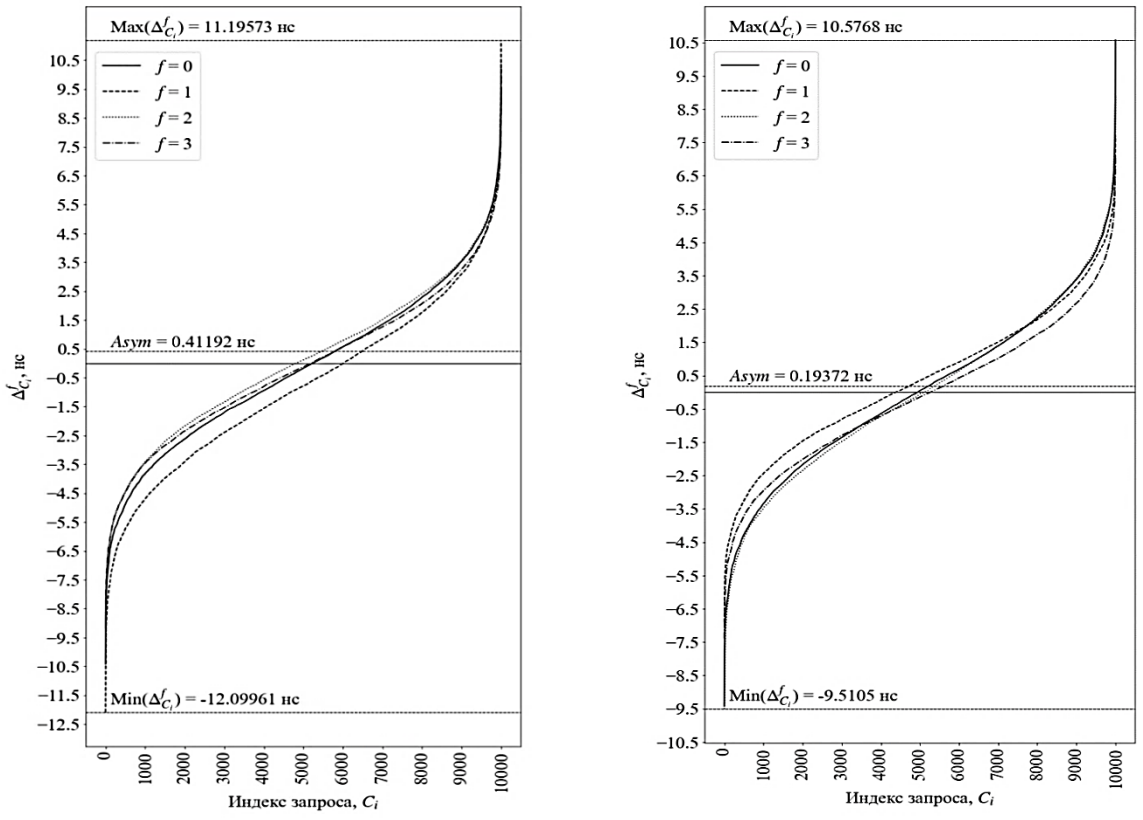
f	Классическая АФНФ *Classic APUF*	Симметричная АФНФ *Symmetrical APUF*
0	–0,1739	0,0212
1	–0,7899	0,3562
2	0,0927	–0,0455
3	–0,1259	–0,1437

Для более детального сравнения реализованных схем АФНФ были определены такие их характеристики, как единообразие (*Un*) и внутрикристальная уникальность (U_{intra}) [16, 22]. Значения единообразия и внутрикристальной уникальности для симметричной АФНФ превышают аналогичные значения для классической АФНФ (табл. 4).

Таблица 4
Усредненные значения *Un* и U_{intra}
Table 4
Average Un and U_{intra}

Тип базового элемента *Base element type*	*Un*	U_{intra}
Классическая схема АФНФ	0,9185	0,7319
Симметричная схема АФНФ	0,9476	0,8162

На рис. 4 приведены значения метрики *Asym*, которая представляет собой среднеквадратическое значение $\sqrt{\sum_{f=0}^{3}\mu^2(\Delta_{C_i}^f)}$, определяющее степень асимметрии множеств нулевых и единичных ответов всех четырех экземпляров схем АФНФ. Так, полученные значения *Asym* для симметричных схем (0,1937 нс) существенно меньше аналогичного значения (0,4119 нс) для классических АФНФ, что подтверждается вычисленными характеристиками единообразия *Un* (табл. 4).



a) *b)*

Рис. 4. Значения величин задержек $\Delta_{C_i}^f$ для классической схемы АФНФ (*a*) и симметричной схемы АФНФ (*b*)

Fig. 4. Delay values $\Delta_{C_i}^f$ *for classical APUF scheme (a) and symmetrical APUF scheme (b)*

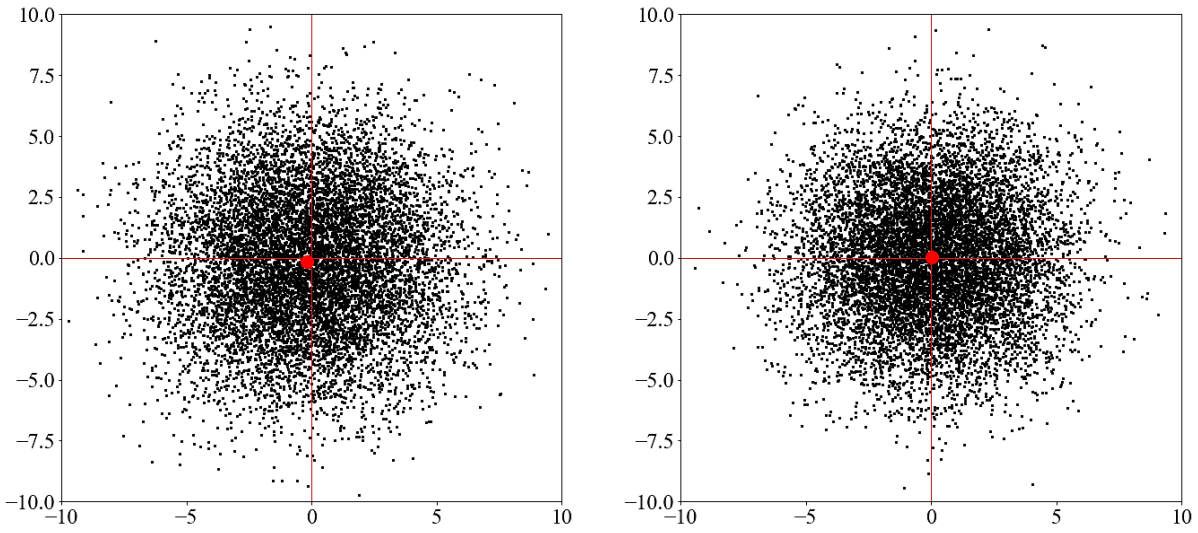


Рис. 5. Графический тест для классической схемы АФНФ (*a*) и симметричной схемы АФНФ (*b*) (*f* = 0)
Fig. 5. Graphical test for classical APUF scheme (a) and symmetrical APUF scheme (b) (f = 0)

Асимметрия множеств значений $\Delta_{C_i}^{f}$ выявляется с помощью графического теста «Распределение на плоскости», результаты которого приведены на рис. 5 и коррелируются со значениями из табл. 3.

Заключение. В статье предложен подход к построению АФНФ, основанный на применении базового элемента, в котором реализованы функции *Generate*, *Select* и *Switch*. Важным отличием С_АФНФ от классических и сбалансированных АФНФ является то, что функции *Generate*, *Select* и *Switch* базового элемента выполняются независимыми его компонентами и задаются разными битами запроса. В отличие от сбалансированных АФНФ С_АФНФ не требуют реализации процедуры балансировки как в процессе изготовления АФНФ, так и при ее применении на практике. Несомненным преимуществом С_АФНФ по отношению к сбалансированным АФНФ, исследованным в работе [24], является достижение симметричности для всего множества запросов C_i. В случае сбалансированных АФНФ симметрия достигается только для сбалансированных запросов C_i, для которых выполняется равенство нулевых и единичных разрядов запроса. Экспериментально подтвержден эффект улучшения характеристик подобных С_АФНФ и в первую очередь их вероятностных свойств, выраженных в равной вероятности ответов, т. е. в отсутствии асимметрии. Перспективным представляется дальнейшее развитие идей построения С_АФНФ, экспериментальное исследование их характеристик, а также анализ устойчивости к различного рода атакам, в том числе и с использованием машинного обучения.

Вклад авторов. *В. Н. Ярмолик* предложил идею построения симметричных физически неклонируемых функций, *А. А. Иванюк* принял участие в обобщении и анализе полученных результатов, а также провел экспериментальные исследования.

Список использованных источников

1. Pappu, R. Physical One-Way Functions: PhD Thesis in Media Arts and Sciences / R. Pappu. – Cambridge : Massachusetts Institute of Technology, 2001. – 154 p.

2. Silicon physical random functions / B. Gassend [et al.] // Proc. of the 9th Computer and Communications Security Conf. (CCS'02), Washington, DC USA, 18–22 Nov. 2002. – Washington, 2002. – P. 148–160.

3. Security with Noisy Data: On Private Biometrics, Secure Key Storage and Anti-Counterfeiting / eds.: P. Tuyls, B. Skoric. – N. Y., USA : Springer, 2007. – 339 p.

4. PUFKY: A fully functional PUF-based cryptographic key generator / R. Maes, A. Van Herrewege, I. Verbauwhede // Proc. of 14th Intern. Workshop on Cryptographic Hardware and Embedded Systems (CHES 2012), Leuven, Belgium, 9–12 Sept. 2012. – Leuven, 2012. – P. 302–319.

5. Robust key extraction from physical uncloneable functions / B. Skoric, P. Tuyls, W. Ophey // Proc. of Intern. Conf. Applied Cryptography and Network Security, N. Y., USA, 7–10 June 2005. – N. Y., 2005. – P. 407–422.

6. Ярмолик, В. Н. Физически неклонируемые функции / В. Н. Ярмолик, Ю. Г. Вашинко // Информатика. – 2011. – № 2(30). – С. 92–103.

7. Suh, G. E. Physical unclonable functions for device authentication and secret key generation / G. E. Suh, S. Devadas // Proc. of Intern. Design Automation Conf., DAC 2007, San Diego, California, USA, 4–8 June 2007. – San Diego, 2007. – P. 9–14.

8. Böhm, C. Physical Unclonable Functions in Theory and Practice / C. Böhm, M. Hofer. – N. Y. : Springer Science + Business Media, 2013. – 270 p.

9. Rührmair, U. Strong PUFs: models, constructions, and security proofs / U. Rührmair, H. Busch, S. Katzenbeisser // Towards Hardware-Intrinsic Security / eds.: A.-R. Sadeghi, D. Naccache. – Berlin, Heidelberg : Springer, 2010. – P. 79–96.

10. A technique to build a secret key in integrated circuits for identification and authentication applications / J. W. Lee [et al.] // Proc. of Intern. Symp. VLSI Circuits (VLSI'04), Honolulu, Hawaii, USA, 7–19 June 2004. – Honolulu, 2004. – P. 176–179.

11. Extracting secret keys from integrated circuits / D. Lim [et al.] // IEEE Transactions on Very Large Scale Integration (VLSI) Systems. – 2005. – Vol. 13, no. 10. – P. 1200–1205.

12. Иванюк, А. А. Физическая криптография и защита цифровых устройств / А. А. Иванюк, С. С. Заливако // Доклады БГУИР. – 2019. – № 2(120). – С. 50–58.

13. Ярмолик, В. Н. Физически неклонируемые функции с управляемой задержкой распространения сигналов / В. Н. Ярмолик, А. А. Иванюк, Н. Н. Шинкевич // Информатика. – 2022. − Т. 19, № 1. – С. 32–49.

14. Using statistical models to improve the reliability of delay-based PUFs / X. Xu, W. Burleson, D. E. Holcomb // Proc. of IEEE Computer Society Annual Symp. on VLSI, Pittsburgh, PA, USA, 11–13 July 2016. – Pittsburgh, 2016. – P. 547–552.

15. An analysis of delay based PUF implementations on FPGA / S. Morozov, A. Maiti, P. Schaumont // Proc. of Intern. Symp. on Applied Reconfigurable Computing: Tools and Applications (ARC 2010), Los Angeles, CA, US, 25–27 Mar. 2010. – Los Angeles, 2010. – P. 382–387.

16. Клыбик, В. П. Метод увеличения стабильности физически неклонируемой функции типа «арбитр» / В. П. Клыбик, С. С. Заливако, А. А. Иванюк // Информатика. – 2017. − № 1(53). – С. 31–43.

17. Ярмолик, В. Н. Физически неклонируемые функции типа арбитр с заведомо асимметричными парами путей / В. Н. Ярмолик, А. А. Иванюк // Доклады БГУИР. – 2022. – № 4(20). – С. 71–79.

18. Secure and reliable XOR arbiter PUF design: An experimental study based on 1 trillion challenge response pair measurements / C. Zhou, K. K. Parhi, C. H. Kim // Proc. of the 54th Annual Design Automation, Austin, TX, USA, 18 June 2017. – Austin, 2017. – P. 18–22.

19. Implementation of double arbiter PUF and its performance evaluation on FPGA / T. Machida [et al.] // Proc. of the 20th Asia and South Pacific Design Automation Conf., Chiba, Japan, 19 Jan. 2015. – Chiba, 2015. – P. 6–7.

20. Side channel modeling attacks on 65nm arbiter PUFs exploiting CMOS device noise / J. Delvaux, I. Verbauwhede // Proc. of IEEE Intern. Symp. on Hardware-Oriented Security and Trust (HOST), Austin, TX, USA, 2–3 June 2013. – Austin, 2013. – P. 137–142.

21. PUF modeling attacks on simulated and silicon data / U. Rührmair [et al.] // IEEE Transactions on Information Forensics and Security. – 2013. – Vol. 11, no. 8. – P. 1876–1891.

22. Шамына, А. Ю. Построение и балансировка путей физически неклонируемой функции типа арбитр на FPGA / А. Ю. Шамына, А. А. Иванюк // Информатика. – 2022. – Т. 19, № 4. – С. 27–41.

23. Ярмолик, В. Н. Двухмерные физически неклонируемые функции типа арбитр / В. Н. Ярмолик, А. А. Иванюк // Информатика. – 2023. – Т. 20, № 1. – С. 7–26.

24. Ярмолик, В. Н. Сбалансированные физически неклонируемые функции типа арбитр / В. Н. Ярмолик, А. А. Иванюк // Безопасность информационных технологий. – 2023. – № 1(30). – С. 92–107.

25. Ярмолик, В. Н. Контроль и диагностика вычислительных систем / В. Н. Ярмолик. – Минск : Бестпринт, 2019. – 387 с.

26. Implementation of pseudo-linear feedback shift register-based physical unclonable functions on silicon and sufficient Challenge-Response pair acquisition using Built-In Self-Test before shipping / Y. Ogasahara [et al.] // Integration, the VLSI J. – 2020. – Vol. 71. – P. 144–153.

27. Evaluation of physical unclonable functions for 28-nm process field-programmable gate arrays / Y. Hori [et al.] // J. of Information Processing. – 2014. – Vol. 22, no. 2. – P. 344–356.

References

1. Pappu R. *Physical One-Way Functions: PhD Thesis in Media Arts and Sciences*. Cambridge, Massachusetts Institute of Technology, 2001, 154 p.

2. Gassend B., Clarke D., Dijk M. S., Devadas S. Silicon physical random functions. *Proceedings of the 9th Computer and Communications Security Conference (CCS'02), Washington, DC USA, 18–22 November 2002.* Washington, 2002, pp. 148–160.

3. Tuyls P., Skoric B. (eds.). *Security with Noisy Data: On Private Biometrics, Secure Key Storage and Anti-Counterfeiting*. New York, USA, Springer, 2007, 339 p.

4. Maes R., Van Herrewege A., Verbauwhede I. PUFKY: A fully functional PUF-based cryptographic key generator. *Proceedings of 14th International Workshop on Cryptographic Hardware and Embedded Systems (CHES 2012), Leuven, Belgium, 9–12 September 2012.* Leuven, 2012, pp. 302–319.

5. Skoric B., Tuyls P., Ophey W. Robust key extraction from physical uncloneable functions. *Proceedings of International Conference Applied Cryptography and Network Security, New York, USA, 7–10 June 2005*. New York, 2005, pp. 407–422.

6. Yarmolik V. N., Vashinko Y. G. *Physical unclonable functions*. Informatika [*Informatics*], 2011, no. 2(30), pp. 92–103 (In Russ.).

7. Suh G. E., Devadas S. Physical unclonable functions for device authentication and secret key generation. *Proceedings of International Design Automation Conference, DAC 2007, San Diego, California, USA, 4–8 June 2007*. San Diego, 2007, pp. 9–14.

8. Böhm C., Hofer M. *Physical Unclonable Functions in Theory and Practice*. New York, Springer Science + Business Media, 2013, 270 p.

9. Rührmair U., Busch H., Katzenbeisser S. Strong PUFs: models, constructions, and security proofs. *Towards Hardware-Intrinsic Security*. In A.-R. Sadeghi, D. Naccache (eds.). Berlin, Heidelberg, Springer, 2010, pp. 79–96.

10. Lee J. W., Lim D., Gassend B., Suh G. E., Van Dijk M., Devadas S. A technique to build a secret key in integrated circuits for identification and authentication applications. *Proceedings of International Symposium VLSI Circuits (VLSI'04), Honolulu, Hawaii, USA, 7–19 June 2004.* Honolulu, 2004, pp. 176–179.

11. Lim D., Lee J. W., Gassend B., Suh G. E., Van Dijk M., Devadas S. Extracting secret keys from integrated circuits. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 2005, vol. 13, no. 10, pp. 1200–1205.

12. Ivaniuk A. A., Zalivaka S. S. *Physical cryptography and security of digital devices.* Doklady Belorusskogo gosudarstvennogo universiteta informatiki i radioèlektroniki [*Reports of the Belarusian State University of Informatics and Radioelectronics*], 2019, no. 2(120), pp. 50–58 (In Russ.).

13. Yarmolik V. N., Ivaniuk A. A., Shynkevich N. N. *Physically unclonable functions with controlled propagation delay.* Informatika [*Informatics*], 2022, vol. 19, no. 1, pp. 32–49 (In Russ.).

14. Xu X., Burleson W., Holcomb D. E. Using statistical models to improve the reliability of delay-based PUFs. *Proceedings of IEEE Computer Society Annual Symposium on VLSI, Pittsburgh, PA, USA, 11–13 July 2016.* Pittsburgh, 2016, pp. 547–552.

15. Morozov S., Maiti A., Schaumont P. An analysis of delay based PUF implementations on FPGA. *Proceedings of International Symposium on Applied Reconfigurable Computing: Tools and Applications (ARC 2010), Los Angeles, CA, US, 25–27 March 2010.* Los Angeles, 2010, pp. 382–387.

16. Klybik V. P., Zalivaka S. S., Ivaniuk A. A. *Reliability enhancement method for "arbiter" physicaly unclonable function*. Informatika [*Informatics*], 2017, no. 1(53), pp. 31–43 (In Russ.).

17. Yarmolik V. N., Ivaniuk A. A. *Arbiter physical unclonable functions with asymmetric pairs of paths*. Doklady Belorusskogo gosudarstvennogo universiteta informatiki i radioèlektroniki [*Reports of the Belarusian State University of Informatics and Radioelectronics*], 2022, no. 4(20), pp. 71–79 (In Russ.).

18. Zhou C., Parhi K. K., Kim C. H. Secure and reliable XOR arbiter PUF design: An experimental study based on 1 trillion challenge response pair measurements. *Proceedings of the 54th Annual Design Automation, Austin, TX, USA, 18 June 2017*. Austin, 2017, pp. 18–22.

19. Machida T., Yamamoto D., Iwamoto M., Sakiyama K. Implementation of double arbiter PUF and its performance evaluation on FPGA. *Proceedings of the 20th Asia and South Pacific Design Automation Conference, Chiba, Japan, 19 January 2015*. Chiba, 2015, pp. 6–7.

20. Delvaux J., Verbauwhede I. Side channel modeling attacks on 65nm arbiter PUFs exploiting CMOS device noise. *Proceedings of IEEE International Symposium on Hardware-Oriented Security and Trust (HOST), Austin, TX, USA, 2–3 June 2013*. Austin, 2005, pp. 137–142.

21. Rührmair U., Sölter J., Sehnke F., Xu X., Mahmoud A., …, Devadas S. PUF modeling attacks on simulated and silicon data. *IEEE Transactions on Information Forensics and Security*, 2013, vol. 11, no. 8, pp. 1876–1891.

22. Shamyna A. Yu., Ivaniuk A. A. *Creating and balancing the paths of arbiter-based physically unclonable functions on FPGA*. Informatika [*Informatics*], 2022, vol. 19, no. 4, pp. 27–41 (In Russ.).

23. Yarmolik V. N., Ivaniuk A. A. *2D physically unclonable functions of the arbiter type*. Informatika [*Informatics*], 2023, vol. 20, no. 1, pp. 7–26 (In Russ.).

24. Yarmolik V. N., Ivaniuk A. A. *Balanced arbiter physical uncloneable functions*. Bezopasnost' informacionnyh tehnologij [*IT Security*], 2023, no. 1(30), pp. 92–107.

25. Yarmolik V. N. Kontrol' i diagnostika vuchislitel'nuh system. *Monitoring and Diagnostics of Computer Systems*. Minsk, Bestprint, 2019, 387 p. (In Russ.).

26. Ogasahara Y., Hori Y., Katashita T., Iizuka T., Awano H., …, Koike H. Implementation of pseudo-linear feedback shift register-based physical unclonable functions on silicon and sufficient Challenge-Response pair acquisition using Built-In Self-Test before shipping. *Integration, the VLSI Journal*, 2020, vol. 71, pp. 144–153.

27. Hori Y., Kang H., Katashita T., Satoh A., Kawamura S., Kobara K. Evaluation of physical unclonable functions for 28-nm process field-programmable gate arrays. *Journal of Information Processing*, 2014, vol. 22, no. 2, pp. 344–356.

Информация об авторах

Ярмолик Вячеслав Николаевич, доктор технических наук, профессор, Белорусский государственный университет информатики и радиоэлектроники.
E-mail: yarmolik10ru@yahoo.com

Иванюк Александр Александрович, доктор технических наук, доцент, профессор кафедры информатики, Белорусский государственный университет информатики и радиоэлектроники.
E-mail: ivaniuk@bsuir.by

Information about the authors

Vyacheslav N. Yarmolik, D. Sc. (Eng.), Prof., Belarusian State University of Informatics and Radioelectronics.
E-mail: yarmolik10ru@yahoo.com

Alexander A. Ivaniuk, D. Sc. (Eng.), Assoc. Prof., Prof. of Computer Science Department, Belarusian State University of Informatics and Radioelectronics.
E-mail: ivaniuk@bsuir.by

ЛОГИЧЕСКОЕ ПРОЕКТИРОВАНИЕ
LOGICAL DESIGN

УДК 519.714.5
https://doi.org/10.37661/1816-0301-2024-21-1-28-47

Оригинальная статья
Original Paper

Технологически независимая оптимизация при реализации в заказных СБИС разреженных систем дизъюнктивных нормальных форм булевых функций

П. Н. Бибило✉, С. Н. Кардаш

Объединенный институт проблем информатики
Национальной академии наук Беларуси,
ул. Сурганова, 6, Минск, 220012, Беларусь
✉*E-mail: bibilo@newman.bas-net.by*

Аннотация

Цели. Рассматривается проблема выбора лучших методов и программ для схемной реализации в заказных цифровых СБИС разреженных систем дизъюнктивных нормальных форм (ДНФ) полностью определенных булевых функций. Для матричных форм разреженных систем ДНФ троичная матрица, задающая элементарные конъюнкции, содержит большую долю неопределенных значений, соответствующих в алгебраической записи отсутствующим литералам булевых входных переменных, а булева матрица, задающая вхождения конъюнкций в ДНФ функций, содержит большую долю нулевых значений.

Методы. Предлагается исследовать различные методы технологически независимой логической оптимизации, выполняемой на первом этапе логического синтеза: совместную минимизацию систем функций в классе ДНФ, раздельную и совместную минимизацию в классах многоуровневых представлений в виде булевых сетей и BDD-представлений с использованием взаимно инверсных кофакторов, разбиение системы функций на подсистемы с ограниченным числом входных переменных, а также метод блочного покрытия систем ДНФ, ориентированный на минимизацию суммарной площади блоков, образующих покрытие.

Результаты. При реализации в заказных СБИС разреженных систем ДНФ булевых функций наряду с традиционными методами совместной минимизации систем функций в классе ДНФ для технологически независимой оптимизации могут применяться методы оптимизации многоуровневых представлений систем булевых функций на основе разложений Шеннона, при этом раздельная минимизация и совместная минимизация всей системы в целом оказываются менее эффективными по сравнению с блочными разбиениями и покрытиями системы ДНФ и последующей минимизацией многоуровневых представлений. Схемы, полученные в результате синтеза по минимизированным представлениям булевых сетей, чаще имеют меньшую площадь, чем схемы, полученные по минимизированным BDD-представлениям.

Заключение. Для проектирования схем заказных цифровых СБИС показана эффективность комбинированного подхода, использующего сначала программы блочного покрытия системы ДНФ с последующим применением программ минимизации многоуровневых представлений блоков в виде булевых сетей, минимизированных на основе разложений Шеннона.

Ключевые слова: система булевых функций, ДНФ, минимизация ДНФ, бинарная диаграмма решений, булева сеть, разложение Шеннона, блочное покрытие системы ДНФ, синтез логической схемы, заказная СБИС, VHDL

Для цитирования. Бибило, П. Н. Технологически независимая оптимизация при реализации в заказных СБИС разреженных систем дизъюнктивных нормальных форм булевых функций / П. Н. Бибило, С. Н. Кардаш // Информатика. – 2024. − Т. 21, № 1. – С. 28–47.
https://doi.org/10.37661/1816-0301-2024-21-1-28-47

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Поступила в редакцию | Received 26.12.2023
Подписана в печать | Accepted 19.01.2024
Опубликована | Published 29.03.2024

Technology independent optimization when implementing sparse systems of disjunctive normal forms of Boolean functions in ASIC

Petr N. Bibilo✉, Sergey N. Kardash

The United Institute of Informatics Problems of the National Academy of Sciences of Belarus, st. Surganova, 6, Minsk, 220012, Belarus
✉*E-mail: bibilo@newman.bas-net.by*

Abstract

Objectives. The problem of choosing the best methods and programs for circuit implementation as part of digital ASIC (Application-Specific Integrated Circuit) sparse systems of disjunctive normal forms (DNF) of completely defined Boolean functions is considered. For matrix forms of sparse DNF systems, the ternary matrix specifying elementary conjunctions contains a large proportion of undefined values corresponding to missing literals of Boolean input variables, and the Boolean matrix specifying the occurrences of conjunctions in DNF functions contains a large proportion of zero values.

Methods. It is proposed to investigate various methods of technologically independent logical optimization performed at the first stage of logical synthesis: joint minimization of systems of functions in the DNF class, separate and joint minimization in classes of multilevel representations in the form of Boolean networks and BDD representations using mutually inverse cofactors, as well as the division of a system of functions into subsystems with a limited number of input variables and the method of block cover of DNF systems, focused on minimizing the total area of the blocks forming the cover.

Results. When implementing sparse DNF systems of Boolean functions in ASIC, along with traditional methods of joint minimization of systems of functions in the DNF class, methods for optimizing multilevel representations of Boolean function systems based on Shannon expansions can be used for technologically independent optimization, while separate minimization and joint minimization of the entire system as a whole turn out to be less effective compared with block partitions and coatings of the DNF system and subsequent minimization of multilevel representations. Schemes obtained as a result of synthesis using minimized representations of Boolean networks often have a smaller area than schemes obtained using minimized BDD representations.

Conclusion. For the design of digital ASIC, the effectiveness of combined approach is shown, when initially the block coverage programs of the DNF system is used, followed by the use of programs to minimize multilevel block representations in the form of Boolean networks minimized based on Shannon expansion.

Keywords: Boolean function system, DNF, DNF minimization, Binary Decision Diagram, Boolean network, Shannon expansion, block cover of the DNF system, logic synthesis, ASIC, VHDL

For citation. Bibilo P. N., Kardash S. N. *Technology independent optimization when implementing sparse systems of disjunctive normal forms of Boolean functions in ASIC*. Informatika [*Informatics*], 2024, vol. 21, no. 1, pp. 28−47 (In Russ.). https://doi.org/10.37661/1816-0301-2024-21-1-28-47

Conflict of interest. The authors declare of no conflict of interest.

Введение. Проблема эффективной схемной реализации цифровых комбинационных блоков в заказных КМОП СБИС (сверхбольших интегральных схемах, выполненных по комплементарной металл-оксид-полупроводник технологии) по-прежнему актуальна при создании средств автоматизированного проектирования цифровых систем. Важным аспектом этой проблемы является то, что современные синтезаторы логических схем чувствительны к форме задания проектной информации, в качестве которой выступают VHDL- либо Verilog-описания [1] моделей функционирования комбинационных схем – те или иные формы задания систем полностью определенных булевых функций. Синтез логических схем выполняется в два этапа: технологически независимая оптимизация представлений систем булевых функций (первый этап) и технологическое отображение в заданный базис (библиотеку) логических элементов заказной СБИС (второй этап). Важнейшим является первый этап, на котором выбирается форма представления системы булевых функций и осуществляется минимизация этой формы. Результат выполнения первого этапа определяет и важнейшие параметры синтезированной на втором этапе логической схемы – площадь, временную задержку и энергопотребление.

Методы и программы технологически независимой оптимизации традиционно развивались для исходных заданий реализуемых систем булевых функций в виде систем ДНФ. Широко известны методы совместной и раздельной минимизации систем булевых функций в классе ДНФ [2, 3], методы факторизации – выделения общих (одинаковых) частей конъюнкций, дизъюнкций и одинаковых подвыражений в скобочных алгебраических представлениях систем булевых функций [4–6], а также многочисленные методы раздельной и совместной функциональной декомпозиции систем булевых функций [7–10 и др.]. В последнее время в качестве методов технологически независимой оптимизации выступают методы минимизации многоуровневых представлений систем функций на основе разложения Шеннона – это методы минимизации BDD (Binary Decision Diagram, бинарная диаграмма решений) [11–17], модификаций BDD [18] и булевых сетей [19]. Предложены также и другие структуры данных [20–24] для представления систем булевых функций и соответствующие методы минимизации.

В настоящей работе рассматриваются разреженные системы ДНФ полностью определенных булевых функций. Для матричных форм таких систем ДНФ троичная матрица, задающая элементарные конъюнкции, содержит большу́ю долю неопределенных значений, а булева матрица, задающая вхождения конъюнкций в ДНФ функций, содержит большу́ю долю нулевых значений и, следовательно, небольшую долю единичных значений. Для разреженных систем ДНФ булевых функций предлагаются алгоритмы их блочного покрытия. Проводится экспериментальное исследование эффективности применения блочных многоуровневых представлений при синтезе комбинационных блоков заказных СБИС в библиотеке КМОП-элементов. Синтезированные схемы сравниваются по площади и временной задержке. Многоблочные многоуровневые представления сравниваются по результатам синтеза с раздельными и совместными многоуровневыми представлениями исходной системы ДНФ и минимизированными двухуровневыми представлениями, под которыми понимаются совместно минимизированные ДНФ. В качестве базовых многоуровневых представлений использованы бинарные диаграммы решений с инверсными кофакторами (BDDI-представления) и булевы сети (Bool-представления). Минимизация BDDI-представлений выполняется по матричным заданиям систем ДНФ булевых функций, Bool-представлений – по логическим уравнениям, задающим те же системы ДНФ. В результате проведенных экспериментов установлено, что для разреженных систем ДНФ наряду с методами минимизации систем функций в классе ДНФ эффективным методом технологически независимой оптимизации является комбинированный метод, включающий блочное покрытие системы ДНФ и последующую минимизацию многоуровневых представлений бло-

ков, при этом функции блоков предпочтительнее минимизировать в классе булевых сетей с использованием разложений Шеннона.

Многоуровневые BDDI-представления систем булевых функций. Под *BDDI-представлением* (BDDI – Binary Decision Diagram with Inverse cofactors) понимается ориентированный бесконтурный граф, задающий последовательные разложения Шеннона булевой функции $f(\boldsymbol{x})=f(x_1,\ldots,x_n)$, $\boldsymbol{x}=(x_1,\ldots,x_n)$, либо системы $\boldsymbol{f}(\boldsymbol{x})=(f^1(\boldsymbol{x}),\ldots,f^m(\boldsymbol{x}))$ булевых функций по всем переменным $x_1, x_2, \ldots, x_n$ при заданном порядке (перестановке) переменных, по которым проводятся разложения, при условии нахождения пар взаимно инверсных кофакторов.

Разложением Шеннона булевой функции $f(\boldsymbol{x})$ по переменной x_i называется представление

$$f(\boldsymbol{x})= \overline{x}_i\, f_0 \vee x_i\, f_1 . \qquad (1)$$

Функции $f_0=f(x_1,\ldots,x_{i-1},0,x_{i+1},\ldots,x_n)$, $f_1=f(x_1,\ldots,x_{i-1},1,x_{i+1},\ldots,x_n)$ в правой части представления (1) называются кофакторами (англ. cofactors) разложения по переменной x_i. Каждый из кофакторов $f(x_1,\ldots,x_{i-1},0,x_{i+1},\ldots,x_n)$, $f(x_1,\ldots,x_{i-1},1,x_{i+1},\ldots,x_n)$ может быть разложен по одной из переменных из множества $\{x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_n\}$. Процесс разложения кофакторов заканчивается, когда все n переменных будут использованы для разложения. BDDI-представлению соответствует совокупность взаимосвязанных формул разложения Шеннона. Сравнение кофакторов на равенство и нахождение взаимно инверсных кофакторов осуществляется с использованием полиномов Жегалкина – канонических представлений булевых функций либо ДНФ, задающих кофакторы.

Минимизация сложности BDDI заключается в нахождении последовательности (перестановки) переменных разложений Шеннона, при которой число кофакторов является наименьшим [18]. Если для каждой функции системы соответствующая ей BDDI строится независимо, то такая минимизация называется *раздельной*. При построении раздельных BDDI могут появляться одинаковые кофакторы в BDDI различных функций системы, однако данный факт при построении BDDI для каждой отдельной функции во внимание не принимается.

Многоуровневые Bool-представления систем булевых функций. *Bool-представление* системы булевых функций соответствует булевой сети (ориентированному бесконтурному графу), функциями вершин которой могут быть логические операции «конъюнкции» либо «дизъюнкции» (возможно с инверсией) над литералами булевых переменных. *Литерал* – это булева переменная либо ее инверсия. Таким образом, вершина булевой сети имеет две заходящие дуги и может иметь одну либо две исходящие дуги, соответствующие прямому и инверсному выходу. При этом предполагается, что доступны как прямые, так и инверсные значения входных переменных [18]. Логическая минимизация булевых сетей на основе разложения Шеннона заключается в поиске такой перестановки переменных разложения, при которой число литералов в булевой сети является наименьшим. В булевой сети разложение Шеннона записывается виде трех формул

$$f(\boldsymbol{x})= w_0 \vee w_1 ;\ w_0 = \overline{x}_i\, f_0 ;\ w_1 = x_i\, f_1 . \qquad (2)$$

Формулы (2) содержат шесть литералов, в то время как формула (1) содержит четыре литерала булевых переменных. Это обстоятельство следует иметь в виду при сравнении сложностей BDDI- и Bool-представлений по числу литералов. Однако как формула (1), так и формулы (2) содержат по три логических оператора – два оператора конъюнкции и один оператор дизъюнкции. После разложения Шеннона по очередной переменной минимизация булевой сети сводится к следующему: ищутся вершины булевой сети, опирающиеся на одинаковые подсети, после чего проводится сокращение сети и находятся уравнения, соответствующие редуцированной сети [19].

Раздельные BDDI- и Bool-минимизации для выделенных подсистем системы булевых функций (либо для отдельных функций системы) заключаются в нахождении своей перестанов-

ки $< x_1, x_2, ..., x_n >$ переменных разложения для каждой из подсистем функций, в то время как при совместной BDDI- и Bool-минимизации используется одна и та же перестановка переменных разложения для всех функций $f^1(\boldsymbol{x}),\ldots,f^m(\boldsymbol{x})$ системы $\boldsymbol{f}(\boldsymbol{x})$. Для некоторых систем функций преимущество при синтезе имеет совместная минимизация, для других – раздельная BDDI- либо Bool-минимизация. Обычно раздельная минимизация позволяет получать схемы, характеризуемые бóльшим быстродействием.

Минимизация в классе ДНФ. Кратчайшей системой ДНФ D_f для системы булевых функций $\boldsymbol{f}(\boldsymbol{x})=(f^1(\boldsymbol{x}),\ldots,f^m(\boldsymbol{x}))$ называется система ДНФ, содержащая минимальное число общих элементарных конъюнкций, на которых заданы ДНФ D_{f^i}, i=1,…, m, всех функций $f^i(\boldsymbol{x})$ системы $\boldsymbol{f}(\boldsymbol{x})=(f^1(\boldsymbol{x}),\ldots,f^m(\boldsymbol{x}))$. Задача совместной минимизации системы булевых функций заключается в нахождении кратчайшей системы ДНФ D_f для заданной системы $\boldsymbol{f}(\boldsymbol{x})=(f^1(\boldsymbol{x}),\ldots,f^m(\boldsymbol{x}))$ булевых функций. Совместная минимизация систем булевых функций в экспериментах выполнялась программой Espresso IIC [3].

Блочное покрытие системы ДНФ булевых функций. Пусть (T^x, B^f) – пара матриц, задающая матричную форму системы ДНФ булевых функций $\boldsymbol{f}(\boldsymbol{x})=(f^1(\boldsymbol{x}),\ldots,f^m(\boldsymbol{x}))$, $\boldsymbol{x}=(x_1,\ldots,x_n)$, где T^x – троичная матрица, задающая общие элементарные конъюнкции, B^f – булева матрица, единичные элементы которой отмечают вхождения элементарных конъюнкций в ДНФ функций [2]. Система ДНФ

$$
\begin{aligned}
f^1 &= x_1\bar{x}_2\bar{x}_3x_4 \vee x_2x_3\bar{x}_5 \vee \bar{x}_1x_4x_5;\\
f^2 &= x_5x_6\bar{x}_7\bar{x}_8\bar{x}_9 \vee \bar{x}_4x_5\bar{x}_6x_8x_9x_{10} \vee x_4x_5x_8;\\
f^3 &= x_5x_6\bar{x}_7\bar{x}_8\bar{x}_9 \vee \bar{x}_4x_5\bar{x}_6x_8x_9x_{10} \vee x_1\bar{x}_2\bar{x}_4x_5 \vee x_2x_3\bar{x}_5 \vee \bar{x}_1x_4x_5;\\
f^4 &= x_5x_6\bar{x}_7\bar{x}_8\bar{x}_9 \vee \bar{x}_4x_5\bar{x}_6x_8x_9x_{10} \vee \bar{x}_1\bar{x}_2x_8x_{10} \vee x_1\bar{x}_8x_{10} \vee x_4x_5x_8;\\
f^5 &= x_2x_8\bar{x}_9 \vee x_1\bar{x}_8x_{10};\\
f^6 &= x_1x_2\bar{x}_8x_9x_{10} \vee \bar{x}_1\bar{x}_2x_8x_{10} \vee x_2x_8\bar{x}_9
\end{aligned}
\qquad (3)
$$

задается парой матриц (T^x, B^f) в табл. 1.

Таблица 1
Система ДНФ булевых функций

Table 1
The DNF system of Boolean functions

Номер строки *Line number*	Троичная матрица T^x *Ternary matrix* T^x										Булева матрица B^f *Boolean matrix* B^f					
	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	f^1	f^2	f^3	f^4	f^5	f^6
1	1	1	–	–	–	–	–	0	1	1	0	0	0	0	0	1
2	–	–	–	–	1	1	0	0	0	–	0	1	1	1	0	0
3	–	–	–	0	1	0	–	1	1	1	0	1	1	1	0	0
4	0	0	–	–	–	–	–	1	–	1	0	0	0	1	0	1
5	1	0	–	0	1	–	–	–	–	–	1	0	1	0	0	0
6	–	1	1	–	0	–	–	–	–	–	1	0	1	0	0	0
7	–	1	–	–	–	–	–	1	0	–	0	0	0	0	1	1
8	1	–	–	–	–	–	–	0	–	1	0	0	0	1	1	0
9	0	–	–	1	1	–	–	–	–	–	1	0	1	0	0	0
10	–	–	–	1	1	–	–	1	–	–	0	1	0	1	0	0

Число n переменных равно 10 (n=10), число функций m равно 6 (m=6), ДНФ заданы на k=10 общих элементарных конъюнкциях, число d операторов дизъюнкции в системе (3) равно 15 (d=15). Площадь $Q(T^x, B^f)$ матриц будем вычислять по формуле (4) и выражать в числе бит:

$$Q(T^x, B^f)=(n+m)k. \quad (4)$$

Рассмотрим систему ДНФ, каждая элементарная конъюнкция которой включает не более t литералов. Это значит, что в каждой строке троичной матрицы T^x находится не более t определенных 0, 1 элементов (остальные элементы равны «–»). Рассмотрим пару (T_{H_i}, B_{H_i}) подматриц, где T_{H_i} – строчная (образованная некоторыми строками матрицы T^x) подматрица матрицы T^x, B_{H_i} – подматрица матрицы B^f, заданная на том же подмножестве строк, что и T_{H_i}. Назовем пару (T_{H_i}, B_{H_i}) *блоком* H_i. Пусть $p \geq t$.

Блок H_i назовем (p,s,q)-*ограниченным* (рис. 1), если одновременно выполняются следующие условия:

– число столбцов T_{H_i}, содержащих определенные элементы 0, 1, не превышает p;

– число ненулевых столбцов матрицы B_{H_i} не превышает q;

– число строк подматриц T_{H_i}, B_{H_i} не превышает s.

Блок (T_{H_i}, B_{H_i}) назовем (p,q)-*ограниченным*, если значение параметра s не ограничивается, т. е. всегда предполагается $s=k$. Блок назовем p-*ограниченным*, если значения параметров s и q не ограничиваются, т. е. всегда предполагается s=k и q=m.

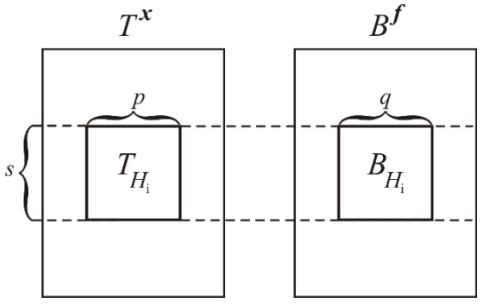


Рис. 1. (p,s,q)-ограниченный блок H_i матричной формы системы ДНФ

Fig. 1. (p,s,q) is a limited H_i block of the matrix form of the DNF system

Видно, что каждой паре (T_{H_i}, B_{H_i}) соответствует своя система ДНФ булевых функций. Рассмотрим три пары подматриц, заданных в табл. 2–4.

Таблица 2
Система ДНФ булевых функций блока H_1

Table 2
DNF system of Boolean block functions H_1

Номер строки *Line number*	Троичная матрица T_{H_1} *Ternary matrix* T_{H_1}	Булева матрица B_{H_1} *Boolean matrix* B_{H_1}
	x_1 x_2 x_8 x_9 x_{10}	f_1^4 f^5 f^6
1	1 1 0 1 1	0 0 1
4	0 0 1 – 1	1 0 1
7	– 1 1 0 –	0 1 1
8	1 – 0 – 1	1 1 0

Таблица 3
Система ДНФ булевых функций блока H_2

Table 3
DNF system of Boolean block functions H_2

Номер строки *Line number*	Троичная матрица T_{H_2} *Ternary matrix* T_{H_2}	Булева матрица B_{H_2} *Boolean matrix* B_{H_2}
	$x_4\ x_5\ x_6\ x_7\ x_8\ x_9$	$f^2\ f_1^3\ f_2^4$
2	– 1 1 0 0 0	1 1 1
3	0 1 0 – 1 1	1 1 1
10	1 1 – – 1 –	1 0 1

Таблица 4
Система ДНФ булевых функций блока H_3

Table 4
DNF system of Boolean block functions H_3

Номер строки *Line number*	Троичная матрица T_{H_3} *Ternary matrix* T_{H_3}	Булева матрица B_{H_3} *Boolean matrix* B_{H_3}
	$x_1\ x_2\ x_3\ x_4\ x_5$	$f^1\ f_2^3$
5	1 0 – 0 1	1 1
6	– 1 1 – 0	1 1
9	0 – – 1 1	1 1

Пара (T_{H_1}, B_{H_1}) является (5,4,3)-ограниченной и имеет площадь 32 бит, пара (T_{H_2}, B_{H_2}) является (6,3,3)-ограниченной и имеет площадь 27 бит, пара (T_{H_3}, B_{H_3}) является (5,3,2)-ограниченной и имеет площадь 21 бит.

Множество $\{H_1,\ldots, H_v\}=\{(T_{H_1}, B_{H_1}),\ldots, (T_{H_v}, B_{H_v})\}$ блоков назовем блочным дизъюнктивным покрытием (далее *блочным покрытием*) пары матриц (T^x, B^f), если каждый единичный элемент матрицы B^f входит только в одну из подматриц B_{H_i}, а каждая строка матрицы T^x входит хотя бы в одну из подматриц T_{H_i}, i=1,…,v.

Блочное покрытие (p,q)-ограниченными блоками будет являться разбиением исходной системы ДНФ на непересекающиеся подсистемы функций, если все функции каждого блока H_i (рис. 2) будут зависеть от одного и того же подмножества переменных, мощность которого не более p. Если же каждая строка троичной матрицы T^x будет входить только в один блок, то блочное покрытие будет задавать разбиение множества строк матрицы T^x на непересекающиеся подмножества T_{H_i}.

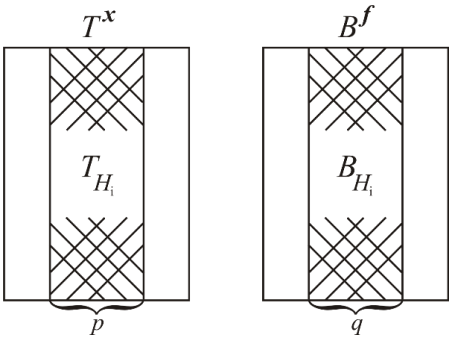


Рис. 2. Блок H_i содержит q функций, зависящих от p переменных

Fig. 2. The H_i block contains q functions depending on p variables

Если же для подсистемы, состоящей из q функций, число ее аргументов превышает число p, то для реализации блочного разложения требуются операции дизъюнкции для некоторых (либо всех) функций нескольких (возможно всех) блоков. На рис. 3 показан этот случай. Для m-входового дизъюнктора площадь будем подсчитывать по формуле $6(m–1)$ бит.

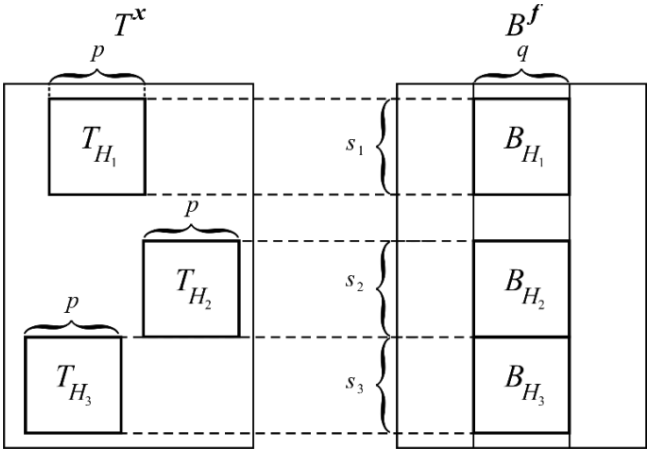


Рис. 3. Подсистема, содержащая q функций, которые зависят более чем от p переменных

Fig. 3. A subsystem containing q functions that depend on more than p variables

Задача 1 блочного покрытия системы ДНФ по критерию минимальности числа блоков: найти покрытие пары (T^x, B^f) возможно меньшим числом (p,q)-ограниченных блоков (T_{H_i}, B_{H_i}), $i=1,\ldots, v$.

Задача 2 блочного покрытия системы ДНФ по критерию минимальности площади блоков: найти покрытие пары (T^x, B^f) возможно меньшим числом p-ограниченных блоков (T_{H_i}, B_{H_i}), $i=1,\ldots, w$, имеющих возможно меньшую суммарную площадь с учетом площадей логических элементов, реализующих дизъюнкции функций.

Алгоритмы и программы решения задач блочного покрытия системы ДНФ описаны в работе [25]. Они являются эвристическими и итерационными – на каждой итерации формируется один блок на основе эвристик выбора столбцов и строк соответствующих подматриц очередного блока. После этого обнуляются те единичные элементы булевой матрицы B^f, которые принадлежат сформированному блоку. Процесс формирования блоков заканчивается, когда все элементы матрицы B^f становятся нулевыми.

Применение программы решения задачи 2 (для $p=6$) блочного покрытия с минимизацией площади блоков для системы ДНФ из табл. 1 позволяет получить три блока H_1, H_2, H_3 (рис. 4), заданных в табл. 2–4 соответственно.

В табл. 3 ДНФ функций f^2, f_2^4 одинаковы, т. е. $f^2 = f_2^4$, в табл. 4 также задаются одинаковые ДНФ $f_1^4 = f_2^3$, поэтому можно сказать, что блочное покрытие является приемом логической оптимизации и позволяет выделять общие подфункции в дизъюнктивных разложениях системы ДНФ. Заметим, что факты равенства ДНФ функций, принадлежащих одному и тому же блоку, устанавливаются при последующей многоуровневой минимизации функций этого блока.

Если не стремиться уменьшить число блоков в блочном покрытии системы ДНФ, то решение задачи 2 по критерию минимальной суммарной площади можно свести к нахождению k-блочного покрытия, где каждый блок формируется по одной из k строк матриц T^x, B^f. В примере таким покрытием будет 10-блочное покрытие с общей суммарной площадью, равной 59 бит, при этом число дополнительных дизъюнкций для представления логической сети будет равно 15, как и в формулах (3).

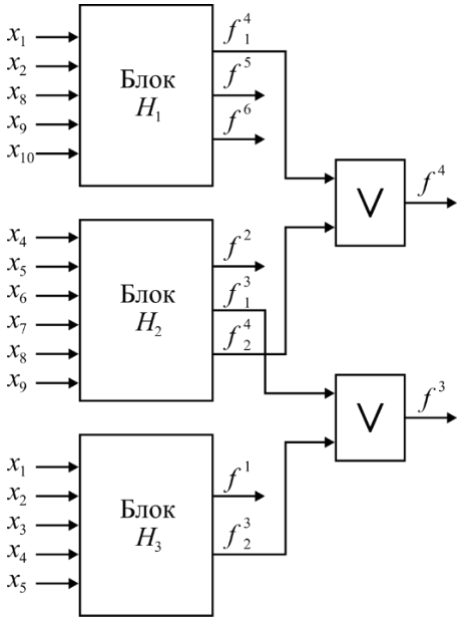


Рис. 4. Логическая схема, соответствующая блочному покрытию (табл. 2–4) системы ДНФ (табл. 1)

Fig. 4. The circuit corresponding to the block cover (Tables 2–4) of the DNF system (Table 1)

Площадь матриц T^x, B^f из табл. 1 составляет $Q(T^x,B^f)=(n+m)k=160$ (бит), суммарная площадь трех блоков, заданных в табл. 2–4, составляет 32+27+21=80 (бит), что в два раза меньше площади матриц из табл. 1, при этом понадобятся только две двухвходовые дизъюнкции (рис. 4). Матричная форма оператора двухвходовой дизъюнкции имеет площадь 6 бит. Поэтому общая площадь составляет 80+12=92 (бит). Если для той же системы ДНФ (см. табл. 1) решить задачу 1, уменьшив значение параметра p, положив p=5, q=3 и выполнив программу блочного покрытия по критерию минимальности числа блоков, то получим четыре блока в покрытии ДНФ с суммарной площадью блоков, равной 77 и 30 бит. В этой логической сети имеется один оператор двухвходовой дизъюнкции и два оператора трехвходовой дизъюнкции с общей площадью 30 бит, так как матричная форма трехвходового оператора дизъюнкции составляет 12 бит.

Таким образом, изменяя параметры p, q, можно получать различные блочные покрытия системы ДНФ, характеризуемые разной суммарной площадью, разным числом блоков и разным числом дизъюнкций, требуемых для формирования выходных функций.

Разреженные системы ДНФ. Под *разреженностью* α *троичной матрицы* T^x будем понимать отношение числа неопределенных элементов «–» к числу всех элементов этой матрицы и выражать это отношение в процентах. Например, троичная матрица T^x (табл. 1) содержит 62 неопределенных значения «–», общее число элементов матрицы T^x равно 100 (матрица состоит из 10 столбцов и 10 строк). Следовательно, α =62 %.

Под *разреженностью* β *булевой матрицы* B^f будем понимать долю числа ее нулевых элементов, выраженную в процентах. В булевой матрице B^f (табл. 1) число нулевых элементов равно 39. Следовательно, β=39/60=0,65, что составляет 65 %. Чем большее значение имеют параметры α и β, тем более разреженной является матричная форма системы ДНФ.

Исходные данные для экспериментов. Системы булевых функций для экспериментов (табл. 5) были заданы в двух формах – матричной и форме логических уравнений. В табл. 5 используются следующие обозначения: n – число аргументов системы ДНФ булевых функций, m – число функций, k – число общих элементарных конъюнкций, d – число дизъюнкций в системе ДНФ булевых функций, α – разреженность (в процентах) троичной матрицы T^x, β –

разреженность (в процентах) булевой матрицы B^f. Примеры Pozd_1, Pozd_2 – это «блочные» системы ДНФ. Параметры трехблочной системы ДНФ Pozd_1 (рис. 5): n_1=n_2=n_3=12, m_1=m_2=m_3=10, k_1=263, k_2=358, k_3=133; параметры трехблочной системы ДНФ Pozd_2: n_1=n_2=n_3=12, m_1=m_2=m_3=10, k_1=263, k_2=137, k_3=205. Соседние блоки в примерах Pozd_1 и Pozd_2 имеют две и четыре общие входные переменные соответственно.

Таблица 5
Исходные данные – разреженные системы ДНФ булевых функций

Table 5
Initial data – sparse DNF systems of Boolean functions

Пример *Example*	*n*	*m*	*k*	*d*	α	β
C8	28	18	70	103	89,5	78,0
DALU	75	16	194	1145	94,0	58,7
LAL	26	19	117	67	83,8	71,8
PM1	16	13	42	27	83,6	70,0
SCT	19	15	64	76	98,9	63,3
TTT2	24	21	222	203	80,7	71,9
Alu4	14	8	1 028	1 020	45,2	40,3
Apex5	117	88	1 227	1 142	95,0	94,3
I2c	147	142	1 357	1 251	98,6	97,8
X1	51	35	324	289	87,0	78,0
X3	135	99	915	523	92,4	93,9
X4	94	71	371	277	94,5	90,8
Blocki1	15	16	355	506	58,1	71,5
Blocki2	15	16	90	101	64,2	75,4
Pozd_1	30	30	754	1 516	79,2	82,5
Pozd_2	30	30	605	1 805	75,2	77,8

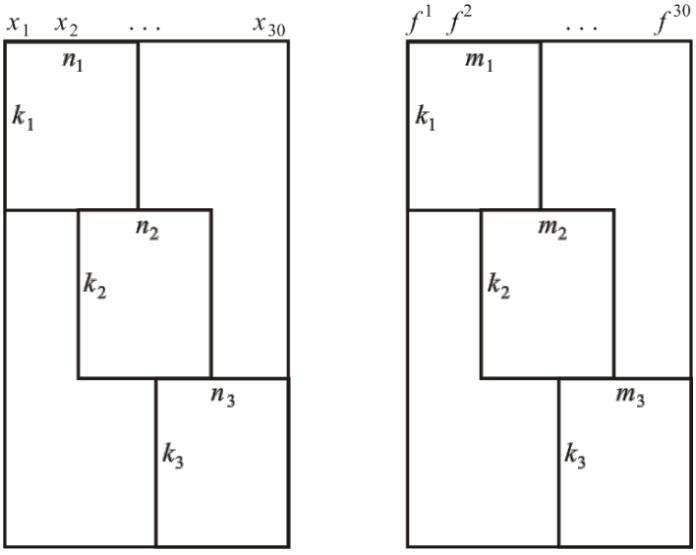


Рис. 5. Структура систем ДНФ примеров Pozd_1 и Pozd_2

Fig. 5. Structure of DNF systems of examples Pozd_1 and Pozd_2

Для примера Blocki1 (двухблочной системы ДНФ) n_1=11, n_2=7, m_1=m_2=8, k_1=255, k_2=100. Для примера Blocki2 (двухблочной системы ДНФ) n_1=10, n_2=7, m_1=m_2=8, k_1=42, k_2=48. Блоки имеют две общие входные переменные для обоих примеров Blocki1 и Blocki2. Пример I2c взят

из библиотеки (http://lsi.epfl.ch/benchmarks) примеров описаний (логических уравнений) графов AIG (And-Inverter Graph). Остальные 11 примеров (табл. 5) – это разреженные системы ДНФ, взятые из библиотеки примеров LGSynth91. Исходные описания примеров Alu4, Apex5 даны в библиотеке в формате PLA, остальные девять примеров – в формате Blif. Все примеры были переведены в матричный формат (SDF) языка SF в системе FLC-2 [26]. Исходные функциональные описания примеров C8, DALU, LAL, PM1, SCT, TTT2, X1, X3, X4 не содержат инверсий литералов входных переменных, т. е. троичные матрицы T^x для этих примеров содержат только 1 и «–». Основными критериями выбора примеров являлись практическая размерность (десятки аргументов и функций) и возможно большая разреженность матричных заданий систем ДНФ. Рассмотрим матричную форму примера Apex5. Троичная матрица T^x включает 117 столбцов (переменных) и 1227 строк (элементарных конъюнкций), площадь этой матрицы 117 × 1227=143 559, она содержит 136 453 неопределенных элемента «–», разреженность этой матрицы 136 453/143 559=0,95, т. е. 95 %. Булева матрица B^f включает 88 столбцов (функций) и 1227 строк, площадь этой матрицы 88 × 1227=107 976, данная матрица содержит 101 759 нулевых элементов, ее разреженность 101 759/107 976=0,943, т. е. 94,3 %. Пример Alu4 является наименее разреженным – доля неопределенных элементов в троичной матрице T^x меньше половины, доля единичных элементов в булевой матрице B^f также меньше половины.

Эксперименты. Всего было проведено 10 экспериментов по выяснению эффективности алгоритмов и программ технологически независимой оптимизации, используемых при синтезе функциональных блоков заказных КМОП СБИС. Этапы экспериментов показаны на рис. 6.

Сначала осуществлялся перевод всех функциональных описаний в стандартные форматы, используемые в системе FLC-2. Затем (кроме эксперимента 1) выполнялась технологически независимая оптимизация, включающая раздельную, совместную либо блочную (комбинированную) логическую минимизацию.

После логической минимизации минимизированные описания представлений систем функций в виде логических уравнений конвертировались в VHDL-описания [1, 27] и подавались на вход синтезатора LeonardoSpectrum. Для всех примеров синтез осуществлялся с одними и теми же опциями управления синтезом и для одной и той же целевой библиотеки синтеза. Синтезатор LeonardoSpectrum [27] имеет свои средства логической минимизации, он перерабатывает входное описание, получает свое (внутреннее) описание, по которому и синтезируется схема. Библиотекой синтеза являлась библиотека проектирования заказных цифровых КМОП СБИС, ее состав приведен в работе [28].

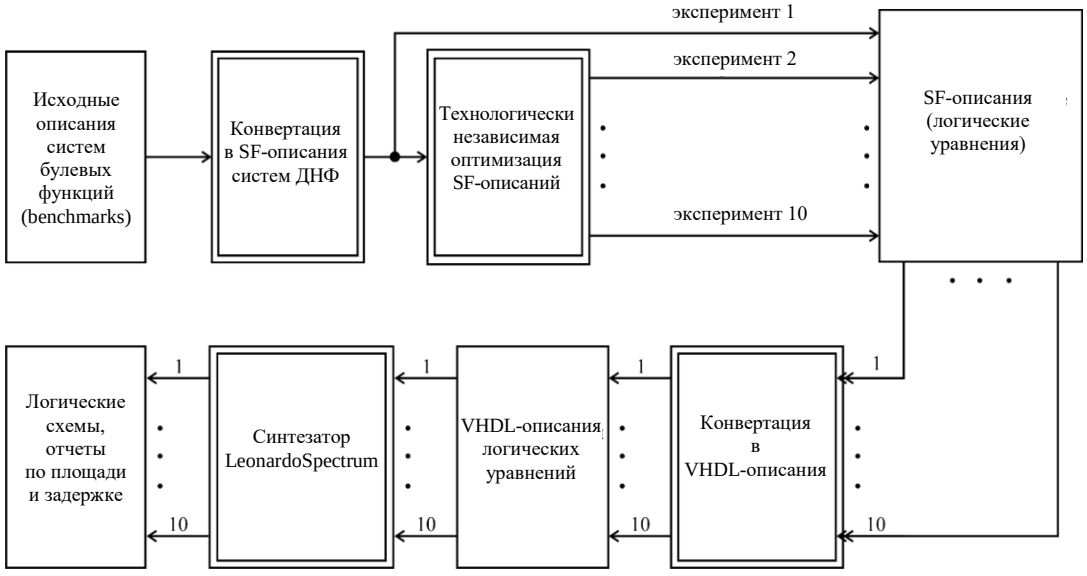


Рис. 6. Этапы экспериментов

Fig. 6. Stages of experiments

BDDI-минимизация выполнялась для матричных форм, Bool-минимизация – для логических уравнений, задающих те же системы функций. Программы блочного покрытия выполнялись только для матричных представлений, обработка многоблочных логических сетей осуществлялась с помощью стратегий системы логической оптимизации FLC2 [26]. Перечислим программы системы FLC2, участвующие в экспериментах. BDDI-минимизация отдельных ДНФ и совместная BDDI-минимизация систем ДНФ выполнялись с помощью программы BDD_Builder [18]; раздельная и совместная Bool-минимизация – с помощью модификации программы BoolNet_Opt [18]. Для решения задачи 1 блочного покрытия систем ДНФ использовалась программа RAZ [25], для решения задачи 2 – программа RAZ_Area [25]. В реализованном варианте программ RAZ, RAZ_Area блочного покрытия предполагается, что каждый единичный элемент матрицы B^f покрывается только одной из подматриц B_H. Улучшение результатов многоуровневой минимизации функций блока (T_{H_i}, B_{H_i}) возможно при реализации многократного покрытия единичного значения в матрице B^f несколькими подматрицами B_{H_i} с целью уменьшения числа различных ДНФ в блоке (T_{H_i}, B_{H_i}). Как уже говорилось, совместная минимизация систем булевых функций в классе ДНФ выполнялась программой Espresso IIC [3].

Эксперимент 1. Логическая оптимизация не выполнялась, исходные матричные описания систем ДНФ переводились в логические уравнения и сразу конвертировались в VHDL-описания. Полученные в эксперименте 1 схемные реализации систем ДНФ названы *базовыми*.

Эксперимент 2. Совместная минимизация систем функций в классе ДНФ с помощью программы Espresso IIC [3].

Эксперимент 3. Разбиение системы на подсистемы, состоящие из отдельных функций, затем для каждой из функций исходной системы, заданной в матричной форме, выполнение раздельной BDDI-минимизации.

Эксперимент 4. Разбиение системы на подсистемы, состоящие из отдельных функций, затем перевод в логические уравнения матричного описания каждой из функций, после этого выполнение для каждой из функций раздельной Bool-минимизации.

Эксперимент 5. Совместная BDDI-минимизация исходной системы ДНФ булевых функций.

Эксперимент 6. Перевод матричного описания системы ДНФ булевых функций в логические уравнения, затем для системы функций, заданных логическими уравнениями, выполнение совместной Bool-минимизации.

Эксперимент 7. Блочное покрытие матричной формы системы ДНФ по критерию минимальности числа (p, q)-ограниченных блоков, затем для функций каждого блока выполнение совместной BDDI-минимизации.

Эксперимент 8. Блочное покрытие матричной формы системы ДНФ по критерию минимальности числа (p, q)-ограниченных блоков, затем для каждого блока выполнение совместной Bool-минимизации.

Эксперимент 9. Блочное покрытие матричной формы системы ДНФ по критерию минимальности суммарной площади блоков, затем для функций каждого блока выполнение совместной BDDI-минимизации.

Эксперимент 10. Блочное покрытие матричной формы системы ДНФ по критерию минимальности суммарной площади блоков, затем для функций каждого блока выполнение совместной Bool-минимизации.

Результаты экспериментов и их обсуждение. Для системы ДНФ функций (см. табл. 1) результаты экспериментов для заказных СБИС даны в табл. 6. Обозначения, используемые в табл. 6, будут пояснены далее. Можно отметить, что результаты экспериментов 1 и 2 для данного примера полностью совпадают. Это связано с тем, что программа Espresso минимизации системы ДНФ (см. табл. 1) не позволяет изменить исходную систему ДНФ – функции являются неминимизируемыми в классе ДНФ.

Таблица 6
Результаты экспериментов для системы ДНФ (табл. 1)

Table 6
Experimental results for the DNF system (Table 1)

Номер эксперимента *Experiment number*	*Z*	*p, q*	*Area*	*Delay*	*r*
1	81	–	8 643	2,74	1
2	81	–	8 643	2,74	1
3	148	–	14 910	***2,53**	6
4	136	–	9 614	3,48	6
5	162	–	9 631	3,06	1
6	80	–	8 643	2,74	1
7	89	5, 3	8 828	2,75	4
8	87	5, 3	9 201	2,78	4
9	80	6, 0	***8 493**	2,64	3
10	81	6, 0	8 643	2,63	3

Результаты экспериментов для примеров из табл. 5 представлены в табл. 7–11, где символом * отмечены лучшие решения для испытываемого примера – меньшие значения параметров площади и временной задержки. При этом сравнение проводилось по всем 10 экспериментам.

Символом # помечены решения, улучшающие базовые решения, т. е. отмечены меньшие значения параметра площади либо задержки, чем соответствующие значения параметров из эксперимента 1.

Таблица 7
Результаты экспериментов 1 и 2

Table 7
Results of experiments 1 and 2

Пример *Example*	Эксперимент 1 (базовые решения) *Experiment 1 (basic solutions)*			Эксперимент 2 *Experiment 2*				
	Z	*Area*	*Delay*	*k*	k_{Espr}	*Z*	*Area*	*Delay*
C8	204	21 500	***2,20**	70	70	204	***#21 494**	2,36
DALU	1 404	106 249	10,79	194	194	1 404	**#47 865**	**#4,60**
LAL	529	26 343	3,80	117	117	529	27 889	**#3,54**
PM1	124	***10 764**	2,58	42	42	124	11 099	**#2,44**
SCT	253	20 406	3,20	64	64	253	**#19 976**	**#3,15**
TTT2	1 263	43 652	***3,85**	222	222	1 263	45 019	4,72
Alu4	7 875	487 848	9,94	1 028	***575***	5 493	**#327 736**	**#8,59**
Apex5	7 106	188 559	8,77	1 227	***1 088***	7 227	***#188 180**	8,80
I2c	7 112	280 601	9,12	1 357	***805***	5 816	***#254 906**	11,70
X1	2 148	67 546	3,99	324	***274***	1 974	***#67 256**	3,99
X3	5 045	203 659	6,75	915	915	5 045	208 151	**#6,47**
X4	2 649	***95 781**	5,38	371	371	2 649	96 400	5,39
Blocki1	3 174	275 607	7,43	355	***240***	3 177	**#243 260**	**#6,61**
Blocki2	648	71 402	4,62	90	***80***	645	**#70 989**	5,07
Pozd_1	9 572	714 586	8,66	754	***555***	9 381	***#615 926**	9,56
Pozd_2	13 543	768 522	9,51	605	***442***	13 497	***#653 089**	10,07
Улучшено базовых решений							11	7

Таблица 8
Результаты экспериментов 3 и 4

Table 8
Results of experiments 3 and 4

Пример *Example*	Эксперимент 3 *Experiment 3*			Эксперимент 4 *Experiment 4*		
	Z	*Area*	*Delay*	*Z*	*Area*	*Delay*
C8	213	22 314	***2,20**	235	21 943	***2,20**
DALU	1 243	**#67 033**	**#6,22**	992	**#52 285**	**#4,89**
LAL	575	26 829	**#3,73**	527	26 377	3,80
PM1	173	11 841	**#2,38**	135	11 082	**#2,38**
SCT	390	**#19 926**	**#3,18**	340	**#20 149**	**#3,18**
TTT2	724	45 683	4,41	764	**#42 631**	4,21
Alu4	3 469	373 492	11,83	3 680	**#378 815**	**#9,16**
Apex5	5 059	196 081	**#7,42**	6 257	277 678	8,86
I2c	4 982	277 320	9,65	7 112	292 398	***#7,24**
X1	1 739	71 720	4,46	1 300	73 835	5,08
X3	3 833	223 222	**#6,51**	3 408	213 625	**#6,66**
X4	3 082	133 959	5,99	2 895	124 373	5,72
Blocki1	6 001	603 358	7,89	4 570	267 243	**#6,80**
Blocki2	1 070	73 338	**#4,20**	1 142	72 730	4,83
Pozd_1	29 913	3 281 191	10,16	15 516	747 480	**#8,39**
Pozd_2	36 180	4 268 153	10,99	19 704	815 160	**#9,49**
Улучшено базовых решений		2	8	–	4	10

Таблица 9
Результаты экспериментов 5 и 6

Table 9
Results of experiments 5 and 6

Пример *Example*	Эксперимент 5 *Experiment 5*			Эксперимент 6 *Experiment 6*		
	Z	*Area*	*Delay*	*Z*	*Area*	*Delay*
C8	225	22 839	***2,20**	269	22 314	***2,20**
DALU	955	**#82 316**	**#7,17**	1 148	**#57 770**	**#6,57**
LAL	405	43 139	5,15	307	27 331	4,12
PM1	161	12 923	**#2,33**	126	***10 764**	2,58
SCT	253	27 744	4,88	289	20 507	***#2,80**
TTT2	506	49 154	5,72	860	44 657	4,58
Alu4	2 129	**#300 835**	**#9,80**	2792	***#252 060**	**#8,21**
Apex5	6 791	221 587	9,11	6999	**#196 583**	**#8,12**
I2c	5245	301 415	**#9,10**	5890	283 765	**#8,24**
X1	2 664	107 912	**#6,45**	2 405	83 527	**#4,70**
X3	4 591	211 633	5,73	4 096	225 895	7,46
X4	4 235	120 031	7,23	3 028	165 759	6,94
Blocki1	5 395	659 534	9,35	2 768	**#274 787**	7,53
Blocki2	1 171	92 472	4,71	742	***#70 542**	4,82
Pozd_1	27 544	3 508 196	11,41	1 417	726 031	**#8,65**
Pozd_2	29 502	3 925 803	11,39	7 104	798 944	***#9,11**
Улучшено базовых решений		2	6	–	6	9

Таблица 10
Результаты экспериментов 7 и 8

Table 10
Results of experiments 7 and 8

Пример *Example*	Эксперимент 7 *Experiment 7*					Эксперимент 8 *Experiment 8*				
	Z	*p,q*	*Area*	*Delay*	*r*	*Z*	*p,q*	*Area*	*Delay*	*r*
C8	262	12,7	21 935	***2,20**	5	279	12,7	22 114	2,98	5
DALU	1391	20,4	**#68 199**	**#4,79**	12	1 228	20,4	**#50 187**	***#4,37**	12
LAL	531	14,8	26 829	**#3,73**	4	515	14,8	***#25 925**	***#3,29**	4
PM1	189	8,4	12 131	**#2,38**	6	173	8,4	11 099	**#2,44**	6
SCT	428	10,5	20 674	**#3,18**	6	335	10,5	20 674	**#3,18**	6
TTT2	958	15,5	64 482	5,11	7	845	15,5	48 457	5,98	7
Alu4	3 327	14,4	**#327 825**	***#7,68**	2	2 784	14,4	**#261 038**	**#8,10**	2
Apex5	6 711	25,20	**#246 005**	**#7,87**	12	6234	25,20	292 900	9,52	12
I2c	4861	24,20	298 497	**#9,09**	20	7 283	24,20	295 400	**#8,71**	20
X1	1 709	25,10	81 652	4,66	7	1 468	25,10	69 867	***#3,93**	7
X3	4 460	25,10	**#198 760**	***#4,98**	14	4 265	25,10	**#201 940**	**#6,43**	14
X4	3 123	15,10	139 238	7,23	21	2 412	15,10	105 473	***#4,31**	21
Blocki1	5 428	10,8	641 739	9,22	10	2 798	10,8	**#274 212**	***#6,58**	10
Blocki2	1 162	10,8	98 264	5,12	2	748	10,8	71 870	***#4,05**	2
Pozd_1	25 322	12,10	3 204 304	11,18	3	7 074	12,10	727 537	8,72	3
Pozd_2	27 065	12,10	3 664 347	10,42	3	7 200	12,10	803 180	**#9,26**	3
Улучшено базовых решений			4	9	–	–	–	5	12	–

Таблица 11
Результаты экспериментов 9 и 10

Table 11
Results of experiments 9 and 10

Пример *Example*	Эксперимент 9 *Experiment 9*					Эксперимент 10 *Experiment 10*				
	Z	*p*	*Area*	*Delay*	*r*	*Z*	*p*	*Area*	*Delay*	*r*
C8	266	12	22 169	***2,20**	5	255	12	22 314	***2,20**	21
DALU	1 305	20	**#68 305**	**#5,33**	12	1 030	20	***#48 044**	**#4,63**	34
LAL	588	14	27 710	**#3,25**	4	523	14	26 829	**#3,73**	27
PM1	198	8	11 127	***#2,20**	6	168	8	11 099	**#2,44**	20
SCT	397	10	***#19 301**	**#2,93**	6	356	10	**#20 116**	**#2,95**	25
TTT2	1 106	15	57 189	4,64	7	1 056	15	***41 415**	5,11	40
Alu4	3 469	14	**#373 492**	11,83	7	3 512	14	**#315 415**	**#7,70**	7
Apex5	6 126	40	195 903	***#6,94**	117	5 949	50	242 981	**#6,98**	107
I2c	6 807	40	**#278 258**	**#6,56**	183	4 859	40	**#259 565**	**#7,63**	88
X1	1 709	25	81 652	4,66	7	1 492	25	73 087	4,71	7
X3	4 238	25	229 427	**#6,74**	135	3 766	25	***#198 252**	**#6,67**	135
X4	3 402	15	123 720	5,75	189	3 030	15	103 799	**#4,75**	189
Blocki1	5 418	10	341 451	**#7,10**	10	4 558	10	***#265 290**	**#6,92**	16
Blocki2	1 070	10	73 338	**#4,20**	2	1 138	10	**#70 609**	**#4,22**	16
Pozd_1	25 975	12	3 233 861	11,15	3	7 892	12	731 611	***#7,94**	6
Pozd_2	27 065	12	3 665 887	11,43	3	7 178	12	803 180	**#9,26**	3
Улучшено базовых решений			4	10	–	–	–	8	14	–

В табл. 6–11 используются следующие обозначения:

Z – число литералов в задании системы булевых функций;

k_{Espr} – число элементарных конъюнкций в совместно минимизированной системе ДНФ с помощью программы Espresso;

Area – суммарная площадь элементов схемы в условных единицах;

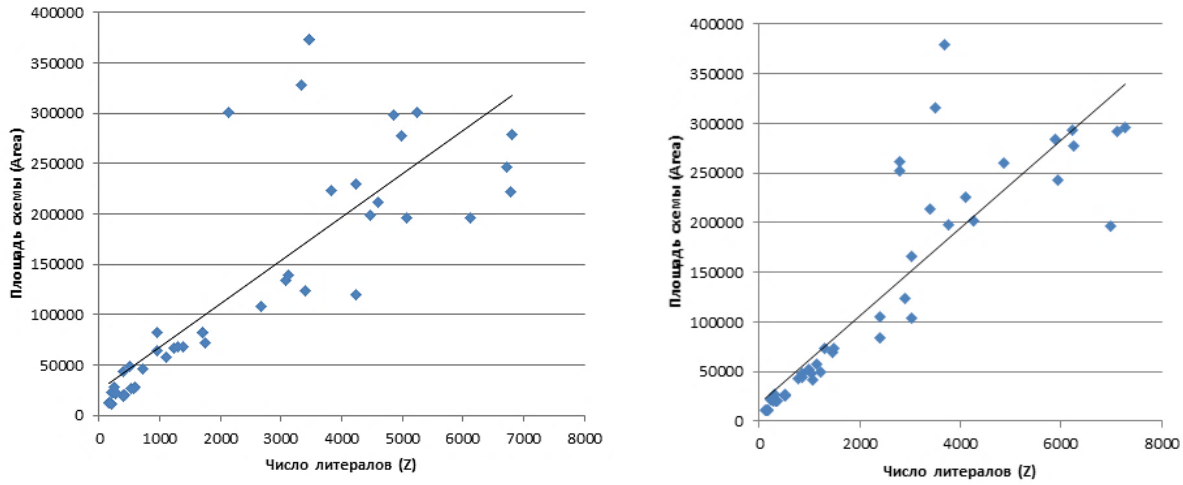
Delay – временная задержка схемы, нс;

p – число входов блока;

q – число выходов блока;

r – число блоков после разбиения системы функций на подсистемы (случай *r*=1 соответствует совместной реализации системы).

Результаты экспериментов позволяют сделать следующие выводы. Для исследованного множества блочных и разреженных систем ДНФ булевых функций значительное преимущество по площади синтезированных многовыходных комбинационных логических схем из библиотечных элементов имеет Bool-минимизация, которая выполняется для систем функций, заданных логическими уравнениями. Синтез схем по матричным представлениям систем функций и последующая BDDI-минимизация являются менее эффективными: для такого вывода достаточно сравнить эксперименты в парах (3, 4), (5, 6), (7, 8), (9, 10). В экспериментах с нечетными номерами используется BDDI-минимизация, в экспериментах с четными номерами – Bool-минимизация. Графики зависимостей площадей схем от числа литералов в BDDI- и Bool-представлениях систем функций показаны на рис. 7. При этом исключены четыре специально сгенерированных примера. График на рис. 7, *a* построен по результатам нечетных (3, 5, 7, 9) экспериментов, график на рис. 7, *b* – по результатам четных (4, 6, 8, 10) экспериментов. Данные графики (тренды) показывают достаточно хорошую линейную зависимость площади схемы от числа литералов в многоуровневом функциональном описании реализуемой системы функций. Поэтому целесообразно в дальнейших исследованиях организовывать переборы блочных покрытий, оценивания их не только по площади, но и по суммарному числу литералов в блоках покрытия. Для разреженных систем ДНФ такой подход может быть более перспективным, чем подход, исследованный в работе [31] и основанный на выделении из формульного задания системы функций таких подсистем, для которых совместная минимизация является более предпочтительной, чем минимизация функций по отдельности либо совместная минимизация всей системы в целом.



a) *b)*

Рис. 7. Зависимость площади схемы (Area) от числа литералов (Z) при BDDI-минимизации (*a*) и Bool-минимизации (*b*)

Fig. 7. Dependence of the area of the scheme (Area) on the number of literals (Z) with BDDI minimization (a) and Bool minimization (b)

Установлено также то, что для разреженных систем ДНФ сравнение кофакторов целесообразно выполнять для их задания матричными формами, так как переход от матричных форм к полиномам Жегалкина [30] становится трудоемким и время BDDI-минимизации возрастает. Сравнение кофакторов в матричном виде позволяет ускорять вычисления.

Алгоритмы и программы [25] блочного покрытия для матричных представлений позволяют выделять блоки в блочных системах ДНФ функций и в разреженных системах ДНФ. Это подтверждается нахождением блочных покрытий для специально сгенерированных примеров Blocki1, Blocki2, Pozd_1, Pozd_2 блочных систем ДНФ. Минимизация блочных покрытий по критерию минимальности числа блоков уступает по результатам синтеза минимизации по критерию (4) минимальности общей площади блоков. Оптимизация блочных покрытий матричных представлений систем ДНФ функций по критерию площади и последующая Bool-минимизация полученных блоков являются достаточно эффективными, так как в большом числе случаев позволяют уменьшить площади схем либо их временные задержки.

Заключение. Система FLC-2 [26] включает разнообразные программы логической минимизации различных форм представлений систем булевых функций. В ней можно провести эффективную технологически независимую оптимизацию разреженных систем ДНФ, используя программы блочного покрытия с последующей минимизацией Bool- либо BDDI-представлений полученных блоков. Это не исключает применения и других программ логической минимизации, особенно в тех случаях, когда система ДНФ не является разреженной. Такая система ДНФ задается матрицей конъюнкций T^x, характеризующейся небольшим процентом неопределенных элементов, а матрица B^f характеризуется небольшим процентом нулевых значений. В этих случаях могут быть эффективными комбинированные методы многоэтапной совместной минимизации многоуровневых представлений [32] либо подходы, основанные на выделении из системы функций таких подсистем, для которых преимущество при синтезе имеет совместная минимизация [31]. После блочного покрытия для логической минимизации могут быть использованы не только представления проектных данных в виде BDDI и Bool, но и другие структуры данных.

Вклад авторов. *С. Н. Кардаш* разработал программные средства для нахождения блочных покрытий систем ДНФ и выбрал лучшую программу блочного покрытия, *П. Н. Бибило* выполнил эксперименты, подготовил текст статьи.

Список использованных источников

1. Тарасов, И. Е. ПЛИС Xilinx. Языки описания аппаратуры VHDL и Verilog, САПР, приемы проектирования / И. Е. Тарасов. – М. : Горячая линия – Телеком, 2020. – 538 с.

2. Закревский, А. Д. Логический синтез каскадных схем / А. Д. Закревский. – М. : Наука, 1981. – 416 с.

3. Logic Minimization Algorithm for VLSI Synthesis / K. R. Brayton [et al.]. – Boston : Kluwer Academic Publishers, 1984. – 193 p.

4. Синтез асинхронных автоматов на ЭВМ / под ред. А. Д. Закревского. – Минск : Наука и техника, 1975. – 184 с.

5. Brayton, R. K. The decomposition and factorization of Boolean expressions / R. K. Brayton, C. T. McMullen // Proc. of IEEE Intern. Symp. on Circuits and Systems (ISCAS 1982), Rome, Italy, 10–12 May 1982. – Rome, 1982. – P. 49–54.

6. MIS: A multiple-level logic optimization systems / R. K. Brayton [et al.] // IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems. – 1987. – Vol. CAD-6, no. 6. – P. 1062–1081.

7. Scholl, C. Functional Decomposition with Applications to FPGA Synthesis / C. Scholl. – Boston : Kluwer Academic Publishers, 2001. – 288 p.

8. Поттосин, Ю. В. Табличные методы декомпозиции систем полностью определенных булевых функций / Ю. В. Поттосин, Е. А. Шестаков. – Минск : Беларус. навука, 2006. – 327 с.

9. Sasao, T. Memory-Based Logic Synthesis / T. Sasao. – N. Y. : Springer, 2011. – 189 p.

10. Бибило, П. Н. Декомпозиция булевых функций на основе решения логических уравнений / П. Н. Бибило. – Минск : Беларус. навука, 2009. – 211 с.

11. Bryant, R. E. Graph-based algorithms for Boolean function manipulation / R. E. Bryant // IEEE Transactions on Computers. – 1986. – Vol. 35, no. 8. – P. 677–691.

12. Drechsler, R. Binary Decision Diagrams: Theory and Implementation / R. Drechsler, B. Becker. – Springer, 1998. – 210 p.

13. Ebendt, R. Advanced BDD Optimization / R. Ebendt, G. Fey, R. Drechsler. – Springer, 2005. – 222 p.

14. Bryant, R. E. Ordered binary decision diagrams / R. E. Bryant, C. Meinel // Logic Synthesis and Verification / eds.: S. Hassoun, T. Sasao, R. K. Brayton. – Kluwer Academic Publishers, 2002. – P. 285–307.

15. Meinel, C. Algorithms and Data Structures in VLSI Design: OBDD – Foundations and Applications / C. Meinel, T. Theobald. – Berlin, Heidelberg : Springer-Verlag, 1998. – 267 p.

16. Кнут, Д. Э. Искусство программирования. Т. 4, А. Комбинаторные алгоритмы. Ч. 1 : пер. с англ. / Д. Э. Кнут. – М. : Вильямс, 2013. – 960 с.

17. Бибило, П. Н. Применение диаграмм двоичного выбора при синтезе логических схем / П. Н. Бибило. – Минск : Беларус. навука, 2014. – 231 с.

18. Бибило, П. Н. Экспериментальное сравнение эффективности алгоритмов оптимизации BDD-представлений систем булевых функций / П. Н. Бибило, Ю. Ю. Ланкевич // Программные продукты и системы. – 2020. – Т. 33, № 3. – С. 449–463.

19. Бибило, П. Н. Логическая минимизация булевых сетей с использованием разложения Шеннона / П. Н. Бибило, Ю. Ю. Ланкевич // Информатика. – 2019. – Т. 16, № 2. – С. 73–89.

20. A novel basis for logic rewriting / W. Haaswijk [et al.] // Proc. of 22nd Asia and South Pacific Design Automation Conf. (ASP-DAC), Chiba, Japan, 16–19 Jan. 2017. – Chiba, 2017. – P. 151–156. https://doi.org/10.1109/ASPDAC.2017.7858312

21. Optimizing majority-inverter graphs with functional hashing / M. Soeken [et al.] // Proc. of the 2016 Design, Automation & Test in Europe Conf. & Exhibition (DATE), Dresden, Germany, 14–18 March 2016. – Dresden, 2016. – P. 1030–1035.

22. Exact synthesis of majority-inverter graphs and its applications / M. Soeken [et al.] // IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems. – 2017. – Vol. 36, no. 11. – P. 1842–1855.

23. Size optimization of MIGs with an application to QCA and STMG technologies / H. Riener [et al.] // Proc. of the 14th IEEE/ACM Intern. Symp. on Nanoscale Architectures, Athens, Greece, 17–19 July 2018. – Athens, 2018. – P. 157–162.

24. Harlecek, I. Are XORs in logic synthesis really necessary? / I. Harlecek, P. Fiser, J. Schmidt // IEEE 20th Intern. Symp. on Design and Diagnostics of Electronic Circuits & Systems (DDECS), Dresden, Germany, 19–21 Apr. 2017. – Dresden, 2017. – P. 134–139.

25. Кардаш, С. Н. Построение блочных разбиений систем булевых функций на основе задачи покрытия булевых матриц / С. Н. Кардаш // BIG DATA и анализ высокого уровня = BIG DATA and Advanced Analytics : сб. науч. ст. IX Междунар. науч.-практ. конф., Минск, 17–18 мая 2023 г. : в 2 ч. – Минск : БГУИР, 2023. – Ч. 2. – С. 326–330.

26. Бибило, П. Н. Система логической оптимизации функционально-структурных описаний цифровых устройств на основе продукционно-фреймовой модели представления знаний / П. Н. Бибило, В. И. Романов // Проблемы разработки перспективных микро- и наноэлектронных систем. – 2020. – Вып. 4. – С. 9–16.

27. Бибило, П. Н. Системы проектирования интегральных схем на основе языка VHDL. StateCAD, ModelSim, LeonardoSpectrum / П. Н. Бибило. – М. : СОЛОН-Пресс, 2005. – 384 с.

28. Авдеев, Н. А. Автоматизированное проектирование цифровых операционных устройств с пониженным энергопотреблением / Н. А. Авдеев, П. Н. Бибило // Программная инженерия. – 2021. – Т. 12, № 2. – С. 63–73.

29. Соловьев, В. В. Архитектуры ПЛИС фирмы Xilinx: FPGA и CPLD 7-й серии / В. В. Соловьев. – М. : Горячая линия – Телеком, 2016. – 392 с.

30. Бибило, П. Н. Использование полиномов Жегалкина при минимизации многоуровневых представлений систем булевых функций на основе разложения Шеннона / П. Н. Бибило, Ю. Ю. Ланкевич // Программная инженерия. – 2017. – № 8. – С. 369–384.

31. Бибило, П. Н. Выделение из многоуровневого представления системы булевых функций подсистем для совместной логической минимизации / П. Н. Бибило, Н. А. Кириенко, В. И. Романов // Программные продукты и системы. – 2023. – Т. 36, № 4. – С. 197–206.

32. Бибило, П. Н. Логическая минимизация многоуровневых представлений систем булевых функций / П. Н. Бибило, Ю. Ю. Ланкевич, В. И. Романов // Информационные технологии. – 2023. – Т. 29, № 2. – С. 59–71.

References

1. Tarasov I. E. PLIS Xilinx. Yazyki opisaniya apparatury VHDL i Verilog, SAPR, priemy proektirovaniya. *XILINX FPGA. Hardware Description Languages VHDL and Verilog, CAD, Design Techniques*. Moscow, Goryachaya liniya – Telekom, 2020, 538 p. (In Russ.).

2. Zakrevskij A. D. Logicheskij sintez kaskadnyh skhem. *Logical Synthesis of Cascading Circuit.* Moscow, Nauka, 1981, 416 p. (In Russ.).

3. Brayton K. R., Hachtel G. D., McMullen C., Sangiovanni-Vincentelli A. L. *Logic Minimization Algorithm for VLSI Synthesis*. Boston, Kluwer Academic Publishers, 1984, 193 p.

4. Zakrevskij A. D. (ed.). Sintez asinhronnyh avtomatov na EHVM. *Synthesis of Asynchronous Automata on a Computer*. Minsk, Nauka i tekhnika, 1975, 184 p. (In Russ.).

5. Brayton R. K., McMullen C. T. The decomposition and factorization of Boolean expressions. *Proceedings of IEEE International Symposium on Circuits and Systems (ISCAS 1982), Rome, Italy, 10–12 May 1982*. Rome, 1982, pp. 49–54.

6. Brayton R. K., Rudell R., Sangiovanni-Vincentelli A. L., Wang A. R. MIS: A multiple-level logic optimization systems. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 1987, vol. CAD-6, no. 6, pp. 1062–1081.

7. Scholl C. *Functional Decomposition with Application to FPGA Synthesis*. Boston, Kluwer Academic Publishers, 2001, 288 p.

8. Pottosin Yu. V., Shestakov E. A. Tablichnye metody dekompozicii sistem polnost'yu opredelennyh bulevyh funkcij. *Tabular Methods for Decomposition of Systems of Completely Defined Boolean Functions*. Minsk, Belaruskaja navuka, 2006, 327 p. (In Russ.).

9. Sasao T. *Memory-Based Logic Synthesis*. New York, Springer, 2011, 189 p.

10. Bibilo P. N. Dekompoziciya bulevyh funkcij na osnove resheniya logicheskih uravnenij. *Decomposition of Boolean Functions Based on the Solution of Logical Equations*. Minsk, Belaruskaja navuka, 2009, 211 p. (In Russ.).

11. Bryant R. E. Graph-based algorithms for Boolean function manipulation. *IEEE Transactions on Computers*, 1986, vol. 35, no. 8, pp. 677–691.

12. Drechsler R., Becker B. *Binary Decision Diagrams: Theory and Implementation*. Springer, 1998, 210 p.

13. Ebendt R., Fey G., Drechsler R. *Advanced BDD Optimization.* Springer, 2005, 222 p.

14. Bryant R. E., Meinel C. Ordered binary decision diagrams. In S. Hassoun, T. Sasao, R. K. Brayton (eds.). *Logic Synthesis and Verification.* Kluwer Academic Publishers, 2002, pp. 285–307.

15. Meinel C., Theobald T. *Algorithms and Data Structures in VLSI Design: OBDD – Foundations and Applications.* Berlin, Heidelberg, Springer-Verlag, 1998, 267 p.

16. Knuth D. E. *The Art of Computer Programming, Volume 4A: Combinatorial Algorithms, Part 1*. Addison-Wesley Professional, 2011, 912 p.

17. Bibilo P. N. Primenenie diagram dvoichnogo vybora pri sinteze logicheskih shem. *Application of Binary Selection Diagrams in the Synthesis of Logic Circuits*. Minsk, Belaruskaja navuka, 2014, 231 p. (In Russ.).

18. Bibilo P. N., Lankevich Yu. Yu. *Experimental investigation of effectiveness of algorithms for minimizing BDD representations of Boolean function systems*, Programmnye produkty i sistemy [*Software & Systems*], 2020, vol. 33, no. 3, pp. 449–463 (In Russ.).

19. Bibilo P. N., Lankevich Yu. Yu. *Logical optimization of Boolean nets using Shannon expansion*. Informatika [*Informatics*], 2019, vol. 16, no. 2, pp. 73–89 (In Russ.).

20. Haaswijk W., Soeken M., Amaru L., Gaillardon P.-E., De Micheli G. A novel basis for logic rewriting. *Proceedings of 22nd Asia and South Pacific Design Automation Conference (ASP-DAC), Chiba, Japan, 16–19 January 2017*. Chiba, 2017, pp. 151–156. https://doi.org/10.1109/ASPDAC.2017.7858312

21. Soeken M., Amaru L. G., Gaillardon P., De Micheli G. Optimizing majority-inverter graphs with functional hashing. *Proceedings of the 2016 Design, Automation & Test in Europe Conference & Exhibition (DATE), Dresden, Germany, 14–18 March 2016*. Dresden, 2016, pp. 1030–1035.

22. Soeken M., Amaru L., Gaillardon P.-E., De Micheli G. Exact synthesis of majority-inverter graphs and its applications. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 2017, vol. 36, no. 11, pp. 1842–1855.

23. Riener H., Testa E., Amaru L., Soeken M., De Micheli G. Size optimization of MIGs with an application to QCA and STMG technologies. *Proceedings of the 14th IEEE/ACM International Symposium on Nanoscale Architectures, Athens, Greece, 17–19 July 2018*. Athens, 2018, pp. 157–162.

24. Harlecek I, Fiser P., Schmidt J. Are XORs in logic synthesis really necessary? *IEEE 20th International Symposium on Design and Diagnostics of Electronic Circuits & Systems (DDECS), Dresden, Germany, 19–21 April 2017*. Dresden, 2017, pp. 134–139.

25. Kardash S. N. *Construction of block partitions of Boolean function systems based on the problem of covering Boolean matrices*. BIG DATA i analiz vysokogo urovnja : sbornik nauchnyh statej IX Mezhdunarodnoj nauchno-prakticheskoj konferencii, Minsk, 17–18 maja 2023 g. : v 2 chastjah. Chast' 2 [*BIG DATA and Advanced Analytics : Collection of Scientific Articles of the IX International Scientific and Practical Conference, Minsk, 17–18 May 2023 : in 2 Parts*]. Minsk, Belorusskij gosudarstvennyj universitet informatiki i radiojelektroniki, 2023, part 2, pp. 326–330 (In Russ.).

26. Bibilo P. N., Romanov V. I. *The system of logical optimization of functional structural descriptions of digital circuits based on production-frame knowledge representation model*. Problemy razrabotki perspektivnyh mikro- i nanoelektronnyh system [*Problems of Developing Promising Micro- and Nanoelectronic Systems*], 2020, iss. 4, pp. 9–16 (In Russ.).

27. Bibilo P. N. Cistemy proektirovaniya integral'nyh skhem na osnove yazyka VHDL. StateCAD, ModelSim, LeonardoSpectrum. *Integrated Circuit Design Systems Based on the VHDL Language. StateCAD, ModelSim, LeonardoSpectrum*. Moscow, SOLON-Press, 2005, 384 p. (In Russ.).

28. Avdeev N. A., Bibilo P. N. *Design of digital operational units with low power consumption*. Programmnaya inzheneriya [*Software Engineering*], 2021, vol. 12, no. 2, pp. 63–73 (In Russ.).

29. Solov'ev V. V. Arhitektury PLIS firmy Xilinx: FPGA i CPLD 7-j serii. *XILINX FPGA Architectures: FPGA and CPLD 7-Series*. Moscow, Goryachaya liniya – Telekom, 2016, 392 p. (In Russ.).

30. Bibilo P. N., Lankevich Yu. Yu. *The use of Zhegalkin polynomials for minimization of multilevel represintations of Boolean functions based on Shannon expansion*. Programmnaya inzheneriya [*Software Engineering*], 2017, no. 8, pp. 369–384 (In Russ.).

31. Bibilo P. N., Kirienko N. A., Romanov V. I. *Extraction from a multilevel representation of a system of Boolean functions of subsystems for joint logical minimization*. Programmnye produkty i sistemy [*Software & Systems*], 2023, vol. 36, no. 4, pp. 197–206 (In Russ.).

32. Bibilo P. N., Lankevich Yu. Yu., Romanov V. I. *Logical minimization of multilevel representations of Boolean function systems*. Informacionnye tekhnologii [*Information Technology*], 2023, vol. 29, no. 2, pp. 59–71 (In Russ.).

Информация об авторах

Бибило Петр Николаевич, доктор технических наук, профессор, Объединенный институт проблем информатики Национальной академии наук Беларуси.
E-mail: bibilo@newman.bas-net.by

Кардаш Сергей Николаевич, кандидат технических наук, Объединенный институт проблем информатики Национальной академии наук Беларуси.
E-mail: kardash77@gmail.com

Information about the authors

Petr N. Bibilo, D. Sc. (Eng.), Prof., The United Institute of Informatics Problems of the National Academy of Sciences of Belarus.
E-mail: bibilo@newman.bas-net.by

Sergey N. Kardash, Ph. D. (Eng.), The United Institute of Informatics Problems of the National Academy of Sciences of Belarus.
E-mail: kardash77@gmail.com

КОСМИЧЕСКИЕ ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ И ГЕОИНФОРМАТИКА

SPACE INFORMATION TECHNOLOGIES AND GEOINFORMATICS

УДК 550.388.2
https://doi.org/10.37661/1816-0301-2024-21-1-48-64

Оригинальная статья
Original Paper

Алгоритм оценки абсолютного полного электронного содержания ионосферы по данным двухчастотных фазовых и дальностных спутниковых измерений

А. С. Шапкин

Объединенный институт проблем информатики
Национальной академии наук Беларуси,
ул. Сурганова, 6, Минск, 220012, Беларусь
E-mail: al_shapkin@newman.bas-net.by, shap1kin2@gmail.com

Аннотация

Цели. Решается задача разработки алгоритма оценки абсолютного полного электронного содержания ионосферы по данным двухчастотных фазовых и дальностных спутниковых измерений для одиночной приемной станции глобальных навигационных спутниковых систем.

Методы. Для получения оценки корректируются данные фазовых измерений методами цифровой обработки сигналов, применяются и комбинируются известные формулы полного электронного содержания для фазовых и дальностных измерений, оценивается дифференциальная кодовая задержка приемной станции методом наименьших квадратов.

Результаты. Показано, что полное электронное содержание, рассчитанное по фазовым измерениям, обеспечивает высокую точность, но с точностью до неизвестной константы, а рассчитанное по дальностным измерениям позволяет получить абсолютное значение, но с большой шумовой составляющей и дифференциальной кодовой задержкой аппаратуры спутника и приемника. Разработан алгоритм оценки абсолютного полного электронного содержания ионосферы, приведены его описание и схема. Алгоритм применен для оценки полного электронного содержания за полгода наблюдений, рассчитана средняя ошибка полученной оценки.

Заключение. Разработанный алгоритм может быть использован для оценки абсолютного полного электронного содержания ионосферы для одиночной приемной станции глобальных навигационных спутниковых систем. В отличие от теоретически известных формул для фазовых и дальностных измерений в настоящей статье содержатся сведения о корректировке фазовых измерений и оценке дифференциальной кодовой задержки приемной станции. Дальнейшие исследования могут быть связаны с адаптивным подбором параметров и тестированием алгоритма для работы с наноспутниками формата CubeSat.

Ключевые слова: ионосфера, полное электронное содержание, глобальная навигационная спутниковая система, показатель преломления, дифференциальная кодовая задержка, метод наименьших квадратов

Благодарности. Работа выполнена в рамках договора № 220/12 «Разработать алгоритмические и программные средства обработки радиотомографических данных низкоорбитального контроля ионосферы» (04.05.2022–31.12.2025 гг.), заключенного с УП «Геоинформационные системы» по проекту «Разработать космическую систему радиометрического контроля околоземного пространства на базе малого космического аппарата и специализированных наземных средств» (мероприятия 8 подпрограммы 6 «Исследование и использование космического пространства в мирных целях» Государственной программы «Наукоемкие технологии и техника» на 2021–2025 гг.), а также в рамках подпрограммы «Разработка аппаратного и программно-алгоритмического комплекса радиометрического анализа динамических состояний ионосферы» научно-технической программы Союзного государства «Комплекс-СГ» (2022–2026).

Для цитирования. Шапкин, А. С. Алгоритм оценки абсолютного полного электронного содержания ионосферы по данным двухчастотных фазовых и дальностных спутниковых измерений / А. С. Шапкин // Информатика. – 2024. − Т. 21, № 1. – С. 48–64. https://doi.org/10.37661/1816-0301-2024-21-1-48-64

Конфликт интересов. Автор заявляет об отсутствии конфликта интересов.

Поступила в редакцию | Received 24.09.2023
Подписана в печать | Accepted 03.01.2024
Опубликована | Published 29.03.2024

Algorithm for estimating the absolute total electron content of the ionosphere from dual-frequency phase and range satellite measurements

Aliaksandr S. Shapkin

The United Institute of Informatics Problems
of the National Academy of Sciences of Belarus,
st. Surganova, 6, Minsk, 220012, Belarus
E-mail: al_shapkin@newman.bas-net.by, shap1kin2@gmail.com

Abstract

Objectives. The problem of developing an algorithm for estimating the absolute total electron content of the ionosphere from dual-frequency phase and range satellite measurements for a single receiving station of global navigation satellite systems is being solved.

Methods. To obtain an estimate the phase measurement data are corrected using digital signal processing methods, well known total electron content formulas for phase and range measurements are applied and combined, and also the differential code bias of the receiving station is estimated using the least squares method.

Results. It is shown that the total electron content calculated from phase measurements provides high accuracy, but up to an unknown constant, but the content calculated from range measurements allows one to obtain the absolute value, but with a large noise component and differential code bias of a satellite and receiver equipment. An algorithm for estimating the absolute total electron content of the ionosphere has been developed, its description and diagram are given. The algorithm was used to estimate the total electronic content within six months of observations, and the average error of the resulting estimate was calculated.

Conclusion. The developed algorithm can be used to estimate the absolute total electron content of the ionosphere for a single receiving station of global navigation satellite systems. In contrast to theoretically known formulas for phase and range measurements, this article contains information about adjusting phase measurements and estimating the differential code delay of receiving station. Further research may be related to the adaptive selection of parameters and testing of the algorithm for working with nanosatellites of the CubeSat format.

Keywords: ionosphere, total electron content, global navigation satellite systems, refractive index, differential code bias, least squares method

Acknowledgements. The work is carried out within the agreement no. 220/12 "Development of algorithmic and software tools for processing radio tomographic data of low-orbit ionosphere monitoring" (05.05.2022–12.31.2025) with UE "Geoinformation Systems" on the project "Develop a space system for radiometric monitoring of near-Earth space based on a small spacecraft and specialized ground facilities" (activities of subprogram 6 "Research and use of outer space for peaceful purposes" of the State Program "Science-intensive technologies and engineering" for 2021–2025), as well as within the framework of the subprogram "Development of hardware and software-algorithmic complex for radiometric analysis of dynamic states of the ionosphere" of the scientific and technical program of the Union State "Complex-SG" (2022–2026).

For citation. Shapkin A. S. *Algorithm for estimating the absolute total electron content of the ionosphere from dual-frequency phase and range satellite measurements*. Informatika [*Informatics*], 2024, vol. 21, no. 1, pp. 48–64 (In Russ.). https://doi.org/10.37661/1816-0301-2024-21-1-48-64

Conflict of interest. The author declares of no conflict of interest.

Введение. Главными факторами, которые влияют на распространение электромагнитных волн, являются концентрация свободных электронов n_e, внешнее магнитное поле B_0 и эффективная частота соударений между электронами и другими частицами ν, а также такие свойства самой волны, как ее частота f, направление распространения $\vec{k}$ и поляризация – направление вращения электрического (или магнитного) вектора волны [1].

Если для ионосферы справедливы следующие условия [1]:

– статистическая однородность распределения зарядов, так что в результате пространственный заряд отсутствует;

– влияние на распространение волны только электронов;

– пренебрежение тепловым движением (холодная плазма);

– магнитные свойства свободного пространства (магнитная проницаемость среды равна единице);

– отсутствие поляризационного члена Лоренца (Лоренцова поправка [2]).

– однородность внешнего магнитного поля $\vec{B}_0$;

– частота электронных соударений ν не зависит от энергии электрона,

то показатель преломления определяется формулой Эпплтона – Лассена [1, 2]:

$$\bar{n}^2 = 1 - \frac{a}{1 - \mathrm{i}c - \frac{b_\perp^2}{2(1-a-\mathrm{i}c)} \pm \sqrt{\frac{b_\perp^4}{4(1-a-\mathrm{i}c)^2} + b_\parallel^2}}; \quad \bar{n} = \mu - \mathrm{i}\chi, \tag{1}$$

где $\mathrm{i} = \sqrt{-1}$ – мнимая единица; $a = \left(\omega_e / \omega\right)^2$, ω – радиальная частота волны, $\omega_e = e\sqrt{\frac{n_e}{\varepsilon_0 m_e}}$ – плазменная частота ионосферы, m_e – масса электрона, e – заряд электрона, ε_0 – диэлектрическая постоянная; $b = \omega_H / \omega$; $b_\perp = b\sin\theta$; $b_\parallel = b\cos\theta$, θ – угол между направлением распространения волны $\vec{k}$ и вектором магнитной индукции $\vec{B}_0$, $\omega_H = \frac{|e| B_0}{m_e}$ – гиромагнитная частота; $c = \nu / \omega$.

Поскольку основной вклад в формулу (1) вносит электрическая составляющая, то на практике влиянием магнитного поля и соударениями можно пренебречь ($b = c = 0$). Получаем известную формулу показателя преломления [1], зависящую только от электронной концентрации:

$$\mu^2 = 1 - a = 1 - \left(\frac{\omega_e}{\omega}\right)^2 = 1 - \frac{e^2 n_e}{4\pi^2 \varepsilon_0 m_e f^2}. \tag{2}$$

Основной характеристикой при изучении ионосферы является полное электронное содержание (ПЭС, англ. *TEC*), определяемое как интеграл от электронной концентрации по пути распространения волны между приемной станцией и спутником:

$$TEC = \int_{L_0(t)}^{L(t)} n_e(z)dz, \quad (3)$$

где $L_0(t)$ – нижняя граница ионосферного слоя; $L(t)$ – расстояние от приемной станции до спутника; $n_e(z)$ – электронная концентрация на трассе вдоль оси z. Таким образом, ПЭС характеризует количество свободных электронов, содержащихся на трассе по пути следования радиоволны с поперечным сечением 1 м^2, и выражается в единицах TECU (1 TECU = 10^{16} эл/м^2).

Существует несколько методов определения ПЭС, которые основаны на допущении, что показатель преломления ионосферы определяется формулой (2). Методы можно разделить по количеству используемых частот сигнала спутника глобальных навигационных спутниковых систем (ГНСС): одночастотные, двухчастотные, трехчастотные и четырехчастотные, а также по параметру, с помощью которого рассчитывается ПЭС: дальностные (по псевдодальности) и фазовые (по фазе несущей).

Одним из основных способов оценки ПЭС является двухчастотный фазовый метод [3, 4], подробное обоснование которого есть, например, в работе [5]:

$$TEC_\psi = \frac{1}{A}\frac{f_1^2 f_2^2}{f_1^2 - f_2^2}\left(O_1\lambda_1 - O_2\lambda_2 + \text{const}_{1,2} + \sigma O\right), \quad (4)$$

где $A = \dfrac{e^2}{8\pi^2\varepsilon_0 m_e} \approx 40{,}308$ м^{3}Гц^2; O_1 и O_2 – фазы несущих (в оборотах) на частотах f_1 и f_2; λ_1 и λ_2 – длины волн; $\text{const}_{1,2}$ – неоднозначность фазовых измерений, вызванная неизвестной начальной фазой; σO – ошибка фазовых измерений. В работе [3] утверждается, что при интервале усреднения в 30 с ошибка расчета не превосходит 0,01 TECU. Несмотря на высокую точность, результаты фазовых измерений являются относительными. В связи с этим требуется применение дополнительных специальных методик для устранения неоднозначности фазовых измерений [6].

Еще одним методом определения ПЭС является двухчастотный дальностный метод [3, 4]:

$$TEC_D = \frac{1}{A}\frac{f_1^2 f_2^2}{f_1^2 - f_2^2}(D_2 - D_1 + \sigma D), \quad (5)$$

где D_1 и D_2 – псевдодальности на частотах f_1 и f_2; σD – ошибка дальностных измерений, связанная главным образом с задержкой радиосигнала в аппаратуре (дифференциальной кодовой задержкой).

В отличие от двухчастотного фазового метода дальностный метод позволяет получить абсолютное значение ПЭС, однако эти результаты будут крайне зашумленными. Согласно [4] уровень шумовой составляющей составляет в среднем от 30 до 50 %, а в некоторых случаях может достигать и 100 %, что затрудняет применение данного метода для определения пространственных и временных возмущений ПЭС. Кроме того, вычисленное таким образом ПЭС также содержит некоторую аддитивную константу, называемую дифференциальной кодовой задержкой (ДКЗ). ДКЗ вызывается частотно-зависимыми задержками в аппаратуре спутника и приемника и может достигать нескольких десятков TECU [7].

Кроме основных двухчастотных методов могут применяться одночастотный [8], трехчастотный [9] и четырехчастотный методы [10]. Последний позволяет довести величину неоднозначности до значений, превосходящих ширину физически возможного диапазона значений ПЭС, т. е. определить ее однозначно. Однако учитывая технические возможности приемопередающей аппаратуры ГНСС, возможным представляется использование одно- и двухчастотных методов. При этом измерения одночастотным методом являются относительными,

а среднеквадратическое отклонение (СКО) вариаций ПЭС не превосходит 0,1 TECU [3], что значительно хуже, чем при использовании формулы (4). В связи с этим использование двухчастотных методов предпочтительнее. На рис. 1 представлены результаты применения различных методов на данных одной реализации спутника ГНСС.

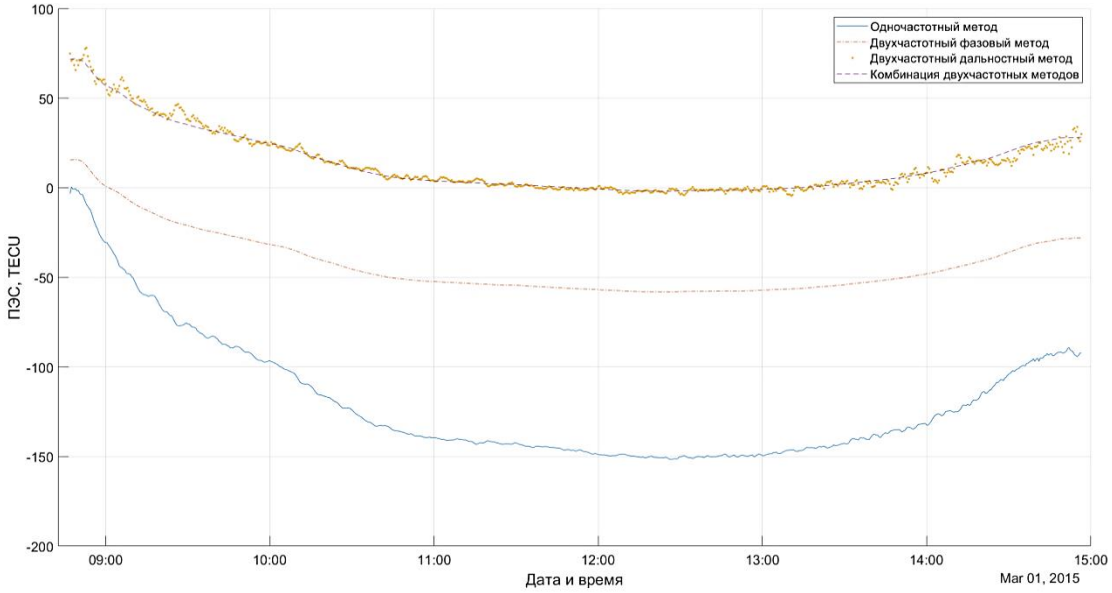


Рис. 1. ПЭС, рассчитанное различными методами на минской станции для одной реализации спутника 7 GPS

Fig. 1. TEC calculated by various methods on Minsk station for single realization of satellite 7 GPS

ПЭС, рассчитанное по фазовым измерениям, обеспечивает высокую точность, однако с точностью до неизвестной константы (которая также может меняться в ходе наблюдения), а ПЭС, рассчитанное по дальностным измерениям, позволяет получить абсолютное значение, но с большой шумовой составляющей и ДКЗ аппаратуры спутника и приемника. В связи с этим возникает потребность получения оценки, сочетающей высокую точность фазовых измерений и абсолютный уровень дальностных.

Целью работы является разработка алгоритма оценки абсолютного ПЭС ионосферы по данным двухчастотных фазовых и дальностных спутниковых измерений для одиночной приемной станции ГНСС. Ниже приводятся описание алгоритма, его упрощенная схема, результаты применения и расчет средней ошибки полученной оценки ПЭС.

Описание алгоритма. Пусть в каждый момент времени известны следующие характеристики:

– фазовые (фаза несущей) O_1 и O_2 и дальностные (псевдодальность) D_1 и D_2 измерения на двух частотах f_1 и f_2;

– угол места спутника ε;

– ДКЗ спутника DCB_s.

Сущность алгоритма состоит в комбинировании расчетов ПЭС по фазовым и дальностным измерениям, а также в оценке и учете ионосферных задержек, вносимых ДКЗ приемной станции и спутников. Алгоритм оценки абсолютного ПЭС состоит из пяти основных шагов.

Шаг 1. Корректировка фазовых измерений.

Данные фазовых измерений, полученные от спутников ГНСС, необходимо «очистить» за счет корректировки разрывов и скачков в значениях [6]. Скачки (cycle slips) и разрывы (gaps) представляют собой кратковременный срыв в фазовой синхронизации приемника на спутниковый сигнал. При фазовых измерениях мгновенное значение фазы можно представить в виде суммы целой (неоднозначности фазового измерения) и дробной частей:

$$\psi = 2\pi(N + l) + \Delta\psi\,, \tag{6}$$

где N – неизвестное целое число, l – целое число. При этом N остается постоянным до потери сигнала. После этого неизвестная константа N переопределяется, что приводит к срыву фазовой синхронизации и скачкообразному изменению принимаемого значения фазы.

Причины возникновения разрывов и скачков могут быть условно разделены на три группы [6]:

1) препятствия на траектории распространения луча: горы, здания, мосты, деревья и т. д.;

2) низкое отношение сигнал/шум, вызванное плохими ионосферными условиями, многолучевостью, низким углом места спутника и т. д.;

3) сбои в работе программного обеспечения приемника или неисправность спутникового генератора.

Для исправления скачка необходимо детектировать его на фоне шумов. Существует достаточно большое количество методов детектирования скачков [6, 11]. Учитывая необходимость проведения точных измерений фазовых характеристик, а также наличие двухчастотного канала приема сигналов ГНСС, целесообразно выбрать метод для одиночной двухчастотный приемной станции. В работе [11] для данных с относительно большим временным шагом наблюдения (не менее 5 с) комбинация методов FBMWA – STPIR (forward and backward moving window averaging – second-order, time-difference phase ionosphere residual) [12] показала наилучший результат, поэтому для детектирования и исправления скачков будет использоваться именно она.

Алгоритм FBMWA (forward and backward moving window averaging) базируется на алгоритме TurboEdit [13] и использует линейную комбинацию Мельбурна – Вюббена:

$$L_{MW} = \frac{f_1 \cdot \lambda_1 O_1 - f_2 \cdot \lambda_2 O_2}{f_1 - f_2} - \frac{f_1 \cdot D_1 - f_2 \cdot D_2}{f_1 + f_2} = \lambda_{WL} N_{WL}\,, \tag{7}$$

где $\lambda_{WL} = c/(f_1 - f_2)$; $N_{WL} = N_1 - N_2$ – разность неоднозначностей фазовых измерений.

Из формулы (7) найдем разность неоднозначностей фазовых измерений:

$$N_{WL} = O_1 - O_2 - \frac{f_1 \cdot \mathrm{D}_1 - f_2 \cdot D_2}{\lambda_{WL}(f_1 + f_2)}\,. \tag{8}$$

В алгоритме TurboEdit далее рекурсивно находятся среднее и дисперсия выражения (8):

$$\bar{N}_{WL}(k) = \bar{N}_{WL}(k-1) + \frac{1}{k}\left[N_{WL}(k) - \bar{N}_{WL}(k-1)\right]; \tag{9}$$

$$\sigma^2(k) = \sigma^2(k-1) + \frac{1}{k}\left[\left(N_{WL}(k) - \bar{N}_{WL}(k-1)\right)^2 - \sigma^2(k-1)\right], \tag{10}$$

где k – номер отсчета внутри одного наблюдения (пролета) спутника.

Считается, что обнаружен скачок, если выполняются следующие условия:

$$\begin{cases} \left|N_{WL}(k) - \bar{N}_{WL}(k-1)\right| \geq 4\sigma(k); \\ \left|N_{WL}(k+1) - N_{WL}(k)\right| \leq 1. \end{cases} \tag{11}$$

Величина скачка находится как

$$\Delta N_1(k) - \Delta N_2(k) = N_{WL}(k) - \bar{N}_{WL}(k-1)\,, \tag{12}$$

где ΔN_1, ΔN_2 – целые величины скачка на частотах f_1 и f_2.

В алгоритме же FBMWA разность неоднозначностей фазовый измерений (8) усредняется:

$$\bar{N}_{WL,B}(k-1)=\frac{1}{m}\sum_{i=k-1}^{k-m}N_{WL}(i); \tag{13}$$

$$\bar{N}_{WL,F}(k)=\frac{1}{n}\sum_{i=k}^{k+n-1}N_{WL}(i); \tag{14}$$

$$\Delta\bar{N}_{WL}(k)=\bar{N}_{WL,F}(k)-\bar{N}_{WL,B}(k-1), \tag{15}$$

где n, m – некоторые целые числа, подбираемые эмпирически. Усреднение (13)–(15) существенно снижает уровень шума и повышает точность алгоритма. Выражение (15) позволяет детектировать скачок, но не дает информации о величине скачка на каждой из частот. Кроме того, алгоритм FMBWA имеет «слепую зону», когда $\Delta N_1=\Delta N_2$. В связи с этим FMBWA используется совместно с алгоритмом STPIR (second-order, time-difference phase ionosphere residual) [12], который является несколько улучшенной версией PIR [14]. STPIR применяет линейную комбинацию фазовых измерений:

$$L_{PIR}=O_1-\frac{f_1}{f_2}O_2=N_1-\frac{\lambda_1}{\lambda_2}N_2+I_{res}, \tag{16}$$

где $I_{res}=\dfrac{f_1^2-f_2^2}{f_2^2\lambda_1}I$, I – задержка сигнала в ионосфере в метрах.

Величина скачка может быть оценена путем нахождения приращения выражения (16) [12]:

$$\Delta N_1(k)-\frac{f_1}{f_2}\Delta N_2(k)=\left(L_{PIR}(k)-2\cdot L_{PIR}(k-1)+L_{PIR}(k-2)\right)-\left(I_{res}(k)-2I_{res}(k-1)+I_{res}(k-2)\right). \tag{17}$$

При этом величина второго слагаемого близка к нулю и на практике ею пренебрегают:

$$\Delta N_1(k)-\frac{f_1}{f_2}\Delta N_2(k)=L_{PIR}(k)-2\cdot L_{PIR}(k-1)+L_{PIR}(k-2)=\Delta L_{PIR}(k). \tag{18}$$

По аналогии с алгоритмом TurboEdit будем считать, что обнаружен скачок, если выполняется одно из условий

$$\left|\Delta\bar{N}_{WL}(k)\right|\geq 4\sigma_{FBMWA}(k); \tag{19}$$

$$\left|\Delta L_{PIR}(k)\right|\geq 4\sigma_{STPIR}(k), \tag{20}$$

где $\sigma_{FBMWA}(k)=\sqrt{\sigma_F^2(k)+\sigma_B^2(k-1)}$ – СКО метода FBMWA; σ_F – СКО переднего усреднения (14); σ_B – СКО заднего усреднения (13); σ_{STPIR} – СКО выражения (18). Определить величину скачка на каждой из частот можно, решив систему

$$\begin{cases}\Delta N_1-\Delta N_2=\left[\Delta\bar{N}_{WL}(k)\right];\\ \Delta N_1-\dfrac{f_1}{f_2}\Delta N_2=\Delta L_{PIR}(k),\end{cases} \tag{21}$$

где $\left[\Delta\bar{N}_{WL}(k)\right]$ – округленное до целого значение $\Delta\bar{N}_{WL}(k)$.

Последующие данные исправляются на округленную величину скачка:

$$\begin{aligned} O_1' &= O_1 - [\Delta N_1]; \\ O_2' &= O_2 - [\Delta N_2]. \end{aligned} \tag{22}$$

С точки зрения временной последовательности разрыв представляет собой кратковременное (до нескольких минут) отсутствие сигнала. Для корректировки разрыва воспользуемся комбинациями (8) и (18) и найдем их средние приращения слева и справа от разрыва:

$$\begin{aligned} \Delta\bar{N}_{WL1} &= \frac{1}{(p-1)^2}\sum_{i=l-p+1}^{l}\left(N_{WL}(i) - N_{WL}(i-p)\right); \\ \Delta\bar{N}_{WL2} &= \frac{1}{(p-1)^2}\sum_{i=r}^{r+p-1}\left(N_{WL}(i+p) - N_{WL}(i)\right); \\ \Delta\bar{L}_{PIR1} &= \frac{1}{(p-1)^2}\sum_{i=l-p+1}^{l}\left(L_{PIR}(i) - L_{PIR}(i-p)\right); \\ \Delta\bar{L}_{WL2} &= \frac{1}{(p-1)^2}\sum_{i=r}^{r+p-1}\left(L_{PIR}(i+p) - L_{PIR}(i)\right), \end{aligned} \tag{23}$$

где m – некоторое целое число, подобранное эмпирически; l – номер отсчета перед разрывом; r – номер отсчета после разрыва.

Величину разрыва определим, решив систему:

$$\begin{cases} \Delta N_1 - \Delta N_2 = \left[\dfrac{1}{2p}\displaystyle\sum_{i=l-2p+1}^{l} N_{WL}(i) - \dfrac{1}{2p}\sum_{i=r}^{r+2p-1} N_{WL}(i) - \left(\dfrac{\Delta\bar{N}_{WL1} + \Delta\bar{N}_{WL2}}{2}\right)(r-l)\right]; \\ \Delta N_1 - \dfrac{f_1}{f_2}\Delta N_2 = \dfrac{1}{2p}\displaystyle\sum_{i=l-2p+1}^{l} L_{PIR}(i) - \dfrac{1}{2p}\sum_{i=r}^{r+2p-1} L_{PIR}(i) - \left(\dfrac{\Delta\bar{L}_{PIR1} + \Delta\bar{L}_{PIR2}}{2}\right)(r-l). \end{cases} \tag{24}$$

Данные справа от разрыва исправляются, как в выражениях (22), а на разрыве заполняются линейно:

$$\begin{aligned} O_1'(i) &= O_1(l) + \left(O_1'(r) - O_1(l)\right)(r-l-1),\ i = \overline{l+1; r-1}; \\ O_2'(i) &= O_2(l) + \left(O_2'(r) - O_2(l)\right)(r-l-1),\ i = \overline{l+1; r-1}. \end{aligned} \tag{25}$$

Шаг 2. Подсчет относительного ПЭС по фазовым измерениям.

В условиях отсутствия априорной информации о сигнале в формуле (4) ошибка определения фазы и неоднозначность фазовых измерений приравниваются к нулю и используется формула

$$TEC_{\psi} = \frac{1}{A}\frac{f_1^2 f_2^2}{f_1^2 - f_2^2}\left(O_1'\lambda_1 - O_2'\lambda_2\right). \tag{26}$$

Шаг 3. Подсчет ПЭС по дальностным измерениям.

Так как ошибка дальностных измерений σD в формуле (5) связана в первую очередь с ДКЗ приемника и передатчика, ее можно записать следующим образом:

$$TEC_D = \frac{1}{A}\frac{f_1^2 f_2^2}{f_1^2 - f_2^2}(D_2 - D_1 + cDCB_s - cDCB_r), \quad (27)$$

где c – скорость света; DCB_s – ДКЗ спутника; DCB_r – ДКЗ приемной станции.

В условиях отсутствия какой-либо априорной информации о сигнале ошибка в расчете приравнивается к нулю. Тем не менее на открытом сайте NASA[1] публикуется ежедневная информация о ДКЗ спутников различных ГНСС (включая GPS и ГЛОНАСС) в наносекундах, поэтому будем считать, что ДКЗ спутников известны. Тогда формулу (27) можно записать в виде, который и будет использоваться в дальнейших расчетах:

$$TEC_D = \frac{1}{A}\frac{f_1^2 f_2^2}{f_1^2 - f_2^2}(D_2 - D_1 + cDCB_s). \quad (28)$$

На рис. 2 представлен результат применения формулы (28).

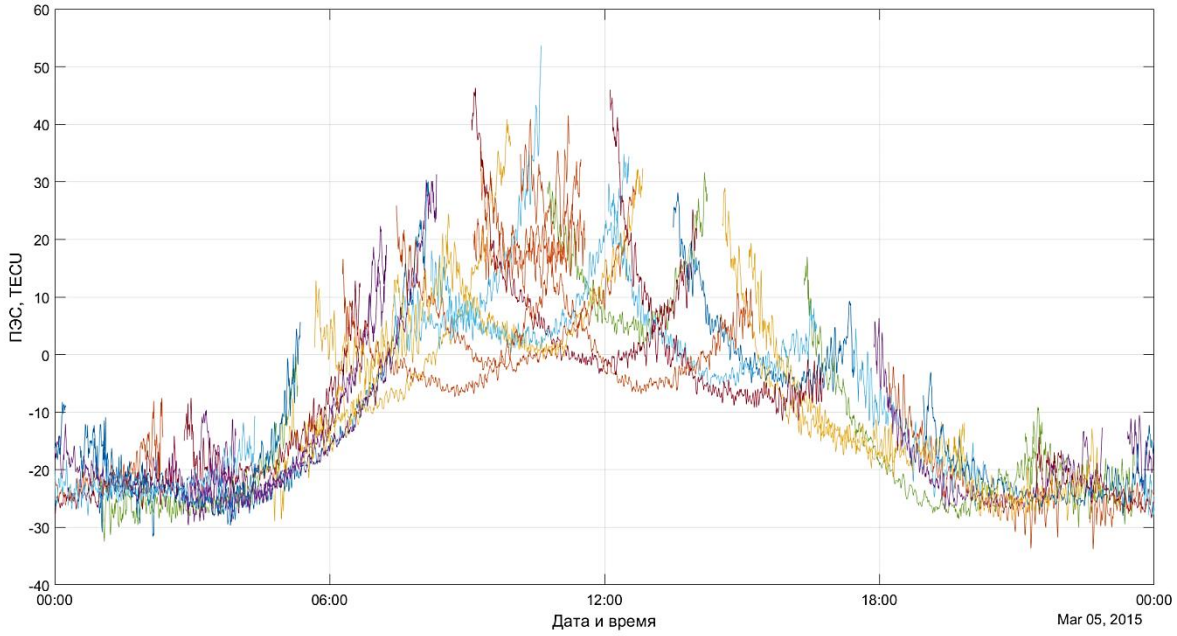


Рис. 2. ПЭС, рассчитанное по формуле (28), в течение дня наблюдений 05.03.2015 г. минской станции

Fig. 2. TEC calculated with formula (28) during the day of observation on 05.03.2015 on Minsk station

Шаг 4. Оценка абсолютного ПЭС с помощью комбинирования оценок ПЭС, полученных по фазовым и дальностным измерениям.

Как уже упоминалось ранее, оценка ПЭС по фазовым измерениям обеспечивает высокую точность, но является относительной, а оценка по дальностным измерениям позволяет получить абсолютное значение, но со значительной шумовой составляющей. Простым решением является подъем относительного ПЭС, посчитанного по фазовым измерениям, до среднего уровня абсолютного ПЭС, посчитанного по дальностным измерениям:

$$TEC_{abs} = TEC_\psi''' + \left\langle TEC_D - TEC_\psi''' \right\rangle. \quad (29)$$

[1]GNSS Differential Code Bias Product [Electronic resource]. – Mode of access: https://cddis.nasa.gov/Data_and_Derived_Products/GNSS/gnss_differential_code_bias_product.html. – Date of access: 14.10.2022.

Результат применения формулы (29) представлен на рис. 3.

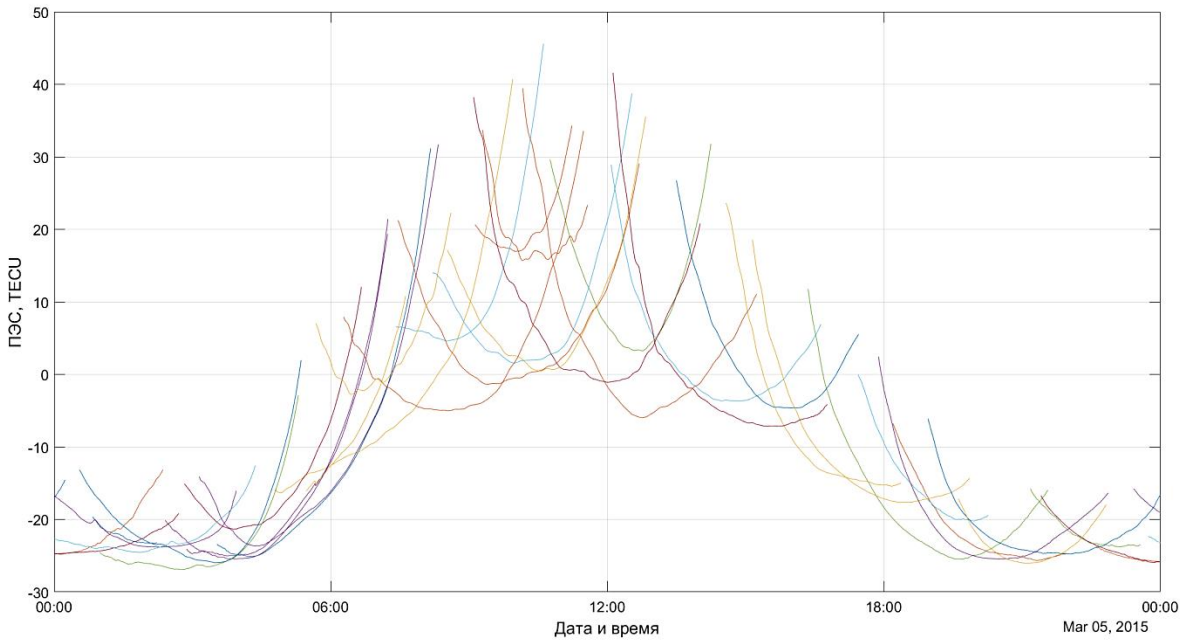


Рис. 3. Абсолютное ПЭС (без учета ДКЗ приемной станции) за день наблюдений 05.03.2015 г. минской станции

Fig. 3. Absolute TEC (excluding DCB of receiving station) for a day of observation on 05.03.2015 on Minsk station

Шаг 5. Уточнение оценки абсолютного ПЭС за счет оценки ДКЗ приемной станции.

Абсолютное ПЭС, рассчитанное по формулам (26), (28) и (29), все еще не является конечным и может быть даже отрицательным, как показано на рис. 3, так как не была принята в расчет ДКЗ приемной станции. Оценим ДКЗ приемной станции на основе метода наименьших квадратов. Для этого для каждой реализации (пролета спутника) минимизируется функция суммы квадратов отклонений вертикального ПЭС:

$$S = \sum_{n=1}^{N} \left(TEC_V^k - \overline{TEC_V^k} \right)^2, \tag{30}$$

где k – номер реализации, TEC_V^k – вертикальное ПЭС, вычисляемое по формуле

$$TEC_V^k = \left(TEC_{abs}^k - c' DCB_r^k \right) M(\varepsilon, h). \tag{31}$$

Здесь $c' = \frac{1}{A} \frac{f_1^2 f_2^2}{f_1^2 - f_2^2} c$, c – скорость света; $M(\varepsilon, h)$ – множитель для отображения наклонного ПЭС в вертикальное:

$$M(\varepsilon, h) = \cos\left(\arcsin\left(\frac{R_e}{R_e + h_{\max}} \cos(\varepsilon) \right) \right), \tag{32}$$

где ε – угол места; $h_{\max}$ – высота точки вхождения луча в ионосферу (однослойная модель ионосферы [15]); R_e – радиус Земли.

Следовательно, необходимо найти такое значение $c'DCB_r^k$, чтобы минимизировать выражение

$$S = \sum_{n=1}^{N}\left(TEC_V^k - \overline{TEC_V^k}\right)^2 \to \min . \tag{33}$$

Преобразуем функцию (30):

$$S = \sum_{n=1}^{N}\left(TEC_V^k - \overline{TEC_V^k}\right)^2 = \sum_{n=1}^{N}\left(\left(TEC_V^k\right)^2 + \left(\overline{TEC_V^k}\right)^2 - 2TEC_V^k\overline{TEC_V^k}\right). \tag{34}$$

Подставим уравнение (31) в (34), для удобства записи сделав замену $x = c'DCB_r^k$:

$$S = \sum_{n=1}^{N}\left(\left(\left[TEC_{abs}^k - x\right]M\right)^2 + \left(\overline{\left[TEC_{abs}^k - x\right]M}\right)^2 - 2\left[TEC_{abs}^k - x\right]M\overline{\left[TEC_{abs}^k - x\right]M}\right). \tag{35}$$

Раскрывая скобки и приводя подобные, можно получить выражения

$$S = \sum_{n=1}^{N}\left(\left[TEC_{abs}^k M - \overline{TEC_{abs}^k M}\right]^2 + 2\left[M - \overline{M}\right]\left[\overline{TEC_{abs}^k M} - TEC_{abs}^k M\right]x + \left[M - \overline{M}\right]^2 x^2\right); \tag{36}$$

$$S' = \sum_{n=1}^{N}\left(2\left[M - \overline{M}\right]\left[\overline{TEC_{abs}^k M} - TEC_{abs}^k M\right] + 2\left[M - \overline{M}^2\right]^2 x\right) = 0; \tag{37}$$

$$x = -\frac{\sum_{n=1}^{N}\left[M - \overline{M}\right]\left[\overline{TEC_{abs}^k M} - TEC_{abs}^k M\right]}{\sum_{n=1}^{N}\left[M - \overline{M}\right]^2}. \tag{38}$$

Так как $\left[M - \overline{M}\right]^2 > 0$, то x – точка минимума.

После получения значения $c'DCB_r^k$ проверяется, лежит ли результат в пределах от –75 до 75 TECU. Если не лежит, то реализация исключается из расмотрения. Значение ДКЗ приемной станции принимается равным среднему по всем рассмотренным реализациям. Полученный результат отнимается от значения абсолютного ПЭС (см. формулу (29)).

Отметим две особенности алгоритма. Во-первых, оценка проводится на основании ночных данных, так как в это время ПЭС изменяется относительно слабо. Во-вторых, минимизация целевой функции (33) проводится по данным вертикального, а не наклонного ПЭС для уменьшения влияния распространения сигналов под разными углами.

Итак, запишем полученный алгоритм:

Шаг 5.1. Для *k*-й реализации находится *M*, $\overline{M}$ $\overline{TEC_{abs}^k M}$.

Шаг 5.2. По формуле (38) находится значение ДКЗ в TECU $c'DCB_r$.

Шаг 5.3. Проверяется условие $-75 < c'DCB_r^k < 75$. Если оно не выполняется, то реализации исключается из рассмотрения.

Шаг 5.4. В качестве ДКЗ приемной станции выбирается среднее ДКЗ для всех реализаций:

$$DCB_r = \frac{1}{K}\sum_{k=1}^{K} c'DCB_r^k , \tag{39}$$

где K – количество рассмотренных реализаций.

Таким образом может быть получено окончательное значение абсолютного ПЭС:

$$TEC'_{abs} = TEC_{abs} - c'DCB_r \,. \tag{40}$$

Упрощенно алгоритм оценки абсолютного ПЭС можно представить в графическом виде (рис. 4).

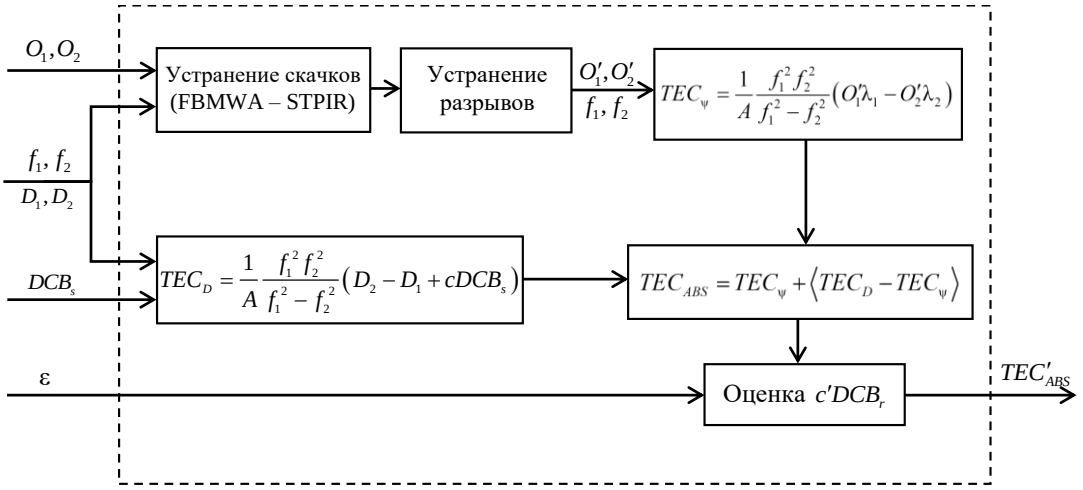


Рис. 4. Упрощенная схема алгоритма оценки абсолютного ПЭС
Fig. 4. Simplified diagram of the algorithm for estimating absolute TEC

Применение алгоритма. Разработанный алгоритм был использован для оценки ПЭС над минской и сокольской станциями, при этом для первой интервал наблюдения был равен $\Delta t = 1c,$ а для второй – $\Delta t = 15c.$ Так как ДКЗ приемных станций, вообще говоря, непостоянны, предлагается оценивать их величину для данных внутри некоторого окна. Центр окна находится в полночь, для которой оценивается ДКЗ, а размер составляет *d* дней. В результате будет сформирована последовательность ДКЗ для каждой ночи $\{DCB_s(k)\}$, где *k* – номер ночи. Найти значение ДКЗ для каждого момента времени *t* можно, использовав соседние ночи:

$$TEC'_{abs}(t) = TEC_{abs}(t) - \left(\frac{DCB_s(t_1)(t-t_2) + DCB_s(t_2)(t-t_1)}{t_2 - t_1}\right),\ t_1 \le t < t_2. \tag{41}$$

Полученные результаты можно перевести в вертикальное ПЭС и усреднить:

$$TEC_V^i(t) = TEC_{abs}^i{}'(t) M\left(\varepsilon^i(t), h\right); \tag{42}$$

$$\overline{TEC}_V(t) = \frac{\sum_{i=1}^{n} w_i(t) TEC_V^i(t)}{\sum_{i=1}^{n} w_i(t)}, \tag{43}$$

где *n* – количество видимых спутников в момент времени *t*; TEC_V^i – вертикальное ПЭС для *i*-го видимого спутника; w_i – коэффициент, определяемый равенством

$$w_i = \sin\left(\frac{\varepsilon_i - \varepsilon_0}{90^\circ - \varepsilon_0} 90^\circ\right). \tag{44}$$

Здесь ε_i – угол места спутника; ε_0 – минимальный рассматриваемый угол места.

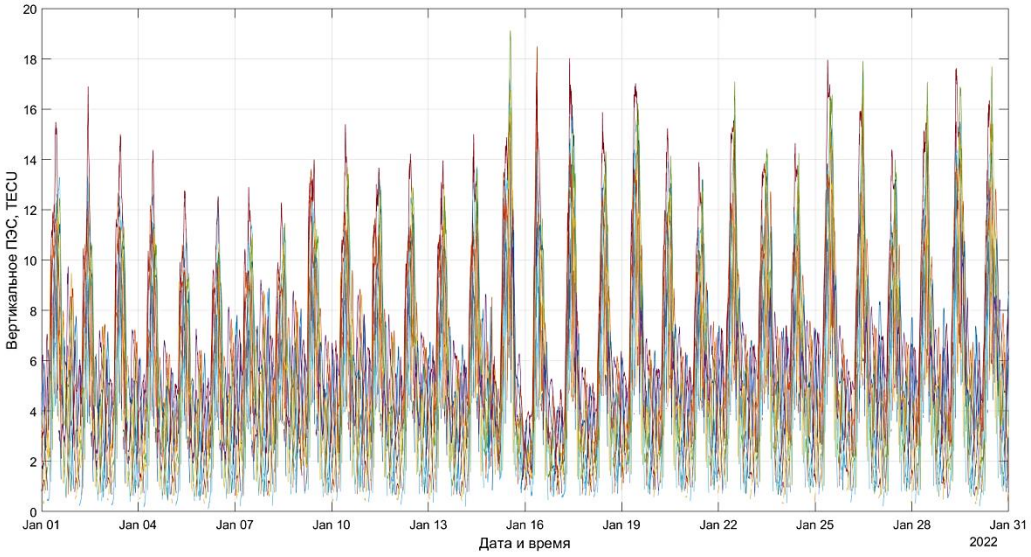
Проведем оценку ПЭС над минской и сокольской станциями при приведенных в табл. 1 значениях параметров.

Таблица 1
Значения параметров при оценке ПЭС

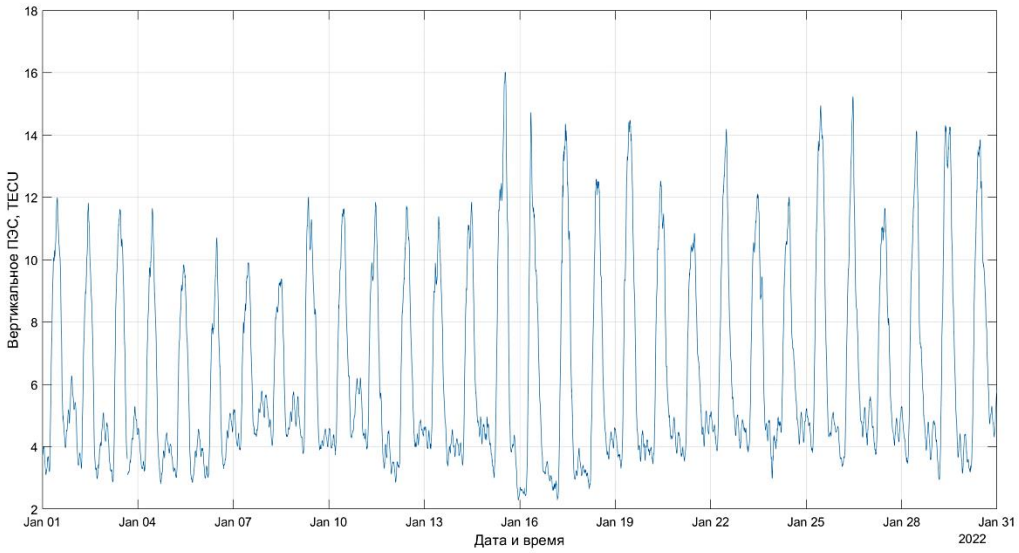
Table 1
Values of parameters when estimating TEC

Станция *Station*	*n*	*m*	*p*	*d*	ε_0
Минская	20	20	10	10	10°
Сокольская	3	20	5	10	10°

Полученные результаты для двух станций представлены на рис. 5 и 6.



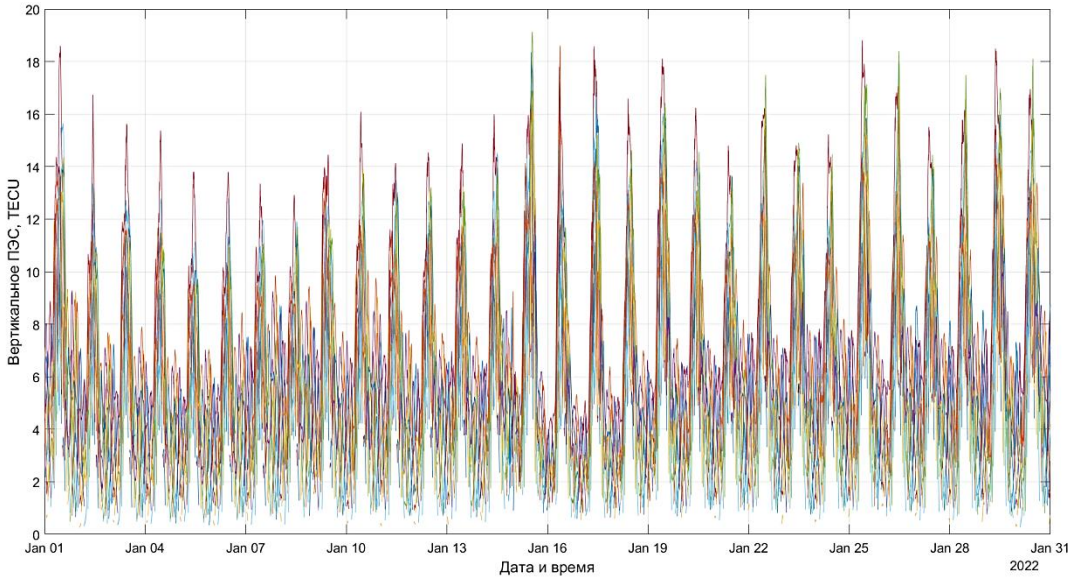
a)



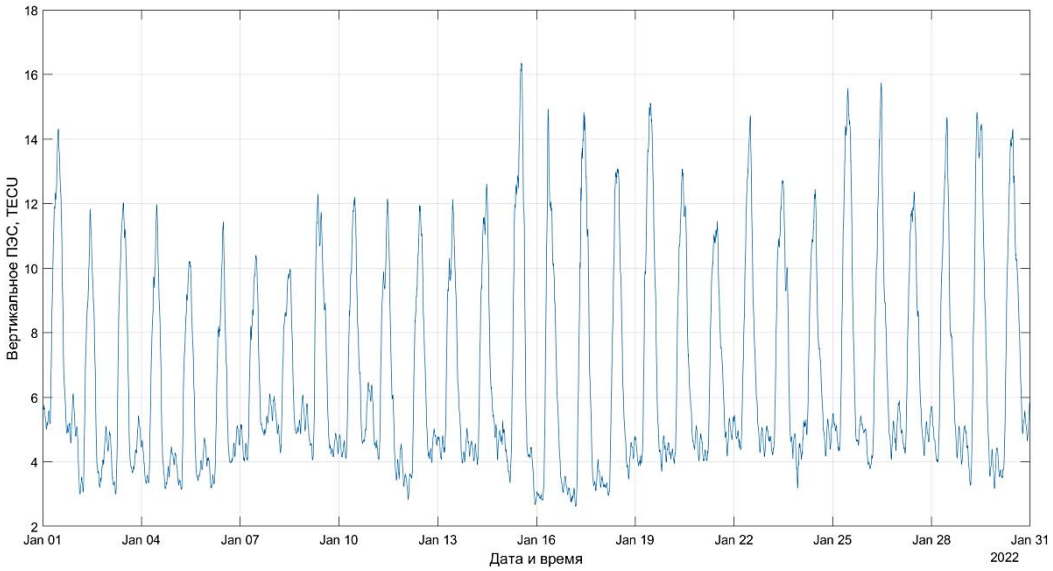
b)

Рис. 5. Результаты оценки ПЭС для минской станции за 1–30 января 2022 г.:
a) все реализации; *b)* взвешенное среднее

Fig. 5. Results of TEC estimation for the Minsk station for January 1–30, 2022:
a) all realizations; b) weighted average



a)



b)

Рис. 6. Результаты оценки ПЭС для сокольской станции за 1–30 января 2022 г.:
a) все реализации; *b)* взвешенное среднее

Fig. 6. Results of TEC estimation for the Sokol station for January 1–30, 2022:
a) all realizations; b) weighted average

Оценка ошибки. Так как оценка влияния скачков и разрывов достаточно сложна и выходит за рамки данной работы, допустим, что все присутствующие в сигналах скачки детектированы и оценены верно. В этом случае ошибка измерений будет такой же, как и при оценке ПЭС по фазовым измерениям. Измерения фазы в системе GPS производятся с высокой степенью точности, так что ошибка в определении ПЭС при 30-секундных интервалах усреднения не превышает 0,01 TECU [3, 6]. При этом существует также средняя ошибка оценки ПЭС, зависящая при указанных выше допущениях главным образом от оценки ДКЗ. Согласно работе [7] ДКЗ приемной станции изменяется с течением года. Считая ДКЗ постоянной в течение месяца, можно оценить стандартную ошибку при определении ДКЗ:

$$SE_{c'DCB} = \frac{\sigma}{\sqrt{n}}, \quad (45)$$

где σ – СКО ДКЗ для отдельного наблюдения спутника; n – среднее количество ночных наблюдений за временное окно в d дней.

Оцененные по месяцам за полгода наблюдений стандартные ошибки среднего для минской и сокольской станций и их усредненные значения за полгода приведены в табл. 2.

Таблица 2
Стандартные ошибки среднего оценки ПЭС

Table 2
Standard error of TEC estimation

Характеристика *Characteristic*	Станция *Station*	Январь *January*	Февраль *February*	Март *March*	Апрель *April*	Май *May*	Июнь *June*	Среднее *Average*
SE	Минская	0,58	0,75	0,75	1,13	1,45	0,90	0,93
	Сокольская	0,60	0,74	0,76	1,18	1,46	0,90	0,95
K	Минская	72,75	73,87	67	78,36	78,93	83,27	75,70
	Сокольская	72,39	70,63	66,33	78,36	78,60	82,60	74,82
σ	Минская	5,00	6,49	6,21	10,05	12,93	8,21	8,15
	Сокольская	5,15	6,28	6,21	10,44	13,01	8,21	8,22

Видно, что результаты для обеих станций близки. Стандартная ошибка среднего при оценке ПЭС составляет 0,94 TECU. При этом стоит учитывать, что ДКЗ меняется в течение года [7] и более высокая ошибка в апреле и мае может быть связана со значительным ее изменением. Тогда предположение о постоянстве ДКЗ в течение месяца является не совсем корректным.

Заключение. В работе показано, что ПЭС, рассчитанное по фазовым измерениям, обеспечивает высокую точность, но с точностью до неизвестной константы, а ПЭС, рассчитанное по дальностным измерениям, позволяет получить абсолютное значение, но с большой шумовой составляющей и дифференциальной кодовой задержкой аппаратуры спутника и приемника. Для решения этих проблем был разработан алгоритм оценки абсолютного ПЭС ионосферы, который может быть использован для одиночной приемной станции ГНСС. Дальнейшая работа может быть посвящена оценке точности, связанной с корректировкой фазовых измерений, улучшением точности алгоритма, адаптивным подбором параметров, тестированием алгоритма для работы с малыми космическими аппаратами (наноспутниками).

Список использованных источников

1. Дэвис, К. Радиоволны в ионосфере : пер. с англ. / К. Дэвис. – М. : Мир, 1973. – 504 с.

2. Ратклифф, Дж. А. Магнито-ионная теория и ее приложения к ионосфере : пер. с англ. / Дж. А. Ратклифф. – М. : Изд-во иностранной литературы, 1962. – 248 с.

3. Афраймович, Э. Л. GPS-мониторинг верхней атмосферы Земли / Э. Л. Афраймович, Н. П. Перевалова. – Иркутск : ГУ НЦ ВСНЦ СО РАН, 2006. – 480 с.

4. Куницын, В. Е. Радиотомография ионосферы / В. Е. Куницын, Е. Д. Терещенко, Е. С. Андреева. – М. : Физматлит, 2007. – 336 с.

5. Способ оценивания полного электронного содержания в ионосфере на основе ретрансляции сигналов глобальной навигационной спутниковой системы GPS / И. В. Белоконов [и др.] // Информатика. – 2023. – Т. 20, № 2. – С. 7–27. https://doi.org/10.37661/1816-0301-2023-20-2-7-27

6. Hofmann-Wellenhof B. Global Positioning System: Theory and Practice / B. Hofmann-Wellenhof, H. Lichtenegger, J. Collins. – N. Y. : Springer-Verlag Wien, 1992. – 327 p.

7. Variability of GPS/GLONASS differential code biases / A. A. Mylnikova [et al.] // Results in Physics. – 2015. – Vol. 5. – P. 9–10.

8. Kunitsyn, V. E. Ionospheric Tomography / V. E. Kunitsyn, E. D. Tereshenko. – Springer, 2003. – 272 p.

9. Atmospheric studies with the tri-band beacon instrument on the COSMIC constellation / P. Bernhardt [et al.] // Terrestrial, Atmospheric and Oceanic Sciences. – 2001. – Vol. 11, no. 1. – P. 291–312. https://doi.org/10.3319/TAO.2000.11.1.291(COSMIC)

10. Романов, А. А. Измерение полного электронного содержания ионосферы Земли с помощью многочастотного когерентного зондирующего сигнала / А. А. Романов, А. В. Новиков // Вопросы электромеханики. Тр. НПП ВНИИЭМ. – 2009. – Т. 111, № 4. – С. 31–36.

11. Ferreira, V. Study on cycle-slip detection and repair methods for a single dual-frequency global positioning system (GPS) / V. Ferreira, X. He, X. Tang // Boletim de Ciensicas. – 2014. – Vol. 20, no. 4. – P. 984–1004.

12. Cycle slip detection and repair for undifferenced GPS observation under high ionospheric activity / C. Cai [et al.] // GPS Solutions. – 2012. – Vol. 17, no. 2. – P. 247–260. https://doi.org/10.1007/s10291-012-0275-7

13. Blewitt, G. An automatic editing algorithm for GPS data / G. Blewitt // Geophysical Research Letters. – 1990. – Vol. 17, no. 3. – P. 199–202.

14. Goad, C. Precise positioning with the global positioning system / C. Goad // Proceedings of the Third Intern. Symp. on Inertial Technology for Surveying and Geodesy, Banff, 16–20 Sept. 1985. – Banff, 1985. – P. 745–756.

15. Ya'acob, N. Determination of GPS total electron content using single layer model (SLM) ionospheric mapping function / N. Ya'acob, M. Abdullah, M. Ismail // Intern. J. of Computer Science and Network Security. – 2008. – Vol. 8, no. 9. – P. 154–160.

References

1. Davies K. *Ionospheric Radio Waves*. Blaisdell Publishing Company, 1969, 460 p.

2. Ratcliffe J. A. *The Magneto-Ionic Theory and its Applications to the Ionosphere*. Cambridge, University Press, 1959, 226 p.

3. Afraimovich E. L., Perevalova N. P. GPS-monitoring verhnej atmosfery Zemli. *GPS Monitoring of the Earth's Upper Atmosphere.* Irkutsk, Gosudarstvennoe uchrezhdenie "Nauchnyj centr Vostochno-Sibirskogo nauchnogo centra Sibirskogo otdelenija Rossijskoj akademii nauk", 2006, 480 p. (In Russ.).

4. Kunitsyn V. E., Tereshchenko E. D., Andreeva E. S. Radiotomografija ionosfery. *Radio Tomography of the Ionosphere*. Moscow, Fizmatlit, 2007, 336 p. (In Russ.).

5. Belokonov I. V., Krot A. M., Kozlov S. V., Kapliarchuk Y. A., Savinykh I. E., Shapkin A. S. *A method for estimating the total electron content in the ionosphere based on the retransmission of signals from the global navigation satellite system GPS*. Informatika [*Informatics*], 2023, vol. 20, no. 2, pp. 7–27 (In Russ.). https://doi.org/10.37661/1816-0301-2023-20-2-7-27.

6. Hofmann-Wellenhof B., Lichtenegger H., Collins J. *Global Positioning System: Theory and Practice*. New York, Springer-Verlag Wien, 1992, 327 p.

7. Mylnikova A. A., Yasyukevich Yu. V., Kunitsyn V. E., Padokhin A. M. Variability of GPS/GLONASS differential code biases. *Results in Physics*, 2015, vol. 5, pp. 9–10.

8. Kunitsyn V. E., Tereshenko E. D. *Ionospheric Tomography*. Springer, 2003, 272 p.

9. Bernhardt P., Selcher C., Basu S., Bust G., Reising S. Atmospheric studies with the tri-band beacon instrument on the COSMIC constellation. *Terrestrial, Atmospheric and Oceanic Sciences*, 2001, vol. 11, no. 1, pp. 291–312. https://doi.org/10.3319/TAO.2000.11.1.291(COSMIC)

10. Romanov A. A., Novikov A. V. *Measurement of the total electron content of the Earth's ionosphere using a multi-frequency coherent sounding signal*. Voprosy jelektromehaniki. Trudy Nauchno-proizvodstvennogo predprijatija Vserossijskogo nauchno-issledovatel'skogo instituta jelektromehaniki [*Questions of Electromechanics. Proceedings of the Research and Production Enterprise of the All-Russian Research Institute of Electromechanics*], 2009, vol. 111, no. 4, pp. 31–36 (In Russ.).

11. Ferreira V., He X., Tang X. Study on cycle-slip detection and repair methods for a single dual-frequency global positioning system (GPS). *Boletim de Ciensicas*, 2014, vol. 20, no. 4, pp. 984–1004.

12. Cai C., Liu Z., Xia P. Dai W. Cycle slip detection and repair for undifferenced GPS observation under high ionospheric activity. *GPS Solutions*, 2012, vol. 17, no. 2, pp. 247–260. https://doi.org/10.1007/s10291-012-0275-7

13. Blewitt G. An automatic editing algorithm for GPS data. *Geophysical Research Letters*, 1990, vol. 17, no. 3, pp. 199–202.

14. Goad C. Precise positioning with the global positioning system. *Proceedings of the Third International Symposium on Inertial Technology for Surveying and Geodesy, Banff, 16–20 September 1985*. Banff, 1985, pp. 745–756.

15. Ya'acob N., Abdullah M., Ismail M. Determination of GPS total electron content using single layer model (SLM) ionospheric mapping function. *International Journal of Computer Science and Network Security*, 2008, vol. 8, no. 9, pp. 154–160.

Информация об авторе

Шапкин Александр Сергеевич, аспирант, лаборатория моделирования самоорганизующихся систем, Объединенный институт проблем информатики Национальной академии наук Беларуси.
E-mail: shap1kin2@gmail.com,
al_shapkin@newman.bas-net.by
https://orcid.org/0009-0009-4947-7313
https://www.researchgate.net/profile/Aliaksandr-Shapkin

Information about the author

Aliaksandr S. Shapkin, Postgraduate Student, Laboratory of Self-organization System Modeling, The United Institute of Informatics Problems of the National Academy of Sciences of Belarus.
E-mail: shap1kin2@gmail.com,
al_shapkin@newman.bas-net.by
https://orcid.org/0009-0009-4947-7313
https://www.researchgate.net/profile/Aliaksandr-Shapkin

БИОИНФОРМАТИКА
BIOINFORMATICS

УДК 519.23
https://doi.org/10.37661/1816-0301-2024-21-1-65-82

Оригинальная статья
Original Paper

Прогнозирование и принятие решений на основе модели нелинейных рисков при лечении рака желудка

О. В. Красько[1]✉, М. Ю. Ревтович[2], А. В. Иванов[3]

[1]*Объединенный институт проблем информатики Национальной академии наук Беларуси, ул. Сурганова, 6, Минск, 220012, Беларусь*
✉*E-mail: krasko@newman.bas-net.by*

[2]*Учреждение образования «Белорусский государственный медицинский университет», пр. Дзержинского, 83, Минск, 220083, Беларусь*
E-mail: mihail_revtovich@yahoo.com

[3]*Республиканский научно-практический центр им. Н. Н. Александрова, аг. Лесной, Минский район, 223040, Беларусь*
E-mail: tennis5000@rambler.ru

Аннотация

Цели. В исследовании ставятся цели разработать модель нелинейных рисков развития неблагоприятных исходов и оценить ее пригодность для прогнозирования в клинической практике.

Методы. Использованы методы анализа выживаемости, регрессионные статистические модели.

Результаты. Предложен практический подход к оценке нелинейных рисков развития неблагоприятных событий на примере лечения рака желудка. Предложена и исследована модель прогнозирования метахронной перитонеальной диссеминации у радикально оперированных по поводу рака желудка пациентов. Оценены риски в различные периоды наблюдения и клиническая пригодность разработанного подхода.

Заключение. В клинической онкологической практике большую роль играет не только своевременное лечение, но и предупреждение неблагоприятных исходов после окончания лечения. Индивидуализация наблюдения за пациентом после лечения снижает риски фатальных исходов, затраты на дополнительные исследования и лечение в случае прогрессирования онкозаболевания. По результатам данного исследования предлагаются решения, которые должны привести к более эффективной и качественной тактике наблюдения после проведенного лечения рака желудка, к выбору оптимальных подходов и получению клинически благоприятных исходов заболевания. Предложенный метод прогноза рисков в конечном итоге приведет к индивидуализированному ведению пациента на основании его данных.

Ключевые слова: нелинейные риски, модель Файна – Грея, рак желудка, перитонеальная диссеминация, прогнозирование

Для цитирования. Красько, О. В. Прогнозирование и принятие решений на основе модели нелинейных рисков при лечении рака желудка / О. В. Красько, М. Ю. Ревтович, А. В. Иванов // Информатика. – 2024. − Т. 21, № 1. – С. 65–82. https://doi.org/10.37661/1816-0301-2024-21-1-65-82

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Поступила в редакцию | Received 08.11.2023
Подписана в печать | Accepted 29.12.2023
Опубликована | Published 29.03.2024

Prediction and decision-making based on nonlinear risks model in stomach cancer treatment

Olga V. Krasko[1]✉, Mikhail Yu. Reutovich[2], Andrey V. Ivanov[3]

[1]The United Institute of Informatics Problems of the National Academy of Sciences of Belarus, st. Surganova, 6, Minsk, 220012, Belarus
✉E-mail: krasko@newman.bas-net.by

[2]Belarusian State Medical University, av. Dzerzhinsky, 83, Minsk, 220083, Belarus
E-mail: mihail_revtovich@yahoo.com

[3]N. N. Alexandrov National Cancer Center of Belarus, Lesnoy, Minsk Region, 223040, Belarus
E-mail: tennis5000@rambler.ru

Abstract

Objectives. The goals are to develop a nonlinear risk model and examine its prediction applicability for clinical use.

Methods. Methods of survival analysis and regression statistical models were used.

Results. A practical approach to assessing nonlinear risks of adverse events using the example of gastric cancer treatment is proposed. A model for predicting the metachronous peritoneal dissemination in patients undergoing radical surgery for gastric cancer was proposed and studied. Assessment of risks for various periods of observation was performed, and the clinical suitability of developed approach was assessed.

Conclusion. In clinical oncological practice, not only timely treatment plays an important role, but also the prevention of adverse outcomes after treatment. Individualization of patient monitoring after treatment reduces the risks of fatal outcomes and the costs of additional research and treatment in the event of cancer progression. Based on the results of this study, we propose solutions that should lead to more effective and high-quality treatment tactics and follow-up after treatment for gastric cancer, also to the selection of optimal approaches and to obtaining clinically favorable outcomes of the disease. The proposed risk prediction method will ultimately lead to individualized patient management based on the results of personal data.

Keywords: nonlinear hazard, Fine – Grey model, gastric cancer, peritoneal dissemination, predict

For citation. Krasko O. V., Reutovich M. Yu., Ivanov A. V. *Prediction and decision-making based on nonlinear risks model in stomach cancer treatment*. Informatika [*Informatics*], 2024, vol. 21, no. 1, pp. 65−82 (In Russ.). https://doi.org/10.37661/1816-0301-2024-21-1-65-82

Conflict of interest. The authors declare of no conflict of interest.

Введение. Анализ выживаемости является основным подходом к анализу данных в клинических онкологических исследованиях и многих других биомедицинских исследованиях [1–4]. Анализ выживаемости, или в более общем смысле анализ времени до события, относится к набору методов анализа промежутка времени до наступления четко определенной конечной точки интереса. Уникальной особенностью данных о выживаемости является то, что обычно не у всех пациентов происходит событие (например, смерть) к концу периода наблюдения, поэтому фактическое время выживания для некоторых пациентов неизвестно. Это явление, называемое цензурой, необходимо учитывать в анализе, чтобы можно было сделать обоснованные выводы.

Во многих приложениях анализа выживаемости интерес сосредотачивается на том, как ковариаты могут повлиять на результат. В клинических исследованиях корректировка эффектов лечения с учетом эффектов других объясняющих переменных может иметь решающее значение, если рандомизированные группы несбалансированы по отношению к важным прогностическим факторам, а в эпидемиологических когортных исследованиях достоверные эффекты воздействия могут быть получены только в том случае, если сделана некоторая корректировка для вмешивающихся переменных. В таких ситуациях полезна регрессионная модель, а наиболее важной моделью для данных о выживаемости является модель регрессии пропорциональных рисков Кокса. Модель предложена в 1972 г. и занимает одно из центральных мест в анализе выживаемости в клинических исследованиях [5, 6]. Она используется для изучения предикторов как благоприятных, так и неблагоприятных исходов во многих областях медицины (онкологии, кардиологии и др.), для прогноза и принятия решений при проведении определенного лечения. Однако на сегодняшний момент только лишь прогнозирование не способствует индивидуализации лечения. Необходимо рассматривать большой пласт вопросов, связанных с принятием решений на основании сделанного прогноза.

Настоящая статья демонстрирует подход от разработки модели до ее клинического использования на примере принятия решения о необходимости и сроках лапароскопии (диагностической операции, которая выполняется для оценки наличия остаточной опухоли у пациентов без клинических проявлений заболевания после ранее проведенного противоопухолевого лечения) в дополнение к стандартному объему обследования радикально оперированных онкологических пациентов.

1. Регрессионные модели анализа времени до события. Опишем математическую нотацию полупараметрической регрессионной модели Кокса.

Пусть вектор $X_i = \left(X_{i1}, \ldots, X_{ip}\right)$ – значения наблюдений p различных ковариат для субъекта i, $i = 1, ..., N$, где N – размер выборки.

Модель Кокса базируется на ключевой концепции оценки функции риска $h(t)$, которая описывает интенсивность наступления событий. Тогда $h(t)dt$ – моментальный риск, вероятность наступления события на интервале $(t + \Delta t)$ при условии, что событие не наступило до момента времени t. Кокс предположил, что существует функция базового уровня $h_0(t)$, которая представляет степень риска для индивидуума, не имеющего воздействия факторов риска или в некоторых случаях имеющего стандартный набор факторов риска, тогда $\log\left(h(t)\right) = \log\left(h_0(t)\right) + \left(\beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_p X_{ip}\right)$. Это ведет к описанию функции риска модели Кокса для субъекта i в виде равенства

$$h\left(t \mid X_i\right) = h_0(t) \exp\left(\beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_p X_{ip}\right) = h_0(t) HR\left(X_i\right) \quad (1)$$

и дает функцию риска в момент времени t для субъекта i, который имеет вектор ковариат X_i. Здесь $h_0(t)$ – базовая функция риска для изучаемой популяции, не зависящая от субъекта i

и изменяющаяся только под воздействием конкретной реализации его вектора $X_i = \left(X_{i1}, \ldots, X_{ip}\right)$;

$h\left(t \mid X_i\right)$ – оценка опасности конкретного субъекта i на протяжении всего времени изучения t;

$HR\left(X_i\right) = \exp\left(\beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_p X_{ip}\right)$ – модификатор базовой функции риска, не зависящий от времени наблюдения.

Кокс показал [5], что для представленной модели соотношение опасностей двух субъектов $i, i = 1, ..., N,$ и $j, j = 1, ..., N,$ является постоянным, т. е. соотношение пропорционально и постоянно, зависит только от значений ковариат субъектов i, j и не зависит от времени:

$$\frac{h\left(t \mid X_i\right)}{h\left(t \mid X_j\right)} = \frac{h_0\left(t\right) HR\left(X_i\right)}{h_0\left(t\right) HR\left(X_j\right)} = \frac{HR\left(X_i\right)}{HR\left(X_j\right)} = const.$$

Выполнение этого предположения является одним из основных условий допустимого применения линейной регрессионной модели пропорциональных рисков Кокса. При нарушении предположения могут быть получены прогнозные значения, далекие от реальной ситуации в когорте, для которой построена модель. При нарушении предположения о пропорциональных рисках существует несколько подходов [7–11], которые различными способами определяют поведение $HR\left(X_i, t\right)$ во времени. Два основных подхода – это моделирование ковариат, зависящих от времени, которое описывается формулой

$$h\left(t \mid X_i\right) = h_0\left(t\right) \exp\left(\boldsymbol{\beta} \times g\left(X_i, t\right)\right), \tag{2}$$

и моделирование коэффициентов регрессии, изменяющихся во времени, которое описывается формулой

$$h\left(t \mid X_i\right) = h_0\left(t\right) \exp\left(g\left(\boldsymbol{\beta}, t\right) \times X_i\right). \tag{3}$$

Первый подход более распространен при ковариатах, которые представлены в количественной шкале, а также ковариатах, которые могу меняться в течение периода наблюдения. Примером таких ковариат являются повторные курсы химиотерапии, повторное хирургическое лечение, повторное измерение некоторого биомаркера и т. п.

В клинических исследованиях большинство ковариат, как правило, измеряются в момент старта периода наблюдения и носят качественный характер. Это, например, общепризнанная система стадирования TNM, где каждая из категорий отражает характеристики опухолевого процесса: T – размеры и распространенность первичной опухоли, N – состояние регионарных лимфоколлекторов, M – наличие или отсутствие отдаленных метастазов. Поэтому последствия, за которые отвечает тот или иной предиктор в момент старта наблюдения, могут быть описаны относительными рисками, меняющимися от года к году, т. е. можно рассматривать регрессионные модели с коэффициентами, изменяющимися во времени. Коэффициенты могут быть кусочно-постоянными:

$$h\left(t \mid X_i\right) = h_0\left(t\right) \exp\left(\begin{array}{l} \sum\limits_j \beta_{1j} I\left(t_j < t < t_{j+1}\right) X_{i1} + \sum\limits_j \beta_{2j} I\left(t_j < t < t_{j+1}\right) X_{i2} + \ldots \\ + \sum\limits_j \beta_{pj} I\left(t_j < t < t_{j+1}\right) X_{i1} X_{ip} \end{array}\right), \tag{4}$$

где j – временные интервалы, на которые разбит весь период наблюдения, $j = 1, \ldots, J$; $I\left(.\right)$ – индикаторная функция.

Основной вопрос, на который отвечают модели времени до события, – это вероятность развития неблагоприятного исхода к определенному моменту времени после старта наблюдения.

В условиях нескольких вариантов неблагоприятных событий функция опасности, связанная с конкретным событием, не имеет прямой интерпретации с точки зрения вероятности выживания для пациента. В этом случае рассматривается оценка кумулятивной инцидентности, связанная с каким-либо событием, поскольку оценка выживаемости в целом определяется несколькими событиями сразу. Для изучения конкретного типа события предлагается использовать модель Файна – Грея [12], которая в некотором смысле является аналогом модели Кокса и дает равномерно непротиворечивую оценку прогнозируемой кумулятивной инцидентности конкретного события для пациента с определенными ковариатами.

Если для модели Кокса функция риска в момент времени t определялась как $h(t) = \lim_{\Delta t \to 0} \frac{\Pr(t \le T < t + \Delta t \mid T \ge t)}{\Delta t}$, то для модели Файна – Грея необходимо учитывать тип события. Пусть имеется несколько вариантов событий $k,\ k = 1,...K,$ тогда для k-го события функция риска имеет вид

$$h_k(t) = \lim_{\Delta t \to 0} \frac{\Pr(t \le T < t + \Delta t, E = k \mid T > t \cup (T < t \cap E \ne k))}{\Delta t}$$

и описывает риск наступления k-го события в момент времени t у субъектов, у которых еще не было ранее события типа k до момента времени t.

Все предположения, связанные с нарушением пропорциональных рисков, сохраняются для модели Файна – Грея [13].

Ниже будет изложен подход, который позволит определить кумулятивную инцидентность наступления неблагоприятного события с учетом изменения во времени коэффициентов ковариат в модели, а также на основе прогноза составить график проведения диагностических мероприятий для своевременного определения развития неблагоприятного события.

Поскольку цель, поставленная в исследовании, достаточно специфична для клинической аналитической эпидемиологии, рассматриваемый подход будет изложен на примере радикального лечения рака желудка.

2. Описательная постановка задачи исследования. У пациентов, радикально оперированных по поводу местнораспространенного рака желудка, имеет место высокий риск развития метахронной перитонеальной диссеминации (МПД), несмотря на проведенное противоопухолевое лечение (радикальную операцию в сочетании с адъювантной полихимиотерапией) [14–16], в том числе и при использовании различных вариантов интраперитонеальной химиотерапии [17]. При развитии прогрессирования опухолевого процесса брюшина может являться не единственным местом локализации отдаленных метастазов, поскольку не исключается лимфогематогенное метастазирование в различные органы с развитием отдаленных лимфогематогенных метастазов различной локализации. Кроме различных вариантов прогрессирования опухолевого процесса (МПД, отдаленные лимфогематогенные метастазы или их сочетание), у пациентов могут наблюдаться и другие неблагоприятные события (развитие метахронной опухоли; смерть по причинам, не связанным с основным заболеванием). Принимая во внимание сложность диагностики МПД с использованием методов неинвазивной интраскопической диагностики [18–20], целесообразно выделить когорту пациентов с высоким риском развития МПД для выполнения так называемых лапароскопий second-look (диагностических операций для оценки наличия остаточной опухоли у пациентов без клинических проявлений заболевания после ранее проведенного противоопухолевого лечения), предполагающих визуальную оценку брюшины с морфологическим исследованием ее биоптатов и позволяющих диагностировать МПД на начальных этапах ее развития.

В ряде исследований обращается внимание на отсутствие в настоящее время объективных показаний для лапароскопии с целью ранней диагностики МПД у радикально оперированных по поводу рака желудка пациентов [21]. В связи с этим более рациональным представляется прогнозирование МПД на основе применения математических моделей. Иначе говоря, после проведенного основного и дополнительного лечения необходимо понять, какие из пациентов нуждаются в персонализированном контроле и в какой период им лучше проводить диагностическую операцию по поводу подозрения на МПД. Таким образом, необходимо определить, кому и когда (в какие сроки наблюдения) проводить диагностическую операцию.

3. Описание исследуемой когорты и основные события. Для решения поставленной задачи были использованы данные радикально оперированных пациентов, которым был проведен различный объем противоопухолевого лекарственного лечения. Подробно виды лечения, прогнозы и иные результаты представлены в работах авторов [17, 22–26]. Необходимо отметить, что из настоящего исследования исключались пациенты с неполным курсом любого из использованных вариантов химиотерапии (системной или интраперитонеальной), т. е. либо химиотерапия была проведена в запланированном объеме, либо она вообще не проводилась. Всего в данном исследовании использовались данные 1311 пациентов, которым в период 2008–2021 гг. было проведено радикальное комбинированное или комплексное лечение по поводу местнораспространенного рака желудка.

Основными предикторами развития МПД являлись: дескрипторы распространенности опухолевого процесса T и N, степень дифференцировки аденокарциномы, форма роста опухоли, вариант проведенного хирургического вмешательства, три вида дополнительного химиотерапевтического лечения (ХТ-1 (адъювантная системная полихимиотерапия), ХТ-2 и ХТ-3 – различные варианты интраперитонеальной химиотерапии).

Для понимания общей картины наступления неблагоприятных событий МПД рассмотрим интенсивность наступления событий во времени в исследуемой когорте, которая показана на рис. 1 в зависимости от наличия или отсутствия любого дополнительного химиотерапевтического лечения.

Интенсивность рассчитывалась как $IR_i = \frac{n_i}{FU_i} \cdot 1000$, где n_i – число событий (пациентов в выявленной диссеминации за период времени (квартал i), FU_i – число человеко-дней наблюдения в соответствующем квартале i, $i = 1, \ldots, 20$ (первые пять лет наблюдения после лечения).

На рис. 1 видно, что у пациентов, получавших дополнительное химиотерапевтическое лечение, наибольшая интенсивность МПД приходится на второй и третий годы наблюдения, в то время как у пациентов, получавших только хирургическое лечение, пик неблагоприятных событий приходится на первый год наблюдения. После трех лет наблюдения интенсивности событий становятся сопоставимыми.

Также можно показать, что, например, пациенты с дескриптором опухоли N3 (массивное метастатическое поражение регионарных лимфоколлекторов) имеют наибольшую интенсивность в первый год наблюдения, сохраняя риск неблагоприятного события достаточно высоким и на третий-четвертый годы наблюдения. Аналогичные зависимости можно продемонстрировать еще по некоторым предикторам риска развития МПД.

Таким образом, необходимо построить модель прогноза неблагоприятного события (МПД) с учетом неоднородности интенсивности событий в первые пять лет наблюдений за пациентами после проведенного лечения, а также с учетом других конкурирующих событий (метахронная опухоль, отдаленные метастазы, смерть от причины, не связанной с основным заболеванием), которые могут происходить в течение периода наблюдения, и определить, каким именно пациентам и в какие сроки проводить дополнительные диагностические исследования.

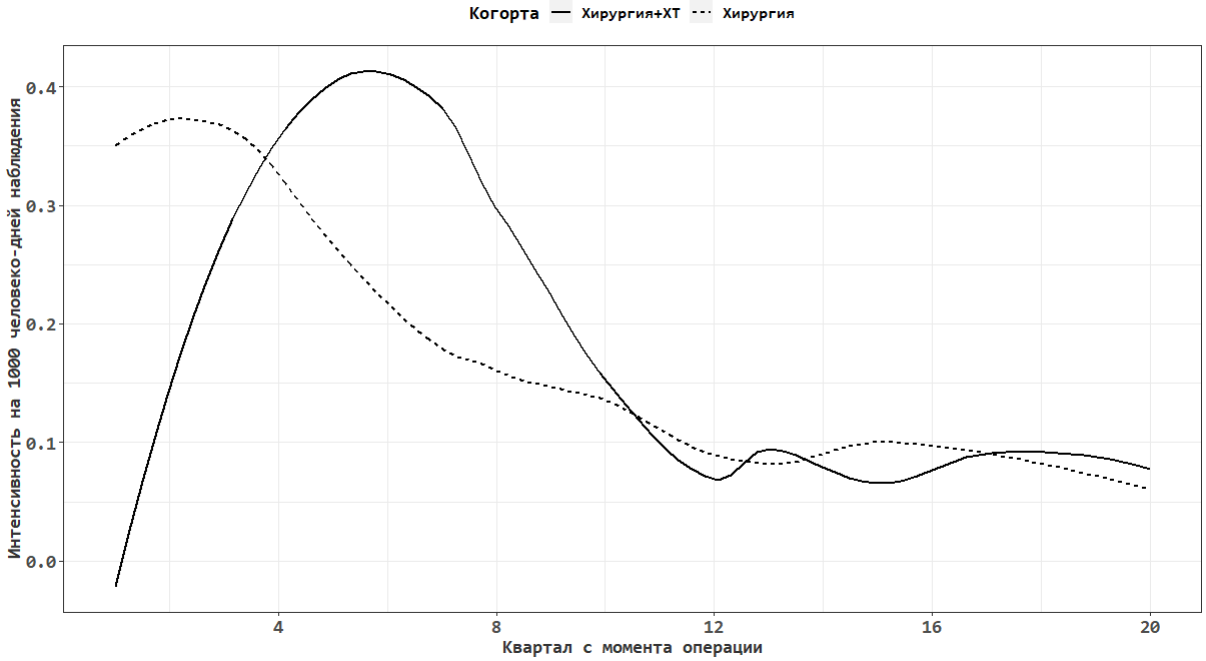


Рис. 1. Интенсивность событий в исследуемой когорте в зависимости от времени и лечения

Fig. 1. Event rates in the study cohort as a function of time and treatment

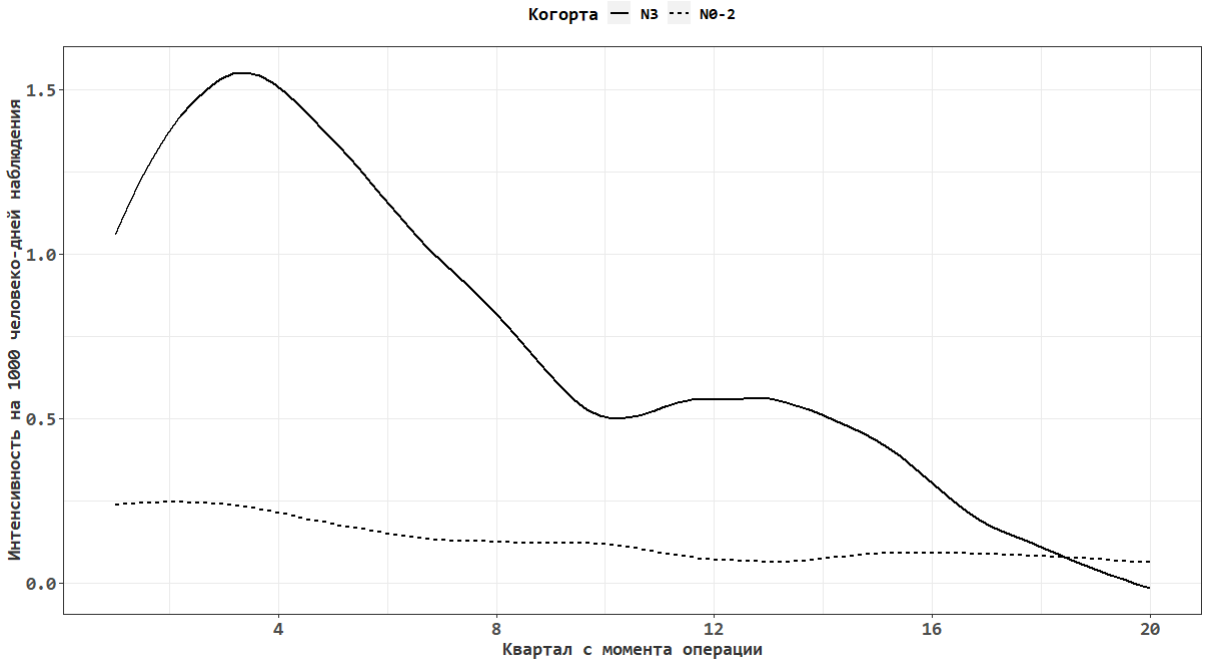


Рис. 2. Интенсивность событий в когорте без дополнительной химиотерапии в зависимости от времени и N-дескриптора опухолевого процесса

Fig. 2. Event rates in the study cohort without additional chemotherapy as a function of time and N-descriptor of tumor process

4. Решение задачи исследования. В исследовании были учтены факторы, представленные в табл. 1.

Таблица 1
Предварительно исследованные предикторы
Table 1
Previously studied predictors

Фактор *Factor*	Описание *Description*	Уровни фактора *Factor levels*
Возраст	Возраст пациента на момент операции	До 56 лет, 56–66 лет, более 65 лет
Форма роста опухоли	Макроскопическая форма роста первичной опухоли, устанавливается при интраоперационной ревизии и (или) во время послеоперационного исследования удаленного макропрепарата желудка	Инфильтративная экзофитная
pT-дескриптор опухолевого процесса	Глубина инвазии первичной опухолью стенки желудка	pT1 pT2 pT3 pT4
pN-дескриптор опухолевого процесса	Степень метастатического поражения регионарных лимфоузлов	pN0 pN1 pN2 pN3
Степень дифференцировки аденокарциномы	Гистологический вариант аденокарциномы, устанавливается при послеоперационном морфологическом исследовании	Некогезивная, high grade (GIII), когезивная, low grade (GI-II)
Операция	Объем операции, определяется в процессе ее проведения	Стандартная комбинированная
Проведение ХТ-1 после хирургического лечения	Адьювантная полихимиотерапия	Да, нет
Проведение ХТ-2 после хирургического лечения	Нормотермическая интраперитонеальная химиотерапия	Да, нет
Проведение ХТ-3 после хирургического лечения	Внутрибрюшная перфузионная интраоперационная термохимиотерапия	Да, нет

После подгонки модели Файна – Грея с конкурирующими рисками без учета изменения рисков во времени было выявлено, что такие факторы, как N-дескриптор опухоли, возраст более 65 лет, форма роста опухоли, степень дифференцировки аденокарциномы, а также ХТ-1, нарушают предположения о пропорциональности рисков [27]. Поэтому была проведена стратификация по временным интервалам наблюдения. Согласно стандартам обследования и лечения рака желудка в Республике Беларусь кратность обследования после проведения радикального лечения определяется следующим образом: первый год – один раз в три месяца, второй год – один раз в шесть месяцев, в последующем пожизненно – один раз в год [28]. Для решения поставленной задачи данные были стратифицированы по годам наблюдения (первый год, вто-

рой год, третий-пятый года) и на каждой страте были подогнаны коэффициенты модели [12], которые менялись в каждом периоде наблюдения для некоторых предикторов с учетом конкурирующих рисков [6–8, 12, 13].

Повторная подгонка модели с коэффициентами регрессии, которые изменяются по временным стратам (кусочно-постоянным) для вышеуказанных предикторов, была также проверена на соответствие предположению о пропорциональных рисках, и результат был удовлетворительным. Результаты подгонки модели приведены в табл. 2. Индекс конкродации составил 0,798 (SE = 0,011).

Таблица 2
Оценка отношения рисков по модели Файна – Грея с кусочно-постоянными коэффициентами

Table 2
Estimation of hazard ratio based on the Fine – Gray model with piecewise constant coefficients

Предиктор *Predictor*	Временная страта *Temporary stratum*	Отношение рисков *Risk ratio*	95 % CI	p-value
Возраст 56–65 vs менее 56	Весь период	0,83	0,62 – 1,10	0,2
Возраст 66+ vs менее 56	1-й год	0,92	0,64 – 1,34	0,7
Возраст 66+ vs менее 56	2-й год	0,53	0,32 – 0,87	0,012
Возраст 66+ vs менее 56	3-5-й года	0,59	0,37 – 0,92	0,021
pT2 vs pT1	Весь период	3,82	1,26 – 11,6	0,018
pT3 vs pT1	Весь период	11,7	4,03 – 33,8	<0,001
pT4 vs pT1	Весь период	18,3	6,44 – 52,1	<0,001
Операция комбинированная vs стандартная	Весь период	1,54	1,22 – 1,94	<0,001
pN1 vs pN0	1-й год	2,44	1,41 – 4,24	0,001
pN1 vs pN0	2-й год	1,10	0,52 – 2,34	0,8
pN1 vs pN0	3-5-й года	1,29	0,78 – 2,14	0,3
pN2 vs pN0	1-й год	2,27	1,27 – 4,05	0,006
pN2 vs pN0	2-й год	1,95	1,01– 3,78	0,048
pN2 vs pN0	3-5-й года	1,16	0,66 – 2,05	0,6
pN3 vs pN0	1-й год	3,96	2,34 – 6,68	<0,001
pN3 vs pN0	2-й год	2,86	1,62 – 5,03	<0,001
pN3 vs pN0	3-5-й года	0,95	0,52 – 1,75	0,9
Инфильтративная vs экзофитная	1-й год	2,65	1,40 – 5,05	0,003
Инфильтративная vs экзофитная	2-й год	6,76	2,11 – 21,7	0,001
Инфильтративная vs экзофитная	3-5-й года	1,34	0,83 – 2,17	0,2

Окончание табл. 2
End of table 2

Предиктор *Predictor*	Временная страта *Temporary stratum*	Отношение рисков *Risk ratio*	95 % CI	p-value
Степень дифференцировки аденокарциномы: некогезивная (high grade, GIII) vs когезивная (low grade, GI-II)	1-й год	0,97	0,66 –1,42	0,9
Степень дифференцировки аденокарциномы: некогезивная (high grade, GIII) vs когезивная (low grade, GI-II)	2-й год	1,93	1,09 – 3,42	0,024
Степень дифференцировки аденокарциномы: некогезивная (high grade, GIII) vs когезивная (low grade, GI-II)	3-5-й года	1,48	0,93 – 2,36	0,10
Проведение ХТ-2 после хирургии	Весь период	0,20	0,09 – 0,45	<0,001
Проведение ХТ-3 после хирургии	Весь период	0,30	0,18 – 0,50	<0,001
Проведение ХТ-1 после хирургии	1-й год	0,21	0,08 – 0,52	<0,001
Проведение ХТ-1 после хирургии	2-й год	0,67	0,33 – 1,34	0,3
Проведение ХТ-1 после хирургии	3-5-й года	0,74	0,32 – 1,71	0,5

Как следует из табл. 2, в возрасте более 65 лет риск развития диссеминации снижается с каждым годом наблюдения, риск при дескрипторе pN1 выше только в первый год наблюдения, риск при дескрипторе pN2 и pN3 выше только в первые два года наблюдения, однако дескриптор pN3 имеет более высокий риск по сравнению с остальными дескрипторами pN0-pN2. Степень дифференцировки аденокарциномы проявляет себя как фактор риска, начиная со второго года наблюдения. Также ХТ-1 «работает» как фактор снижения риска развития МПД только в первый год после проведенного лечения. Последнее объясняет результаты ранее проведенных исследований, продемонстрировавших недостаточную эффективность данного вида химиотерапии для предупреждения развития МПД у радикально оперированных по поводу рака желудка пациентов [14–16].

Достаточно сложное пересечение факторов риска, каждого со своим влиянием на изменение кумулятивной инцидентности и выживаемости, приводит к тому, что имеет смысл рассмотреть результат пятилетнего прогноза как новый комбинированный предиктор (КП) МПД, который определит, насколько пациент подвержен риску МПД.

Подгонка модели с новым единственным КП МПД продемонстрировала значение индекса конкродации 0,785 (SE = 0,011). Проверка на пропорциональность рисков не выявила нарушения для данной модели с МПД. Внутренняя валидация модели с новым единственным КП МПД методом бутстрепа (5000 повторений) показала высокие параметры производительности модели [29], которые представлены в табл. 3.

Кроме оценки производительности, также были построены ROC-кривые для каждого года наблюдения и рассчитаны величины AUC [30], показанные на рис. 3. В диагностической медицине считается, что кривые ROC имеют несколько привлекательных особенностей: (a) ROC-кривая описывает внутреннюю способность модели к дискриминации, не привязывая ее к какому-либо конкретному порогу; (b) площадь под кривой ROC (AUC) можно интерпретировать как вероятность того, что результат прогноза для случайно выбранного индивидуума с событием превысит результат прогноза для случайно выбранного индивидуума без события; (c) значение AUC не зависит от распространенности исходов в популяции [29].

Таблица 3
Параметры производительности модели
Table 3
Model performance indices

Показатель* *Index**	На полном наборе *On full set*	По обучающей выборке *Based on training set*	По тестовой выборке *Based on test sample*	Оптимизм *Optimism*	Скорректированный индекс *Adjusted index*
Somers' Dxy ранговой корреляции	0,567	0,566	0,567	–0,0006	0,567
Индекс уклона/ калибровки	1,0000	1,0000	1,0006	–0,0006	1,0006
g-индекс	0,964	0,961	0,959	0,0013	0,963

Примечание: *Somers' Dxy ранговой корреляции связан с индексом конкордации и определяет, насколько хорошо модель различает пациентов с высоким и низким риском развития МПД; индекс уклона/калибровки – насколько хорошо калибрована модель; g-индекс – это индекс Джини по логарифмической шкале относительного риска. Согласно работе [30] модель представляется пригодной для прогноза.

Note: *Somers' Dxy rank correlation is related to the concordance index and determines how well the model distinguishes patients with high and low risk of developing IVD; slope/calibration index – how well the model is calibrated; g-index is Gini index on a logarithmic scale of relative risk. According to [30], the model seems suitable for forecasting.

По результатам анализа AUC КП МПД показывает высокую способность к дискриминации на любом сроке наблюдения (рис. 3). Данный КП МПД может быть принят в качестве прогностического фактора. По результатам моделирования можно определить уровень повышенного риска неблагоприятного исхода. На основании значения выше и ниже данного уровня принимается решение об индивидуальном контроле пациента по вопросу возможной МПД.

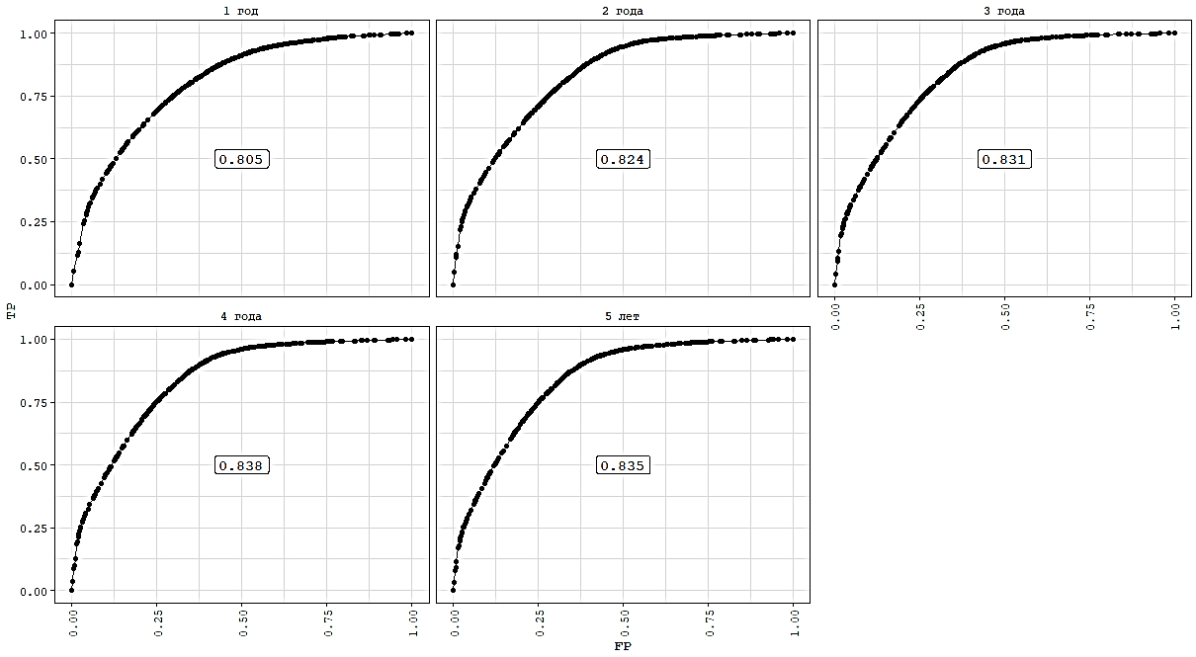


Рис. 3. AUC для модели с КП МПД
Fig. 3. AUC for model with complex predictor of metachronous peritoneal dissemination

Таким образом, можно получить ответ на первый вопрос: кого контролировать? Для ответа на второй вопрос – когда контролировать? – предлагается использовать значения прогноза риска развития МПД на год, два, три, четыре и пять лет, используя свойства кумулятивного накопления на основе модели табл. 1. Пусть вероятность не наступления неблагоприятного события за первый год составила $P1$, за два года $P2$, за три – $P3$ и т. д. Тогда вероятность наступления события в первый год $p_1 = 1 - P1$, во второй $p_2 = 1 - P2/P1$, в третий $p_3 = 1 - P3/P2$ и т. д. Таким образом, можно выявить максимальное значение вероятности наступления события в течение года наблюдения (первого, второго и т. д.) и соотнести с порядковым номером года наблюдения: $I = \arg\max_{year}\left(p_{year}\right)$, $year = 1,...,5$. Этот год будет являться годом контроля, возможно с применением диагностической операции.

5. Клиническая полезность разработанной модели. Традиционные статистические меры для оценки моделей прогнозирования, новых маркеров и диагностических тестов включают чувствительность, специфичность, площадь под кривой, калибровку, индекс конкордации и другие меры производительности модели [29, 31]. Однако эти меры не дают ответа на вопрос, следует ли использовать модель (маркер, тест) в клинической практике, насколько она улучшит текущую ситуацию. Анализ решений (Decision Curve Analysis, DCA) [32–34] дает возможность рассмотреть полезность модели, учитывая клинические последствия использования модели, маркера или теста. Ключевой концепцией анализа решений является идея компромисса между различными конечными точками.

Применительно к рассматриваемой задаче можно определить соотношение, что вред от задержки диагностики МПД в q раз больше, чем вред от ненужной лапароскопии, т. е. одна обнаруженная МПД равна q ненужным лапароскопиям, как, например 1:3.

Чистая польза для модели рассчитывается следующим образом: польза – (вред · обменный курс), и полностью алгоритм расчета чистой пользы выглядит следующим образом:

1. Выбираем порог КП МПД, когда пациент считается под риском.

2. Рассчитываем число пациентов с таким же риском и выше, которые реально имели неблагоприятное событие (True Positive, TP), и число пациентов с таким же риском и выше, которые не имели события (False Positive, FP), но были отнесены к группе риска.

Пусть размер выборки равен N, чистая польза в случае проведения лапароскопии всем пациентам может быть рассчитана следующим образом: если D – число случаев диссеминации в исследуемой когорте, то чистая польза составит $B_{all} = D/N - \left(1 - D/N\right) \times \left(1:q\right)$, q в анализе решений называется обменным курсом. Если использовать модель, то чистая польза составит $B_{model} = TP/N - \left(1 - FP/N\right) \times \left(1:q\right)$.

Фактически «обменный курс» является порогом принятия решения по модели. Модель дает результат прогноза от 0 до 1, который трактуется как риск развития события в течение пяти лет. Тогда при установке порога принятия решения на уровне p_{pr} «обменный курс» составит $\frac{p_{pr}}{1 - p_{pr}}$. Имея в виду различный «обменный курс», можно построить график чистой пользы для разных значений p_{pr}. Данный график изображен на рис. 4, значения p_{pr} приведены в процентах. Значения $p_{pr} > 0{,}5$ (т. е., исходя из результатов моделирования, более 50 % пациентов будут иметь неблагоприятные события) не рассматриваются, поскольку в этом случае клиническая стратегия сводится к рекомендации, что дополнительная лапароскопическая диагностика показана всем.

Для DCA-анализа разработанной модели использованы результаты прогноза (КП МПД) по данным 1311 пациентов, у которых в период 2008–2021 гг. было проведено радикальное комбинированное или комплексное лечение по поводу местнораспространенного рака желудка. Полученный прогноз (вероятность иметь событие МПД в первые пять лет наблюдения после

лечения) сравнивался с имеющимися данными продолжительного наблюдения за пациентами. Рассматривались три варианта принятия решений:

1. Не использовать лапароскопию. В этом случае имеем пятилетнюю кумулятивную инцидентность 22,9 %. Иными словами, почти четверть исследуемой популяции получит событие МПД в течение пяти лет после проведенного лечения.

2. Проверять всех. В этом случае три четверти популяции будет подвергнута процедуре (лапароскопии), которая им не нужна. Более того, нагрузка на клинику возрастает даже при условии, что вред от лапароскопии нулевой.

3. Проверять выборочно на основании расчета по модели.

Таблица 4
Расчет чистой пользы по различным вариантам принятия решений об определении МПД

Table 4
Net benefit for different strategies for detection of metachronous peritoneal dissemination (MPD)

Обменный курс, $p_{pr} \cdot 100\,\%$ *Exchange rate,* $p_{pr} \cdot 100\,\%$	Чистая польза, варианты *Net benefit, options*			Распределение событий в когорте после пяти лет наблюдения, *n* (%) *Distribution of events in the cohort after five years of follow-up, n (%)*			
	1	2	3	Без события, N=515 *No event,* N=515	МПД, N=316 *MPD,* N=316	Отдаленные метастазы, N=167 *Distant metastases,* N=167	Второй рак, N=55 *Second cancer,* N=55
5 % (1:19)	0	0,188	0,218	327 (63,5)	310 (98,1)	150 (89,8)	29 (52,7)
10 % (1:9)	0	0,143	0,195	244 (47,4)	299 (94,6)	139 (83,2)	21 (38,2)
20 % (1:4)	0	0,036	0,144	148 (28,7)	271 (85,8)	101 (60,5)	14 (25,5)
25 % (1:3)	0	–0,028	0,115	112 (21,7)	238 (75,3)	84 (50,3)	13 (23,6)
40 % (1:1.5)	0	–0,285	0,056	30 (5,8)	146 (46,2)	45 (26,9)	6 (10,9)
50 % (1:1)	0	–0,542	0,029	12 (2,3)	95 (30,1)	22 (13,2)	2 (3,6)

Из табл. 4 видно, что принимать решение по модели оправдано при любом «обменном курсе», чистая польза всегда больше в варианте 3. Более того, может быть обследована часть пациентов, у которых возможно развитие отдаленных метастазов, что также является дополнительным положительным эффектом принятия решений на основе модели по варианту 3. На рис. 4 видно, что модель дает чистую пользу при любом обменном курсе.

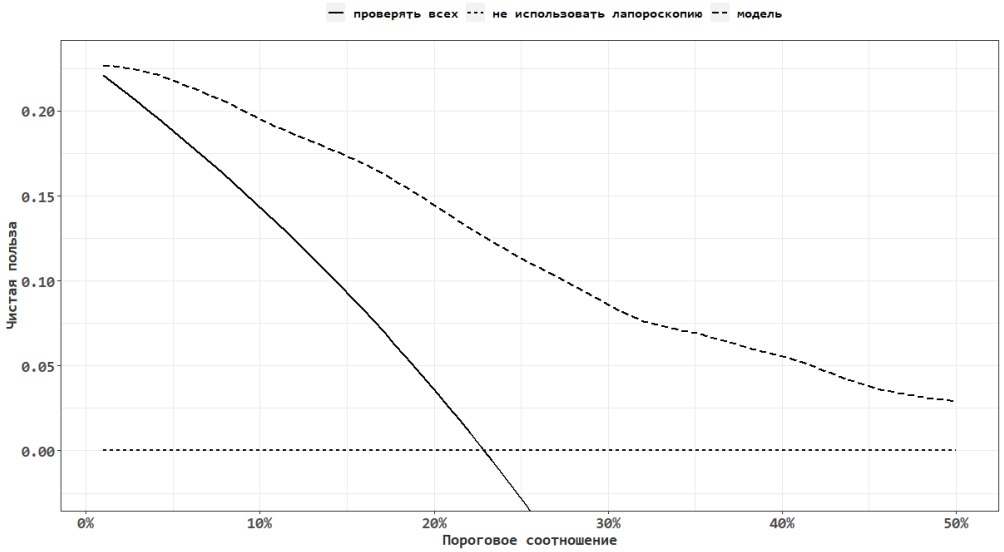


Рис. 4. Чистая польза при различных стратегиях выявления МПД

Fig. 4. DCA for different strategies for detection of metachronous peritoneal dissemination

Таким образом, можно утверждать, что разработанная модель дает преимущество в клинической практике, позволяя определять необходимость контроля МПД для пациентов с различными исходными характеристиками опухолевого процесса, различным дополнительным химиотерапевтическим лечением, а также с учетом возраста (см. табл. 1).

Заключение. В настоящем исследовании рассмотрен подход к построению модели прогноза и принятию решений на ее основе. В отличие от многих публикаций, которые исследуют только построение и валидацию модели, в работе учтены и проанализированы все аспекты – от построения модели до оценки ее пригодности в клинической практике. Основной акцент сделан на нелинейности рисков, которая присуща многим вариантам лечения, однако часто игнорируется из-за малого объема выборки, когда нелинейность рисков сложно оценить статистически. Иногда в публикациях игнорируется оценка условия пропорциональности рисков. При опубликовании модели прогноза на основе регрессии Кокса крайне редко сообщается о проверке предположений, лежащих в основе регрессии Кокса (регрессии Файна – Грея).

В клинической онкологической практике большую роль играет не только своевременное лечение, но и предупреждение неблагоприятных исходов после окончания лечения. Индивидуализация наблюдения за пациентом после лечения снижает риски фатальных исходов, затрат на дополнительные исследования и лечение в случае прогрессирования онкозаболевания. По результатам исследования принимаются решения, которые должны привести к более эффективной и качественной тактике лечения, к выбору оптимальных подходов и получению клинически благоприятных исходов заболевания. Изучаемые методы прогноза рисков должны в конечном итоге привести к индивидуализированному ведению пациента по результатам его данных.

Таким образом, в публикации не только преследуется цель разработать конкретную модель, но и демонстрируются пути продвижения модели – от построения и анализа ее валидности до клинической пригодности.

Вклад авторов. *О. В. Красько* – концепция и дизайн исследования, редактирование, анализ данных, написание текста; *М. Ю. Ревтович* – концепция и дизайн исследования, сбор материала, обработка и написание текста; *А. В. Иванов* – обработка и написание текста.

Список использованных источников

1. Alonzo, T. A. Clinical prediction models: a practical approach to development, validation, and updating: by Ewout W. Steyerberg / T. A. Alonzo // American J. of Epidemiology. – 2009. – Vol. 170, iss. 4. – P. 528. https://doi.org/10.1093/aje/kwp129

2. A novel clinical risk prediction model for sudden cardiac death in hypertrophic cardiomyopathy (HCM risk-SCD) / C. O'Mahony [et al.] // European Heart J. – 2014. – Vol. 35, no. 30. – P. 2010–2020.

3. Scrucca, L. Competing risk analysis using R: an easy guide for clinicians / L. Scrucca, A. Santucci, F. Aversa // Bone Marrow Transplantation. – 2007. – Vol. 40, no. 4. – P. 381–387.

4. Prognostic models with competing risks: methods and application to coronary risk prediction / M. Wolbers [et al.] // Epidemiology. – 2009. – Vol. 20, iss. 4 – P. 555–561.

5. Cox, D. R. Regression models and life-tables / D. R. Cox // J. of the Royal Statistical Society: Series B (Methodological). – 1972. – Vol. 34, no. 2. – P. 187–202.

6. Hosmer, Jr. D. W. Applied Survival Analysis: Regression Modeling of Time-to-Event Data / Jr. D. W. Hosmer, S. Lemeshow, S. May. – John Wiley & Sons, 2011. – 416 p.

7. Therneau, T. Using time dependent covariates and time dependent coefficients in the cox model / T. Therneau, C. Crowson, E. Atkinson // Survival Vignettes. – 2017. – Vol. 2, no. 3. – P. 1–25.

8. Murphy, S. A. Time-dependent coefficients in a Cox-type regression model / S. A. Murphy, P. K. Sen // Stochastic Processes and their Applications. – 1991. – Vol. 39, no. 1. – P. 153–180.

9. Thomas, L. Tutorial: survival estimation for Cox regression models with time-varying coefficients using SAS and R / L. Thomas, E. M. Reyes // J. of Statistical Software. – 2014. – Vol. 61. – P. 1–23.

10. Redmond, C. The methodologic dilemma in retrospectively correlating the amount of chemotherapy received in adjuvant therapy protocols with disease-free survival / C. Redmond, B. Fisher, H. S. Wieand // Cancer Treatment Reports. – 1983. – Vol. 67, no. 6. – P. 519–526.

11. Suissa, S. Immortal time bias in pharmacoepidemiology / S. Suissa // American J. of Epidemiology. – 2008. – Vol. 167, no. 4. – P. 492–499.

12. Fine, J. P. A proportional hazards model for the subdistribution of a competing risk / J. P. Fine, R. J. Gray // J. of the American Statistical Association. – 1999. – Vol. 94, no. 446. – P. 496–509.

13. Li, J. Checking Fine and Gray subdistribution hazards model with cumulative sums of residuals / J. Li, T. H. Scheike, M. J. Zhang // Lifetime Data Analysis. – 2015. – Vol. 21, no. 2. – P. 197–217.

14. A detailed analysis of the recurrence timing and pattern after curative surgery in patients undergoing neoadjuvant therapy or upfront surgery for gastric cancer / A. Agnes [et al.] // J. of Surgical Oncology. – 2020. – Vol. 122, no. 2. – P. 293–305.

15. Incidence, time course and independent risk factors for metachronous peritoneal carcinomatosis of gastric origin – a longitudinal experience from a prospectively collected database of 1108 patients / F. Seyfried [et al.] // BMC Cancer. – 2015. – Vol. 15. – P. 1–10.

16. Lauren histologic type is the most important factor associated with pattern of recurrence following resection of gastric adenocarcinoma / J. H. Lee [et al.] // Annals of Surgery. – 2018. – Vol. 267, no. 1. – P. 105.

17. Reutovich, M. Y. Hyperthermic intraperitoneal chemotherapy in prevention of gastric cancer metachronous peritoneal metastases: a systematic review / M. Y. Reutovich, O. V. Krasko, O. G. Sukonko // J. of Gastrointestinal Oncology. – 2021. – Vol. 12, suppl. 1. – P. S5–S17. https://doi.org/10.21037/jgo-20-129

18. Analysis and external validation of a nomogram to predict peritoneal dissemination in gastric cancer / X. Chen [et al.] // Chinese J. of Cancer Research. – 2020. – Vol. 32, no. 2. – P. 197–207.

19. Staging of peritoneal carcinomatosis: enhanced CT vs. PET/CT / C. Dromain [et al.] // Abdominal Imaging. – 2008. – Vol. 33. – P. 87–93.

20. Added value of pretreatment 18F-FDG PET/CT for staging of advanced gastric cancer: comparison with contrast-enhanced MDCT / Y. Kawanaka [et al.] // European J. of Radiology. – 2016. – Vol. 85, no. 5. – P. 989–995.

21. Peritoneal recurrence in gastric cancer following curative resection can be predicted by postoperative but not preoperative biomarkers: a single-institution study of 320 cases / F. Wu [et al.] // Oncotarget. – 2017. – Vol. 8, no. 44. – P. 78120.

22. Ревтович, М. Ю. Местнораспространенный рак желудка: современные направления радикального лечения и прогнозирование отдаленных результатов : монография / М. Ю. Ревтович, О. В. Красько. – Минск : БелМАПО, 2022. – 217 с.

23. Результаты радикального лечения инфильтративных форм рака желудка с применением перфузионной термохимиотерапии / М. Ю. Ревтович [и др.] // Евразийский онкологический журнал. – 2022. – Т. 10, № 2. – С. 107–117.

24. Ревтович, М. Ю. Интраоперационная оценка риска развития канцероматоза после радикального хирургического лечения рака желудка / М. Ю. Ревтович, О. В. Красько // Онкология и радиология Казахстана. – 2020. – № 2(56). – С. 26–30. https://doi.org/10.52532/2521-6414-2020-2-56-26-30

25. Reutovich, M. Prophylactic hyperthermic intraperitoneal chemotherapy in gastric cancer management: short- and long-term outcomes of a prospective randomized study / M. Reutovich, O. Krasko // Oncology in Clinical Practice. – 2021. – Vol. 17, no. 5. – P. 187–193. https://doi.org/10.5603/OCP.2021.0028

26. Reutovich, M. Yu. Efficacy of adjuvant systemic chemotherapy combined with radical surgery and hyperthermic intraperitoneal chemotherapy in gastric cancer treatment / M. Yu. Reutovich, O. V. Krasko, O. G. Sukonko // Indian J. of Surgical Oncology. – 2020. – Vol. 11. – P. 337–343. https://doi.org/10.1007/s13193-020-01102-w

27. Schoenfeld, D. Partial residuals for the proportional hazards regression model / D. Schoenfeld // Biometrika. – 1982. – Vol. 69, no. 1. – P. 239–241.

28. Алгоритмы диагностики и лечения злокачественных новообразований : клинический протокол : утв. Постановлением М-ва здравоохранения Респ. Беларусь № 60 от 06.07.2018 г. / под ред. О. Г. Суконко, С. А. Красного. – Минск : Профессиональные издания, 2019. – С. 97–110.

29. Heagerty, P. J. Time-dependent ROC curves for censored survival data and a diagnostic marker / P. J. Heagerty, T. Lumley, M. S. Pepe // Biometrics. – 2000. – Vol. 56, no. 2. – P. 337–344.

30. Harrell, F. E. Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis / F. E. Harrell. – N. Y. : Springer, 2001. – 600 p.

31. Steyerberg, E. W. Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating / E. W. Steyerberg. – Springer, 2009. – 528 p.

32. Vickers, A. J. Decision curve analysis: a novel method for evaluating prediction models / A. J. Vickers, E. B. Elkin // Medical Decision Making. – 2006. – Vol. 26, no. 6. – P. 565–574.

33. Vickers, A. J. Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests [Electronic resource] / A. J. Vickers, B. Van Calster, E. W. Steyerberg // BMJ. – 2016. – Vol. 352. – Mode of access: https://www.bmj.com/content/bmj/352/bmj.i6.full.pdf. – Date of access: 12.09.2023.

34. Extensions to decision curve analysis, a novel method for evaluating diagnostic tests, prediction models and molecular markers / A. J. Vickers [et al.] // BMC Medical Informatics and Decision Making. – 2008. – Vol. 8. – P. 1–17.

References

1. Alonzo T. A. Clinical prediction models: a practical approach to development, validation, and updating: by Ewout W. Steyerberg. *American Journal of Epidemiology*, 2009, vol. 170, iss. 4, p. 528. https://doi.org/10.1093/aje/kwp129

2. O'Mahony C., Jichi F., Pavlou M., Monserrat L., Anastasakis A., …, Elliott P. M. A novel clinical risk prediction model for sudden cardiac death in hypertrophic cardiomyopathy (HCM risk-SCD). *European Heart Journal*, 2014, vol. 35, no. 30, pp. 2010–2020.

3. Scrucca L., Santucci A., Aversa F. Competing risk analysis using R: an easy guide for clinicians. *Bone Marrow Transplantation*, 2007, vol. 40, no. 4, pp. 381–387.

4. Wolbers M., Koller M. T., Witteman J. C. M., Steyerberg E. W. Prognostic models with competing risks: methods and application to coronary risk prediction. *Epidemiology*, 2009, vol. 20, iss. 4, pp. 555–561.

5. Cox D. R. Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 1972, vol. 34, no. 2, pp. 187–202.

6. Hosmer Jr. D. W., Lemeshow S., May S. Applied Survival Analysis: Regression Modeling of Time-to-Event Data. John Wiley & Sons, 2011, 416 p.

7. Therneau T., Crowson C., Atkinson E. Using time dependent covariates and time dependent coefficients in the cox model. *Survival Vignettes*, 2017, vol. 2, no. 3, pp. 1–25.

8. Murphy S. A., Sen P. K. Time-dependent coefficients in a Cox-type regression model. *Stochastic Processes and their Applications*, 1991, vol. 39, no. 1, pp. 153–180.

9. Thomas L., Reyes E. M. Tutorial: survival estimation for Cox regression models with time-varying coefficients using SAS and R. *Journal of Statistical Software*, 2014, vol. 61, pp. 1–23.

10. Redmond C., Fisher B., Wieand H. S. The methodologic dilemma in retrospectively correlating the amount of chemotherapy received in adjuvant therapy protocols with disease-free survival. *Cancer Treatment Reports*, 1983, vol. 67, no. 6, pp. 519–526.

11. Suissa S. Immortal time bias in pharmacoepidemiology. *American Journal of Epidemiology*, 2008, vol. 167, no. 4, pp. 492–499.

12. Fine J. P., Gray R. J. A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association*, 1999, vol. 94, no. 446, pp. 496–509.

13. Li J., Scheike T. H., Zhang M. J. Checking Fine and Gray subdistribution hazards model with cumulative sums of residuals. *Lifetime Data Analysis*, 2015, vol. 21, no. 2, pp. 197–217.

14. Agnes A., Biondi A., Laurino A., Strippoli A., Ricci R., …, D'Ugo D. A detailed analysis of the recurrence timing and pattern after curative surgery in patients undergoing neoadjuvant therapy or upfront surgery for gastric cancer. *Journal of Surgical Oncology*, 2020, vol. 122, no. 2, pp. 293–305.

15. Seyfried F., Von Rahden B. H., Miras A. D., Gasser M., Maeder U., …, Kerscher A. G. Incidence, time course and independent risk factors for metachronous peritoneal carcinomatosis of gastric origin – a longitudinal experience from a prospectively collected database of 1108 patients. *BMC Cancer*, 2015, vol. 15, pp. 1–10.

16. Lee J. H., Chang K. K., Yoon C., Tang L. H., Strong V. E., Yoon S. S. Lauren histologic type is the most important factor associated with pattern of recurrence following resection of gastric adenocarcinoma. *Annals of Surgery*, 2018, vol. 267, no. 1, p. 105.

17. Reutovich M. Y., Krasko O. V., Sukonko O. G. Hyperthermic intraperitoneal chemotherapy in prevention of gastric cancer metachronous peritoneal metastases: a systematic review. *Journal of Gastrointestinal Oncology*, 2021, vol. 12, suppl. 1, pp. S5–S17. https://doi.org/10.21037/jgo-20-129

18. Chen X., Chen S., Wang X., Nie R., Chen D., …, Peng J. Analysis and external validation of a nomogram to predict peritoneal dissemination in gastric cancer. *Chinese Journal of Cancer Research*, 2020, vol. 32, no. 2, pp. 197–207.

19. Dromain C., Leboulleux S., Auperin A., Goere D., Malka D., …, Elias D. Staging of peritoneal carcinomatosis: enhanced CT vs. PET/CT. *Abdominal Imaging*, 2008, vol. 33, pp. 87–93.

20. Kawanaka Y., Kitajima K., Fukushima K., Mouri M., Doi H., …, Hirota S. Added value of pretreatment 18F-FDG PET/CT for staging of advanced gastric cancer: comparison with contrast-enhanced MDCT. *European Journal of Radiology*, 2016, vol. 85, no. 5, pp. 989–995.

21. Wu F., Shi C., Wu R., Huang Z., Chen Q. Peritoneal recurrence in gastric cancer following curative resection can be predicted by postoperative but not preoperative biomarkers: a single-institution study of 320 cases. *Oncotarget*, 2017, vol. 8, no. 44, p. 78120.

22. Reutovich M. Yu., Krasko O. V. Mestnorasprostranennyj rak zheludka: sovremennye napravlenija radikal'nogo lechenija i prognozirovanie otdalennyh rezul'tatov. *Locally Advanced Gastric Cancer: Modern Directions of Radical Treatment and Prediction of Long-Term Results*. Minsk, Belorusskaja medicinskaja akademija poslediplomnogo obrazovanija, 2022, 217 p. (In Russ.).

23. Reutovich M. Yu., Krasko O. V., Malkevich V. T., Patseika A. I. *Results of radical treatment of infiltrative gastric cancer using perfusion thermochemotherapy*. Evrazijskij onkologicheskij zhurnal [*Eurasian Journal of Oncology*], 2022, vol. 10, no. 2, pp. 107–117 (In Russ.).

24. Reutovich M. Yu., Krasko O. V. *Intraoperative risk assessment of carcinomatosis development after radical surgery for gastric cancer*. Onkologija i radiologija Kazahstana [Oncology and Radiology of Kazakhstan], 2020, no. 2(56), pp. 26–30 (In Russ.). https://doi.org/10.52532/2521-6414-2020-2-56-26-30

25. Reutovich M., Krasko O. Prophylactic hyperthermic intraperitoneal chemotherapy in gastric cancer management: short- and long-term outcomes of a prospective randomized study. *Oncology in Clinical Practice*, 2021, vol. 17, no. 5, pp. 187–193. https://doi.org/10.5603/OCP.2021.0028

26. Reutovich M. Yu., Krasko O. V., Sukonko O. G. Efficacy of adjuvant systemic chemotherapy combined with radical surgery and hyperthermic intraperitoneal chemotherapy in gastric cancer treatment. *Indian Journal of Surgical Oncology*, 2020, vol. 11, pp. 337–343. https://doi.org/10.1007/s13193-020-01102-w

27. Schoenfeld D. Partial residuals for the proportional hazards regression model. *Biometrika*, 1982, vol. 69, no. 1, pp. 239–241.

28. Sukonko O. G., Krasnogo S. A. (eds.). Algoritmy diagnostiki i lechenija zlokachestvennyh novoobrazovanij : klinicheskij protokol : utverzhden Postanovleniem Ministerstva zdravoohranenija Respubliki Belarus' № 60 ot 06.07.2018 g. *Algorithms for the Diagnosis and Treatment of Malignant Neoplasms : Clinical Protocol : Approved by Resolution of the Ministry of Health of the Republic of Belarus no. 60 of 07.06.2018*. Minsk, Professional'nye izdanija, 2019, pp. 97–110 (In Russ.).

29. Heagerty P. J., Lumley T., Pepe M. S. Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics*, 2000, vol. 56, no. 2, pp. 337–344.

30. Harrell F. E. *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis*. New York, Springer, 2001, 600 p.

31. Steyerberg E. W. *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating*. Springer, 2009, 528 p.

32. Vickers A. J., Elkin E. B. Decision curve analysis: a novel method for evaluating prediction models. *Medical Decision Making*, 2006, vol. 26, no. 6, pp. 565–574.

33. Vickers A. J., Van Calster B., Steyerberg E. W. Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests. *BMJ*, 2016, vol. 352. Available at: https://www.bmj.com/content/bmj/352/bmj.i6.full.pdf (accessed 12.09.2023).

34. Vickers A. J., Cronin A. M., Elkin E. B., Gonen M. Extensions to decision curve analysis, a novel method for evaluating diagnostic tests, prediction models and molecular markers. *BMC Medical Informatics and Decision Making*, 2008, vol. 8, pp. 1–17.

Информация об авторах

Красько Ольга Владимировна, кандидат технических наук, доцент, ведущий научный сотрудник, Объединенный институт проблем информатики Национальной академии наук Беларуси.
E-mail: krasko@newman.bas-net.by
https://orcid.org/0000-0002-4150-282X

Information about the authors

Olga V. Krasko, Ph. D. (Eng.), Assoc. Prof., Leading Researcher, The United Institute of Informatics Problems of the National Academy of Sciences of Belarus.
E-mail: krasko@newman.bas-net.by
https://orcid.org/0000-0002-4150-282X

Ревтович Михаил Юрьевич, доктор медицинских наук, доцент, декан лечебного факультета, Белорусский государственный медицинский университет.
E-mail: mihail_revtovich@yahoo.com
https://orcid.org/0000-0001-7202-6902

Иванов Андрей Владимирович, аспирант Республиканского научно-практического центра онкологии и медицинской радиологии им. Н. Н. Александрова.
E-mail: tennis5000@rambler.ru
https://orcid.org/0009-0005-1288-2121

Mikhail Yu. Reutovich, D. Sc. (Med.), Assoc. Prof., Dean of the Faculty of General Medicine Belarusian State Medical University.
E-mail: mihail_revtovich@yahoo.com
https://orcid.org/0000-0001-7202-6902

Andrey V. Ivanov, Postgraduate Student N. N. Alexandrov National Cancer Center of Belarus.
E-mail: tennis5000@rambler.ru
https://orcid.org/0009-0005-1288-2121

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ
INFORMATION TECHNOLOGIES

УДК 004.89
https://doi.org/10.37661/1816-0301-2024-21-1-83-104

Оригинальная статья
Original Paper

Классификация займа с использованием нейронной сети прямого распространения

В. И. Бегунков

E-mail: vbegunkov@gmail.com

Аннотация

Цели. Целью исследования являются построение и изучение использования нейронной сети прямого распространения для решения задачи классификации займа, а также проведение сравнительного анализа подхода на основе нейронной сети с существующим подходом, основанным на логистической регрессии.

Метод. На базе нейронной сети прямого распространения с использованием исторических данных по выданным займам вычисляются следующие метрики: стоимостная функция, *Accuracy, Precision, Recall* и мера F_1, рассчитанная на основе значений *Precision* и *Recall*. Полиномиальные параметры и метод главных компонент применяются для определения оптимального модифицированного набора входных данных для исследуемой нейронной сети.

Результаты. Проанализировано воздействие нормализации исходных данных на конечный результат, оценено влияние количества элементов скрытого уровня на конечный результат при помощи двухэтапного метода и метода Монте-Карло, определено воздействие использования сбалансированных данных, рассчитано оптимальное граничное значение для выходного уровня рассматриваемой нейронной сети, найдена оптимальная функция активации для элементов скрытого уровня, изучено влияние увеличения количества входных показателей путем заполнения отсутствующих значений и использования полиномов разной степени, а также проанализирован на избыточность имеющийся набор входных показателей.

Заключение. По итогам исследования можно сделать вывод, что применение сети прямого распространения для решения задач классификации займа является целесообразным. В сравнении с логистической регрессией реализация решения с использованием нейронной сети прямого распространения требует больше времени и вычислительных ресурсов. Однако полученные наиболее важные значения *Accuracy* и меры F_1 выше, чем те, которые были рассчитаны с применением логистической регрессии [1].

Ключевые слова: классификация займа, скоринг, нейронная сеть прямого распространения, машинное обучение, нормализация данных

Для цитирования. Бегунков, В. И. Классификация займа с использованием нейронной сети прямого распространения / В. И. Бегунков // Информатика. – 2024. – Т. 21, № 1. – С. 83–104.
https://doi.org/10.37661/1816-0301-2024-21-1-83-104

Конфликт интересов. Автор заявляет об отсутствии конфликта интересов.

Поступила в редакцию | Received 02.11.2023
Подписана в печать | Accepted 08.01.2024
Опубликована | Published 29.03.2024

Loan classification using a feed-forward neural network

Uladzimir I. Behunkou

E-mail: vbegunkov@gmail.com

Abstract

Objectives. The purpose of the study is to construct and study the use of a feed-forward neural network to solve the problem of loan classification, as well as to conduct a comparative analysis of the neural network-based approach with the existing approach based on logistic regression.

Methods. Based on a feed-forward neural network using historical data on loans issued, the following metrics are calculated: cost function, *Accuracy*, *Precision*, *Recall*, and F_1 measure, calculated on Precision and Recall values. Polynomial parameters and the principal component method are used to determine the optimal set of input data for the studied neural network.

Results. The impact of data normalization on the final result was analyzed, the influence of the number of units in the hidden layer on the outcome was evaluated using a two-stage method and the Monte Carlo method, the effect of balanced data use was determined, the optimal threshold value for output layer of the neural network under investigation was calculated, the optimal activation function for the hidden layer units was found, the effect of increasing input indicators through missing values imputation and the use of polynomials of varying degrees was studied and the redundancy in the existing set of input indicators was analyzed.

Conclusion. Based on the results of the research, we can conclude that the use of a direct distribution network to solve problems of loan classification is appropriate. Compared to logistic regression, implementing a solution using a feed-forward neural network requires more time and computing resources. However, the obtained most important values of *Accuracy* and F_1 measure are higher than those calculated using logistic regression [1].

Keywords: loan classification, scoring, feed-forward neural network, machine learning, data normalization

For citation. Behunkou U. I. *Loan classification using a feed-forward neural network*. Informatika [*Informatics*], 2024, vol. 21, no. 1, pp. 83−104 (In Russ.). https://doi.org/10.37661/1816-0301-2024-21-1-83-104

Conflict of interest. The author declares of no conflict of interest.

Введение. Эффективное распределение денежных активов между субъектами хозяйствования посредством кредитования является важной задачей для экономики. Кроме того, данное направление очень привлекательно для финансовых организаций в связи с тем, что высокомаржинально. Если рассматривать Европу, то потребительское кредитование является наиболее интересным сектором, так как позволяет акционерам получить годовой доход в размере 11,5 % [2], что существенно выше, чем величина 7,4 % в сегменте корпоративного банкинга. Также стоит отметить, что сектор потребительского кредитования занимает существенную долю и растет быстрыми темпами. Например, сумма выданных займов в секторе потребительского кредитования в США выросла на 118,3 % с $ 829 млрд[1] в январе 2010 г. до $1810 млрд[2] в сентябре 2022 г. Пропорционально с ростом объема выданных потребительских кредитов растет и ответственность, лежащая на финансовых институтах, за успешное предоставление таких займов.

Изначально решение о выдаче займа принималось ответственным лицом в финансовом институте субъективно на основе имеющегося опыта проведения предыдущих сделок [3]. Однако в условиях существенного роста рынка кредитования появилась необходимость в применении более надежных методов и инструментов для принятия решений по выдаче займов. Учитывая существенное развитие информационных технологий, многие финансовые институты стали использовать сложные статистические модели для решения задачи по выдаче займов.

[1]Assets and Liabilities of Commercial Banks in the United States – H.8 [Electronic resource]. – Mode of access: https://www.federalreserve.gov/releases/h8/20100108/. – Date of access: 02.09.2019.

[2]Assets and Liabilities of Commercial Banks in the United States – H.8 [Electronic resource]. – Mode of access: https://www.federalreserve.gov/releases/h8/20180928/. – Date of access: 02.09.2019.

В дополнение прослеживается существенное изменение поведения клиентов финансовых организаций: согласно наблюдаемой тенденции клиенты финансовых институтов все больше предпочитают использовать онлайн-банкинг, так как он позволяет осуществлять операции круглосуточно и облегчает процесс управления финансами [4].

В свою очередь Хэнд и Хэнли не только представили задачу классификации займа как бинарную [3], разделив заемщиков на два класса в соответствии с вероятностью погашения займа на хороших (без дефолта) и плохих (с дефолтом), но и определили кредитный скоринг как более формальный процесс по расчету вероятности дефолта по платежам у заемщиков на основе статистических моделей, которые используют независимые переменные для получения оценки вероятности дефолта. Также они предложили детальный обзор статистических методов, которые к тому моменту использовались на практике для кредитного скоринга, и пришли к выводу, что не существует одного лучшего метода. С их точки зрения определение лучшего метода возможно только для конкретного примера задачи классификации займа в зависимости от его входных данных.

В опубликованной в 2015 г. статье [5] в финальную выборку для исследования кроме рассмотренного ранее [1] индивидуального классификатора на основе логистической регрессии (logistic regression, LR) попал и второй индивидуальный классификатор, базирующийся на искусственной нейронной сети (artificial neural network, ANN).

Целью настоящей работы является изучение возможности эффективного применения нейронной сети прямого распространения для решения задачи классификации займа и сравнение результатов со значениями, полученными при использовании логистической регрессии.

Описание данных. Для решения задачи все данные можно разделить на три группы: входные данные, настраиваемые параметры рассматриваемых методов и выходные данные.

Входные данные. В качестве данных для настройки параметров и проведения экспериментов с рассматриваемыми методами используются исторические данные по выданным на платформе для кредитования от человека человеку Lending Club займам[3], состоящие из 2 260 668 строк (займы, выданные за период с апреля 2016 по сентябрь 2018 г.). Перечень входных показателей и принцип преобразования входных данных аналогичны тем, которые были описаны ранее при рассмотрении логистической регрессии [1]. Таким образом, финальный набор входных данных состоит из m = 1 221 731 позиции и n = 54 входных показателей.

Предполагается, что значения данных показателей были известны до принятия решения о выдаче соответствующего займа. Так же, как в работе [1], обозначим значение показателя j в займе i из исходного набора данных через элемент $x_j^{(i)}$ матрицы X размером $m \times n$, где $j = 1, \ldots, n$, $i = 1, \ldots, m$. Обозначим через x_j столбец матрицы X, а через $x^{(i)}$ – строку матрицы X, которая содержит значения независимых показателей в отдельной позиции (займе) i набора данных. Кроме того, в качестве исходных данных используются целевые значения $y^{(i)}$ (итоговый результат по займу i, где $i = 1, \ldots, m$), которые определены в поле *loan_status* исходного набора данных. Показатель $y^{(i)}$ принимает два значения:

1. Возвратный займ (*Fully Paid*). Такие займы были погашены. Соответствует значению $y^{(i)} = 1$.

2. Невозвратный займ (*Charged Off* или *Default*). Займы, по которым был объявлен дефолт или погашение займа просрочено более чем на 180 дней. Соответствует значению $y^{(i)} = 0$.

Займы со значениями *Current, In Grace period, Late* (16–30 *days*) и *Late* (31–120 *days*) исключаются из анализа, так как однозначно нельзя определить, будут они возвратными или невозвратными.

Также весь набор входных данных разделяется на тренировочный, включающий 0,7 m, и тестовый, включающий 0,3 m займов. Такое разделение необходимо, чтобы была возможность проверить точность прогнозирования на данных, которых нейронная сеть еще не видела.

[3]All Lending Club loan data [Electronic resource]. – Mode of access: https://www.kaggle.com/datasets/wordsforthewise/lending-club. – Date of access: 04.09.2019.

Параметры используемого алгоритма. В рассматриваемом алгоритме используется ряд настраиваемых параметров:

1. $w^{(l)}$, l = 1, …, L, – матрица весов нейронной сети, где вектор-строка $w_k^{(l)}$, k = 1, …, $K^{(l)}$, в свою очередь содержит коэффициенты (числа) $w_{kh}^{(l)}$, h = 1, …, $H^{(l-1)}$, L – количество уровней сети, равное двум (учитываются скрытый и выходной уровни), $K^{(l)}$ – количество нейронов в уровне l, а H – количество элементов в уровне l–1. Отметим, что $H^{(0)} = n$.

2. $b^{(l)}$ – вектор-столбцы, состоящие из значений $b_k^{(l)}$ коэффициентов взвешенного набора сигналов нейрона k, k = 1, …, $K^{(l)}$.

3. Функции активации $a_k^{(l)}\left(x^{(i)}\right)$ элементов (нейронов), k = 1, …, $K^{(l)}$.

Выходные данные. Выходными данными исследуемой бинарной задачи классификации (т. е. определения займа как возвратного или потенциально невозвратного) являются величины $\hat{y}^{(i)} \in \{0, 1\}$, где 1 соответствует возвратному, а 0 – потенциально невозвратному займу i, $i \in \{1, \ldots, m\}$.

Постановка задачи. Нейронная сеть прямого распространения (рис. 1) состоит из входного (нулевого), скрытого (первого) и выходного (второго) уровней. Нулевой уровень характеризуется входными показателями x_1,..,x_n. Скрытый уровень характеризуется набором нейронов и выходной уровень – одним нейроном. Матрицы $w^{(l)}$ и вектор-столбцы $b^{(l)}$ контролируют функциональное преобразование из уровня l в уровень l+1. Так как решаемая задача относится к бинарной классификации, то выходной уровень состоит из одного элемента, который рассчитывает значение функции активации $a_1^{(2)}\left(x^{(i)}\right)$.

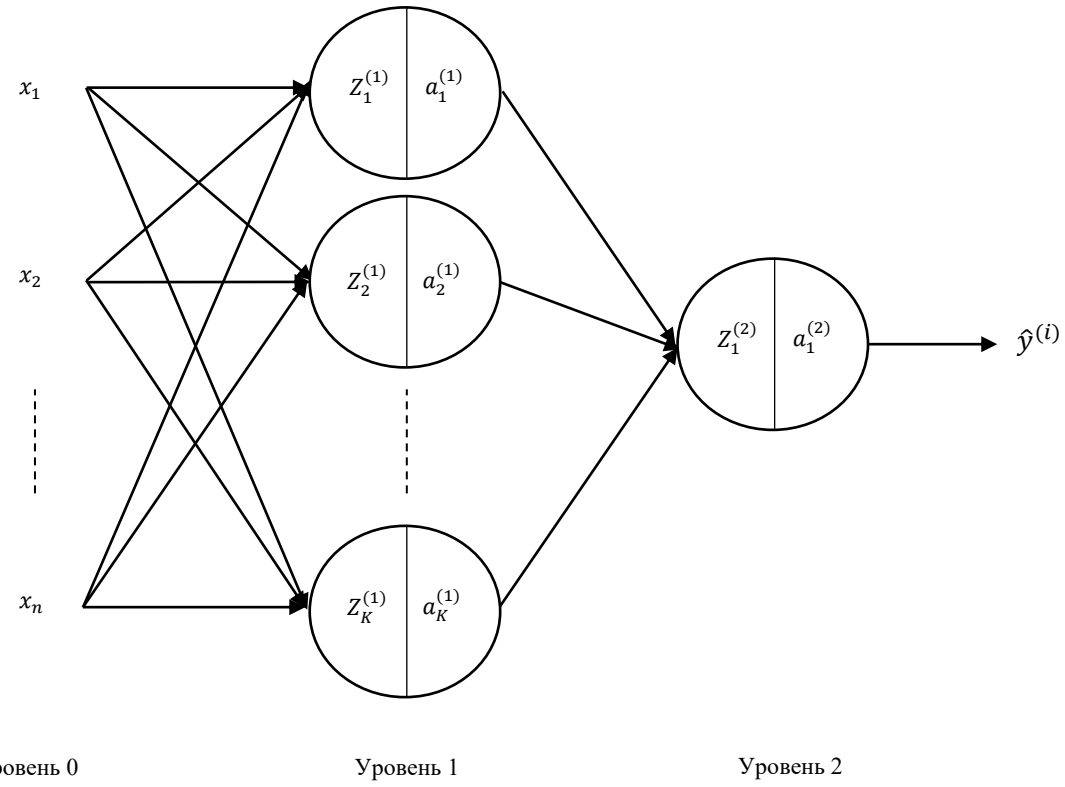


Рис. 1. Нейронная сеть прямого распространения (аргументы в функциях опущены)

Fig. 1. Feed-forward neural network (arguments in functions are omitted)

Принцип действия отдельного нейрона на примере первого нейрона скрытого уровня является следующим [6]: на вход подается набор воздействий и определяется взвешенная сумма $z_1^{(1)}\left(x^{(i)}\right)$ данных сигналов в виде линейной функции с помощью формулы

$$z_1^{(1)}\left(x^{(i)}\right) = b_1^{(1)} + w_{11}^{(1)} \cdot x_1^{(i)} + \cdots + w_{1H}^{(1)} \cdot x_n^{(i)}. \quad (1)$$

На следующем шаге осуществляется расчет функции активации $a_1^{(1)}\left(x^{(i)}\right)$. При использовании логистической регрессии в качестве активации принята сигмовидная функция [1]. В данном случае для нейронов скрытого уровня применяется функция активации гиперболического тангенса $tanh(z)$, которую обозначим как $g(z)$. Данная функция схожа с логистической, но находится в диапазоне от –1 до 1 и пересекает ось координат в значении 0 (рис. 2). Преимуществом данной функции является ее центрирование возле 0, а не около 0,5 в случае с логистической регрессией. Это очень часто приводит к упрощению обучения на следующем уровне. Для расчета данной функции и ее производной используются следующие формулы [7]:

$$g(z) = tanh(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}; \quad (2)$$

$$g'(z) = 1 - g(z)^2. \quad (3)$$

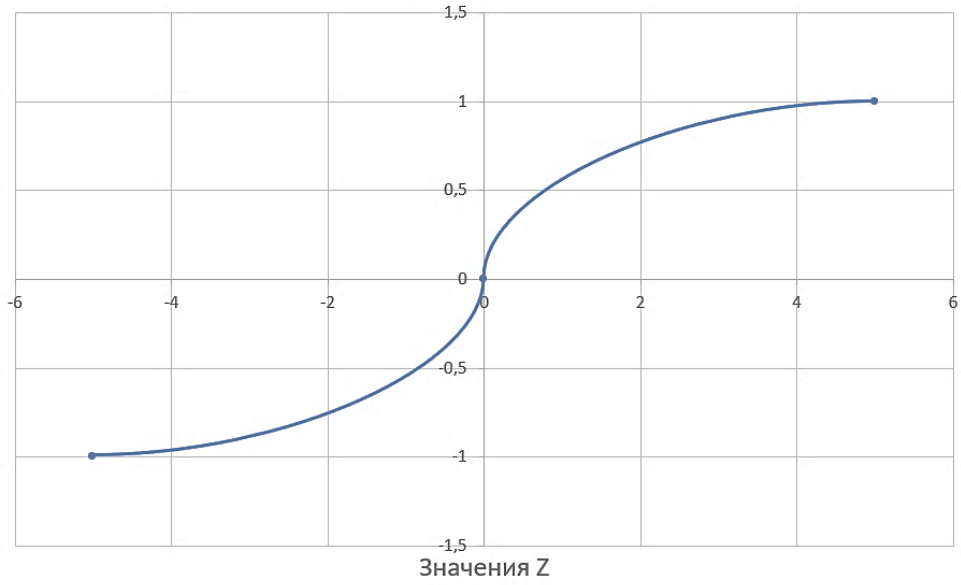


Рис. 2. Функция активации гиперболического тангенса

Fig. 2. Hyperbolic tangent activation function

Функцию активации нейрона k скрытого уровня можно представить в виде

$$a_k^{(1)}\left(x^{(i)}\right) = g\left(Z_k^{(1)}\left(x^{(i)}\right)\right) = \frac{e^{Z_k^{(1)}\left(x^{(i)}\right)} - e^{-Z_k^{(1)}\left(x^{(i)}\right)}}{e^{Z_k^{(1)}\left(x^{(i)}\right)} + e^{-Z_k^{(1)}\left(x^{(i)}\right)}}. \quad (4)$$

В нейроне выходного уровня функция активации определяется иначе – по аналогии с формулой на основе сигмоидной функции σ, которая применяется при использовании логис-

тической регрессии, так как диапазон определяемых значений 0 или 1 является наиболее удобным для выполнения задачи бинарной классификации [8]:

$$a_1^{(2)}\left(x^{(i)}\right) = \sigma\left(Z_1^{(2)}\left(x^{(i)}\right)\right) = \frac{1}{1+e^{-Z_1^{(2)}\left(x^{(i)}\right)}}. \quad (5)$$

Производная данной функции активации выражается формулой [9]

$$\sigma'\left(Z_1^{(2)}\left(x^{(i)}\right)\right) = \sigma\left(Z_1^{(2)}\left(x^{(i)}\right)\right).\left(1 - \sigma\left(Z_1^{(2)}\left(x^{(i)}\right)\right)\right) = a_1^{(2)}\left(x^{(i)}\right).\left(1 - a_1^{(2)}\left(x^{(i)}\right)\right). \quad (6)$$

Таким образом, расчет функции активации нейрона состоит из двух шагов, но с использованием различных функций активации на скрытом и выходном уровнях. Имея более детальное представление о функционировании нейрона, обучение нейронной сети можно представить в виде следующей последовательности шагов:

1. Задание начальных значений весов $w_{kh}^{(l)}$ нейронов и величин $b_k^{(l)}$. В отличие от $b_k^{(l)}$, которые можно принять равными 0, веса $w_{kh}^{(l)}$ не могут быть изначально равны 0, так как в этом случае значения функций активации нейронов на одном уровне будут равны, например, $a_1^{(1)}\left(x^{(i)}\right) = a_2^{(1)}\left(x^{(i)}\right)$ для любого i. Таким образом, нейронная сеть становится неэффективной, так как скрытый уровень осуществляет расчет одной и той же функции активации вне зависимости от их количества в скрытом уровне. Во избежание такой ситуации значения весов $w_{kh}^{(l)}$ задаются как малые величины произвольным образом [10] на основе стандартного нормального распределения со средним значением, равным 0, и стандартным отклонением, равным 1, а также как величины, уменьшенные в 10^{-2} раз.

2. Реализация прямого распространения данных в нейронной сети для расчета $a_1^{(2)}\left(x^{(i)}\right)$. По аналогии с расчетами для отдельного нейрона расчеты, выполняемые на первом уровне нейронной сети (см. рис. 1), можно представить в векторном виде, где T означает транспонирование:

$$z_k^{(1)}\left(x^{(i)}\right) = w_k^{(1)} \cdot x^{(i)\,T} + b_k^{(1)};\; a_k^{(1)}\left(x^{(i)}\right) = g\left(z_k^{(1)}\left(x^{(i)}\right)\right). \quad (7)$$

Данные формулы можно представить и в матричном виде. Если использовать матрицу $w^{(1)}$ и вектор-столбец $b^{(1)}$, то для первого уровня расчет векторов $z^{(1)}\left(x^{(i)}\right)$ и $a^{(1)}\left(x^{(i)}\right)$ выглядит следующим образом:

$$z^{(1)}\left(x^{(i)}\right) = w^{(1)} \cdot x^{(i)T} + b^{(1)};\; a^{(1)}\left(x^{(i)}\right) = g\left(z^{(1)}\left(x^{(i)}\right)\right). \quad (8)$$

Аналогичным образом на втором уровне, содержащем один нейрон, двухшаговый расчет выполняется с помощью формул

$$z_1^{(2)}\left(x^{(i)}\right) = w_1^{(2)} \cdot a^{(1)}\left(x^{(i)}\right) + b_1^{(2)};\; a_1^{(2)}\left(x^{(i)}\right) = \sigma\left(z_1^{(2)}\left(x^{(i)}\right)\right). \quad (9)$$

В связи с тем что второй, выходной, уровень содержит один нейрон, представление в матричном виде данных формул нецелесообразно. Таким образом определена функция активации нейронной сети для одного займа. Аналогично выполняется расчет функций активации $a_1^{(2)}\left(x^{(i)}\right)$ для всех i элементов набора для обучения, где $i \in \{1, \ldots, m\}$.

3. Расчет функции потерь, подлежащей минимизации, от несоответствия значений, рассчитанных функцией активации $a_1^{(2)}(x^{(i)})$, значениям $y^{(i)}$. Данный расчет осуществляется с использованием стоимостной функции нейронной сети по следующей формуле [6]:

$$J(w,b) = -\frac{1}{m}\sum_{i=1}^{m}\left[y^{(i)} \cdot ln\left(a_1^{(2)}(x^{(i)})\right) + (1-y^{(i)}) \cdot ln\left(1 - a_1^{(2)}(x^{(i)})\right)\right]. \quad (10)$$

4. Реализация алгоритма обратного распространения ошибки для вычисления градиента с целью минимизации стоимостной функции $J(w,b)$. Суть алгоритма состоит в расчете дельты (ошибки) при активации каждого нейрона в каждом уровне сети. Для минимизации ошибки используется производная от функции $J(w,b)$, так как она определяет, каким образом изменить входные параметры для требуемого изменения соответствующей функции [11]. В данном случае рассчитывается степень изменения параметров $w_{kh}^{(l)}$ и $b_k^{(l)}$ для получения минимального значения стоимостной функции $J(w,b)$.Так как в общем виде $J(w,b)$ является функцией переменных двух типов, то фактически необходимо определить частные производные двух типов $\frac{\partial J}{\partial w}$ и $\frac{\partial J}{\partial b}$. Стоит отметить, что функция $J(w,b)$ является сложной, поскольку напрямую зависит от функции активации a, в свою очередь зависящей от линейной функции z, которая согласно формуле (1) уже напрямую зависит от w и b. Поэтому обозначенные выше частные производные могут быть найдены с помощью цепного правила [11]:

$$\begin{aligned} \frac{\partial J}{\partial z} &= \frac{\partial J}{\partial a} \cdot \frac{\partial a}{\partial z}; \\ \frac{\partial J}{\partial w} &= \frac{\partial J}{\partial z} \cdot \frac{\partial z}{\partial w}; \\ \frac{\partial J}{\partial b} &= \frac{\partial J}{\partial z} \cdot \frac{\partial z}{\partial b}. \end{aligned} \quad (11)$$

Таким образом, в обратном порядке, начиная с последнего до первого уровня, рассчитываются частные производные функции $J(w,b)$ по a, z, w и b для каждого нейрона в каждом уровне. Как следует из работы [12], частная производная $\frac{\partial J}{\partial a_1^{(2)}(x^{(i)})}$ выражается формулой

$$\frac{\partial J}{\partial a_1^{(2)}(x^{(i)})} = \frac{a_1^{(2)}(x^{(i)}) - y^{(i)}}{a_1^{(2)}(x^{(i)}) \cdot (1 - a_1^{(2)} \cdot (x^{(i)}))}. \quad (12)$$

На основании формулы (6) находится частная производная

$$\frac{\partial a_1^{(2)}(x^{(i)})}{\partial z_1^{(2)}(x^{(i)})} = a_1^{(2)}(x^{(i)}) \cdot (1 - a_1^{(2)} \cdot (x^{(i)})). \quad (13)$$

Далее, используя цепное правило, обозначенное выше, а также формулы (12) и (13), определяется частная производная

$$\frac{\partial J}{\partial z_1^{(2)}(x^{(i)})} = \frac{\partial J}{\partial a_1^{(2)}(x^{(i)})} \cdot \frac{\partial a_1^{(2)}(x^{(i)})}{\partial z_1^{(2)}(x^{(i)})} = a_1^{(2)}(x^{(i)}) - y^{(i)}. \quad (14)$$

Из формулы (9) следует, что частная производная $\frac{\partial z_1^{(2)}(x^{(i)})}{\partial w_1^{(2)(i)}}$ равняется $a^{(1)}\left(x^{(i)}\right)$. Снова используя цепное правило, найдем частную производную

$$\frac{\partial J}{\partial w_1^{(2)(i)}} = \frac{\partial J}{\partial z_1^{(2)}(x^{(i)})} \cdot \frac{\partial z_1^{(2)}(x^{(i)})}{\partial w_1^{(2)(i)}} = a^{(1)}\left(x^{(i)}\right) \cdot \left(a_1^{(2)}\left(x^{(i)}\right) - y^{(i)}\right). \quad (15)$$

Таже из формулы (9) определяется $\frac{\partial z_1^{(2)}(x^{(i)})}{\partial b_1^{(2)(i)}} = 1$. Это позволяет найти частную производную

$$\frac{\partial J}{\partial b_1^{(2)(i)}} = \frac{\partial J}{\partial z_1^{(2)}(x^{(i)})} \cdot \frac{\partial z_1^{(2)}(x^{(i)})}{\partial b_1^{(2)(i)}} = a_1^{(2)}\left(x^{(i)}\right) - y^{(i)}. \quad (16)$$

Таким образом, найдены искомые частные производные $\frac{\partial J}{\partial w_1^{(2)(i)}}$ и $\frac{\partial J}{\partial b_1^{(2)(i)}}$ для второго, выходного, уровня.

Аналогичным образом определяются частные производные $\frac{\partial J}{\partial w^{(1)(i)}}$ и $\frac{\partial J}{\partial b^{(1)(i)}}$ на скрытом уровне. Для этого сразу можно найти частную производную $\frac{\partial J}{\partial z^{(1)}(x^{(i)})}$ также на основе цепного правила:

$$\frac{\partial J}{\partial z^{(1)}(x^{(i)})} = \frac{\partial J}{\partial a_1^{(2)}(x^{(i)})} \cdot \frac{\partial a_1^{(2)}(x^{(i)})}{\partial z_1^{(2)}(x^{(i)})} \cdot \frac{\partial z_1^{(2)}(x^{(i)})}{\partial a^{(1)}(x^{(i)})} \cdot \frac{\partial a^{(1)}(x^{(i)})}{\partial z^{(1)}(x^{(i)})}. \quad (17)$$

В формуле (17) произведение первых двух множителей равняется частной производной $\frac{\partial J}{\partial z_1^{(2)}(x^{(i)})}$, которая была определена в формуле (14). Из формулы (9) следует, что $\frac{\partial z_1^{(2)}(x^{(i)})}{\partial a^{(1)}(x^{(i)})}$ равняется $w_1^{(2)}$. Однако при обратном распространении используются транспонированные матрицы и вектор-строки весов [11], т. е. в данном случает $w_1^{(2)T}$. Из формулы (8) следует, что $\frac{\partial a^{(1)}(x^{(i)})}{\partial z^{(1)}(x^{(i)})}$ уже была представлена ранее в формуле (3). В результате частная производная $\frac{\partial J}{\partial z^{(1)}(x^{(i)})}$ определяется следующим образом:

$$\frac{\partial J}{\partial z^{(1)}(x^{(i)})} = [a_1^{(2)}\left(x^{(i)}\right) - y^{(i)}) \cdot w_1^{(2)T}] * [(1 - a^{(1)}\left(x^{(i)}\right)^2]. \quad (18)$$

Здесь $C * D$ означает поэлементное произведение вектор-столбцов C и D. Используя $\frac{\partial J}{\partial z^{(1)}(x^{(i)})}$ и определив на основе формулы (8) частную производную $\frac{\partial z^{(1)}(x^{(i)})}{\partial w^{(1)(i)}}$ как $x^{(i)}$, находим частную производную

$$\frac{\partial J}{\partial w^{(1)(i)}} = \frac{\partial J}{\partial z^{(1)}(x^{(i)})} \cdot \frac{\partial z^{(1)}(x^{(i)})}{\partial w^{(1)(i)}} = \frac{\partial J}{\partial z^{(1)}(x^{(i)})} \cdot x^{(i)}. \quad (19)$$

Из формулы (8) следует, что $\frac{\partial z^{(1)}(x^{(i)})}{\partial b^{(1)(i)}} = 1$. Соответственно, $\frac{\partial J}{\partial b^{(1)(i)}}$ рассчитывается как

$$\frac{\partial J}{\partial b^{(1)(i)}} = \frac{\partial J}{\partial z^{(1)}(x^{(i)})} \cdot \frac{\partial z^{(1)}(x^{(i)})}{\partial b^{(1)(i)}} = \frac{\partial J}{\partial z^{(1)}(x^{(i)})}. \quad (20)$$

При этом для нулевого уровня производные не рассчитываются, так как входные показатели x_j были заданы и принимаются неизменными.

Таким образом определены значения частных производных $\frac{\partial J}{\partial w_1^{(2)(i)}}$, $\frac{\partial J}{\partial b_1^{(2)(i)}}$, $\frac{\partial J}{\partial w^{(1)(i)}}$ и $\frac{\partial J}{\partial b^{(1)(i)}}$ для одного займа. Аналогично выполняется расчет значений этих частных производных для всех i займов, $i \in \{1, \ldots, m\}$. В конце рассчитываются средние значения $\frac{\partial J}{\partial w_1^{(2)}}$, $\frac{\partial J}{\partial b_1^{(2)}}$, $\frac{\partial J}{\partial w^{(1)}}$ и $\frac{\partial J}{\partial b^{(1)}}$.

5. Использование метода градиентного спуска для нахождения оптимального значения. Все последующие параметры нейронной сети обновляются одновременно с использованием следующих формул на основе данного метода:

$$\begin{aligned} w^{(1)} &:= w^{(1)} - \alpha \cdot \frac{\partial J}{\partial w^{(1)}}; \\ b^{(1)} &:= b^{(1)} - \alpha \cdot \frac{\partial J}{\partial b^{(1)}}; \\ w_1^{(2)} &:= w_1^{(2)} - \alpha \cdot \frac{\partial J}{\partial w_1^{(2)}}; \\ b_1^{(2)} &:= b_1^{(2)} - \alpha \cdot \frac{\partial J}{\partial b_1^{(2)}}. \end{aligned} \quad (21)$$

Здесь параметр α определяет размер шага градиентного спуска.

6. Обучение нейронной сети на тренировочном наборе данных путем многократного (от 1000 до 10 000 в зависимости от эксперимента) повторения шагов 2–5. При обучении на каждой из итераций значение стоимостной функции должно быть меньше, чем на предыдущей. В результате определяются оптимальные значения $w_{kh}^{(l)}$ и $b_k^{(l)}$, а также минимальное значение стоимостной функции $J(w, b)$.

После завершения обучения нейронной сети необходимо рассчитать ее точность при прогнозировании. Для этого с помощью оптимальных величин $w_{kh}^{(l)}$ и $b_k^{(l)}$ и метода прямого распространения нейронной сети на тестовых данных рассчитываются значения $a_1^{(2)}(x^{(i)})$ и определяются $\hat{y}^{(i)} \in \{0, 1\}$ для всех займов с учетом симметричности логистической функции относительно значения 0,5 [6]:

$$a^{(2)}(x^{(i)}) \geq 0{,}5 \rightarrow \hat{y}^{(i)} = 1;$$

$$a^{(2)}(x^{(i)}) < 0{,}5 \rightarrow \hat{y}^{(i)} = 0.$$

Далее требуется провести оценку эффективности данной нейронной сети. Для этого необходимо, используя $\hat{y}^{(i)}$ и $y^{(i)}$, рассчитать четыре основные метрики аналогично подходу, задействованному при использовании логистической регрессии [1]: коэффициенты эффективности *Accuracy* (A), *Precision* (P), *Recall* (R) и меру F_1.

Нормализация исходных данных. Как отмечалось в работе [1], нормализация исходных данных привела к улучшению точности прогнозирования при использовании логистической регрессии. Для оценки эффективности применения того же инструмента средней нормализации для решения задачи классификации займа было проведено обучение исследуемой нейронной сети с использованием 10 000 итераций при равных ее параметрах (с 10 элементами в скрытом уровне и $\alpha = 1$), но с использованием нормализованных и ненормализованных данных, а также произведена оценка эффективности на тестовых данных (табл. 1 и 2).

Таблица 1
Результаты эксперимента при использовании нормализации

Table 1
Experiment results when using normalization

Исследуемый параметр *Parameter under study*	Значение без нормализации *Value without normalization*	Значение с нормализацией *Normalized value*
Средняя длительность обучения одной итерации алгоритма, с	0,31 652	0,27 986
Среднее значение стоимостной функции	0,50 517	0,45 088
Accuracy training, %	79,65 285	80,17 588
Accuracy testing, %	79,72 852	80,21 663

Таблица 2
Ключевые метрики при использовании нормализации на тестовых данных

Table 2
Key metrics when using normalization on test data

Класс *Class*	*Precision*	*Recall*	Мера F_1 *Measure* F_1
Невозвратные займы	0,56 843	0,10 000	0,17 008
Возвратные займы	0,81 081	0,98 070	0,88 770
Средневзвешенное	0,76 168	0,80 217	0,74 223

Из полученных результатов видно, что нормализация привела к улучшению по всем исследуемым параметрам. В дальнейшем целесообразно использовать нормализованные входные данные для исследования нейронной сети прямого распространения.

Классификация займов при разном количестве элементов скрытого уровня и значении коэффициента скорости обучения. Значение коэффициента α влияет на скорость реализации градиентного спуска, что может привести к различным результатам при решении текущей задачи. Однако оптимальное значение также зависит и от количества элементов в скрытом уровне. Поэтому требуется провести обучение данной нейронной сети с одновременным изменением значения параметра α и использованием разного количества элементов скрытого уровня. С учетом имеющихся вычислительных мощностей диапазон исследования для α составит от 1e-4 до 10, а количество элементов скрытого уровня – от 1 до 100. Так как количество уникальных комбинаций 10^7 является очень большим для простого перебора с учетом вычислительных ограничений, то данное исследование целесообразно провести на основе следующих подходов:

1. Двухшаговое исследование. На первом шаге для каждого варианта по количеству элементов в скрытом слое (от 1 до 100) предполагается использовать α из множества (0,0005, 0,005, 0,05, 0,5, 5), т. е. на основе логарифмической шкалы, по которой следующее число получается умножением предыдущего на 10. Логарифмический масштаб выбран в связи с тем, что изменение α с 0,0005 на 0,005 окажет существенно большее влияние на результат обучения нейронной сети, чем изменение α с 0,0005 до 0,0006. При этом выбираются средние значения на каждом из отрезков. По результатам выполнения первого шага выявляется оптимальное количество нейронов в скрытом слое и оптимальная величина α, характеризующая оптимальный отрезок, т. е. комбинация, при которой была получена минимальная стоимостная функция по результатам 10 000 итераций обучения сети. На втором шаге при неизменном и определенном ранее количестве нейронов более детально исследуется отрезок, которому принадлежит оптимальное значение α с целью поиска более оптимальной величины. Данный отрезок делится на 90 равных частей, в результате чего получается новый набор α для исследования. Например, если на первом шаге оптимальной была определена α = 0,5, то на втором шаге будут исследованы па-

раметры α из множества от 0,1 до 1 с шагом 0,01. Таким образом, данный подход изначально определяет диапазон значений, в котором может находиться оптимальное решение, а на втором шаге более детально его исследует. Итоги эксперимента приведены в табл. 3 и 4.

Таблица 3
Результаты двухшагового эксперимента
Table 3
Two-step experiment results

Исследуемый параметр *Parameter under study*	Значение *Value*
Оптимальное количество нейронов в скрытом слое	93
Оптимальный α	0,87
Средняя длительность обучения одной итерации алгоритма, с	2,46 785
Среднее значение стоимостной функции	0,44 926
Accuracy training, %	80,20 301
Accuracy testing, %	80,26 165

Таблица 4
Ключевые метрики при использовании двухшагового подхода на тестовых данных
Table 4
Key metrics when using a two-step approach on test data

Класс *Class*	*Precision*	*Recall*	Мера F_1 *Measure* F_1
Невозвратные займы	0,58 408	0,09 135	0,15 799
Возвратные займы	0,80 977	0,98 346	0,88 820
Средневзвешенное	0,76 402	0,80 262	0,74 018

2. Использование метода Монте-Карло [13]. Суть использования метода для решения текущей задачи состоит в генерации двух случайных чисел для определения количества элементов скрытого уровня и величины α, обучении нейронной сети с помощью сгенерированных параметров и нахождении оптимальных значений по итогам 10 000 имитаций данного процесса. Таким образом, для определения количества элементов разыгрывается случайное число в диапазоне от 0,5 до 100,4 999 и округляется до целого. В результате количество элементов случайным образом определяется из диапазона от 1 до 100. В свою очередь, α определяется также на основе логарифмической шкалы 10^r, где r – также случайно разыгранное число в диапазоне от –4 до 1. В результате α произвольным образом будет присвоено значение из диапазона от 0,0001 до 1. Далее в рамках одной имитации с использованием определенным случайным образом количеством элементов скрытого уровня и α осуществляются обучение нейронной сети на основе 1000 итераций алгоритма градиентного спуска и расчет стоимостной функции. В результате имитационного моделирования, состоящего из 10 000 таких имитаций, определяется оптимальное количество элементов скрытого уровня и α, которые будут соответствовать полученному минимальному значению стоимостной функции. При этом стоит отметить, что чем больше количество имитаций метода Монте-Карло и количество итераций алгоритма градиентного спуска при обучении нейронной сети, тем точнее полученный результат. Значения 10 000 имитаций и 1000 итераций выбраны с учетом имеющихся вычислительных мощностей. Результаты реализации данного подхода отражены в табл. 5 и 6.

Таблица 5
Результаты эксперимента при использовании метода Монте-Карло

Table 5
Experiment results when using Monte-Carlo method

Исследуемый параметр *Parameter under study*	Значение *Value*
Оптимальное количество нейронов в скрытом слое	20
Оптимальный α	1,27 003
Средняя длительность обучения одной итерации алгоритма, с	0,58 571
Среднее значение стоимостной функции	0,45 282
Accuracy training, %	80,08 983
Accuracy testing, %	80,17 107

Таблица 6
Ключевые метрики при использовании метода Монте-Карло на тестовых данных

Table 6
Key metrics when using Monte-Carlo method on test data

Класс *Class*	*Precision*	*Recall*	Мера F_1 *Measure* F_1
Невозвратные займы	0,57 468	0,08 400	0,14 657
Возвратные займы	0,80 864	0,98 419	0,88 782
Средневзвешенное	0,76 121	0,80 171	0,73 756

Проведенное исследование показало, что оптимальными (имеют минимальную стоимостную функцию) являются α=0,87 и количество нейронов в скрытом уровне, равное 93, которые были определены с помощью двухшагового подхода. Однако при этом стоит отметить, что использование метода Монте-Карло потенциально может привести к более оптимальному решению при увеличении количества имитаций и итераций алгоритма градиентного спуска в рамках одной имитации.

Влияние сбалансированности исторических целевых значений на классификацию займов. Как отмечалось выше, набор входных данных содержит 1 221 731 позицию (заявки на займ). Однако возвратным займам соответствует 973 421 (~ 79,7 %) позиция, а невозвратным – 248 310 (~ 20,3 %) позиций. Очевидно, что набор входных данных не сбалансирован по целевым значениям и содержит большинство позиций с возвратными займами.

Так как наличие несбалансированности может влиять на результаты обучения при использовании нейронной сети прямого распространения, необходимо провести исследование влияния сбалансированности входных данных на результаты прогнозирования в рамках текущей задачи. По аналогии с подходом, примененным при исследовании логистической регрессии [1], из входного набора данных создается подмассив данных [14, с. 148], состоящий из всех 248 310 позиций входных данных, соответствующих невозвратным займам, и как следствие – только из 248 310 позиций, соответствующих возвратным займам. В результате набор входных данных в подмассиве будет сбалансирован, но общее количество позиций снизится до 496 620.

Итоги компьютерного исследования, состоящего из 10 000 итераций градиентного спуска при α= 0,87, представлены в табл. 7 и 8.

Таблица 7
Итоги исследования при сбалансированности исторических целевых значений
Table 7
Research results when the historical target values are balanced

Результаты 10 000 итераций при α = 0,87 *Results of 10 000 iterations with* α = 0,87	Значения *Values*
Оптимальное количество нейронов в скрытом слое	93
Средняя длительность обучения одной итерации алгоритма, с	1,03 051
Среднее значение стоимостной функции	0,60 709
Accuracy training, %	66,51 794
Accuracy testing, %	66,12 701

Таблица 8
Ключевые метрики при использовании сбалансированных данных
Table 8
Key metrics when using balanced data

Результаты 10 000 итераций при α = 0,87 *Results of 10 000 iterations with* α = 0,87	*Precision*	*Recall*	Мера F_1 *Measure* F_1
Невозвратные займы	0,67 295	0,63 072	0,65 116
Возвратные займы	0,65 092	0,69 197	0,67 082
Средневзвешенное	0,66 196	0,66 127	0,66 096

Как видно из полученных результатов, абсолютная сбалансированность не привела к улучшению значения стоимостной функции, величин A и меры F_1 модели при использовании нейронной сети прямого распространения в задаче классификации займа. Уменьшение точности прогнозирования и увеличение значения стоимостной функции вызваны значительным уменьшением набора входных позиций с 1 221 731 до 496 620 в связи с намерением сбалансировать набор входных данных. Если рассматривать метрики P, R и F_1 для невозвратных займов, то стоит отметить их улучшение. Из этого следует, что применять сбалансированные входные данные целесообразно в случае, когда точность прогнозирования невозвратных займов более важна, чем возвратных. Учитывая, что величины A и F_1 всей модели оказались хуже значений, полученных при отсутствии сбалансированности, в дальнейшем будет использован весь набор входных данных, состоящий из 1 221 731 позиции.

Классификация займов при различных граничных значениях. Так как в элементе выходного уровня в качестве функции активации используется логистическая регрессия, то изначально значение 0,5 было выбрано для классификации займа как возвратного или невозвратного [6]. Однако, как было отмечено ранее, оптимальное граничное значение может несколько отличаться от 0,5. Для определения лучшего пограничного значения требуется провести расчет влияния разных граничных значений от 0,01 до 1 с шагом 0,01 на точность прогнозирования на основе сравнения A при классификации займа с помощью нейронной сети прямого распространения. Предполагается проведение 10 000 итераций обучения данной нейронной сети и нахождение оптимального пограничного значения при одинаковом определенном ранее значении стоимостной функции. По результатам обучения находится оптимальное пограничное значение, которому соответствует наибольшая точность прогнозирования, выраженная значением A на тестовых данных. В результате оптимальное (максимальное) значение A получено при граничном значении 0,51 (табл. 9).

Таблица 9
Итоги исследования при оптимальном граничном значении

Table 9
Research results at the optimal boundary value

Результаты 10 000 итераций при α = 0,87 *Results of 10 000 iterations with* α = 0,87	Значения *Values*
Accuracy training, %	80,21 938
Accuracy testing, %	80,26 520

Как следует из результатов исследования, полученная на тестовых данных точность в некоторой степени больше точности 80,26 165 %, рассчитанной при использовании пограничного значения 0,5 с помощью двухшагового исследования. Поэтому при дальнейшем анализе данного алгоритма машинного обучения будет применяться граничное значение 0,51.

При этом данные из табл. 10 показали, что значения метрик P, R и F_1 также изменились. В частности, значение F_1 всей модели улучшилось до 0,74 276 по сравнению с 0,74 018.

Таблица 10
Ключевые метрики при оптимальном граничном значении

Table 10
Key metrics at the optimal boundary value

Результаты 10 000 итераций при α = 0,87 *Results of 10 000 iterations with* α = 0,87	*Precision*	*Recall*	Мера F_1 *Measure* F_1
Невозвратные займы	0,57 557	0,10 082	0,17 159
Возвратные займы	0,81 101	0,98 110	0,88 798
Средневзвешенное	0,76 328	0,80 265	0,74 276

Таким образом, обнаружено, что исследование различных граничных значений в задаче классификации займа при использовании нейронной сети прямого распространения является целесообразным.

Классификация займов с использованием разных функций активации в скрытых уровнях. Как отмечено в постановке текущей задачи, до текущего момента во всех элементах скрытого уровня исследуемой нейронной сети использовалась функция активации гиперболического тангенса. Однако часто альтернативные функции активации могут привести к лучшим результатам. При этом в задачах бинарной классификации в выходном уровне логистическая функция активации остается неизменной. В качестве альтернативной наиболее часто используется функция активации ReLu, которую обозначим как $f(z)$ (рис. 3). Функция является очень популярной, особенно при обучении глубоких нейронных сетей, и имеет ряд преимуществ над логистической и функцией гиперболического тангенса:

1. Производная функции рассчитывается проще и равняется 1 для положительных значений $z_k^{(1)}\left(x^{(i)}\right)$ и 0 для отрицательных.

2. При достаточно больших значениях $z_k^{(1)}\left(x^{(i)}\right)$ крутизна логистической функции и гиперболического тангенса приближается к 0, что существенно замедляет выполнение алгоритма градиентного спуска. Напротив, функция ReLu для всех значений $z_k^{(1)}\left(x^{(i)}\right)$ больше 0 имеет крутизну (в то же время и производную) функции, равную 1, и только для значений меньше 0 крутизна функции равняется 0 (рис. 3). Такой подход на практике существенно ускоряет процесс выполнения алгоритма градиентного спуска.

Функция ReLu рассчитывается на основе формул [15]

$$f(z) := \begin{cases} 0 \text{ для } Z < 0, \\ Z \text{ для } Z \geq 0; \end{cases} \quad (22)$$

$$f(z)' = \begin{cases} 0 \text{ для } Z < 0, \\ 1 \text{ для } Z \geq 0. \end{cases} \quad (23)$$

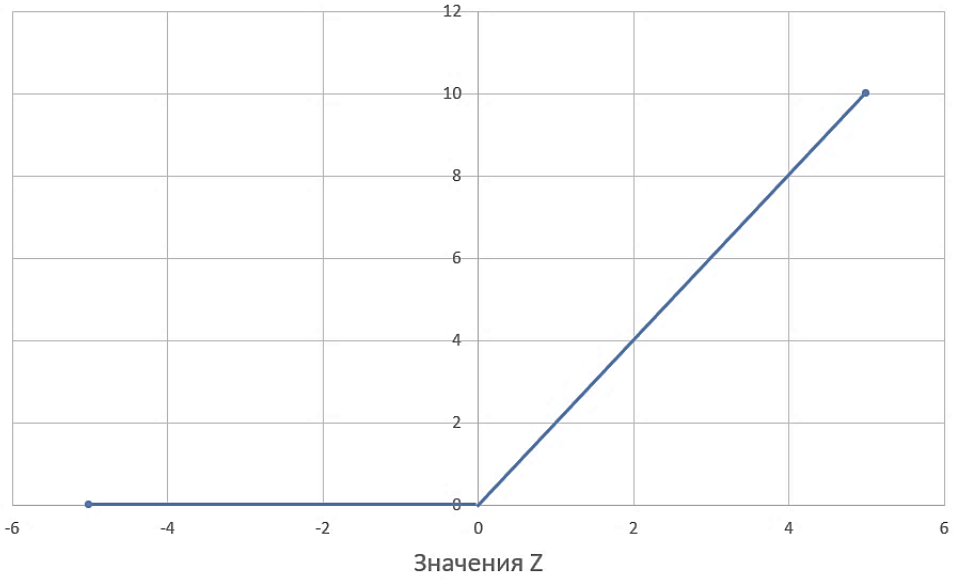


Рис. 3. Функция активации ReLu

Fig. 3. ReLu activation function

Несмотря на имеющиеся преимущества функции активации ReLu над функцией гиперболического тангенса, при решении конкретной задачи оптимальной может оказаться любая из них. Поэтому необходимо провести анализ влияния функции активации ReLu на результаты с ее использованием в элементах скрытого уровня при решении задачи классификации займа. В результате проведенного эксперимента получены следующие данные (табл. 11 и 12).

Таблица 11
Итоги исследования при использовании функции активации ReLu

Table 11
Research results when using ReLu activation function

Результаты 10 000 итераций при α = 0,87 *Results of 10 000 iterations with* α = 0,87	Значения *Values*
Средняя длительность обучения одной итерации алгоритма, с	1,70 897
Среднее значение стоимостной функции	0,44 745
Accuracy training, %	80,30 416
Accuracy testing, %	80,27 420

Таблица 12
Ключевые метрики при использовании функции активации ReLu
Table 12
Key metrics when using ReLu activation function

Результаты 10 000 итераций при α = 0,87 *Results of 10 000 iterations with* α = 0,87	*Precision*	*Recall*	Мера F_1 *Measure* F_1
Невозвратные займы	0,56 557	0,11 610	0,19 265
Возвратные займы	0,81 304	0,97 733	0,88 765
Средневзвешенное	0,76 287	0,80 274	0,74 676

Из табл. 11 и 12 следует, что среднее значение стоимостной функции 0,44 745, полученное при использовании ReLu в качестве функции активации, ниже (лучше) значения 0,44 926, полученного при использовании функции гиперболического тангенса. Также коэффициент эффективности *A* на тестовых данных показал большее (лучшее) значение. Поэтому в дальнейших исследованиях целесообразно применять именно ReLu вместо функции гиперболического тангенса в элементах скрытого уровня.

Классификация займов при увеличении входных показателей. Как было отмечено в предыдущем исследовании [1], набор входных показателей можно расширить путем включения в исследуемый набор данных тех показателей, которые имеют до 30 % отсутствующих значений во всем перечне выданных займов, а недостающие значения заполнить поочередно на модальное, среднее и медианное значение соответствующего параметра. Реализация данного подхода привела к улучшению результатов прогнозирования при использовании логистической регрессии. В связи с этим целесообразно провести аналогичные исследования при обучении нейронной сети прямого распространения.

Такими дополнительными входными показателями являются:

1. *mths_since_last_delinq* – количество месяцев с момента последней просрочки.
2. *mths_since_last_record* – количество месяцев с момента последней публичной записи.
3. *open_acc_6m* – количество открытых кредитных счетов за последние шесть месяцев.
4. *open_act_il* – количество текущих активных счетов с рассрочкой платежа.
5. *open_il_12m* – количество счетов с рассрочкой платежа, открытых за последние 12 месяцев.
6. *open_il_24m* – количество счетов с рассрочкой платежа, открытых за последние 24 месяца.
7. *mths_since_rcnt_il* – количество месяцев с момента открытия последнего счета с рассрочкой платежа.
8. *total_bal_il* – текущий баланс по всем счетам с рассрочкой платежа.
9. *il_util* – соотношение суммарного текущего баланса к кредитному лимиту по всем счетам с рассрочкой.
10. *open_rv_12m* – количество револьверных счетов, открытых за последние 12 месяцев.
11. *open_rv_24m* – количество револьверных счетов, открытых за последние 24 месяца.
12. *max_bal_bc* – максимальный текущий баланс задолженности по всем револьверным счетам.
13. *all_util* – соотношение баланса к кредитному лимиту по всем счетам.
14. *inq_fi* – количество персональных финансовых запросов.
15. *total_cu_tl* – количество финансовых счетов.
16. *inq_last_12m* – количество запросов на кредит за последние 12 месяцев.
17. *mo_sin_old_il_acct* – количество месяцев со времени открытия самого старого счета с рассрочкой платежа.
18. *mths_since_recent_inq* – количество месяцев с момента последнего запроса.
19. *percent_bc_gt_75* – процент всех счетов по банковским картам, которые превышают 75 % лимита.

При проведении исследования были использованы модальные, средние и медианные значения соответствующих дополнительных параметров для устранения пустых позиций и рас-

считана точность прогнозирования для каждого случая при решении задачи классификации займа. Результаты представлены в табл. 13 и 14.

Таблица 13
Результаты исследования при увеличении количества входных показателей
Table 13
Research results with an increase in the number of input features

Результаты 10 000 итераций при α = 0,87 *Results of 10 000 iterations with* α = 0,87	Заполнение модальными значениями *Filling with modal values*	Заполнение средними значениями *Filling with averages*	Заполнение медианными значениями *Filling with median values*
Средняя длительность обучения одной итерации алгоритма, с	1,72 914	1,70 839	1,71 910
Среднее значение стоимостной функции	0,44 311	0,44 328	0,44 348
Accuracy training, %	80,42 752	80,43 290	80,43 009
Accuracy testing, %	80,30 721	80,27 475	80,32 468

Таблица 14
Ключевые метрики при заполнении медианными значениями
Table 14
Key metrics when filled with median values

Результаты 10 000 итераций при α = 0,87 *Results of 10 000 iterations with* α = 0,87	*Precision*	*Recall*	Мера F_1 *Measure* F_1
Невозвратные займы	0,56 451	0,12 867	0,20 957
Возвратные займы	0,81 481	0,97 476	0,88 764
Средневзвешенное	0,76 407	0,80 325	0,75 018

Из полученных результатов видно, что увеличение входных показателей положительно влияет на улучшение результатов обучения глубокой нейронной сети: среднее значение стоимостной функции уменьшилось при заполнении модальными, медианными и средними значениями по сравнению со значением 0,44 745, определенным ранее. При этом увеличение входных данных с помощью использования модальных значений приводит к получению минимальной стоимостной функции. Однако значение A на тестовых данных является максимальным при использовании медианных значений. Так как максимизация значения A на тестовых данных является более важной, чем минимизация стоимостной функции, то в дальнейших исследованиях будет использован увеличенный с помощью медианных значений набор входных данных. При этом стоит отметить, что при выборе иного параметра для оптимизации, например меры F_1, оптимальным вариантом может быть использование увеличенного с помощью средних или модальных значений набора входных данных.

Классификация займов при использовании полиномиальных показателей. Как описано в исследовании [1], дополнительным вариантом расширения количества параметров во входных данных является использование полиномиальных показателей. Реализация данного подхода привела к улучшению результатов прогнозирования для логистической регрессии. Следовательно, целесообразно провести аналогичные исследования при решении текущей задачи с использованием нейронной сети прямого распространения. Для этого также требуется расширить набор входных показателей $x_j^{(i)}$ с использованием полинома от второй до четвертой степени.

Результаты текущего эксперимента, состоящего из 10 000 итераций для каждой степени полинома, представлены в табл. 15 и 16.

Таблица 15
Результаты исследования при использовании полиномиальных показателей
Table 15
Research results when using polynomial features

Результаты 10 000 итераций при α = 0,87 *Results of 10 000 iterations with* α = 0,87	При полиноме второй степени *With a polynomial of the second degree*	При полиноме третьей степени *With a polynomial of the third degree*	При полиноме четвертой степени *With a polynomial of the fourth degree*
Средняя длительность обучения одной итерации алгоритма, с	2,65 637	2,70 170	3,05 421
Среднее значение стоимостной функции	0,44 208	0,47 962	0,50 518
Accuracy training, %	80,43 231	80,02 072	79,65 286
Accuracy testing, %	80,26 602	80,00 409	79,72 853

Таблица 16
Ключевые метрики при использовании полинома второй степени
Table 16
Key metrics when using of the second degree polynomial

Результаты 10 000 итераций при α = 0,87 *Results of 10 000 iterations with* α = 0,87	*Precision*	*Recall*	Мера F_1 *Measure* F_1
Невозвратные займы	0,58 538	0,09 089	0,15 735
Возвратные займы	0,80 972	0,98 363	0,88 824
Средневзвешенное	0,76 424	0,80 266	0,74 008

Из данных табл. 15 и 16 следует, что использование полиномиальных показателей привело к более оптимальному (меньшему) значению 0,44 208 стоимостной функции при полиноме второй степени по сравнению с величиной, полученной при использовании других полиномов и определенной ранее. Однако, как видно из эксперимента, значение коэффициента эффективности A, рассчитанное на тренировочных данных, и мера F_1 при полиноме второй степени хуже значений, полученных без использования полиномов. Так как данные величины являются более важными, чем стоимостная функция, то использование полиномов в дальнейших исследованиях будет нецелесообразным при α = 0,87. Между тем при других значениях α результаты могут измениться.

Использование метода главных компонент для задачи классификации займа. Проведенные исследования позволяют сделать заключение, что увеличение количества входных показателей с 54 до 73 улучшило метрики исследуемой модели. Однако дальнейшее увеличение количества показателей с применением полиномов выразилось лишь в улучшении стоимостной функции только при полиноме второй степени, но привело к ухудшению коэффициента эффективности A на тестовых данных и меры F_1. Поэтому целесообразно провести анализ имеющегося набора входных показателей на избыточность. Для этого, как и в случае исследования логистической регрессии [1] ранее, будет использован метод главных компонент (principal component analysis) [14, с. 269–279; 16]. Основой данного метода является уменьшение линейной размерности с использованием разложения по сингулярным значениям. С целью расчета главных компонент (начиная с первого и до заданного количества) будет применен класс *sklearn.decomposition.PCA*[4]. В рамках данного эксперимента главные компоненты используются в диапазоне от 1 до 73 включительно с целью выявления оптимального количества, которое обеспечит максимальное значение коэффициента эффективности A на тестовых данных.

Согласно полученным результатам оптимальное количество главных компонент было 69, значения остальных метрик представлены в табл. 17 и 18.

[4]Sklearn.decomposition.PCA [Electronic resource]. – Mode of access: https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.PCA.html. – Date of access: 22.12.2022.

Таблица 17
Результаты исследования при использовании метода главных компонент

Table 17
Results of the study using the method of principal components analysis

Результаты 10 000 итераций при α = 0,87 *Results of 10 000 iterations with* α = 0,87	Значения *Values*
Средняя длительность обучения одной итерации алгоритма, с	6,96 205
Среднее значение стоимостной функции	0,44 407
Accuracy training, %	80,39 618
Accuracy testing, %	80,30 640

Таблица 18
Ключевые метрики при использовании метода главных компонент

Table 18
Key metrics when using principal component analysis

Результаты 10 000 итераций при α = 0,87 *Results of 10 000 iterations with* α = 0,87	*Precision*	*Recall*	Мера F_1 *Measure* F_1
Невозвратные займы	0,56 286	0,12 762	0,20 806
Возвратные займы	0,81 464	0,97 480	0,88 755
Средневзвешенное	0,76 360	0,80 306	0,74 981

Как следует из полученных результатов, применение метода главных компонент не привело к увеличению значения коэффициента эффективности $A = 80{,}30\,640\,\%$. Данное значение меньше полученного ранее 80,32 468 % при заполнении медианными значениями. Поэтому можно сделать вывод, что использование метода главных компонент нецелесообразно, если приоритетом является оптимизация коэффициента эффективности A. Однако ситуация может измениться, если в качестве оптимизации будет выбрана другая метрика.

Сравнение результатов при использовании нейронной сети прямого распространения и логистической регрессии для решения задачи классификации займа. В рамках настоящего исследования наибольшее значение коэффициента эффективности A на тестовых данных было получено при увеличении входных параметров с использованием медианных значений (см. табл. 13). Целесообразно сравнить полученные результаты со значениями, рассчитанными при использовании логистической регрессии [1] и соответствующими максимальному значению коэффициента эффективности A на тестовых данных, который был получен с помощью метода главных компонент (табл. 19).

Таблица 19
Сравнение оптимальных результатов при применении нейронной сети прямого распространения и логистической регрессии

Table 19
Comparison of optimal results when applying a feed-forward neural network and logistic regression

Результаты 10 000 итераций *Results of 10 000 iterations*	При нейронной сети прямого распространения *With a feed-forward neural network*	При логистической регрессии *With logistic regression*
Средняя длительность обучения одной итерации алгоритма, с	1,71 910	0,20 632
Среднее значение стоимостной функции	0,44 348	0,45 706
Accuracy training, %	80,43 009	79,93 408
Accuracy testing, %	80,32 468	80,04 065

Из табл. 19 следует, что с помощью нейронной сети прямого распространения были получены бóльшие (лучшие) значения коэффициента эффективности A и меньшее (лучшее) среднее значение стоимостной функции. Однако стоит отметить, что средняя длительность обучения одной итерации алгоритма нейронной сети прямого распространения существенно выше аналогичной величины для логистической регрессии.

Значения меры F_1 (табл. 20) выше при использовании нейронной сети прямого распространения по невозвратным и возвратным займам, а также при расчете средневзвешенной величины.

Таблица 20
Мера F_1 при применении нейронной сети прямого распространения и логистической регрессии

Table 20
Measure F_1 when applying a feed-forward neural network and logistic regression

Результаты 10 000 итераций *Results of 10 000 iterations*	Мера F_1 при нейронной сети прямого распространения *F_1 measure with a feed-forward neural network*	Мера F_1 при логистической регрессии *F_1 measure with logistic regression*
Невозвратные займы	0,20 957	0,15 201
Возвратные займы	0,88 764	0,88 689
Средневзвешенное	0,75 018	0,73 792

Таблица 21
Метрика *Precision* при применении нейронной сети прямого распространения и логистической регрессии

Table 21
Precision metric when applying a feed-forward neural network and logistic regression

Результаты 10 000 итераций *Results of 10 000 iterations*	*Precision* при нейронной сети прямого распространения *Precision with a feed-forward neural network*	*Precision* при логистической регрессии *Precision measure with logistic regression*
Невозвратные займы	0,56 451	0,54 779
Возвратные займы	0,81 481	0,80 894
Средневзвешенное	0,76 407	0,75 600

Таблица 22
Метрика *Recall* при применении нейронной сети прямого распространения и логистической регрессии

Table 22
Recall metric when applying a feed-forward neural network and logistic regression

Результаты 10 000 итераций *Results of 10 000 iterations*	*Recall* при нейронной сети прямого распространения *Recall measure with a feed-forward neural network*	*Recall* при логистической регрессии *Recall measure with logistic regression*
Невозвратные займы	0,12 867	0,08 825
Возвратные займы	0,97 476	0,98 148
Средневзвешенное	0,80 325	0,80 041

Как следует из табл. 21, значение метрики *Precision* улучшилось наиболее существенно по невозвратным займам при использовании нейронной сети. В то же время, исходя из данных табл. 22, значение метрики *Recall* улучшилось только по невозвратным займам, но на 45,80 169 %.

Заключение. В работе исследовано применение нейронной сети прямого распространения для решения задачи классификации займа. Обнаружено, что использование нормализации улучшает точность прогнозирования. Также выявлено, что поиск оптимального числа нейронов

в скрытом уровне и оптимального значения α привел к улучшению стоимостной функции и коэффициента эффективности A. Однако абсолютная сбалансированность целевых значений не привела к улучшению конечных результатов выбранных метрик. При этом установлено, что оптимальным граничным значением для выходного уровня данной нейронной сети является 0,51 вместо используемого по умолчанию 0,5. Было определено, что применение функции активации ReLu привело к улучшению результатов классификации. Обнаружено также, что увеличение показателей в наборе входных данных и их преобразование с помощью средних, модальных и медианных значений послужили улучшению стоимостной функции и коэффициента эффективности A. В то же время применение полиномов и метода главных компонент не привело к увеличению коэффициента эффективности A на тестовых данных. Оптимальные значения коэффициента эффективности A, меры F_1 и метрики *Precision* оказались выше значений, рассчитанных с помощью логистической регрессии.

Список использованных источников

1. Бегунков, В. И. Классификация займов с использованием логистической регрессии / В. И. Бегунков, М. Я. Ковалев // Информатика. – 2023. − Т. 20, № 1. – С. 55–74. https://doi.org/10.37661/1816-0301-2023-20-1-55-74

2. Murati, A. Disruption in European consumer finance: Lessons from Sweden [Electronic resource] / A. Murati, O. Skau, Z. Taraporevala // McKinsey Quarterly. – 2018. – Mode of access: https://www.mckinsey.com/industries/financial-services/our-insights/disruption-in-european-consumer-finance-lessons-from-sweden. – Date of access: 01.06.2021.

3. Hand, D. J. Statistical classification methods in consumer credit scoring: a review / D. J. Hand, W. E. Henley // J. of the Royal Statistical Society: Series A (Statistics in Society). – 1997. – Vol. 160, no. 3. – P. 523–541.

4. Kombe, S. K. Effects of internet banking on the financial performance of commercial banks in Kenya a case of Kenya Commercial Bank / S. K. Kombe, M. K. Wafula // Intern. J. of Scientific and Research Publications. – 2015. – Vol. 5, no. 5. – P. 1–10.

5. Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research / S. Lessmann [et al.] // European J. of Operational Research. – 2015. – Vol. 247, iss. 1. – P. 124–136.

6. Geron, A. Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow / A. Geron. – 2nd ed. – O'Reilly Media, 2019. – P. 205–207, 370–371.

7. Oldham, K. An Atlas of Functions: with Equator, the Atlas Function Calculator / K. Oldham, J. Myland, J. Spanier. – 2nd ed. – Springer Science + Business Media, 2009. – P. 289–290.

8. Shalev-Shwartz, S. Understanding Machine Learning: From Theory to Algorithms / S. Shalev-Shwartz, S. Ben-David. – Cambridge University Press, 2014. – P. 125–127.

9. Raschka, S. Python Machine Learning / S. Raschka, V. Mirjalili. – 3d ed. – Packt Publishing Ltd., 2019. – P. 65, 415–421.

10. Rumelhart, D. Learning representations by back-propagating errors / D. Rumelhart, G. Hinton, R. Williams // Nature. – 1986. – Vol. 323. – P. 533–536.

11. Goodfellow, I. Deep Learning / I. Goodfellow, Y. Bengio, A. Courville. – MIT Press, 2016. – P. 82–86, 205.

12. Bishop, C. Neural Networks for Pattern Recognition / C. Bishop. – Clarendon Press Oxford, 1995. – P. 231.

13. Metropolis, N. The Monte Carlo method / N. Metropolis, S. Ulam // J. of the American Statistical Association. – 1949. – Vol. 44, no. 247. – P. 335–341.

14. Harrington, P. Machine Learning in Action / P. Harrington. – 1st ed. – Manning Publication Co., 2012. – P. 148, 269–279.

15. Glorot, X. Deep sparse rectifier neural networks / X. Glorot, A. Bordes, Y. Bengio // Proc. of the 14th Intern. Conf. on Artificial Intelligence and Statistics, Fort Lauderdale, FL, USA, 11–13 Apr. 2011. – Fort Lauderdale, 2011. – Vol. 15. – P. 315–323.

16. Murphy, K. P. Machine Learning: A Probabilistic Perspective (Adaptive Computation and Machine Learning series) / K. P. Murphy. – The MIT Press, 2012. – P. 387–407.

References

1. Behunkou U. I., Kovalyov M. Y. *Loan classification using logistic regression*. Informatika [*Informatics*], 2023, vol. 20, no. 1, pp. 55–74 (In Russ.). https://doi.org/10.37661/1816-0301-2023-20-1-55-74

2. Murati A., Skau O., Taraporevala Z. Disruption in European consumer finance: Lessons from Sweden. *McKinsey Quarterly*, 2018. Available at: https://www.mckinsey.com/industries/financial-services/our-insights/disruption-in-european-consumer-finance-lessons-from-sweden (accessed 01.06.2021).

3. Hand D. J., Henley W. E. Statistical classification methods in consumer credit scoring: a review. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 1997, vol. 160, no. 3, pp. 523–541.

4. Kombe S. K., Wafula M. K. Effects of internet banking on the financial performance of commercial banks in Kenya a case of Kenya Commercial Bank. *International Journal of Scientific and Research Publications*, 2015, vol. 5, no. 5, pp. 1–10.

5. Lessmann S., Baesens B., Seow H.-V., Thomas L. C. Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. *European Journal of Operational Research*, 2015, vol. 247, iss. 1, pp. 124–136.

6. Geron A. *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow*, 2nd edition. O'Reilly Media, 2019, pp. 205–207, 370–371.

7. Oldham K., Myland J., Spanier J. *An Atlas of Functions: with Equator, the Atlas Function Calculator*, 2nd edition. Springer Science + Business Media, 2009, pp. 289–290.

8. Shalev-Shwartz S., Ben-David S. *Understanding Machine Learning: From Theory to Algorithms*. Cambridge University Press, 2014, pp. 125–127.

9. Raschka S., Mirjalili V. *Python Machine Learning*, 3d edition. Packt Publishing Ltd., 2019, pp. 65, 415–421.

10. Rumelhart D., Hinton G., Williams R. Learning representations by back-propagating errors. *Nature*, 1986, vol. 323, pp. 533–536.

11. Goodfellow I., Bengio Y., Courville A. *Deep Learning*, MIT Press, 2016, pp. 82–86, 205.

12. Bishop C. *Neural Networks for Pattern Recognition*. Clarendon Press Oxford, 1995, p. 231.

13. Metropolis N., Ulam S. The Monte Carlo method. *Journal of the American Statistical Association*, 1949, vol. 44, no. 247, pp. 335–341.

14. Harrington P. *Machine Learning in Action*, 1st edition. Manning Publication Co., 2012, pp. 148, 269–279.

15. Glorot X., Bordes A., Bengio Y. Deep sparse rectifier neural networks. *Proceedings of the 14th International Conference on Artificial Intelligence and Statistics, Fort Lauderdale, FL, USA, 11–13 April 2011*. Fort Lauderdale, 2011, vol. 15, pp. 315–323.

16. Murphy K. P. *Machine Learning: A Probabilistic Perspective (Adaptive Computation and Machine Learning series)*. The MIT Press, 2012, pp. 387–407.

Информация об авторе

Бегунков Владимир Иванович, магистр технических наук.
E-mail: vbegunkov@gmail.com

Information about the author

Uladzimir I. Behunkou, M. Sc. (Eng.).
E-mail: vbegunkov@gmail.com

УДК 519.684.6;004.021
https://doi.org/10.37661/1816-0301-2024-21-1-105-120

Оригинальная статья
Original Paper

Система комплексного анализа данных тематических сайтов ИСКАД ИИ

И. И. Пилецкий✉, М. П. Батура, Н. А. Волорова, П. А. Зорко, А. О. Кулевич

Белорусский государственный университет информатики и радиоэлектроники,
ул. П. Бровки, 6, Минск, 220013, Беларусь
✉*E-mail: ianmenski@gmail.com*

Аннотация

Цели. В настоящее время основным источником получения информации является Интернет. Огромный объем информации, доступной в сети, делает актуальной задачу всестороннего анализа данных из открытых интернет-источников. Цель работы заключается в создании многоцелевого, модифицируемого кластера для глубокого анализа данных интернет-источников, основными задачами которого являются выявление наиболее важных публикаций в некоторой предметной области и их тематический анализ, определение лидера научного направления и тенденций развития направлений деятельности и взаимодействия групп людей.

Методы. Для решения поставленной задачи была разработана методология построения многоцелевого кластера с использованием технологий быстрого построения тематической графовой базы данных, графа знаний, методов и моделей машинного обучения для глубокого анализа данных.

Результаты. Разработана Система комплексного анализа данных тематических сайтов ИСКАД ИИ, апробированы методология быстрого построения тематической графовой базы данных и комплексная технология глубокого анализа данных интернет-источников и известных мировых сайтов.

Заключение. Создана среда информационных технологий для быстрого построения тематических графовых баз данных. Результаты применения технологии быстрого построения графовых баз данных показаны на примерах работы ИСКАД ИИ.

Ключевые слова: тематические сайты, большие данные, метод машинного обучения, анализ данных, графовая база данных, граф знаний, база данных Neo4j

Для цитирования. Система комплексного анализа данных тематических сайтов ИСКАД ИИ / И. И. Пилецкий [и др.] // Информатика. – 2024. – Т. 21, № 1. – С. 105–120.
https://doi.org/10.37661/1816-0301-2024-21-1-105-120

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Поступила в редакцию | Received 08.12.2023
Подписана в печать | Accepted 12.01.2024
Опубликована | Published 29.03.2024

System of complex data analysis of thematic sites ISCAD IS

Ivan I. Piletski✉, Michal P. Batura, Natalia A. Volorova, Polina A. Zorko, Alexei O. Kulevich

Belarusian State University
of Informatics and Radioelectronics,
st. P. Brovki, 6, Minsk, 220013, Belarus
✉*E-mail: ianmenski@gmail.com*

Abstract

Objectives. Currently, the main source of information is the Internet. The huge amount of information available on the Internet makes it urgent to comprehensively analyze data from open Internet sources. The goal of this work is to create a multi-purpose, modifiable cluster for in-depth analysis of data from Internet sources, the main objectives of which are to identify the most important publications in a certain subject area, thematic analysis of these publications, identifying the leader of a scientific direction and determining trends in the development of areas and interaction of groups of people.

Methods. To solve this problem, a methodology was developed for constructing a multi-purpose cluster using technologies for quickly constructing a thematic graph database, a knowledge graph, methods and models of machine learning for in-depth analysis of data.

Results. A system for comprehensive analysis of data from thematic sites ISKAD IS has been developed, a methodology for quickly constructing a thematic graph database and a comprehensive technology for in-depth analysis of data from Internet sources and analysis of data from the most important well-known world sites have been tested.

Conclusion. An IT environment has been created for the rapid construction of thematic graph databases. The results of using the technology for quickly constructing graph databases are shown using examples of the work of ISKAD IS.

Keywords: thematic sites, Big Data, machine learning method, analysis of data, graph database, knowledge graph, database Neo4j

For citation. Piletski I. I., Batura M. P., Volorova N. A., Zorko P. A., Kulevich A. O. *System of complex data analysis of thematic sites ISCAD IS*. Informatika [*Informatics*], 2024, vol. 21, no. 1, pp. 105–120 (In Russ.). https://doi.org/10.37661/1816-0301-2024-21-1-105-120

Conflict of interest. The authors declare of no conflict of interest.

Введение. В настоящей работе используются материалы по разработке программных комплексов анализа данных из интернет-источников, полученные авторами ранее [1–3] при реализации Системы комплексного анализа данных тематических сайтов ИСКАД ИИ. Все работы выполнялись в Белорусском государственном университете информатики и радиоэлектроники на протяжении нескольких лет.

По данным Gartner group за 2023 г., большинство известных ИТ-компаний разрабатывали или имели аналитические средства анализа данных. Проблемами существующих средств являются их тяжеловесность, сложность модифицирования и адаптации к изменению тематики области применения. Последние данные трендов Gartner в области ИТ "Gartner Identifies Top 10 Data and Analytics Technology Trends for 2021" показывают возрастающую роль графовых технологий. Так, к 2025 г. графовые технологии будут использоваться в 80 % инноваций в области данных и аналитики по сравнению с 10 % в 2021 г., что даст возможность быстро принимать решения в организации[1].

Одним из сложных современных направлений является представление знаний с помощью специальных глобальных словарей предметных областей, метаописаний и специальных языков, а также методологий их применения. Многие важные мировые тематические сайты, такие как EBSCO, ScienceDirect, SpringerLink, ACM Digital Library, IEEE Xplore, CiteSeerX, Google

[1]Gartner [Electronic resource]. – Mode of access: https://www.gartner.com/en/newsroom/press-releases/2021-03-16-gartner-identifies-top-10-data-and-analytics-technologies-trends-for-2021/. – Date of access: 18.10.2023.

Scholar, Semantic Scholar, libgen: Library Genesis, Medium, КиберЛенинка, SpringerOpen, Wikipedia, Wikidata и др., используют специальную технику описания ресурса RDF (Resource Description Framework, среда описания ресурса)[2].

RDF представляет собой абстрактную модель, обеспечивающую способ разбиения знаний на дискретные части и позволяющую обмениваться информацией. RDF – это модель обмена данными, которая описывает, как данные сериализуются и как ими обмениваются. Модель RDF не описывает, как данные хранятся и организуются, она предназначена для обмена информацией (импорта и экспорта). Такой подход позволяет описывать знания в тематических предметных словарях и обмениваться этими знаниями с другими сайтами. Словари RDF и онтологии OWL (Web Ontology Language, язык представления веб-онтологий) применяют абстрактные модели RDF и RDFS описания ресурса. Онтология – это конкретное формальное представление того, что означают термины в той области, в которой они используются. Данные для импорта и экспорта на RDF-сайтах могут быть представлены в нескольких форматах: JSON-LD, Turtle, N-Triples, RDF/XML, TriG и N-Quads, TriG.

Разработанная методология описания сайтов позволяет применять специальную технологию построения (генерации) графовой БД из описания RDF-данных. Такая тематическая графовая БД содержит базу знаний сайта в виде графа знаний, что дает возможность применять различные аналитические алгоритмы ML для более глубокого анализа данных сайта [4–6].

Целью настоящей работы являются апробация и тестирование методологии [2] разработки многоцелевого, модифицируемого кластера (семейства программного обеспечения) для анализа данных интернет-источников (например, научных публикаций, социальных сетей, СМИ). Такой анализ позволяет выявлять наиболее важные публикации в некоторой предметной области (например, в космических исследованиях, здравоохранении, социальной сфере), определять тематику этих публикаций, выявлять лидера научного направления, предсказывать тенденции развития направлений и взаимодействия групп людей.

Разработанное программное обеспечение Системы комплексного анализа данных тематических сайтов ИСКАД ИИ позволит реализовать поставленные цели и выполнить анализ публикаций в предметной области. В качестве предметной области для Системы комплексного анализа данных тематических сайтов ИСКАД ИИ могут быть использованы важные мировые сайты, в которых применяется специальная техника описания ресурса RDF.

1. Методы построения тематических сайтов

1.1. Среда описания ресурса. Одним из способов формализации знаний является применение какого-либо доступного стандартного языка. Для формального описания знаний в тематических словарях наиболее широко используются схема RDF, язык веб-онтологий (Web Ontology Language, OWL) для онтологий и простая система организации знаний (Simple Knowledge Organization System, SKOS) для схем таксономической классификации. Каждый из словарей допускает разные уровни выразительности: от базового определения категорий и отношений до таксономий и более сложных конструкций, например сложных классов. Такой подход позволяет применять различное программное обеспечение для реализации методологии [7].

Множество RDF-утверждений образует ориентированный граф, в котором вершинами являются субъекты и объекты, а ребра отображают отношения.

Для доступа к данным мировых тематических сайтов можно использовать специально разработанный язык SPARQL Protocol and RDF Query Language. ИТ-специалистами разработано множество различных редакторов, помогающих строить простые и сложные запросы на языке SPARQL [7]. Применение языка SPARQL позволяет получать результат в виде простого скалярного или несложного структурированного значения. Однако этот язык не дает возможность решить главную задачу, которая заключается том, что по множеству описаний RDF необходимо построить тематическую графовую БД с целью дальнейшего глубокого анализа данных сайта и длительного исследования данных в БД.

---

[2]Среда описания ресурса (RDF): понятия и абстрактный синтаксис [Электронный ресурс]. – Режим доступа: https://www.w3.org/2007/03/rdf_concepts_ru/Overview.html. – Дата доступа: 18.10.2023.

1.2. Правила преобразования троек RDF в графовую БД. Граф RDF – это набор триплетов или операторов (субъект, предикат, объект), где и субъект, и предикат являются ресурсами, а объект может быть либо другим ресурсом, либо литералом. Литералы не могут быть предметом других утверждений. Ресурсы однозначно идентифицируются URI. Существуют три основных правила сериализации троек RDF (субъект – предикат – объект) в графовую БД[3]. Данные преобразования являются частью методологии [2] построения ИСКАД ИИ и реализуются с помощью ИТ-среды в графовой БД Neo4j.

Правило 1. Узел в Neo4j, представляющий ресурс RDF, помечен :Resource и будет иметь свойство uri с URI-ресурса:

(S,P,O) => (:Resource {uri:S}).

Правило 2. Предикаты троек отображаются в свойствах узла в Neo4j, если объект тройки является литералом:

(S,P,O) && isLiteral(O) => (:Resource {uri:S, P:O}).

Правило 3. Предикаты троек отображаются на отношения в Neo4j, если объект тройки является ресурсом:

(S,P,O) && !isLiteral(O) => (:Resource {uri:S})-[:P]->(:Resource {uri:O}).

1.3. Графовые технологии и машинное обучение. Основа совместного применения графовых технологий и методов машинного обучения, используемых в ИСКАД ИИ, описана в работах [1, 3]. Графовая БД (узлы и отношения) содержит неструктурированные данные, так как они представлены в реальном мире, но для решения задач с помощью машинного обучения нужно преобразовать пространство, где находится граф, в другое пространство для машинного обучения – векторное, для которого применимы известные алгоритмы машинного обучения (например, node2vec или GraphSAGE). Данное преобразование выполняется с помощью сложной методологии выделения вектора свойств, называемого включением (embedding) [8]. Графовые включения – это представление узлов и отношений в графе как вектора свойств. В качестве значений вектора свойств могут быть выбраны некоторые атрибуты вершин и отношений. Такая методология совместного применения графовых технологий и методов машинного обучения позволяет создавать технологии глубокого анализа данных интернет-источников. Конкретные технологические решения и примеры применения их в ИСКАД ИИ приведены в разд. 2.

2. Результаты построения системы ИСКАД ИИ

2.1. Система комплексного анализа данных тематических сайтов ИСКАД ИИ. Существуют различные варианты архитектурных решений построения систем анализа данных интернет-источников. Так, в публикациях [1, 2] при разработке интеллектуальной системы комплексного анализа данных интернет-источников выполнен ретроспективный анализ трех вариантов архитектурных решений, определены структурные компоненты и их функции. Основное архитектурное решение заключается в том, что система должна состоять из следующих компонентов: сбора данных, фильтрации данных и составления «мешка слов» из N-грамм (векторизации), библиотеки аналитических модулей, хранилища данных, графовой БД и графа знаний, аналитического компонента, обеспечивающего взаимодействие с пользователем и подготовку выдачи результата, клиентского модуля и универсальной интеграционной шины (управляющего компонента). При необходимости набор модулей и компонентов может быть расширен, а некоторые модули заменены новыми. Были последовательно разработаны и реализованы три варианта данной системы. В третьем варианте приняты важные дополнительные архитектурные решения: все компоненты функционируют как постоянно работающие самостоятельные серверы; в качестве хранилища скаченных данных использовалась БД лидера хранилища типа «семейство столбцов» Cassandra; для анализа данных применялась графовая БД, моделирующая предметную область (данные поступали из хранилища). Для обеспечения взаимодействия компонентов использовался управляющий компонент, который выполнял роль

[3]Neosemantics (n10s): Neo4j RDF & Semantics toolkit [Electronic resource]. – Mode of access: https://neo4j.com/labs/neosemantics/. – Date of access: 18.10.2023.

интеграционной шины (разработан на базе интеграционной шины Kafka). При такой архитектуре остановка работы одного из компонентов не приводила к остановке работы всего комплекса и можно было легко выполнять его модернизацию. Однако данное фундаментальное архитектурное решение не позволяет быстро перестроить систему на новую тематику и построить многоцелевой, модифицируемый кластер семейства тематических графовых БД.

Анализ аналогов известных сайтов: КиберЛенинка, Semantic Scholar, SpringerOpen, Medium – позволил принять решение о разработке мультиплатформенного решения для анализа данных различных предметных областей, с помощью которого можно быстро получать информацию о предметной области с мировых крупных сайтов на базе быстрого построения графовой(ых) БД предметной области и выполнять более глубокий анализ данных предметной области.

2.2. Архитектура ИСКАД ИИ. Система комплексного анализа данных тематических сайтов ИСКАД ИИ позволяет анализировать данные с различных тематических сайтов и предназначена для сбора информации о научных публикациях, генерации графовой БД, построения графа знаний, преобразования свойств узлов и отношений графовой БД в векторное представление с целью применения алгоритмов ML для более глубокого анализа данных. Такой комплексный подход к анализу данных дает возможность определять передовые научные направления и экспертов в предметных областях, тематику их работ и взаимосвязи.

Основными компонентами ИСКАД ИИ являются: получение данных из интернет-источников; графовая БД и граф знаний; извлечение свойств из графовой БД и их анализ с помощью алгоритмов ML, а также интеграция (специальный веб-сайт ИСКАД ИИ). Компонент «извлечение свойств из графовой БД» обладает дополнительной функциональностью и может использовать технологию включений [8], что позволяет строить векторы свойств меньшей размерности для более глубокого анализа данных. Общая функциональная архитектура ИСКАД ИИ показана на рис. 1.

Компонент «получение данных из интернет-источников» выполняет различные операции над данными, включая очистку, структурирование, нормализацию и приведение их к единому формату. Это позволяет обеспечить качество и консистентность данных, а также подготовить их для дальнейшего анализа. Получение данных осуществляется в специальной ИТ-среде и обеспечивает быстрое построение тематических графовых БД, использующих RDF-описание данных в различных форматах (рис. 1, п. 1 и п. 2).

Компонент «графовая БД и граф знаний» реализуется с помощью ИТ-среды и графовой БД Neo4j (рис. 1, п. 3), которая является лидером среди графовых БД на протяжении последних 10 лет[4]. Она обладает исключительными свойствами горизонтального масштабирования, с ростом данных не деградирует, работает в десятки и сотни раз быстрее, чем реляционная БД, и обеспечивает требования ACID (atomicity, consistency, isolation, durability) и соответствие спецификациям JTA, JTS и XA. Графовая БД Neo4j обеспечивает работу с миллионами узлов и отношений, а также доступ к данным как на языке запросов Cypher (соответствует требованиям CRUD (create, read, update, delete)), так и на популярных языках программирования. Она позволяет получать графы знаний и графически визуализировать наборы данных и результаты запросов. В процессе доступа скачивания данных с сайтов компонентом «получение данных» графовая БД строится и модифицируется автоматически.

Компонент «извлечение свойств из графовой БД и их анализ с помощью алгоритмов ML» (рис. 1, п. 4–6) позволяет реализовывать запросы пользователей, выдавать тематические графы знаний и обеспечивает анализ данных в графовых БД, преобразовывая свойства в узлах и ребрах (отношениях) в векторное представление с целью применения для дальнейшего анализа данных алгоритмов ML [8, 9].

Компонентом «интеграция» является веб-сайт ИСКАД ИИ (рис. 1, п. 7), который дает возможность работать другим компонентам системы и предоставляет доступ к получению аналитических данных пользователям системы. Веб-сайт обеспечивает пользователей возможностью создавать, обновлять и взаимодействовать с тематическими графовыми БД. Он

[4]Neo4j Graph Database [Electronic resource]. – Mode of access: https://neo4j.com/product/neo4j-graph-database. – Date of access: 18.10.2023.

предоставляет удобный интерфейс для выполнения запросов к графу знаний, получения и выдачи репортов. Веб-сайт способствует интеграции с различными источниками данных в системе, такими как БД, файловые хранилища или внешние API, чтобы получать необходимую информацию для отчетов.

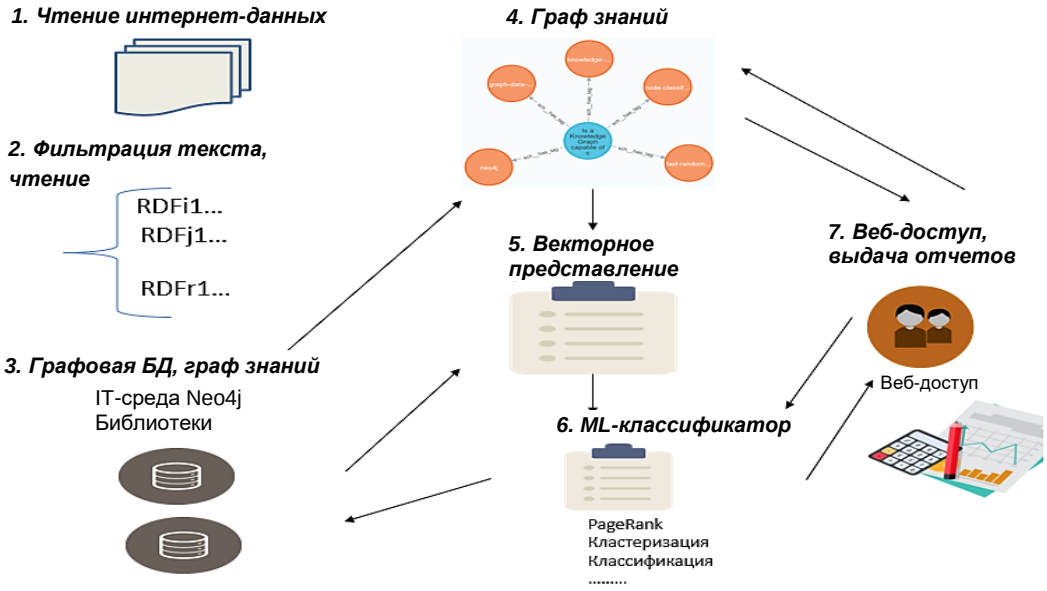


Рис. 1. Общая функциональная архитектура ИСКАД ИИ

Fig. 1. General functional architecture of ISKAD IS

Зарегистрированный пользователь может иметь доступ к просмотру публикаций различных предметных областей, поиску наиболее цитируемых авторов предметной области, просмотру параметров некоторой предметной области и различной информации об авторе публикации. Пользователь может видеть имя, количество публикаций и сами публикации автора, а также гистограммы по авторам и статьям. Гистограммы по авторам и статьям строятся с помощью алгоритма PageRank. При этом пользователь сам выбирает, какая гистограмма ему нужна и какое количество статей или авторов должно в нее входить.

Администратор веб-сайта может добавлять или удалять пользователей и предоставлять им расширенные права, управлять состоянием БД, менять структуру и содержимое БД и делать замену на сайте одной БД на другую.

Веб-сайт является центральным компонентом, который облегчает управление данными, взаимодействие пользователей и получение информации из системы. Клиент-серверная архитектура была использована для реализации компонента. Она является распространенным подходом к разработке веб-приложений. Такой подход обеспечивает интеграцию и эффективность в работе с пользователями и компонентами системы. Данное решение и представленная методология реализованы при разработке проекта БГУИР «ГПНИ по теме "Интеллектуальная система комплексного анализа данных интернет-источников (ИСКАД ИИ)"». В настоящей работе продемонстрирована реализация основных технических решений.

2.3. ИТ-платформа для ИСКАД ИИ. В качестве основных компонентов ИТ-среды для построения ИСКАД ИИ используются графовая СУБД Neo4j Desktop, специальные библиотеки (плагины), расширяющие возможности анализа данных в графовой БД (Neosemantics (n10s), библиотека APOC (Awesome Procedures on Cypher), библиотека Neo4j Graph Data Science (GDS)) и фреймворки для разработки веб-сайта ИСКАД ИИ.

2.4. Инфологическая модель графовой БД. В ИСКАД ИИ разработаны БД и ее инфологическая модель. При создании данной модели за основу бралась предметная область проекта, которая во многом совпадает с предметной областью, приведенной в работе [2]. Важно отметить, что шаблон графовой БД может быть определен исходя из назначения БД для анализа данных предметной области и он может отличаться от приведенного в настоящей статье. Сама графовая БД может модифицироваться в процессе своей работы. Главными сущностями и атрибутами предметной области ИСКАД ИИ являются: User (User) – пользователь ИСКАД ИИ, Article (sch__Article) – статья или публикация, Author (sch__Person) – автор статьи или публикации, Tag (sch__Tag) – ключевые слова, которые относятся к статье, List (sch__List) url – ссылка на список со статьями. Общая схема БД, сущности и их атрибуты представлены на рис. 2.

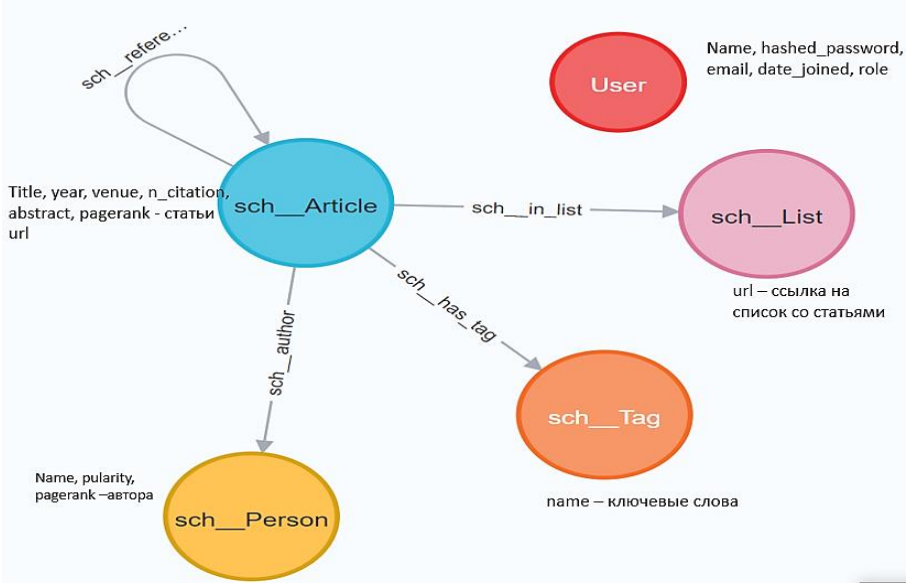


Рис. 2. Общая схема БД

Fig. 2. General Database Schema

2.5. Компонент «получение данных из интернет-источников». Для реализации методологии ИСКАД ИИ в графовую БД системы последовательно были загружены данные с сайтов SpringerOpen, Semantic Scholar, КиберЛенинка и Medium.

Извлечение информации с веб-страницы и RDF выполняется процедурами библиотек APOC n10s apoc.load.html и n10s.rdf.import. Для этого в командах следует указать URL-адрес страницы и CSS-подобный селектор, чтобы выбрать конкретный требуемый элемент. Визуализация и анализ RDF выполняются процедурами плагина Neosemantics. Код типичных команд загрузки данных с сайта SpringerOpen представлен ниже. Объем загружаемых и преобразуемых данных определяется кодом команд:

```
CALL apoc.load.html("https://journalofcloudcomputing.springeropen.com/articles/10.1186/2192-113X-1 18" , { jsonld: 'head script[type="application/ld+json"]'})
 YIELD value
UNWIND ["https://cybersecurity.springeropen.com/articles/10.1186/s42400-023-00144-1", "https://cybersecurity.springeropen.com/articles/10.1186/s42400-023-00141-4", "https://cybersecurity.springeropen.com/articles/10.1186/s42400-023-00140-5", "https://cybersecurity.springeropen.com/articles/10.1186/s42400-023-00138-z", "https://cybersecurity.springeropen.com/articles/10.1186/s42400-020-00050-w"] as page
CALL apoc.load.html(page, { jsonld: 'head script[type="application/ld+json"]'}) YIELD value
CALL n10s.rdf.import.inline(value.jsonld[0].data, "JSON-LD")  yield terminationStatus, triplesLoaded, triplesParsed, extraInfo
RETURN page, terminationStatus, triplesLoaded, triplesParsed, extraInfo
```

На рис. 3 представлена графовая БД ИСКАД ИИ, полученная с помощью технологии быстрого построения тематической графовой БД.

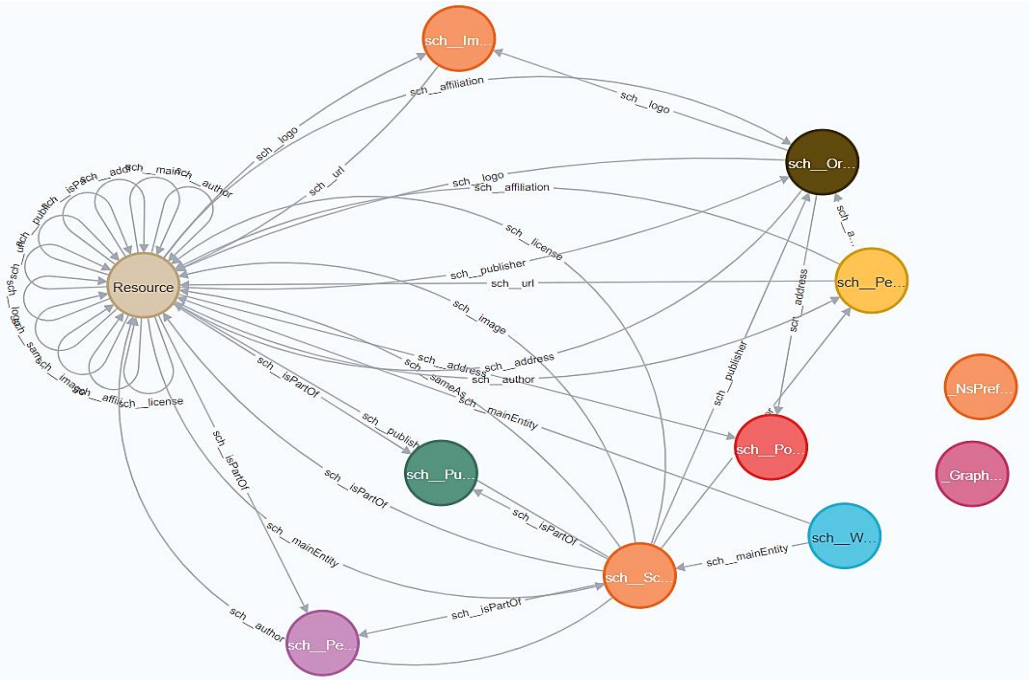


Рис. 3. Общее представление графовой БД сайта SpringerOpen

Fig. 3. General presentation of the graph database of the SpringerOpen site

С помощью описанной технологии в графовую БД ИСКАД ИИ были загружены данные с сайтов SpringerOpen, Semantic Scholar, КиберЛенинка и Medium. Важно отметить, что данные с различных сайтов были объединены автоматически в одной графовой БД (для тестирования загружено около 250 000 документов).

2.6. Компонент «графовая БД и граф знаний». Графы знаний – это особый тип графов с упором на контекстное понимание. Они представляют собой взаимосвязанные наборы фактов, которые описывают объекты, события или вещи реального мира и их взаимосвязи в формате, понятном человеку и машине. В графах знаний используется принцип организации, позволяющий пользователю (или компьютерной системе) рассуждать о лежащих в их основе данных. Принцип организации дает дополнительный уровень метаданных, который добавляет связанный контекст для поддержки рассуждений и получения знаний. Принцип организации делает сами данные более интеллектуальными, а не блокирует инструменты для понимания данных внутри кода приложения. В свою очередь, это одновременно упрощает системы и поощряет их широкое повторное использование [8]. Граф знаний (KG, Knowledge Graph) – ориентированный граф, узлы которого представляют собой сущности и литеральные значения (литералы), а ребра – отношения между этими сущностями [9].

Приведем примеры применения графов знаний. В них используется техника выполнения запросов drill-down, позволяющая уточнять полученные данные. Результаты выполнения запросов представлены на рис. 4–6.

Пример 1. Найдем все статьи одного из главных и самых успешных исследователей БД Neo4j и графов знаний Томаза Братанича (Tomaz Bratanic). Запрос будет иметь ограничение в 10 узлов из-за большого количества статей, написанных автором (рис. 4):

```
MATCH (a:sch__Article)-[:sch__author]->(au:sch__Person)
WHERE au.name = "Tomaz Bratanic"
RETURN a, au LIMIT 10
```

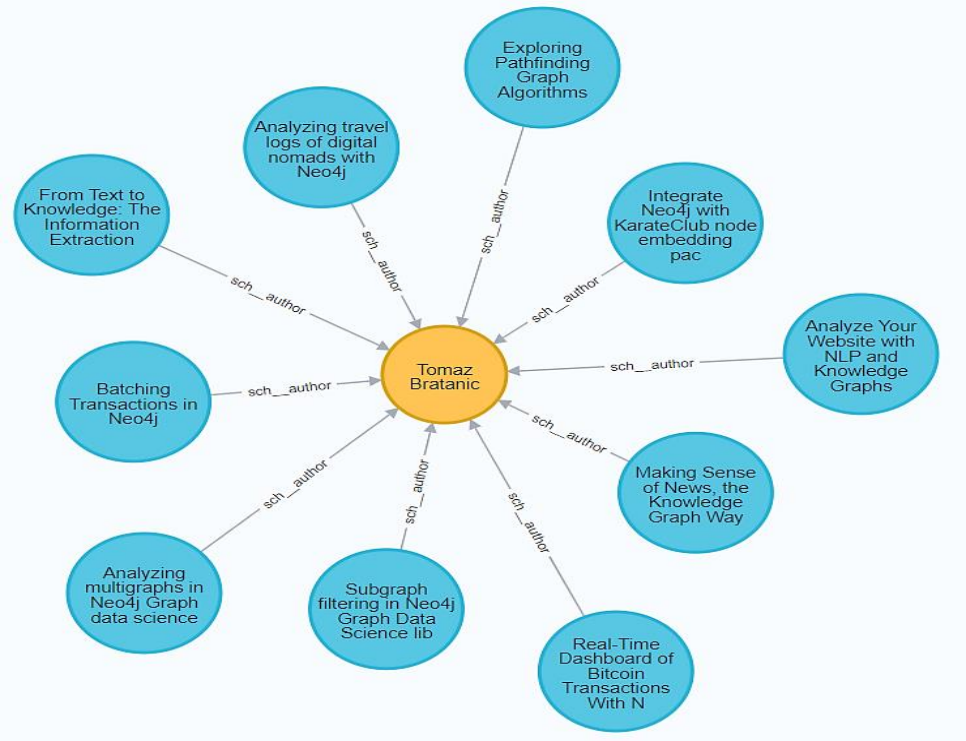


Рис. 4. Результат выполнения запроса из примера 1

Fig. 4. Example 1 of query result

На данном этапе можно отметить главную тематику статей автора: все, что связано с анализом данных, графами и графовыми алгоритмами, особенно с БД Neo4j.

Пример 2. Найдем авторов, которые пишут статьи на тему Neo4j (рис. 5):

```
MATCH (t:sch__Tag)<-[:sch__has_tag]-(a:sch__Article)-[:sch__author]->(au:sch__Person)
WHERE t.name = "neo4j"
RETURN t, au LIMIT 10
```

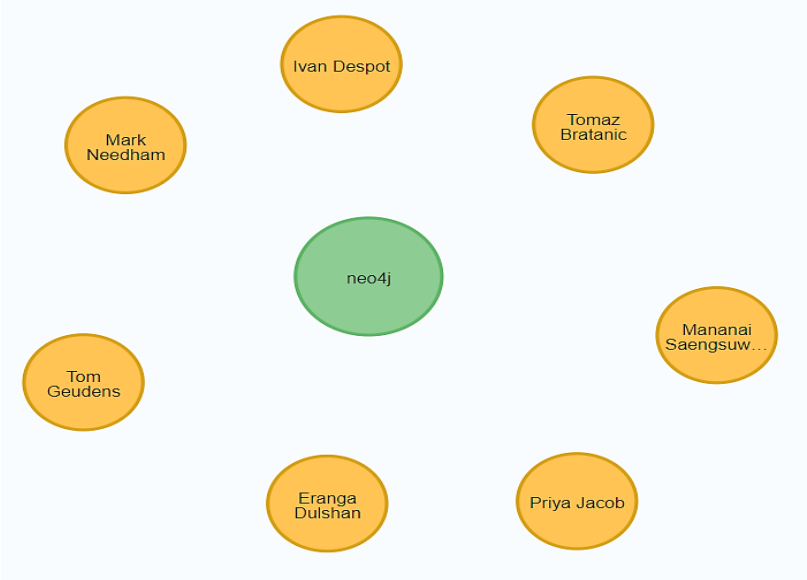


Рис. 5. Результат выполнения запроса из примера 2

Fig. 5. Example 2 of query result

Пример 3. Найдем статьи Томаза Братанича в соавторстве с еще одним исследователем использования графовых БД Марком Нидхемом (Mark Needham) (рис. 6):

```
MATCH (au1:sch__Person)<-[:sch__author]-(a:sch__Article)-[:sch__author]->(au2:sch__Person)
WHERE au1.name = "Tomaz Bratanic" AND au2.name = "Mark Needham"
RETURN au1, au2, a
```

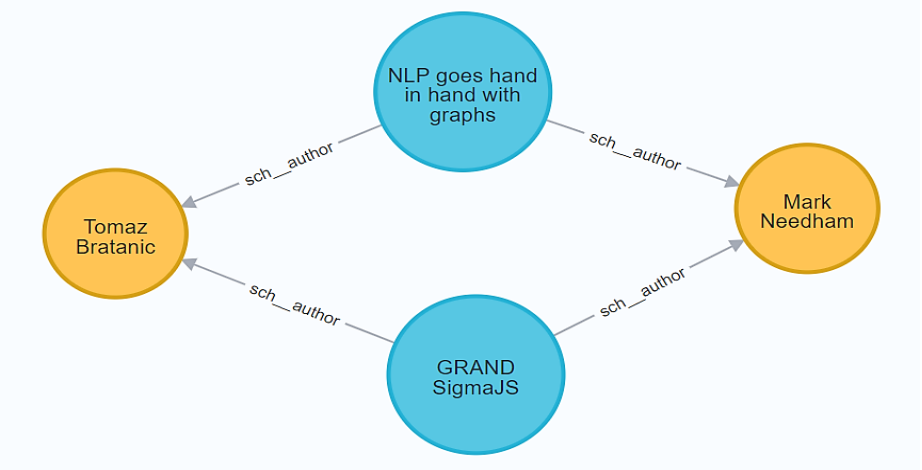


Рис. 6. Результат выполнения запроса из примера 3

Fig. 6. Example 3 of query result

В рассмотренной графовой БД нашлись две статьи, которые были написаны совместно этими авторами. Дополнительно можно уточнить ключевые слова, которые относятся к данным статьям.

2.7. Компонент «извлечение свойств из графовой БД и их анализ с помощью алгоритмов ML». Данный компонент состоит из двух модулей: модуля создания модели машинного обучения предметной области и модуля анализа свойств графовой БД с помощью алгоритмов ML, подготовки и выдачи репортов. Модель машинного обучения позволяет распознавать закономерности взаимодействия авторов публикаций, их рейтинг, рейтинг статей и предметных областей. Рейтинги определяются с помощью специального алгоритма PageRank. Однако в БД предметной области отсутствуют некоторые важные свойства узлов и отношений между ними (специфика исходных интернет-сайтов), которые не позволяют применять алгоритмы ML.

Одна из проблем заключается в отсутствии связей между соавторами. Для создания таких связей объединяются те авторы, которые совместно написали статьи. Это задача прогнозирования существования связи между двумя объектами, и она функционально решается в модуле создания модели машинного обучения. С помощью модели машинного обучения с предсказанием связей производится прогнозирование соавторства.

Другая проблема заключается в том, что не все загруженные данные отражают тематику опубликованных статей, около половины статей не содержат теги (Tag (sch__Tag) – ключевые слова статьи). Данные с сайтов SpringerOpen, Semantic Scholar, КиберЛенинка и Medium неоднородные и неотфильтрованные, что затрудняет применение общих решений алгоритмов ML.

Сама графовая БД (узлы и отношения) содержит набор разнотипных данных, к которым невозможно применить алгоритмы ML. Данные свойства в узлах и ребрах (отношениях) были преобразованы в векторное представление, что позволило использовать алгоритмы ML.

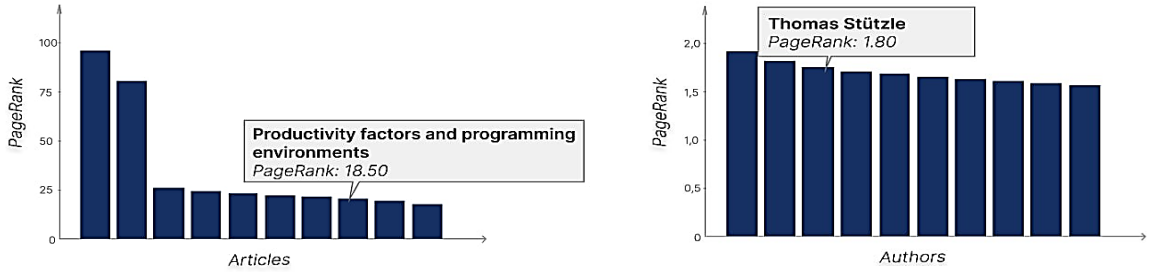
Все указанные проблемы были решены в модуле создания модели машинного обучения предметной области с помощью специально разработанных функций. Полученная обновленная модель графовой БД была протестирована и подготовлена для анализа данных. Модуль анализа

свойств графовой БД с помощью алгоритмов ML состоит из многих разработанных функций, которые применяются для анализа данных, содержащихся в подготовленной и модифицированной графовой БД ИСКАД ИИ.

Расчет PageRank каждой статьи, содержащейся в графовой БД, выполнен с помощью процедур, предоставляемых библиотекой Graph Data Science Library. Первый шаг – это выполнение процедуры построения графа in-memory из целевых сущностей статей, второй шаг – вычисление PageRank статей на основе собранных данных in-memory.

Так как набор данных не имеет между авторами прямых связей, необходимых для выполнения алгоритма PageRank, для расчет PageRank каждого автора публикаций, содержащихся в графовой БД, применяется алгоритм создания связей между авторами, аналогичный приведенному выше. Первый шаг – это выполнение процедуры создания графа in-memory из целевых сущностей авторов публикаций, для которых будет определяться PageRank, второй шаг – вычисление PageRank статей авторов на основании данных in-memory.

Диаграммы популярности статей и авторов (рис. 7) получены в модуле при настройке компонента для работы в комплексе ИСКАД ИИ. Основной поток получения отчетов в ИСКАД ИИ предусмотрен через компонент «интеграция» и веб-доступ ИСКАД ИИ (см. разд. 2.8).



a) *b)*

Рис. 7. Популярность статей (*a*); авторов (*b*)

Fig. 7. Articles (a) and authors (b) popularity

Анализ графа знаний выполнен с помощью алгоритмов вычисления графовых включений и алгоритмов машинного обучения.

Векторное представление или графовые включения (graph embedding) – технология, которая отображает узлы в графе в виде плотных векторов низкой размерности и позволяет аналогичным узлам в исходном графе (разные методы имеют разные определения сходства) быть похожими в пространстве выражений низкой размерности. Полученные векторы используются в различных алгоритмах ML для более глубокого анализа данных в графовой БД и графах знаний, например для решения таких задач, как классификация узлов, прогнозирование ссылок, визуализация или реконструкция исходного графа, и других алгоритмах.

В ИСКАД ИИ решена задача создания алгоритма, предсказывающего теги для статей, которые в графовой БД их еще не имеют. Анализ качества документов, полученных из интернет-источников, не позволяет выдавать полную и достоверную информацию по предметным областям, по публикациям и их авторам. Часть документов не содержит теги. Анализ данных графовой БД показал, что примерно половина узлов вообще не содержит теги.

Для прогнозирования тегов с использованием библиотеки GraphSAGE и языка программирования Python построена модель преобразования графа в векторное представление с помощью алгоритма генерации графовых включений. Входными данными для GraphSAGE является вектор свойств узлов. GraphSAGE поддерживает графы с несколькими типами узлов, где каждый тип узла имеет разные представляющие его функции. В настоящей статье применен алгоритм моночастичной проекции. Моночастичная проекция позволяет на основе графа с двумя типами узлов вывести из него граф с одним типом узлов. Библиотека Neo4j предоставляет пользователям алгоритм построения моночастичной проекции с помощью алгоритма

сходства узлов Node Similarity из библиотеки GDS. Алгоритм Node Similarity сравнивает наборы узлов на основе узлов, которые связаны между собой. Два узла считаются похожими, если у них много общих соседей.

Как и ранее для определения PageRank, для реализации алгоритма создается граф in-memory, который будет содержать свойство (атрибут) узла openaiEmbedding включения слов, созданного на основе данных свойства (атрибута) title статьи с помощью алгоритма text-embedding-ada-002. Построение графа in-memory выполняется с помощью специального кода. К полученному графу применяется алгоритм сходства узлов с установленным значением topK, равным 1000, для создания связи с как можно большим количеством статей в графе.

Алгоритм Node Similarity создает несколько компонентов связности. Для нахождения большего из них, который содержит практически все необходимые узлы, применяется алгоритм Weakly Connected Components.

Для реализации алгоритма классификации требуется выбрать теги, которые необходимо предсказать. Выберем теги, которые встречаются как минимум в 200 статьях, и добавим свойство target в выбранные узлы.

Анализ результатов работы алгоритмов позволяет сделать выводы, что полученные теги действительно отражают суть выбранных статей. Ниже приведен пример выдачи тегов:

1. Introduction to Data Mesh adoption in Adidas – motivation and takeaways --- [data],
2. Things to Do When You Feel Ruled by Time --- [productivity],
3. A Data Science project start to finish --- [coding, programming, python, python3, software-development],
4. Time series anomaly detection – in the era of deep learning --- [data-science, machine-learning],
5. How to Optimize Your Apache Spark Application with Partitions --- [spark],
6. Rule Execution with SHACL --- [knowledge-graph],
7. Language & Cognition: re-reading Jerry Fodor --- [data-science, machine-learning],
8. The Jobs Of The Future --- [leadership].

2.8. Компонент «интеграция и веб-доступ ИСКАД ИИ». Веб-сайт является центральным компонентом, который облегчает управление данными, взаимодействие пользователей и получение информации из системы. Все компоненты ИСКАД ИИ взаимодействуют через веб-сайт, который предоставляет единый интерфейс для пользователей и обеспечивает согласованность данных и операций между компонентами системы.

Веб-сайт обеспечивает взаимодействие с компонентами «получение данных из интернет-источников», «графовая БД и граф знаний», «извлечение свойств из графовой БД и их анализ с помощью алгоритмов ML». Для пользователей системы веб-сайт реализует функцию регистрации пользователей; предоставляет доступ к просмотру публикаций различных предметных областей, поиску наиболее цитируемых авторов предметной области, просмотру параметров некоторой предметной области, просмотру различной информации об авторе публикации. Пользователь может получать гистограммы по авторам и статьям с применением алгоритма PageRank и др.

В системе также есть администраторы, которые обладают функциями управления ею. Разработка веб-сайта выполнена по классической двухзвенной клиент-серверной архитектуре, в которой клиентский компьютер взаимодействует напрямую с сервером без участия промежуточных узлов или компонентов. Клиент-серверная архитектура является распространенным подходом к разработке веб-приложений. Она представляет собой модель, в которой приложение разделяется на две основные составляющие: клиентскую и серверную.

2.9. Примеры работы веб-сайта. При входе на сайт пользователь с ролью «гость» попадает на стартовую страницу, которая содержит надпись BSUIR Science Work. В начале страницы есть кнопки перехода на страницу регистрации и авторизации. После регистрации и авторизации пользователь получает доступ к режимам выдачи отчетов и управления работой системы с помощью кнопок DATA, STATISTICS и DATA MANIPULATION. При нажатии кнопки DATA пользователь переходит на страницу со всеми статьями и авторами, которые находятся в графовой БД ИСКАД ИИ.

На страницах выдачи информации есть функция фильтрации по авторам и статьям, пользователь также может искать статью по ее названию, введя нужный текст в поле search. На рис. 8

отображены такие данные по статьям как, как заголовок статьи, год издания, краткое описание статьи, PageRank, издание, цитирование, на рис. 9 – информация о конкретной выбранной статье. Для просмотра авторов пользователь должен нажать на вкладку Author. Изначально показывается 10 авторов. Для получения более подробной информации об авторе нужно выделить его, и в сплывающем окне появится необходимая информация, например как на рис. 10, где указано, сколько статей из тех, которые имеются на сайте (в рассматриваемом случае одна), написал именно этот автор, и ниже они приведены. Чтобы убрать всплывающее окно с данными о статье, можно нажать на кнопку Close или на любое место затемненной области вокруг окна. Нажав на кнопку STATISTICS, пользователь попадает на страницу для сбора статистики по имеющимся на сайте данным.

uid	title	year	url	abstract	pagerank	venue	n_citation
97961	The multinotch, part IV: Extra precision via Δ coefficients	2022		In [1], we presented a new digital filter architecture, the multinotch, which minimized the computational latency while preserving numerical accuracy even in the presence of severe quantization. While this method is far more accurate than discretizing polynomial filters, it can still be susceptible to problems caused by a sample rate which is significantly higher than the frequencies of the features that the filter is trying to implement. This paper presents a modification, called Δ coefficients, which preserve all the positive properties of the multinotch while dramatically increasing the numerical accuracy over a large frequency range.	1.585258472082883	advances in computing and communications	50
132663	The multinotch, part X: A low latency, high numerical fidelity filter for mechatronic control systems	2023		Control of lightly damped mechatronic systems is often accomplished in practice with a PID-like controller in series with a filter to limit the effects of high frequency resonances. The high frequency filtering is often limited	1.585258472082883	advances in computing and communications	8

Рис. 8. Фильтрованные данные по статьям

Fig. 8. Filtered data by article

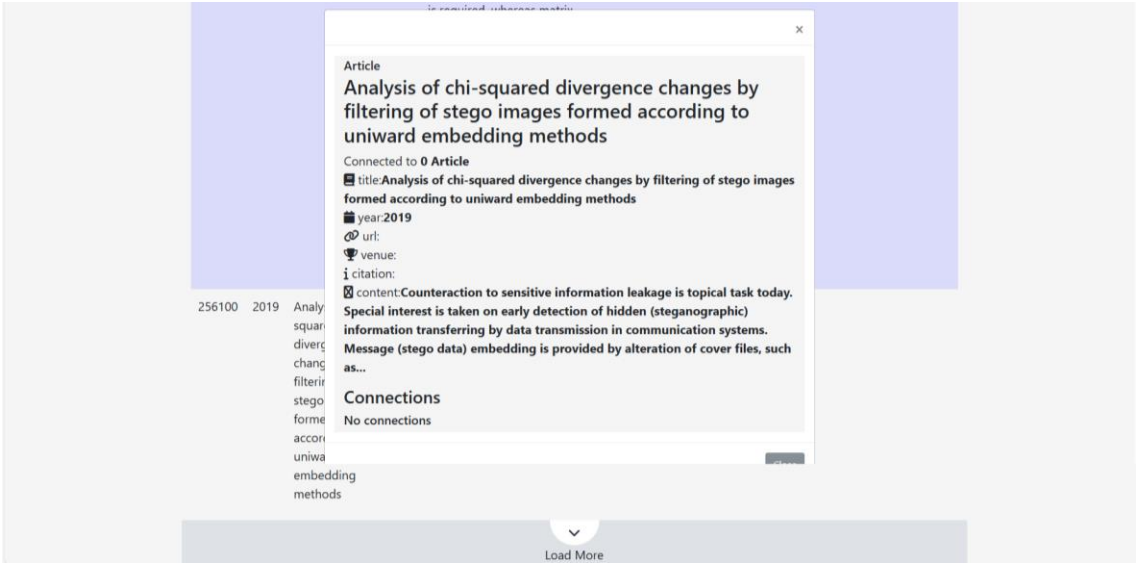


Рис. 9. Информация о выбранной статье

Fig. 9. Information about the selected article

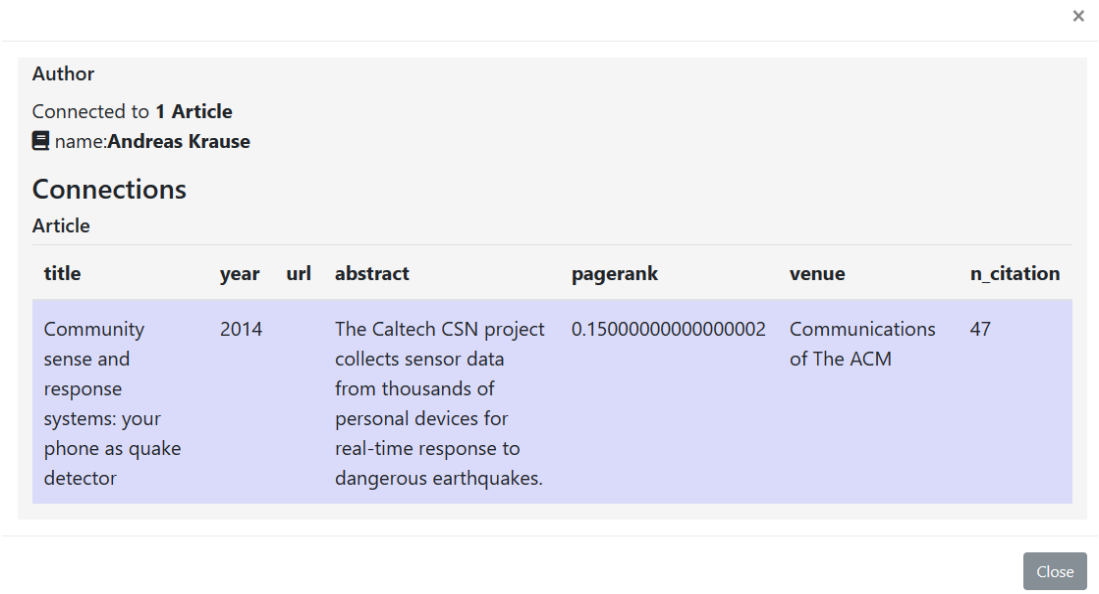


Рис. 10. Фильтрованные данные по автору

Fig. 10. Filtered data by author

Загружать данные пользователь может двумя способами. Первый способ – нажать на кнопку DATA MANIPULATION, затем на кнопку Choose file, выбрать нужный файл с расширением txt, где находятся ссылки на статьи, которые пользователь хочет добавить в графовую БД проекта, и нажать на кнопку SUBMIT. Второй способ – во второе поле вставить готовые ссылки на статьи, которые пользователь хочет добавить в графовую БД проекта, и нажать на кнопку SUBMIT.

Заключение. В статье разработана и апробирована комплексная технология последовательного применения взаимосвязанных методов, методологий и инструментов по построению графовой БД, графа знаний, анализа данных с использованием векторного преобразования графовых данных, методов и моделей машинного обучения и предоставления аналитических результатов пользователям. Создана и апробирована ИТ-среда для быстрого построения тематической графовой БД из данных сайтов и продемонстрировано применение графа знаний. Использована технология преобразования графов (графовых данных) в непрерывное низкоразмерное векторное представление, что позволяет анализировать содержимое графовых БД с помощью алгоритмов ML.

Представленная технология реализована в ИСКАД ИИ и применяется в БГУИР при анализе публикаций известных мировых сайтов, а также при проведении занятий с магистрантами. В дальнейшем при использовании ИСКАД ИИ необходимо предварительно проводить анализ загружаемых данных в графовую БД на их полноту. Не следует совмещать данные с различной структурой в одной графовой БД.

Вклад авторов. *И. И. Пилецкий* выполнил анализ предметной области, разработал методику и технологию быстрого протопирования тематических графовых БД, а также методологию углубленного анализа графовой БД. *М. П. Батура* руководил выполнением всего проекта, проанализировал полученные результаты на соответствие функциональным требованиям ИСКАД ИИ. *Н. А. Волорова* выполнила анализ интернет-источников предметной области, подготовила требования к разработке ИСКАД ИИ, организовала технологию реализации и тестирование системы. *П. А. Зорко* разработала компонент извлечения свойств из графовой БД и осуществила их анализ с помощью алгоритмов ML. *А. О. Кулевич* разработал ПО получения данных из интернет-источников, графовую БД и граф знаний.

Список использованных источников

1. Интеллектуальная система комплексного анализа данных интернет-источников / М. П. Батура [и др.] // BIG DATA and Advanced Analytics = BIG DATA и анализ высокого уровня : сб. материалов VI Междунар. науч.-практ. конф., Минск, 20–21 мая 2020 г. : в 3 ч. Ч. 1 / редкол.: В. А. Богуш [и др.]. – Минск : Бестпринт, 2020. – С. 220–241.

2. Пилецкий, И. И. Графовые технологии в интеллектуальной системе комплексного анализа данных интернет-источников / И. И. Пилецкий, М. П. Батура, Л. Ю. Шилин // Доклады БГУИР. – 2020. – Т. 18, № 5. – С. 89–97.

3. Граф знаний и машинное обучение как ИТ-среда интеллектуального анализа данных интернет-источников / М. П. Батура [и др.] // BIG DATA and Advanced Analytics = BIG DATA и анализ высокого уровня : сб. науч. ст. VIII Междунар. науч.-практ. конф., Минск, 11–12 мая 2022 г. / Бел. гос. ун-т информатики и радиоэлектроники ; редкол.: В. А. Богуш [и др.]. – Минск, 2022. – С. 330–344.

4. Diestel, R. Graph Theory / R. Diestel. – Berlin : Springer-Verlag, 2017. – 448 p.

5. Needham, M. Graph Algorithms / M. Needham, A. E. Hodler. – Sebastopol : O'Reilly Media, 2019. – 265 p.

6. Hamilton, W. L. Representation learning on graphs: Methods and applications / W. L. Hamilton, R. Ying, J. Leskovec // IEEE Data Engineering Bulletin. – 2017. – Vol. 40, no. 3. – P. 52–74.

7. Ovcinnikova, J. Visual diagrammatic queries in ViziQuer: Overview and implementation / J. Ovcinnikova, A. Sostaks, K. Cerans // Baltic J. of Modern Computing. – 2023. – Vol. 11, no. 2. – P. 317–350.

8. Portisch, J. Knowledge graph embedding for data mining vs. knowledge graph embedding for link prediction – two sides of the same coin? / J. Portisch, N. Heist, H. Paulheim // Semantic Web. – 2022. – Vol. 13, no. 3. – P. 399–422. https://doi.org/10.3233/SW-212892

9. Barrasa, J. Knowledge Graphs / J. Barrasa, A. E. Hodler, J. Webber. – Sebastopol : O'Reilly Media, 2021. – 85 p.

References

1. Batura M. P., Piletski I. I., Prytkov V. A., Volorova N. A. *Intelligent system for comprehensive analysis of data from Internet sources.* BIG DATA i analiz vysokogo urovnja : sbornik materialov VI Mezhdunarodnoj nauchno-prakticheskoj konferencii, Minsk, 20–21 maja 2020 g. : v 3 chastjah. Chast' 1 [*BIG DATA and Advanced Analytics : Collection of Materials of the VI International Scientific and Practical Conference, Minsk, 20–21 May 2020 : in 3 Parts. Part 1*]. Ed. board: V. A. Bogush [et al.]. Minsk, Bestprint, 2020, pp. 220–241 (In Russ.).

2. Piletski I. I., Batura M. P., Shilin L. Yu. *Graph technologies in an intelligent system for complex analysis of data from Internet sources*. Doklady Belorusskogo gosudarstvennogo universiteta informatiki i radioèlektroniki [*Doklady BGUIR*], 2020, vol. 18, no. 5. pp. 89–97 (In Russ.).

3. Batura M. P., Piletsky I. I., Volorova N. A., Zorko P. A., Kulevich A. O. *Knowledge graph and machine learning as an IT environment for mining data from Internet sources.* BIG DATA i analiz vysokogo urovnja : sbornik nauchnyh statej VIII Mezhdunarodnoj nauchno-prakticheskoj konferencii, Minsk, 11–12 maja 2022 g. [*BIG DATA and Advanced Analytics : Collection of Scientific Articles of the VIII International Scientific and Practical Conference, Minsk, 11–12 May 2022*]. Ed. board: V. A. Bogush [et al.]. Minsk, 2022, pp. 330–344 (In Russ.).

4. Diestel R. *Graph Theory*. Berlin, Springer-Verlag, 2017, 448 p.

5. Needham M., Hodler A. E. *Graph Algorithms*. Sebastopol, O'Reilly Media, 2019, 265 p.

6. Hamilton W. L., Ying R., Leskovec J. Representation learning on graphs: Methods and applications. *IEEE Data Engineering Bulletin*, 2017, vol. 40, no. 3, pp. 52–74.

7. Ovcinnikova J., Sostaks A., Cerans K. Visual diagrammatic queries in ViziQuer: Overview and implementation. *Baltic Journal of Modern Computing*, 2023, vol. 11, no. 2, pp. 317–350.

8. Portisch J., Heist N., Paulheim H. Knowledge graph embedding for data mining vs. knowledge graph embedding for link prediction – two sides of the same coin? *Semantic Web*, 2022, vol. 13, no. 3, pp. 399–422. https://doi.org/10.3233/SW-212892

9. Barrasa J., Hodler A. E., Webber J. *Knowledge Graphs*. Sebastopol, O'Reilly Media, 2021, 85 p.

Информация об авторах

Пилецкий Иван Иванович, кандидат физико-математических наук, доцент кафедры информатики Белорусского государственного университета информатики и радиоэлектроники.

Батура Михаил Павлович, доктор технических наук, профессор, заведующий лабораторией «Новые обучающие технологии» Белорусского государственного университета информатики и радиоэлектроники.

Волорова Наталья Алексеевна, кандидат технических наук, доцент, старший научный сотрудник лаборатории «Новые обучающие технологии» Белорусского государственного университета информатики и радиоэлектроники.

Зорко Полина Александровна, магистрант кафедры информатики Белорусского государственного университета информатики и радиоэлектроники.

Кулевич Алексей Олегович, магистрант кафедры информатики Белорусского государственного университета информатики и радиоэлектроники.

Information about the authors

Ivan I. Piletski, Ph. D. (Phys.-Math.), Assoc. Prof. of the Department of Informatics of Belarusian State University of Informatics and Radioelectronics.

Michal P. Batura, D. Sc. (Eng.), Prof., Head of the Laboratory "New Educational Technologies" of Belarusian State University of Informatics and Radioelectronics.

Natalia A. Volorova, Ph. D. (Eng.), Assoc. Prof., Senior Researcher of the Laboratory "New Educational Technologies" of Belarusian State University of Informatics and Radioelectronics.

Polina A. Zorko, Master's Student of the Department of Informatics of Belarusian State University of Informatics and Radioelectronics.

Alexei O. Kulevich, Master's Student of the Department of Informatics of Belarusian State University of Informatics and Radioelectronics.

Правила для авторов

Редакция журнала «Информатика» просит авторов руководствоваться приведенными ниже правилами.

I. Статьи принимаются в редакцию через электронную систему подачи по адресу http://inf.grid.by в формате файлов текстовых редакторов Microsoft Word. Объем оригинальной статьи – от 8 до 16 стр., включая рисунки, таблицы и достаточное количество наиболее актуальных ссылок; объем обзорной статьи – от 16 до 32 стр., включая все основные ссылки. Текст набирается с переносами, шрифт Times New Roman 11 пт, интервал между строками одинарный, абзацный отступ 0,5 см, поля по 2,5 см со всех сторон.

Материал статьи должен быть четко структурированным: Введение; основные разделы, в которых изложены цели и задачи, методы, результаты; Заключение (выводы).

II. Статьи о результатах работ, проведенных в научных учреждениях, должны иметь разрешение на публикацию (сопроводительное письмо за подписью руководителя или выписку из заседания ученого совета, отдела или кафедры, акт экспертизы).

III. Статьи в обязательном порядке должны включать аннотацию, ключевые слова, список литературы, информацию об авторах на русском и английском языках.

На титульной странице располагаются следующие метаданные:

1. Индекс по универсальной десятичной классификации (УДК); *на русском и английском языках* тип статьи (оригинальная или обзорная), название статьи, инициалы и фамилии всех авторов, полное наименование учреждений, где работают авторы, с указанием почтового адреса, при наличии указывается ученая степень и ORCID, e-mail ответственного лица.

2. Аннотация (Abstract) объемом 150–250 слов в оригинальной статье должна быть структурирована отдельными подразделами: Цели, Методы, Результаты, Заключение, а также максимально характеризовать содержательную часть рукописи. Сюда не следует включать впервые введенные термины, аббревиатуры (за исключением общеизвестных), ссылки на литературу.

3. Ключевые слова (Keywords) – наиболее значимые слова или словосочетания по теме работы, отражающие специфику темы, объекты и результаты исследования; перечень ключевых слов должен содержать 5–10 слов.

4. В разделе Благодарности (Acknowledgements) указываются все источники финансирования исследования, а также благодарности людям, которые участвовали в работе над статьей.

5. Автор обязан уведомить редакцию о реальном или потенциальном конфликте интересов, включив информацию в раздел Конфликт интересов (Conflict of interest).

6. Формулы, рисунки, таблицы в статье нумеруются в соответствии с порядком их упоминания в тексте. Ссылки на рисунки и таблицы в тексте обязательны. Рисунки должны быть выполнены с хорошим разрешением в масштабе, позволяющем четко различать надписи и обозначения. Цветные иллюстрации печатаются только в том случае, когда это необходимо для понимания излагаемого материала. Подрисуночные подписи с расшифровкой всех позиций, представленных на рисунке, и названия таблиц набираются шрифтом гарнитуры основного текста размером 9 пт. Перевод подрисуночной подписи и пояснений к рисунку, а также перевод названия таблицы, заголовки строк или столбцов располагаются курсивом после русскоязычной версии.

7. Набор формул выполняется в формульном редакторе Microsoft Equation или Math Type. Прямым шрифтом набираются: греческие и русские буквы; математические символы (sin, lg, ∞); символы химических элементов (C, Cl, CHCl3); цифры (римские и арабские); индексы (верхние и нижние), являющиеся сокращениями слов. Курсивом набираются латинские буквы, символы физических величин (в том числе и в индексе).

8. Список использованной литературы оформляется в соответствии с требованиями Высшей аттестационной комиссии Республики Беларусь (ГОСТ 7.5–2008). Номер литературной ссылки в тексте дается порядковым номером в квадратных скобках. Ссылаться на неопубликованные работы не допускается.

9. Отдельно оформляется References со следующей структурой: авторы (транслитерация), транслитерированное название монографии, *Перевод названия монографии на английский язык*. Выходные данные с обозначениями на английском языке. От транслитераций названий статей можно отказаться.

10. Ссылки на учебно-методическую литературу, ГОСТы, авторефераты, статистические отчеты в список не включаются, а оформляются в виде сносок (с подробными рекомендациями можно ознакомиться на сайте журнала в разделе Правила для авторов).

11. В разделе Информация об авторах (Information about the authors) приводятся ФИО авторов полностью, ученая степень, звание, должность, название организации, ORCID (при наличии).

IV. Все поступающие в редакцию рукописи проходят предварительную проверку на соответствие Правилам для авторов. Статья может быть возвращена автору на доработку с просьбой устранить недостатки или дополнить информацию. После проверки на соответствие правилам статья направляется рецензенту с указанием сроков рецензирования.

V. При наличии замечаний рецензента автору предоставляется определенное время на доработку рукописи. Статьи, направляемые на доработку, должны быть возвращены в исправленном виде с ответами на все замечания. Окончательное решение о публикации или отклонении рукописи принимается редколлегией журнала. При положительном заключении рецензента статья передается редактору для подготовки к печати. Редакция оставляет за собой право на редакционные изменения, не искажающие основное содержание статьи.

VI. Редакция журнала предоставляет возможность первоочередного опубликования статей, представленных лицами, которые осуществляют послевузовское обучение (аспирантура, докторантура, соискательство) в год завершения обучения.

VII. Авторы несут ответственность за направление в редакцию статей, уже опубликованных ранее или принятых к публикации другими изданиями.

ИНДЕКСЫ

00827
для индивидуальных
подписчиков

008272
для предприятий
и организаций

ИНФОРМАТИКА ■ INFORMATICS ТОМ ■ VOL. 21 1 | 2024

ISSN 1816-0301 (Print)
9 771816 030000