

p-ISSN 1607-3274
e-ISSN 2313-688X

Радіоелектроніка
Інформатика
Управління

Радиоэлектроника
Информатика
Управление

Radio Electronics
Computer Science
Control

2016/2

ISSN 1607-3274



9 771607 327005 62>





Запорізький національний технічний університет

Радіоелектроніка, інформатика, управління

Науковий журнал

Виходить чотири рази на рік

№ 2(37) 2016

Заснований у січні 1999 року.

Засновник і видавець – Запорізький національний технічний університет.

ISSN 1607-3274 (друкований), ISSN 2313-688X (електронний).

Запоріжжя

ЗНТУ

2016

Запорожский национальный технический университет

Радиоэлектроника, информатика, управление

Научный журнал

Выходит четыре раза в год

№ 2(37) 2016

Основан в январе 1999 года.

Основатель и издатель – Запорожский национальный технический университет.

ISSN 1607-3274 (печатный), ISSN 2313-688X (электронный).

Запорожье

ЗНТУ

2016

Zaporizhzhya National Technical University

Radio Electronics, Computer Science, Control

The scientific journal

Published four times per year

№ 2(37) 2016

Founded in January 1999.

Founder and publisher – Zaporizhzhya National Technical University.

ISSN 1607-3274 (print), ISSN 2313-688X (on-line).

Zaporizhzhya

ZNTU

2016

Науковий журнал «Радіоелектроніка, інформатика, управління» (скорочена назва – РІУ) видається Запорізьким національним технічним університетом (ЗНТУ) з 1999 р. періодичністю чотири номери на рік.

Зареєстрований Державним комітетом інформаційної політики, телебачення та радіомовлення 29.01.2003 р. Свідцтво про державну реєстрацію друкованого засобу масової інформації серія KB №6904.

ISSN 1607-3274 (друкований), ISSN 2313-688X (електронний).

Наказом Міністерства освіти і науки України № 1328 від 21.12.2015 р. «Про затвердження рішень Атестаційної колегії Міністерства щодо діяльності спеціалізованих вчених рад від 15 грудня 2015 року» **журнал включений до переліку наукових фахових видань України**, в яких можуть публікуватися результати дисертаційних робіт на здобуття наукових ступенів доктора і кандидата фізико-математичних та технічних наук.

В журналі безкоштовно публікуються наукові статті англійською, російською та українською мовами.

Правила оформлення статей подано на сайті: <http://ric.zntu.edu.ua/information/authors>.

Журнал забезпечує **безкоштовний відкритий он-лайн доступ** до повнотекстових публікацій.

Журнал дозволяє авторам мати авторські права і зберігати права на видання без обмежень. Журнал дозволяє користувачам читати, завантажувати, копіювати, поширювати, друкувати, шукати або посилатися на повні тексти своїх статей. Журнал дозволяє повторне використання його вмісту у відповідності з CC ліцензією CC-BY.

Опублікованими статтям присвоюється унікальний ідентифікатор цифрового об'єкта DOI.

Журнал реферується та індексується у провідних міжнародних та національних реферативних журналах і наукометричних базах даних, а також розміщується у цифрових архівах та бібліотеках з безкоштовним доступом у режимі on-line (у т. ч. DOAJ, DOI, CrossRef, EBSCO, eLibrary.ru / РИНЦ, Google Scholar, Index Copernicus, INSPEC, ISSN, Ulrich's Periodicals Directory, WorldCat, BИHIT, Джерело), повний перелік яких подано на сайті: <http://ric.zntu.edu.ua/about/editorialPolicies#custom-0>.

Журнал розповсюджується за Каталогом періодичних видань України (передплатний індекс – 22914).

Тематика журналу містить: радіофізику, мікро-, нано- і радіоелектроніку, апаратне і програмне забезпечення комп'ютерної техніки, комп'ютерні мережі і телекомунікації, теорію алгоритмів і програмування, оптимізацію і дослідження операцій, міжмашинну і людино-машинну взаємодію, математичне і комп'ютерне моделювання, обробку даних і сигналів, управління в технічних системах, штучний інтелект, включаючи системи, засновані на знаннях, і експертні системи, інтелектуальний аналіз даних, розпізнавання образів, штучні нейронні і нейро-нечіткі мережі, нечітку логіку, колективний інтелект і мультиагентні системи, гібридні системи.

Усі статті, пропоновані до публікації, одержують **об'єктивний розгляд**, що оцінюється за суттю без урахування раси, статі, віросповідання, етнічного походження, громадянства або політичної філософії автора(ів).

Усі статті проходять двоступінчасте закрите (анонімне для автора) **рецензування** штатними редакторами і незалежними рецензентами – провідними вченими за профілем журналу.

РЕДАКЦІЙНА КОЛЕГІЯ

Головний редактор – Погосов В. В., д-р фіз.-мат. наук, Україна

Заст. головного редактора – Субботін С. О., д-р. техн. наук, Україна

Члени редколегії:

Андролідакіс Й., д-р філософії, Греція

Безрук В. М., д-р техн. наук, Україна

Бодяньський С. В., д-р техн. наук, Україна, редактор розділу з управління

Васильєв С. М., д-р фіз.-мат. наук, академік РАН, Росія

Гімплевич Ю. Б., д-р техн. наук, Україна

Горбань О. М., д-р фіз.-мат. наук, Великобританія

Дробахін О. О., д-р фіз.-мат. наук, Україна

Зайцева О. М., канд. фіз.-мат. наук, Словаччина

Камеяма М., д-р техн. наук, Японія

Карпуков Л. М., д-р техн. наук, Україна

Корніч Г. В., д-р фіз.-мат. наук, Україна, редактор розділу з радіофізики

Кулік А. С., д-р техн. наук, Україна

Лебедєв Д. В., д-р техн. наук, Україна, редактор розділу з управління

Левашенко В. Г., канд. фіз.-мат. наук, Словаччина

Лиснянський А., канд. техн. наук, Ізраїль

Марковська-Качмар У., д-р наук, Польща

Олещук В. О., канд. фіз.-мат. наук, Норвегія, редактор розділу з радіоелектроніки

Онуфрієнко В. М., д-р фіз.-мат. наук, Україна

Папшицький М., д-р філософії, Польща

Піза Д. М., д-р техн. наук, Україна

Рубель О. І., канд. техн. наук, Канада

Хаханов В. І., д-р техн. наук, Україна, редактор розділу з інформатики

Шарпанських О. А., д-р філософії, Нідерланди, редактор розділу з інформатики

Рекомендовано до видання вченою радою ЗНТУ, протокол № 13 від 01.07.2016.

Журнал зверстаний редакційно-видавничим відділом ЗНТУ.

Веб-сайт журналу: <http://ric.zntu.edu.ua>.

Адреса редакції: Редакція журналу «РІУ», Запорізький національний технічний університет, вул. Жуковського, 64, м. Запоріжжя, 69063, Україна.

Тел: (061) 769-82-96 – редакційно-видавничий відділ

Факс: (061) 764-46-62

E-mail: rvv@zntu.edu.ua

Научный журнал «Радиоэлектроника, информатика, управление» (сокращенное название – РИУ) издается Запорожским национальным техническим университетом (ЗНТУ) с 1999 г. периодичностью четыре номера в год.

Зарегистрирован Государственным комитетом информационной политики, телевидения и радиовещания 29.01.2003 г. (Свидетельство о государственной регистрации печатного средства массовой информации серия КВ №6904).

ISSN 1607-3274 (печатный), ISSN 2313-688X (электронный).

Приказом Министерства образования и науки Украины № 1328 от 21.12.2015 г. «Об утверждении решений Аттестационной коллегии Министерства относительно деятельности специализированных ученых советов от 15 декабря 2015 года» **журнал включен в перечень научных профессиональных изданий Украины**, в которых могут публиковаться результаты диссертационных работ на соискание ученых степеней доктора и кандидата физико-математических и технических наук.

В журнале бесплатно публикуются научные статьи на английском, русском и украинском языках.

Правила оформления статей представлены на сайте: <http://ric.zntu.edu.ua/information/authors>.

Журнал обеспечивает **бесплатный открытый он-лайн доступ** к полнотекстовым публикациям. Журнал разрешает авторам иметь авторские права и сохранять права на издание без ограничений. Журнал разрешает пользователям читать, загружать, копировать, распространять, печатать, искать или ссылаться на полные тексты своих статей. Журнал разрешает повторное использование его содержания в соответствии с CC лицензией CC-BY.

Опубликованным статьям присваивается уникальный идентификатор цифрового объекта DOI.

Журнал реферируется и индексируется в ведущих международных и национальных реферативных журналах и наукометрических базах данных, а также размещается в цифровых архивах и библиотеках с бесплатным доступом on-line (в т.ч. DOAJ, DOI, CrossRef, EBSCO, eLibrary.ru / РИНЦ, Google Scholar, Index Copernicus, INSPEC, ISSN, Ulrich's Periodicals Directory, WorldCat, ВИНИТИ, Джэрло), полный перечень которых представлен на сайте: <http://ric.zntu.edu.ua/about/editorialPolicies#custom-0>.

Журнал распространяется по Каталогу периодических изданий Украины (подписной индекс – 22914).

Тематика журнала включает: радиофизику, микро-, нано- и радиоэлектронику, аппаратное и программное обеспечение компьютерной техники, компьютерные сети и телекоммуникации, теорию алгоритмов и программирования, оптимизацию и исследование операций, межмашинное и человеко-машинное взаимодействие, математическое и компьютерное моделирование, обработку данных и сигналов, управление в технических системах, искусственный интеллект, включая системы, основанные на знаниях, и экспертные системы, интеллектуальный анализ данных, распознавание образов, искусственные нейронные и нейро-нечеткие сети, нечеткую логику, коллективный интеллект и мультиагентные системы, гибридные системы.

Все статьи, предлагаемые к публикации, получают **объективное рассмотрение**, которое оценивается по существу без учета расы, пола, вероисповедания, этнического происхождения, гражданства или политической философии автора(ов).

Все статьи проходят двухступенчатое закрытое (анонимное для автора) **рецензирование** штатными редакторами и независимыми рецензентами – ведущими учеными по профилю журнала.

РЕДАКЦИОННАЯ КОЛЛЕГИЯ

Главный редактор – Погосов В. В., д-р физ.-мат. наук, Украина

Зам. главного редактора – Субботин С. А., д-р техн. наук, Украина

Члены редколлегий:

Андропулидакис И., д-р философии, Греция

Безрук В. М., д-р техн. наук, Украина

Бодянский Е. В., д-р техн. наук, Украина, редактор раздела по управлению

Васильев С. Н., д-р физ.-мат. наук, академик РАН, Россия

Гимпилович Ю. Б., д-р техн. наук, Украина

Горбань А. Н., д-р физ.-мат. наук, Великобритания

Дробахин О. О., д-р физ.-мат. наук, Украина

Зайцева Е. Н., канд. физ.-мат. наук, Словакия

Камеяма М., д-р техн. наук, Япония

Карпуков Л. М., д-р техн. наук, Украина

Корнич Г. В., д-р физ.-мат. наук, Украина, редактор раздела по радиофизике

Кулик А. С., д-р техн. наук, Украина

Лебедев Д. В., д-р техн. наук, Украина, редактор раздела по управлению

Левашенко В. Г., канд. физ.-мат. наук, Словакия

Лиснянский А., канд. техн. наук, Израиль

Марковска-Качмар У., д-р наук, Польша

Олещук В. А., канд. физ.-мат. наук, Норвегия, редактор радиоэлектроники

Онуфриенко В. М., д-р физ.-мат. наук, Украина

Папшицкий М., д-р философии, Польша

Пиза Д. М., д-р техн. наук, Украина

Рубель О. И., канд. техн. наук, Канада

Хаханов В. И., д-р техн. наук, Украина, редактор раздела по информатике

Шарпанских А. А., доктор философии, Нидерланды – редактор раздела по информатике

Рекомендовано к изданию ученым советом ЗНТУ, протокол № 13 от 01.07.2016.

Журнал сверстан редакционно-издательским отделом ЗНТУ.

Веб-сайт журнала: <http://ric.zntu.edu.ua>.

Адрес редакции: Редакция журнала «РИУ», Запорожский национальный технический университет, ул. Жуковского, 64, г. Запорожье, 69063, Украина.

Тел.: +38-061-769-82-96 – редакционно-издательский отдел

Факс: (061) 764-46-62

E-mail: rvv@zntu.edu.ua

© Запорожский национальный технический университет, 2016

The scientific journal «Radio Electronics, Computer Science, Control» is published by the Zaporizhzhya National Technical University (ZNTU). since 1999 with periodicity four numbers per year.

The journal is registered by the State Committee for information policy, television and radio broadcasting of Ukraine in 29.01.2003. The journal has a State Registration Certificate of printed mass media (series KB №6904).

ISSN 1607-3274 (print), **ISSN 2313-688X** (on-line).

By the Order of the Ministry of Education and Science of Ukraine from 21.12.2015 № 1328 “On approval of the decision of the Certifying Collegium of the Ministry on the activities of the specialized scientific councils dated 15 December 2015” **journal is included in the list of scientific specialized periodicals of Ukraine**, where the results of dissertations for Doctor of Science and Doctor of Philosophy in Mathematics and Technical Sciences may be published.

The journal publishes scientific articles in English, Russian, and Ukrainian free of charge.

The **article formatting rules** are presented on the site: <http://ric.zntu.edu.ua/information/authors>.

The journal provides policy of **on-line open (free of charge) access** for full-text publications. The journal allow the authors to hold the copyright without restrictions and to retain publishing rights without restrictions. The journal allow readers to read, download, copy, distribute, print, search, or link to the full texts of its articles. The journal allow reuse and remixing of its content, in accordance with a CC license CC-BY.

Published articles have a unique digital object identifier (DOI).

The journal is abstracted and indexed in leading international and national abstracting journals and scientometric databases, and also placed to the digital archives and libraries with a free on-line access (including DOAJ, DOI, CrossRef, EBSCO, eLibrary.ru / РИИЦ, Google Scholar, Index Copernicus, INSPEC, ISSN, Ulrich's Periodicals Directory, WorldCat, VINITI (All-Russian Institute of scientific and technical information), Djerelo), full list of which is presented on the site: <http://ric.zntu.edu.ua/about/editorialPolicies#custom-0>.

The journal is distributed by the Catalogue of Ukrainian periodicals (the catalog number is 22914).

The journal scope: radio physics, micro-, nano- and radio electronics, computer hardware and software, computer networks and telecommunications, algorithm and programming theory, optimization and operations research, machine-machine and man-machine interfacing, mathematical modeling and computer simulation, data and signal processing, control in technical systems, artificial intelligence, including knowledge-based and expert systems, data mining, pattern recognition, artificial neural and neuro-fuzzy networks, fuzzy logics, swarm intelligence and multiagent systems, hybrid systems.

All articles proposed for publication receive an **objective review** that evaluates substantially without regard to race, sex, religion, ethnic origin, nationality, or political philosophy of the author(s).

All articles undergo a two-stage **blind peer review** by the editorial staff and independent reviewers – the leading scientists on the profile of the journal.

EDITORIAL BOARD

Editor-in-Chief: V. V. Pogosov, Doctor of Science in Physics and Mathematics, Ukraine

Deputy Editor-in-Chief: S. A. Subbotin, Doctor of Science in Engineering, Ukraine

Members of Editorial Board:

I. Androulidakis, Ph. D, Greece

V. M. Bezruk, Doctor of Science in Engineering, Ukraine

Ye. V. Bodyanskiy, Doctor of Science in Engineering, Ukraine, Control section editor

O. O. Drobakhin, Doctor of Science in Physics and Mathematics

Yu. B. Gimpilevich, Doctor of Science in Engineering, Ukraine

A. N. Gorban, Doctor of Science in Physics and Mathematics, United Kingdom

V. I. Hahanov, Doctor of Science in Engineering, Ukraine, Computer Science section editor

M. Kameyama, Doctor of Science, Japan

L. M. Karpukov, Doctor of Science in Engineering, Ukraine

G. V. Kornich, Doctor of Science in Physics and Mathematics, Ukraine, Radio Physics section editor

A. S. Kulik, Doctor of Science in Engineering, Ukraine

D. V. Lebedev, Doctor of Science in Engineering, Ukraine, Control section editor

V. G. Levashenko, Ph.D, Slovakia

A. Lisnianski, Ph.D, Israel

U. Markowska-Kaczmar, Doctor of Science, Poland

V. A. Oleshchuk, Ph.D in Physics and Mathematics, Norway, Radio Electronics section editor

V. M. Onufrienko, Doctor of Science in Physics and Mathematics, Ukraine

M. Paprzycki, Ph.D, Poland

D. M. Piza, Doctor of Science in Engineering, Ukraine

O. I. Rubel, Ph.D, Canada

A. A. Sharpanskykh, Ph.D, Netherlands, Computer Science section editor

S. N. Vassilyev, Doctor of Science in Physics and Mathematics, Academician of Russian Academy of Sciences, Russia

E. N. Zaitseva, Ph.D, Slovakia

Recommended for publication by the Academic Council of ZNTU, protocol № 13 dated 01.07.2016.

The journal is imposed by the editorial-publishing department of ZNTU.

The journal web-site is <http://ric.zntu.edu.ua>.

The address of the editorial office: Editorial office of the journal «Radio Electronics, Computer Science, Control», Zaporizhzhia National Technical University, Zhukovskiy street, 64, Zaporizhzhya, 69063, Ukraine.

Tel.: +38-061-769-82-96 – the editorial-publishing department.

Fax: +38-061-764-46-62

E-mail: rvv@zntu.edu.ua

ЗМІСТ

РАДІОФІЗИКА.....	7
<i>Ivanov I. A., Korolev P. S., Polesskiy S. N., Zhadnov V. V.</i> THE DESIGN PROCEDURE OF SPECIFIED OPERATING LIFE OF FIBER-OPTIC CABLES.....	7
РАДІОЕЛЕКТРОНІКА ТА ТЕЛЕКОМУНІКАЦІЇ.....	15
<i>Осадчук О. В., Коваль К. О., Припула М. О.</i> РАДІОВИМІРЮВАЛЬНИЙ ПЕРЕТВОРЮВАЧ МАГНІТНОГО ПОЛЯ НА ТРАНЗИСТОРНИЙ СТРУКТУРІ.....	15
МАТЕМАТИЧНЕ ТА КОМП'ЮТЕРНЕ МОДЕЛЮВАННЯ.....	20
<i>Прохоров А. В., Кузнецова Ю. А.</i> АГЕНТНОЕ МОДЕЛИРОВАНИЕ МУЛЬТИКЛАСТЕРНЫХ ТЕХНОЛОГИЧЕСКИХ КОМПЛЕКСОВ НАНОЭЛЕКТРОННЫХ ПРОИЗВОДСТВ.....	20
НЕЙРОІНФОРМАТИКА ТА ІНТЕЛЕКТУАЛЬНІ СИСТЕМИ.....	29
<i>Бісікало О. В., Васілевський О. М.</i> ОЦІНКА НЕВИЗНАЧЕНОСТІ ВИМІРЮВАННЯ СЕНСУ ПРИРОДНО-МОВНИХ КОНСТРУКЦІЙ.....	29
<i>Бодянский С. В., Тищенко О. К., Бойко О. О.</i> ЕВОЛЮЦІЙНА КАСКАДНА СИСТЕМА НА ОСНОВІ НЕЙРО-ФАЗЗИ ВУЗЛІВ.....	40
<i>Левин В. И.</i> РАСЧЕТ И АНАЛИЗ ПОВЕДЕНИЯ НЕПОЛНОСТЬЮ ОПРЕДЕЛЕННЫХ ФУНКЦИЙ МЕТОДОМ ДЕТЕРМИНИЗАЦИИ.....	46
<i>Скрупский С. Ю.</i> ЭКСПЕРИМЕНТАЛЬНОЕ ИССЛЕДОВАНИЕ МЕТОДА СИНТЕЗА НЕЙРО-НЕЧЕТКИХ МОДЕЛЕЙ В ПАРАЛЛЕЛЬНОЙ КОМПЬЮТЕРНОЙ СИСТЕМЕ.....	56
<i>Субботин С. А.</i> МЕТОД СИНТЕЗА ДИАГНОСТИЧЕСКИХ МОДЕЛЕЙ НА ОСНОВЕ РАДИАЛЬНО-БАЗИСНЫХ НЕЙРОННЫХ СЕТЕЙ С ПОДДЕРЖКОЙ ОБОБЩАЮЩИХ СВОЙСТВ.....	64
<i>Штовба С. Д., Галушак А. В.</i> КРИТЕРІЙ НАВЧАННЯ НЕЧІТКОГО КЛАСИФІКАТОРА НА ОСНОВІ ВІДСТАНІ МІЖ ГОЛОВНИМИ КОНКУРЕНТАМИ.....	70
ПРОГРЕСИВНІ ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ.....	77
<i>Гальченко А. В., Козіна Г. Л.</i> МОДИФІКАЦІЯ АЛГОРИТМУ ЗАПЕРЕЧУВАНОВОГО ШИФРУВАННЯ МЕНГА.....	77
<i>Гороховатский В. А., Берестовский А. Е., Передрий Е. О.</i> СИСТЕМАТИЗАЦИЯ ПРОСТРАНСТВА СТРУКТУРНЫХ ПРИЗНАКОВ НА ОСНОВЕ МЕТОДОВ САМООБУЧЕНИЯ В ЦЕЛЯХ РЕЗУЛЬТАТИВНОГО РАСПОЗНАВАНИЯ ИЗОБРАЖЕНИЙ.....	87
<i>Дичка І. А., Онай М. В., Дрозда Т. П.</i> МОДИФІКОВАНИЙ ВІКОННИЙ МЕТОД ОДНОКРАТНОГО МНОЖЕННЯ ТОЧКИ ЕЛІПТИЧНОЇ КРИВОЇ НА СКАЛЯР У ПОЛІ $GF(P)$	95
<i>Нич Л. Я., Шаховська Н. Б., Камінський Р. М.</i> ОЦІНЮВАННЯ ЕФЕКТИВНОСТІ ІНФОРМАЦІЙНОГО ПОШУКУ В СИСТЕМАХ КОНСОЛІДОВАНОЇ ІНФОРМАЦІЇ.....	103
УПРАВЛІННЯ У ТЕХНІЧНИХ СИСТЕМАХ.....	110
<i>Kołowrocki K., Kwiatkowska-Sarnecka B., Soszyńska-Budny J.</i> RELIABILITY AND RISK OPTIMIZATION OF MULTISTATE SYSTEMS WITH APPLICATION TO PORT TRANSPORTATION SYSTEM.....	110
<i>Лахно В. А.</i> МОДЕЛЬ ІНТЕЛЕКТУАЛЬНОЇ СИСТЕМИ УПРАВЛІННЯ ГОРОДСКИМИ АВТОБУСНИМИ ПЕРЕВОЗКАМИ.....	119

CONTENTS

RADIOPHYSICS.....	7
<i>Ivanov I. A., Korolev P. S., Polesskiy S. N., Zhadnov V. V.</i> THE DESIGN PROCEDURE OF SPECIFIED OPERATING LIFE OF FIBER-OPTIC CABLES.....	7
RADIO ELECTRONICS AND TELECOMMUNICATIONS.....	15
<i>Osadchuk O. V., Koval K. O., Prytula M. O.</i> THE RADIOMEASURING CONVERTER OF THE MAGNETIC FIELD BASED ON TRANSISTORS STRUCTURE.....	15
MATHEMATICAL AND COMPUTER MODELLING.....	20
<i>Prokhorov A. V., Kuznetsova Yu. A.</i> AGENT-BASED MODELING OF MULTICLUSTER TECHNOLOGICAL COMPLEXES OF NANOELECTRONIC MANUFACTURING.....	20
NEUROINFORMATICS AND INTELLIGENT SYSTEMS.....	29
<i>Bisikalo O. V., Vasilevskiy O. M.</i> EVALUATION OF UNCERTAINTY MEASURING OF SENSE OF THE NATURAL LANGUAGE CONSTRUCTS.....	29
<i>Bodyanskiy Ye. V., Tyshchenko O. K., Boiko O. O.</i> AN EVOLVING CASCADE SYSTEM BASED ON NEURO-FUZZY NODES.....	40
<i>Levin V. I.</i> CALCULATION AND ANALYSIS OF INCOMPLETELY DEFINED FUNCTIONS BY DETERMINATION METHOD.....	46
<i>Skrupsky S. Yu.</i> EXPERIMENTAL INVESTIGATION OF METHOD FOR THE SYNTHESIS OF NEURO-FUZZY MODELS IN A PARALLEL COMPUTER SYSTEM.....	56
<i>Subbotin S. A.</i> THE METHOD OF DIAGNOSTIC MODEL SYNTHESIS BASED ON RADIAL BASIS NEURAL NETWORKS WITH THE SUPPORT OF GENERALIZATION PROPERTIES.....	64
<i>Shtovba S. D., Galushchak A. V.</i> FUZZY CLASSIFIER LEARNING BASED ON DISTANCE BETWEEN THE MAIN COMPETITORS.....	70
PROGRESSIVE INFORMATION TECHNOLOGIES.....	77
<i>Galchenko A. V., Kozina G. L.</i> MODIFICATION OF MENG'S DENIABLE ENCRYPTION ALGORITHM.....	77
<i>Gorokhovatsky V. A., Berestovskiy A. E., Peredrii E. O.</i> SYSTEMATIZATION OF SPACE OF STRUCTURAL FEATURES BASED ON SELF-LEARNING METHODS FOR EFFECTIVE IMAGE RECOGNITION.....	87
<i>Dychka I. A., Onai M. V., Drozda T. P.</i> MODIFIED METHOD FOR ELLIPTIC CURVE SCALAR POINT MULTIPLICATION OVER GF(P).....	95
<i>Nych L. Ya., Kaminskyj R. M., Shakhovska N. B.</i> EFFECTIVENESS EVALUATION OF SEARCH IN INFORMATION SYSTEMS WITH CONSOLIDATED INFORMATION.....	103
CONTROL IN TECHNICAL SYSTEMS.....	110
<i>Kołowrocki K., Kwiatuszewska-Sarnecka B., Soszyńska-Budny J.</i> RELIABILITY AND RISK OPTIMIZATION OF MULTISTATE SYSTEMS WITH APPLICATION TO PORT TRANSPORTATION SYSTEM.....	110
<i>Lakhno V. A.</i> MODEL OF INTELLIGENT MANAGEMENT SYSTEM OF CITY BUS TRANSPORTATIONS.....	119

РАДИОФИЗИКА

РАДИОФИЗИКА

RADIOPHYSICS

UDC 621.396.6.019

Ivanov I. A.¹, Korolev P. S.², Polesskiy S. N.³, Zhadnov V. V.⁴

¹PhD, Senior teacher of department of Electronic Engineering, National research university 'Higher school of economics', Moscow, Russia

²BS, Master's student of department of Electronic Engineering, National research university 'Higher school of economics', Moscow, Russia

³PhD, Assistant professor of department of Computer Engineering, National research university 'Higher school of economics', Moscow, Russia

⁴PhD, Associate Professor, Professor of department of Electronic Engineering, National research university 'Higher school of economics', Moscow, Russia

THE DESIGN PROCEDURE OF SPECIFIED OPERATING LIFE OF FIBER-OPTIC CABLES

The task of prediction of fiber cable operating life for automating of design study of reliability of optical-fiber transmission system was solved in the paper. The paper offers generation method of mathematical models of complex coefficients, which brings out from the input models failure rates the coefficients, which considers the effects of regimes and conditions of usage of fiber cables on the durability indices.

The developed procedure does not require carrying out the experimental investigations and tests, based on the using of standardized models of failure rates and durability characteristics of fiber cables. Herewith for the prescribed modes and terms of usage of fiber cables automatically detected the right process of degradation, which determines values of operating life.

In contrast to the prediction technique of equipment operating life on the base of the probabilistic-physical failure patterns, the offered method operates with the dates, which are stated in the standard-technical documentation on the fiber cables, and allows to increase forecast precision of the operating life in comparison with the methods, which is recommended in the branch standards.

In the paper were developed software, which implements offered mathematical tool, by which were solved real-world forecasting problem of the operating live of the cable type OK-IIH-01-5-60.

Keywords: optical-fiber transmission system, optical-fiber cable, reliability, durability, operating life, failure rate.

NOMENCLATURE

$d_{\max}$ is a maximal value of the damping constant in the fiber cable, by which provides operation of fiber-optic systems of information transmission;

I is a quantity of self-contained failure flows of the component parts of electro-radio elements;

t is a quantity of operating temperature (except the maximal allowed according to the technical requirements);

J_i is a quantity of factors, which is considered in i failure flow;

L_k is a length of fiber cable;

N is a quantity of the bends (back winding, etc.) of the cable per time of its operation;

m is a quantity of the optic fibers in the cable;

t is a operating time of the cable;

$t(\text{work})$ is a cable temperature in the operating mode, °C;

t_i is a total interval of operating time of the cable by the temperature T_i ;

$t_{T_{\max}}$ is a total interval of operating time of the cable by the maximal temperature according to the technical requirements;

$T_{\max}$ is a maximal allowed cable temperature according to the technical requirements;

λ_{b1} is a base failure rate of the optical fiber during the process of its operating time, referred to the 1m cable length;

λ_{b2} is a base sudden failure rate of the optical fiber composed of fiber cables during the process of its repeated back winding referred to the 1m cable length;

λ_{b3} is a base sudden failure rate of structure of cables during the process of its operating time, referred to the 1m cable length;

λ_{b4} is a base sudden failure rate of structure of cables during the process of its repeated back winding, referred to the 1m cable length;

λ_{b5} is a base gradual failure rate of fiber cables during the process of its operating time;

λ_{bi} is a base failure rates i failure flow;

λ_e is a failure rate of the fiber cable on the maximal regime and conditions of usage according to the technical requirements;

$\lambda_e(\text{work})$ is a failure rate of the fiber cable on the operating mode;

v is a constant variation of the resource; $\chi_\gamma = 95\%$, $\chi_\gamma = 99,9\%$ – fractile of the normalized normal distribution

for the probability $\gamma = 95\%$ и $\gamma = 99,9\%$ accordingly;

K_{ry}^* (work) is a value of the standardized coefficient for the operating mode of the cable;

K_{ij} is a coefficient, which considers j factor in i failure flow;

K_{xy} is a coefficient, which depends on the mode and conditions of usage of cable;

K_E is a coefficient, which depends on energy of activation of degradation processes;

K_{or} is a coefficient of cable operation rate in the regime of application of communication line;

CS_1 is a serviceability criterion;

K_{cs1} is a coefficient of serviceability criterion of fiber cables by quantity of attenuation;

K_c is a cable capacity factor (by critical parameter);

$K_{av,b}$ is an average quantity of the bends (back winding, etc.) of the cable per time unit of its operating;

K_{T1} is a temperature coefficient of degradation speed of statical, mechanical strength of optical fiber;

K_{T2} is a temperature coefficient of change of dynamic, mechanical strength of optical fiber and cable sheath;

K_{T3} is a temperature coefficient of degradation speed of safety and reinforcing elements of structure of cable;

K_{T4} is a temperature coefficient of speed change of tensile strength of cable serving;

K_{T7} is a coefficient, which characterize maximal reversible changes of the damping constant in the fiber cable in over the range negative operating temperature;

K_e is a stiffness coefficient of operating conditions;

$H(\text{work})$ is a load of cable (by critical parameter) in the mode of operation;

$H(TR)$ is a maximum permissible load of the cable (by critical parameter) according to the technical requirements;

$T_{o,k}$ is a total operating time of the cable in a time $T_{o,1}$;

$T_{o,1}$ is a total operating time of the communication line during the year;

$T_{r,\gamma}$ is a gamma-percentile operating life before the writing of the cable in all regimes and terms of operation under the technical requirements;

T_{eq} is an operative temperature of components operation;

ε is a mean-root square error.

INTRODUCTION

In the period of intensive development of information and communication means the important place in our life plays systems of transfer, processing and storing of information. In spite of all-round introduction of wireless technologies, the most widespread has still the systems, which are constructed on the wire communication lines. In the systems of information transmission, where it is necessary to provide high speed of information exchange, quality and reliability, mainly will be applied the fiber-optic cables.

Obviously, by the design of such type communication lines the durability characteristics of fiber cables (minimal time between failures, gamma-percentile operating life, life time) will determine extensively indices of its durability. Consequently for increasing of the accuracy and validity of calculation value of durability characteristics of communication lines on the base of fiber cables it is

necessary to use not only data, which is stated in the technical conditions for the cable, but also features of exploitation patterns composed of communication line, procedure and conditions of application.

The object of research is the forecasting method of optical cables resource in the reference data and the parameters of the modes and conditions of their application.

Calculated methods of operating life prediction [1, 2], usually, characterizes low self-descriptiveness [2], and also considerable time spent for the experimental investigation [1]. In the early design stage, the usage of such methods is difficult by the task solution of provision of the necessary level of durability indices. This situation brings about the necessity of models design, which does not demand carrying out of the tests and more fully considered the influence of conditions and terms of usage of fiber cables.

Subject of investigation are the methods and models, which is used for analysis of the design level of durability of fiber cables.

The goal of this work is design of methodology, which allows to increase the accuracy of prediction of durability indices of fiber cables and ensure the probability calculation of conditions of usage, in the frame of the approved at the present time calculation methodology of the operating life of electronic means.

1 PROBLEM STATEMENT

We consider characteristics of cable durability $T_{p,\gamma TY} = \{T_{p,\gamma 1TY}, T_{p,\gamma 2TY}, \dots, T_{p,\gamma I TY}\}$ – set I gamma-percentile operating in maximum and lite conditions under technical requirements and corresponding to its condition use $y = \{R_{1TY}, R_{2TY}, \dots, R_{ITY}\}$ – set I vectors параметров parameter conditions, and also exploitation pattern optical-fiber transmission system $t = \{t_1, t_2, \dots, t_k\}$ – set K residence time of optical-fiber transmission system in different operation conditions, waiting and storing and respective application condition, waiting and storing $R = \{R_1, R_2, \dots, R_K\}$.

In that case forecasting problem of gamma-percentile operating life of fiber cables will be concerned with determination of relation $T_{p,\gamma} = F(\Pi_1, \Pi_2, \dots, \Pi_L)$ and parameter value (Π) .

For quality rating of formed model can be used the criteria of mean-root square error [3]:

$$\varepsilon = \sum_{i=1}^L \left[T_{p,\gamma_i TY} - F(\Pi_1, \Pi_2, \dots, \Pi_{L_i}) \right]^2 \rightarrow 0 \text{ при } L \rightarrow \max. \quad (1)$$

2 LITERATURE REVIEW

Calculated methods of forecasting of operating life can be divided into probabilistic and determinate.

Probabilistic methods [1, 4–5] assumes pilot researches of degradation process of electro-radio equipment, according to which forms the functions of distribution of accrued operating time and determines parameters. Yet even if prefer the hypothesis of one or other kind of cumulative distribution curve, then determination of its parameters according to the technical requirements will be nontrivial task, let alone about receiving of relations from the modes and operating conditions of electro-radio equipment[6]. Though, that the probabilistic methods found their way in a number of national standards

of CIS countries (for example, [7]), in the engineering practical work still use determinate methods [2].

Basic data for calculation of durability indices of electro-radio equipment according to the methodology, stated in [2] are the normalized in technical requirements durability characteristics of electro-radio equipment and parameters of the model of communication line.

Value of gamma-percentile operating life of electro-radio equipment in operating mode is calculated under the mathematical model:

$$T_{r,\gamma}(\text{work}) = \frac{T_{r,\gamma}}{K_c \cdot K_{o,r}}, \quad (2)$$

where

$$K_{o,r} = \frac{T_{o,k}}{T_{o,l}}, \quad K_c = \frac{H(\text{work})}{H(TR)}.$$

Values $T_{r,\gamma}$ stated in the data book [8]. Fig. 1 shows a fragment with data of reliability characteristics of fiber cables.

Справочник "Надежность ЭРИ"

Компоненты волоконно-оптических систем
передачи информации

Таблица 8

**Характеристика надежности и справочные данные
отдельных марок оптических кабелей**

Марка кабеля	d ₁ , шт	λ ₆₁ , 1/ч.м	d ₂ , шт	λ ₆₂ , 10 ⁶ , 1/пер. м	d ₃ , шт	λ ₆₃ , 1/ч.м	d ₄ , шт	λ ₆₄ , 10 ⁶ , 1/пер. м	d ₅ , шт	λ ₆₅ , 1/ч	T _{н.м} , тыс.ч (при t _{раб.макс} , °C)	T _{р,γ} , тыс.ч (γ=95%)
Оптические кабели												
<i>Монтажные и для подвижных объектов</i>												
ОК-БС01	0	3,25· 10 ⁻¹²	-	-	0	4,44· 10 ⁻¹⁰	-	-	7	4,12· 10 ⁻⁹	10 (70)	11,2
ОК-БС06	0	6,58· 10 ⁻¹⁰	-	-	0	3,72· 10 ⁻¹⁰	-	-	0	9,03· 10 ⁻⁸	10 (85)	15
ОК-БС07	0	1,94· 10 ⁻¹¹	-	-	0	4,25· 10 ⁻¹⁰	-	-	0	2,93· 10 ⁻⁹	10 (85)	15
ОК-МС01	0	3,41· 10 ⁻¹²	-	-	0	3,54· 10 ⁻¹⁰	-	-	0	4,67· 10 ⁻¹⁰	10 (70)	11,2
ОК-МС06	0	7,67· 10 ⁻¹⁰	-	-	0	3,27· 10 ⁻¹⁰	-	-	0	1,06· 10 ⁻⁷	10 (85)	15
ОК-МС09	0	1,21· 10 ⁻⁹	-	-	0	5,42· 10 ⁻¹⁰	-	-	0	3,55· 10 ⁻⁷	10 (85)	15
ОК-МС11	0	1,17· 10 ⁻⁸	-	-	0	5,03· 10 ⁻⁹	-	-	0	6,38· 10 ⁻⁶	10 (85)	15
СБ-50-2	0	3,56· 10 ⁻¹⁵	-	-	0	1,02· 10 ⁻¹⁴	-	-	8	5,46· 10 ⁻¹²	1 (200)	1,5
СБ-200-2	0	8,64· 10 ⁻¹⁵	-	-	0	2,16· 10 ⁻¹⁴	-	-	0	1,84· 10 ⁻¹¹	1 (200)	1,5
<i>Полевые и для стационарных объектов и сооружений</i>												
ОК-СС01	0	4,62· 10 ⁻¹⁰	-	-	0	7,88· 10 ⁻¹⁰	-	-	0	6,32· 10 ⁻⁸	5 (85)	10
ОК-СС02	0	9,31· 10 ⁻¹⁰	-	-	0	5,34· 10 ⁻⁹	-	-	1	8,74· 10 ⁻⁷	150 (50)	187,5
ОК-СС03	0	4,66· 10 ⁻¹⁰	-	-	0	5,36· 10 ⁻⁹	-	-	0	2,61· 10 ⁻⁷	150 (50)	187,5
ОК-ПН-01	0	9,05· 10 ⁻⁹	1	0,0098	3	2,47· 10 ⁻⁸	3	2,22	6	1,74· 10 ⁻⁵	30 (70)	60

Figure 1 – Table fragment from the data book "Reliability of electro-radio product"

As fig. 1 shows according to the classification of the data book [1] fiber cables are the electro-radio equipment, which are related to the group “Fiber cables” type “Components of optical fiber systems of information transmission”, for which standardized $T_{m,h}$ и $T_{p,\gamma}$.

Consider the value calculation $T_{r,\gamma}$ (work) by the example of fiber cable type OK-ПН-01-5-60. The cable is intended for operating in the communication lines by conditions of stationary, nonstationary and air laying and operation in the field conditions with the repeated laying.

Construction of the cable consists of quartz optic fibers, longitudinal laid together with thread CBM-K in cover made from the epoxyacrylate (fig. 2).

According to the technical requirements [9] are indicated following values of no failure operating time (minimal) of the cable for the application mode:

- to 35°C – 200000 h. (light mode);
- to 55°C – 100000 h. (light mode);
- 70°C – 30000 h (maximum mode).

As follows from the fig. 2 and [9] “critical” parameter for the cable type OK-ПН-01 is the temperature. In that case

$$K_c = \frac{t(work)}{t_{work_max}}. \quad (3)$$

Find the values K_c by formula (1) by $K_{or} = 1$ and by formula (2). The results of calculation are shown in the table 1.

As follows from the table 1 by using formula (2) if in the technical requirements are not indicated data for the facilitated regimes, above considered method of standard [2] gives underestimation of durability characteristics of the fiber cables.

Consequently, for improving the accuracy of prediction of operating life of fiber cables it is necessary to find mathematical model, which adequate describes relation of operating life from the modes and conditions usage of cables, and determine parameters and coefficients.

3 MATERIALS AND METHODS

For receiving of calculation correlation of estimation of gamma-percentile operating life of fiber cables on the facilitated regimes we will take principle of duality [10], reasoning from we can state, that [11, 12]:

$$\frac{T_{r,\gamma} work}{T_{r,\gamma}} \sim \frac{\lambda_e}{\lambda_e(work)}. \quad (4)$$



Figure 2 – Sketch of construction cable type OK-ПН-01-5-6/0

Table 1 – The results of calculation of capacity factor

№	$t(work), ^\circ C$	Capacity factor	
		Model (1)	Model (2)
1	2	3	4
1	35	0,15	0,5
2	55	0,3	0,786
3	70	1,0	1,0

Mathematical calculation modes of failure rates of fiber cables are stated in the data book [8]. As the fiber cables refers to the group of complex items, cumulative, which compounded of self-contained failure flows of its component parts, mathematical calculation mode of its failure rates has the following view:

$$\lambda_e = \sum_{i=1}^I \lambda_{e_i} = \sum_{i=1}^I \left[\lambda_{b_i} \cdot \prod_{j=1}^{J_i} (K_{i,j}) \right]. \quad (5)$$

Mathematical modes of the failure rates (λ_{e_i}) has the following view:

$$\lambda_{e_1} = \lambda_{b_1} \cdot m \cdot K_{T1} \cdot L_K \cdot K_e; \quad (6)$$

$$\lambda_{e_2} = \lambda_{b_2} \cdot m \cdot K_{av,b} \cdot K_{T2} \cdot L_K \cdot K_e; \quad (7)$$

$$\lambda_{e_3} = \lambda_{b_3} \cdot K_{T3} \cdot L_K \cdot K_e; \quad (8)$$

$$\lambda_{e_4} = \lambda_{b_4} \cdot K_{av,b} \cdot K_{T4} \cdot L_K \cdot K_e; \quad (9)$$

$$\lambda_{e_5} = \lambda_{b_5} \cdot m \cdot K_{T1} \cdot K_{cs1}. \quad (10)$$

Coefficient values K_{T1} and K_{T3} calculated by the formula:

$$K_T = e^{-K_E \left(\frac{1}{T_{eq}} - \frac{1}{298} \right)}. \quad (11)$$

Value T_{eq} determines by the formula:

$$T_{eq} = \left[\frac{1}{T_{max}} + \frac{1}{K_E} \cdot \ln \left(\frac{t}{t_{T_{max}} + \sum_{i=1}^I t_i^*} \right) \right]^{-1}. \quad (12)$$

Value t_i^* determines by the formula:

$$t_i^* = t_i \cdot e^{-K_E \left(\frac{1}{T_i} - \frac{1}{T_{max}} \right)}, \quad (13)$$

where:

$$t = t_{T_{max}} + \sum_{i=1}^I t_i.$$

Value $K_{av,b}$ determines by the formula:

$$K_{av,b} = \frac{N}{t}. \quad (14)$$

Value of the serviceability criterion CS_1 , by which defines value of coefficient K_{cs1} , determines by the formula:

$$CS_1 = \frac{d_{max}}{K_{T7}}. \quad (15)$$

Since the negative temperature is the factor, which reduces the rate of chemical reaction, then in the calculation durability value models (7) and (9) can be disregarded. In addition, analysis of the formulas (6), (8) and (10)–(15) and the tables of the data book [1] shows, that the parameters of

the modes and conditions of usage of fiber cables are K_e , N , t , $T_{\max}$, $t_{T\max}$, set of values T_p , set of values t_i and I . Remaining parameters and coefficients of the formulas (6), (8) and (10)–(15), determines in fact values λ_{bi} , as it depends on the features technological structure of the cable.

According to the methodology of the standard [2], by the calculation of the durability values of the equipment, by which total failure flow rate consists of self-contained failure flows of its component parts, the value of the resource determines in the following way:

$$T_{r,\gamma} = \min_{i=1,3} \{T_{r,\gamma_1}, T_{r,\gamma_3}, T_{r,\gamma_5}\}. \quad (16)$$

Hence, using (4) find the formula for calculation of the coefficients, which depends on the mode and conditions of usage of cable, for each self-contained failure flow:

$$K_{H1} = \frac{\lambda_{e1}(\text{work})}{\lambda_{e1}} = \frac{K_{T1}(\text{work}) \cdot K_e(\text{work})}{K_{T1} \cdot K_e}; \quad (17)$$

$$K_{H3} = \frac{\lambda_{e3}(\text{work})}{\lambda_{e3}} = \frac{K_{T3}(\text{work}) \cdot K_e(\text{work})}{K_{T3} \cdot K_e}; \quad (18)$$

$$K_{H5} = \frac{\lambda_{e5}(\text{work})}{\lambda_{e5}} = K_{T1}(\text{work}). \quad (19)$$

In formulas (17)–(19) indicate:

$$\prod_{j=1}^{J_i} K_{i,j} = K_{r,\gamma_i}.$$

4 EXPERIMENTS

To verify the proposed method find the value $K_{r,\gamma}$ for maximal the mode and conditions of usage of cable type

OK-III-01-5-6/0 according to the data, indicated in the data book [8] and technical requirement [9]. Table 2 shows the results of calculation.

In a similar way calculate the value $K_{r,\gamma}$ for the temperature $T_{\max} = 35^\circ\text{C}$ and $T_{\max} = 55^\circ\text{C}$ and divide it into the limit value from the table 2. Table 3 shows the results of calculation.

Data, which is stated in the table can be considered as functions $T_{o,m}(K_{r,\gamma}^*)$, defined of its values. Besides, as the table 3 shows, functions $T_{o,m}(K_{r,\gamma_1}^*)$ and $T_{o,m}(K_{r,\gamma_5}^*)$ are congruent. So the form (16) has the following view:

$$T_{r,\gamma} = \min \{T_{r,\gamma_1}, T_{r,\gamma_3}\}.$$

Find the value 95% of the cable resource for the temperature 35°C and 55°C . For that using the formula, which is stated in the standard [2]:

$$T_{r,\gamma} = \frac{1 - v \cdot \chi_{\gamma=95\%}}{1 - v \cdot \chi_{\gamma=99,9\%}} \cdot T_{o,m}. \quad (20)$$

Find the value of the constant variation, solve (20) relatively v by values $T_{o,m}$ and $T_{r,\gamma}$, stated on the picture 2:

$$v = \frac{T_{r,\gamma} - T_{o,m}}{\chi_{\gamma=99,9\%} \cdot T_{r,\gamma} - \chi_{\gamma=95\%} \cdot T_{o,m}} = 0,22.$$

By that value constant variation 95% cable resource by the temperature 35°C is 400000 h., by temperature 55°C – 200000 h.

5 RESULTS

For the approximation of the relationship $T_{r,\gamma}(K_{r,\gamma}^*)$ can be used piecewise linear function. Besides, in the technical requirements [3] values $T_{o,m}$ and for normal temperature (25°C) are not stated, let take, in the temperature range 25 – 35°C value is constant and equal to 400000 h.

Table 2 – Values of the coefficient $K_{p,\gamma i}$

№	Design formula	Value	Reference
1	2	3	4
1	$K_{r,\gamma_1} = K_{T1} \cdot K_e$	371.3	$I = 0$; $t_i = 0$; $t_{T\max} = t$; $t = 30000$ h – minimal operating time according to the technical requirements [3]; $T_{\max} = 70^\circ\text{C}$ (343°K) – according to the technical requirements [3]; $K_{E1} = 13.44 \cdot 10^3$ – according to the table in the data book [2]; $K_e = 1$ (group 1.1) – according to the table in the data book [2]
2	$K_{r,\gamma_3} = K_{T3} \cdot K_e$	34.6	$I = 0$; $t_i = 0$; $t_{T\max} = t$; $t = 30000$ h – minimal operating time according to the technical requirements [3]; $T_{\max} = 70^\circ\text{C}$ (343°K) – according to the technical requirements [3]; $K_{E3} = 8.05 \cdot 10^3$ – according to the table in the data book [2]; $K_e = 1$ (group 1.1) – according to the table in the data book [2]
3	$K_{r,\gamma_5} = K_{T1}$	371.3	$I = 0$; $t_i = 0$; $t_{T\max} = t$; $t = 30000$ h – minimal operating time according to the technical requirements [3]; $T_{\max} = 70^\circ\text{C}$ (343°K) – according to the technical requirements [3]; $K_{E1} = 13.44 \cdot 10^3$ – according to the table in the data book [2]

Table 3 – Value of the standardized coefficients $K_{r,\gamma i}^*$

№	Temperature	K_{r,γ_1}^*	K_{r,γ_3}^*	K_{r,γ_5}^*	$T_{o,m}$, thousand. h.
1	2	3	4	5	6
1	35°C	$1,16 \cdot 10^{-2}$	$6,94 \cdot 10^{-2}$	$1,16 \cdot 10^{-2}$	200,0
2	55°C	0,167	0,341	0,167	100,0
3	70°C	1,0	1,0	1,0	30,0

Having absorbed the disclaimers of the relationship $T_{r,\gamma}$ from K_{ry}^* can be represented in the form:

$$T_{r,\gamma}(\text{work}) = \begin{cases} 400000 \text{ npu} & 0 \leq K_{r,y}^*(\text{work}) \leq K_{r,y_1}^* \\ T_{r,\gamma_n} + \frac{T_{r,\gamma_{n+1}} - T_{r,\gamma_n}}{K_{r,y_{n+1}}^* - K_{r,y_n}^*} \cdot [K_{r,y}^*(\text{work}) - K_{r,y_n}^*] & \text{для } K_{r,y_n}^* < K_{r,y}^*(\text{work}) \leq K_{r,y_{n+1}}^* \end{cases}$$

Diagrams of relationship $T_{r,\gamma}(\text{work})$ from $K_{ry1}^*(\text{work})$ and $K_{ry3}^*(\text{work})$ are stated on the fig. 3.

Fig. 4 shows relation $K_{ry1}^*(\text{work})$ from the temperature T_i and $(\frac{\Delta t}{t})K_i$, i.e. interval size of cable operation by the temperature T_i to the total hours of service t , which was received with the help of designed software in the MathCad programming language [13].

Calculation of the values $T_{r,\gamma}(\text{work.g})$ for the group of equipment, which uses cable, distinct from the group 1.1 calculated by the formula [5]:

$$T_{r,\gamma}(\text{work.g}) = \frac{T_{r,\gamma}(\text{work})}{K_e}$$

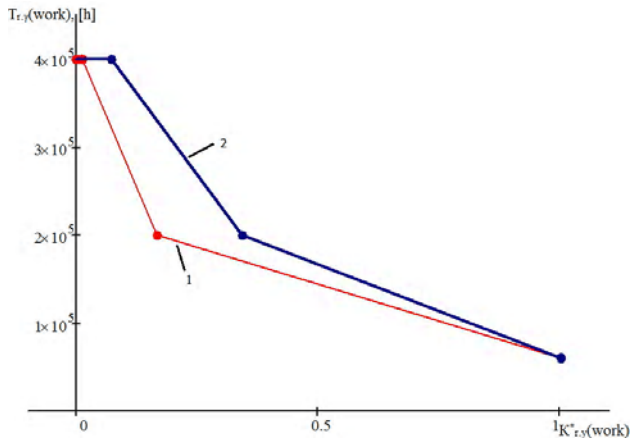


Figure 3 – Relationship of gamma-percentile operating life from the standardized coefficient: 1 – from $K_{ry1}^*(\text{work})$; 2 – $K_{ry3}^*(\text{work})$

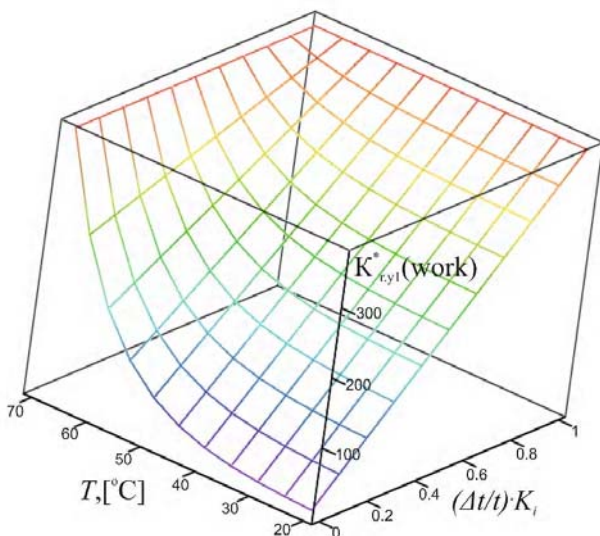


Figure 4 – Relation $K_{ry1}^*(\text{work})$ from the temperature T_i and interval size of cable operation by the temperature T_i to the total hours of service t

Moreover, it should also be noted, that in this methodology value $K_{o,r}$ is not used, as the parameters of the exploitation pattern explicitly enters into the formula (11). Therefore the received value will be numerically equal the service life of the cable.

6 DISCUSSION

Suggested methodology of forecasting of operating life of fiber cables in comparison with methodology [2] provides substantially larger calculation accuracy. However suggested methodology demands high volume of basic data and higher labor-intensiveness of calculation, than methodology [2].

In comparison with methodologies, which are based on the probabilistic methods [1, 4–5] suggested methodology does not requires experimental researches and tests, at the same time provides comparable accuracies of forecasting of durability indices.

Weakness of the suggested methodology is using piecewise-linear approximation, thus complicating carrying out of calculation.

Efficient use of the developed methodology will be higher with large quantity of operating life values stated in the technical requirements for the fiber cable. In case in the technical requirements is stated only one value, the effect from the developed methodology will be insignificant.

CONCLUSIONS

In the paper were solved the task of forecasting of operating life of fiber cables according to the reference data and characteristics of the modes and its conditions of usage.

Scientific novelty of the results, which were received in the article is that for the first time were suggested the methodology of forecasting of durability indices of fiber cables, which allows automatically select from the totality of failure rate the right failure rate, which determines the value of its durability indices, and also making the registration of influence of the modes and conditions usage on is, which allows to increase calculation accuracy and also detect the causes, which has an influence upon the level of calculated values.

Practical significance of the received results is that were developed software, which realizes the suggested method, on the base of which was solved practical task of forecasting of operating life of cable type OK-ПН-01-5-6/0.

Prospects of further investigations are to research performance capability of the suggested methodology for forecasting of durability indices for other groups of components of fiber optic systems of information transmission, such as fiber connectors, fiber switchers, fiber couplers and splitters and optic-electronic modules.

ACKNOWLEDGEMENTS

The paper was conducted within the framework of research project “Development of the methods and

methodologies of forecasting of electronic means durability at all stages of life cycle” (№ 15-05-0029), which was performed with the support of the HSE Academic Fund in 2015.

REFERENCES

1. Комплексная система контроля качества. Аппаратура, приборы, устройства и оборудование военного назначения. Типовые методики оценки показателей безотказности и ремонтпригодности расчетно-экспериментальными методами. РДВ 319.01.16-98. [дата введения 1999-07-01]. – М. : МО РФ, 1998. – 55 с. (Руководящий документ).
2. Аппаратура радиоэлектронная. Определение показателей долговечности. ОСТ 4.012.013-84. [дата введения 1985-07-01]. – М. : ВНИИ, 1985. – 15 с. (Отраслевой стандарт)
3. Маслов А. Я. Оптимизация радиоэлектронной аппаратуры / А. Я. Маслов, А. А. Чернышов, В. В. Ведерников ; под. ред. А. Я. Маслова, А. А. Чернышева. – М. : Радио и связь, 1982. – 200 с.
4. Дружинин Г. В. Надежность автоматизированных систем. Изд. 3-е, перераб. и доп. / Г. В. Дружинин. – М. : Энергия, 1977. – 536 с.
5. Грибов В. М., Оценивание и прогнозирование надежности бортового аэрокосмического оборудования / В. М. Грибов, Ю. Н. Кофанов, В. П. Стрельников. – М. : НИУ ВШЭ, 2013. – 495 с.
6. Черкесов Г. Н. Надежность аппаратно-программных комплексов : учебное пособие. / Г. Н. Черкесов, Г. Н. Черкесов. – СПб. : Питер, 2005. – 479 с.
7. Надежность в технике. Модели отказов. Основные положения. ДСТУ 3433-96 [дата введения 1999-01-01]. – К. : Держстандарт Украины, 1999. – 98 с. (Национальный стандарт Украины).
8. Надежность ЭРИ : справочник. – М. : МО РФ, 2006. – 641 с.
9. ТУ 16.К71-026-88. Технические условия. Оптические кабели для полевых и стационарных объектов и сооружений.
10. Маклейн С. Категории для работающего математика: Пер. с англ. / С. Маклейн ; под ред. В. А. Артамонова. – М. : Физматлит, 2004. – 352 с.
11. Жаднов В. В. Расчетная оценка показателей долговечности электронных средств космических аппаратов и систем / В. В. Жаднов // Надежность и качество сложных систем. – 2013. – № 2. – С. 65–73.
12. Zhadnov V. V. Method the Estimated Life of Electronic Devices for Control and Communication Systems / V. V. Zhadnov // 2015 International Siberian Conference on Control and Communications (SIBCON). Proceedings. – Omsk : Omsk State Technical University, 2015.
13. Воскобойников Ю. Е. Основы вычислений и программирования в пакете MathCAD : учеб. пособие / Ю. Е. Воскобойников, А. Ф. Задорожный, Л. А. Литвинов ; под ред. Ю. Е. Воскобойникова. – Новосибирск : НГАСУ (Сибстрин), 2012. – 212 с.

Article was submitted 27.01.2016.

After revision 11.04.2016.

Иванов И. А.¹, Королев П. С.², Полесский С. Н.³, Жаднов В. В.⁴

¹Канд. техн. наук, старший преподаватель департамента Электронной инженерии, Национальный исследовательский университет «Высшая школа экономики», Москва, Россия

²Бакалавр, магистрант департамента Электронной инженерии, Национальный исследовательский университет «Высшая школа экономики», Москва, Россия

³Канд. техн. наук, доцент департамента Компьютерной инженерии, Национальный исследовательский университет «Высшая школа экономики», Москва, Россия

⁴Канд. техн. наук, доцент, профессор департамента Электронной инженерии, Национальный исследовательский университет «Высшая школа экономики», Москва, Россия

МЕТОДИКА РАСЧЕТА УТОЧНЕННОГО РЕСУРСА ОПТИЧЕСКИХ КАБЕЛЕЙ

Решена задача прогнозирования ресурса оптических кабелей для автоматизации проектных исследований надежности волоконно-оптических систем передачи информации. Предложена методика формирования математических моделей комплексных коэффициентов, которая позволяет выделить из исходных моделей интенсивностей отказов коэффициенты, учитывающие влияние режимов и условий применения оптических кабелей на их показатели долговечности.

Разработанная методика не требует проведения экспериментальных исследований и испытаний, базируясь на использовании стандартизованных моделей интенсивностей отказов и нормированных характеристиках долговечности оптических кабелей. При этом для заданных режимов и условий применения оптических кабелей автоматически выявляется тот процесс деградации, который определяет значения их ресурса.

В отличие от методик прогнозирования ресурса изделий на основе вероятностно-физических моделей отказов предложенный метод оперирует только с данными, приведенными в нормативно-технической документации на оптические кабели, и позволяет повысить точность прогнозирования ресурса по сравнению с методиками, рекомендованными в отраслевых стандартах.

Разработано программное обеспечение, реализующее предложенный математический аппарат, с помощью которого решена практическая задача прогнозирования ресурса кабеля марки ОК-ПН-01-5-60.

Ключевые слова: волоконно-оптическая система передачи информации, оптоволоконный кабель, надежность, долговечность, ресурс, интенсивность отказов.

Иванов И. А.¹, Корольов П. С.², Поліський С. Н.³, Жаднов В. В.⁴

¹Канд. техн. наук, старший викладач департаменту Електронної інженерії, Національний дослідницький університет «Вища школа економіки», Москва, Росія

²Бакалавр, магістрант департаменту Електронної інженерії, Національний дослідницький університет «Вища школа економіки», Москва, Росія

³Канд. техн. наук, доцент департаменту Комп'ютерної інженерії, Національний дослідницький університет «Вища школа економіки», Москва, Росія

⁴Канд. техн. наук, доцент, професор департаменту Електронної інженерії, Національний дослідницький університет «Вища школа економіки», Москва, Росія

МЕТОДИКА РОЗРАХУНКУ УТОЧНЕНОГО РЕСУРСУ ОПТИЧНИХ КАБЕЛІВ

Вирішено задачу прогнозування ресурсу оптичних кабелів для автоматизації проектних досліджень надійності волоконно-оптичних систем передачі інформації. Запропоновано методику формування математичних моделей комплексних коефіцієнтів, яка дозволяє виділити з вихідних моделей інтенсивностей відмов коефіцієнти, що враховують вплив режимів і умов застосування оптичних кабелів на їх показники довговічності.

Розроблена методика не вимагає проведення експериментальних досліджень і випробувань, базуючись на використанні стандартизованих моделей інтенсивностей відмов і нормованих характеристик довговічності оптичних кабелів. При цьому для заданих режимів і умов застосування оптичних кабелів автоматично виявляється той процес деградації, який визначає значення їх ресурсу.

На відміну від методик прогнозування ресурсу виробів на основі імовірно-фізичних моделей відмов запропонований метод оперує тільки з даними, наведеними в нормативно-технічній документації на оптичні кабелі, і дозволяє підвищити точність прогнозування ресурсу порівняно з методиками, рекомендованими в галузевих стандартах.

Розроблено програмне забезпечення, що реалізує запропонований математичний апарат, з допомогою якого вирішена практична задача прогнозування ресурсу кабелю марки ОК-ПН-01-5-60.

Ключові слова: волоконно-оптична система передачі інформації, оптоволоконний кабель, надійність, довговічність, ресурс, інтенсивність відмов.

REFERENCES

1. Kompleksnaya sistema kontrolya kachestva. Apparatura, pribory, ustroystva i oborudovanie voennogo naznacheniya. Tipovye metodiki otsenki pokazateley bezotkaznosti i remontprigodnosti raschetno-eksperimental'nymi metodami. RDV 319.01.16-98. [data vvedeniya 1999-07-01]. Moscow, MO RF, 1998, 55 p. (Rukovodyashchiy dokument).
2. Apparatura radioelektronnaya. Opredelenie pokazateley dolgovechnosti. OST 4.012.013-84. [data vvedeniya 1985-07-01]. Moscow, VNII, 1985, 15 p. (Otraslevoy standart)
3. Maslov A. Ya., Chernyshov A. A., Vedernikov V. V. Optimizatsiya radioelektronnay apparatury, pod. red. A. Ya. Maslova, A. A. Chernysheva. Moscow, Radio i svyaz', 1982, 200 p.
4. Druzhinin G. V. Nadezhnost' avtomatizirovannykh sistem. Izd. 3-e, pererab. i dop. Moscow, Energiya, 1977, 536 p.
5. Gribov V. M., Kofanov Yu. N., Strel'nikov V. P. Otsenivanie i prognozirovaniye nadezhnosti bortovogo aerokosmicheskogo oborudovaniya. Moscow, NIU VShE, 2013, 495 p.
6. Cherkasov G. N., Cherkasov G. N. Nadezhnost' apparatno-programmnykh kompleksov: Uchebnoe posobie. Sankt-Peterburg, Piter, 2005, 479 p.
7. Nadezhnost' v tekhnike. Modeli otkazov. Osnovnye polozheniya. DSTU 3433-96 [data vvedeniya 1999-01-01]. Kiev, Derzhstandart Ukrainy, 1999, 98 p. (Natsional'nyy standart Ukrainy).
8. Nadezhnost' ERI: spravochnik. Moscow, MO RF, 2006, 641 p.
9. TU 16.K71-026-88. Tekhnicheskie usloviya. Opticheskie kabeli dlya polevykh i statsionarnykh ob'ektov i sooruzheniy.
10. Makleyn S. Pod red. V. A. Artamonova Kategorii dlya rabotayushchego matematika : Per. s angl. Moscow, Fizmatlit, 2004, 352 p.
11. Zhadnov V. V. Raschetnaya otsenka pokazateley dolgovechnosti elektronnykh sredstv kosmicheskikh apparatov i sistem, *Nadezhnost' i kachestvo slozhnykh sistem*, 2013, No. 2, pp. 65–73.
12. Zhadnov V. V. Method the Estimated Life of Electronic Devices for Control and Communication Systems. *2015 International Siberian Conference on Control and Communications (SIBCON). Proceedings*. Omsk, Omsk State Technical University, 2015.
13. Voskoboynikov Yu. E., Zadorozhnyy A. F., Litvinov L. A. Pod red. Yu. E. Voskoboynikova Osnovy vychisleniy i programmirovaniya v pakete MathCAD: ucheb. posobie. Novosibirsk: NGASU (Sibstrin), 2012, 212 p.

РАДИОЕЛЕКТРОНИКА ТА ТЕЛЕКОМУНІКАЦІЇ

РАДИОЕЛЕКТРОНИКА ТА ТЕЛЕКОММУНІКАЦИИ

RADIO ELECTRONICS AND TELECOMMUNICATIONS

УДК 621.383

Осадчук О. В.¹, Коваль К. О.², Притула М. О.³

¹Д-р. техн. наук, професор, завідувач кафедри радіотехніки Вінницького національного технічного університету, Вінниця, Україна

²Канд. техн. наук, доцент, доцент кафедри радіотехніки Вінницького національного технічного університету, Вінниця, Україна

³Аспірант кафедри радіотехніки Вінницького національного технічного університету, Вінниця, Україна

РАДІОВИМІРЮВАЛЬНИЙ ПЕРЕТВОРЮВАЧ МАГНІТНОГО ПОЛЯ НА ТРАНЗИСТОРНІЙ СТРУКТУРІ

Вирішено завдання вимірювання напруженості магнітного поля радіовимірювальним перетворювачем магнітного поля на біполярних та польовому транзисторах. Запропоновано схемотехнічне рішення радіовимірювального перетворювача, який забезпечує перетворення напруженості магнітного поля в частоту електричного коливання, що дозволяє підвищити чутливість пристрою до напруженості магнітного поля. Проведено експеримент з дослідження розробленого радіовимірювального перетворювача магнітного поля. У результаті проведених експериментів досліджено залежності вихідної частоти та чутливості пристрою від напруженості магнітного поля.

Ключові слова: радіовимірювальний перетворювач, магнітотранзистор, напруженість магнітного поля, чутливість.

ВСТУП

Вимірювання магнітного поля знаходить практичне застосування при дослідженні властивостей матеріалів, у магнітній дефектоскопії, при вивченні магнітного поля Землі, при контролі магнітного поля в різних установках та інше [1].

Загальна тенденція в розвитку пристроїв вимірювання магнітного поля, зумовлена збільшенням вимог точності до них при одночасному ускладненні експлуатаційних умов. Все це обумовлює проводити пошуки і розробку нових фізичних явищ, нових матеріалів, нових методів вимірювання, що дозволило б покращити метрологічні та економічні показники пристроїв вимірювання магнітного поля.

Проблемою високоточного вимірювання магнітних полів займаються як українські (Вікуліна І. М., Готра З. Ю., Голяка Р. Л., Кухарчук В. В., Осадчук В. С., Осадчук О. В. та інші) так і закордонні вчені (Хомерікі О. К., Аша Ж., Віглеба Г., Егіазаряна Г. А., Новицького П. В., Стафеева В. І. та інші).

В сучасній техніці існують надчутливі датчики магнітного поля, які використовують явище надпровідності та ефект Джозефсона. Але конструктивні особливості та-

ких сенсорів дозволяють їх застосовувати лише в лабораторних умовах [2]. Використання інших видів датчиків магнітного поля у складі сучасних інтегральних пристроїв в даний час пов'язано з низкою проблем, серед яких головними є низька чутливість датчика до магнітного поля і відсутність загальної технології виготовлення функціональної інтегральної схеми і датчика. Водночас вдосконалюються алгоритми та схеми перетворення інформативних сигналів датчиків магнітного поля в вимірювальну величину [3].

Перспективним науковим напрямком є розробка та створення первинних радіовимірювальних перетворювачів, які реалізують принцип перетворення «магнітна індукція – частота», на основі реактивних властивостей напівпровідникових структур з від'ємним опором. Використання таких приладів виключає з їх конструкції аналого-цифрові перетворювачі, що дозволяє знизити собівартість систем контролю та управління, а також створити «інтелектуальні» високоточні радіовимірювальні перетворювачі магнітного поля в результаті поєднання на одному кристалі схем обробки інформації та сенсора магнітного поля [4].

Таким чином, актуальною науковою задачею є створення високочутливого радіовимірювального перетво-

рювача магнітного поля, який легко може бути реалізований в сучасній інтегральній технології.

Постановка задачі: розробити електричну схему радіовимірювального перетворювача магнітного поля з підвищеною чутливістю та провести її експериментальне дослідження.

Метою статті є розробка електричної схеми радіовимірювального перетворювача магнітного поля з підвищеною чутливістю та її експериментальне дослідження на основі теоретичних засад роботи двоколекторного магнітотранзистора.

АНАЛІЗ ПРИНЦИПУ ДІЇ ДВОКОЛЕКТОРНОГО МАГНІТОТРАНЗИСТОРА

Двоколекторний магнітотранзистор являє собою біполярний транзистор, колектор у якого розділений на дві частини. За відсутності магнітного поля інжектовані емітером носії заряду розподіляються рівномірно між колекторами і їхні струми рівні. Відповідно потенціали колекторів однакові (при рівних колекторних опорах R_K), і різниця напруг між колекторами дорівнює нулю ($U = 0$). Поперечне магнітне поле відхиляє потік носіїв у бік одного з колекторів (K2), що приводить до збільшення його струму і зменшення струму іншого колектора (K1). З цієї причини потенціал одного колектора зменшується, а іншого збільшується і напруга між колекторами росте зі збільшенням індукції магнітного поля. При зміні напрямку магнітного поля змінюється і знак напруги між колекторами [3].

Крім основного ефекту – перерозподілу інжектованих носіїв між колекторами у такому магнітотранзисторі одночасно діє й ефект зміни ефективної довжини бази. Явище зміни ефективної довжини бази знижує чутливість двоколекторного магнітотранзистора, тому що він збільшує струм колектора K1 і зменшує струм колектора K2, тобто його дія протилежна дії ефекту перерозподілу носіїв між колекторами [4].

Проведено розрахунок коефіцієнтів передачі струму двоколекторного магнітотранзистора h_{21B1} і h_{21B2} у колектори K1 і K2 на основі структури рис. 1, де колектори розташовані по всій поверхні його бічних граней. Спо-

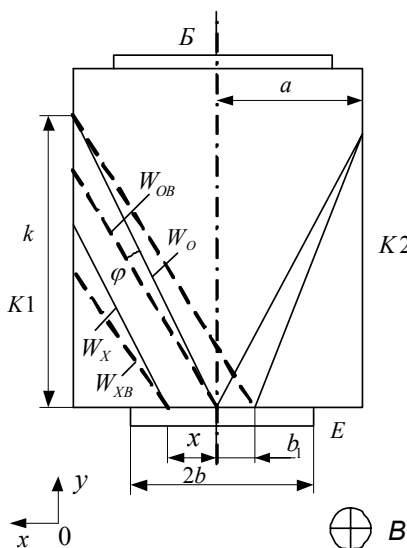


Рисунок 1 – Структура двоколекторного магнітотранзистора

чатку будемо вважати емітер близьким до точкового і розглянемо рух носіїв у бік колектора K1. Інкеровані в базу носії рухаються в напрямку y зі швидкістю $v = \mu E$ (напруженість електричного поля досить велика, так що швидкість дрейфу в цьому напрямку набагато більша швидкості дифузії).

За відсутності магнітного поля $W = W_0$, у магнітному полі носії відхиляються у бік колектора на кут Холла φ і траєкторія їхнього руху скорочується (штрихова лінія на рис. 1). На основі простих геометричних розрахунків, вважаючи $W_0 \approx h$ при невеликих φ ($\tan \varphi \ll 1$) маємо вираз для ефективної довжини бази [3]

$$W_{0B} = W_0 a / (h \cdot \tan \varphi + a) = a^2 E \chi (1 - a \chi E \cdot \tan \varphi), \quad (1)$$

де E – напруженість електричного поля; φ – кут Холла; $\chi = q / (2kT)$; h – відстань, яку проходять інжектовані в базу носії за час дифузії.

Коефіцієнт передачі транзистора та, з врахуванням $\tan \varphi = \mu B$, одержимо

$$h_{21B1} = 1 - (a^2 \chi / \mu \tau) (1 - a \chi E \mu B). \quad (2)$$

У реальній конструкції емітер витягнутий у напрямку x , отже, довжина траєкторії (відстань від емітера до колектора) для носіїв, інжектованих з різних частин емітера, неоднакова. Тому формулою (2) можна користатися лише для малої ділянки dx , розташованої на відстані x від осі симетрії транзистора. Відповідно

$$h_{21B1}(x) = 1 - [(a - x)^2 \chi / \mu \tau] [1 - (a - x) \chi E \mu B].$$

Струм через відповідну ділянку колектора K1

$$dI_{K1}(I_E / 2b) h_{21B1}(x) dx,$$

а через весь колектор

$$I_{K1} = (I_E / 2b) \int_0^b h_{21B1}(x) dx.$$

Тоді, коефіцієнт передачі струму дорівнює

$$h_{21B1} = (1 / 2b) \int_0^b h_{21B1}(x) dx. \quad (3)$$

Підставивши рівняння (2) в (3) та проінтегрувавши, одержимо вираз [3]

$$h_{21B1} = \frac{1}{2b} \left(b - \frac{\chi}{3\mu\tau} [a^3 - (a-b)^3] + \frac{\chi^2}{4\tau} BE [a^4 - (a-b)^4] \right). \quad (4)$$

Формула (4) враховує тільки зміну ефективної довжини бази магнітотранзистора. Як видно з рис. 1, дія ефекту перерозподілу носіїв між колекторами полягає в тому, що носії, інжектовані з частини емітера $0 - b_1$, під час відсутності магнітного поля попадають у колектор K2, а під дією магнітного поля відхиляються у бік першого колектора K1. Це еквівалентно тому, що в магнітному полі значення a і b з боку K1 збільшилися на величину b_1 . У малих магнітних полях

$$b_1 \approx W_0 \sin \varphi \approx h \mu B = a^2 \chi E \mu B.$$

У реальних схемах вихідним параметром є напруга між колекторами $U = R_H (I_{K1} - I_{K2}) = R_H I_E (h_{21B1} - h_{21B2})$. Враховуючи (4),

$$h_{21B1} - h_{21B2} = [\chi^2 BE / (4b\tau)] [a^4 - (a-b)^4]. \quad (5)$$

Таким чином, в слабких магнітних полях основний внесок у магніточутливість дає зміна ефективної довжини бази. З метою збільшення чутливості в структуру двоколекторного магнітотранзистора можна ввести додатковий базовий контакт. При проходженні струму основних носіїв через базові електроди Б2 і Б1 у базі виникає ЕРС Холла, яка спрямована так, що відхиляє інжектовані носії у одному напрямку з силою Лоренца. Отже, перерозподіл інжектованих носіїв між колекторами збільшується. Крім того, прикладена до контактів Б2 і Б1 напруга збільшує напруженість електричного поля в базі, що приводить до збільшення швидкості руху інжектованих носіїв, а відповідно і сили Лоренца.

Відповідно до принципу дії біполярного транзистора включення контакту Б2 еквівалентно шунтуванню емітерного $p-n$ переходу і зменшенню струму через нього. При однакових робочих струмах чутливість двоколекторних магнітотранзисторів на два-три порядки вище чутливості сенсорів Холла [5]. Збільшена чутливість і лінійність характеристики при малих магнітних полях дозволяють використовувати магнітотранзистори як сенсори слабких магнітних полів (відтворюючи магнітні голівки, електронні компаси тощо). При цьому в значних магнітних полях чутливість зменшується тому, що всі носії вже перерозподілені, і напруга збільшується виключно через зменшення ефективної довжини бази. В даному радіовимірювальному частотному перетворювачі магнітного поля на транзисторній структурі використовуємо двоколекторний магнітотранзистор в якості чутливого елемента.

РОЗРОБКА РАДІОВІМІРЮВАЛЬНОГО ПЕРЕТВОРЮВАЧА МАГНІТНОГО ПОЛЯ

Розроблено структурну схему радіовимірювального частотного перетворювача магнітного поля з частотним

виходом (рис. 2), в якій використовується двоколекторний магнітотранзистор та частотний перетворювач на основі двох біполярних та одного польового транзисторів.

В магніточутливому радіовимірювальному перетворювачі відбуваються три етапи перетворення енергії: енергія магнітного поля перетворюється в енергію зміни опору; енергія зміни опору перетворюється в енергію зміни еквівалентної ємності; енергія зміни еквівалентної ємності в енергію зміни частоти.

Електрична схема розробленого радіовимірювального частотного перетворювача для вимірювання напруженості магнітного поля наведена на рис. 3.

Як датчик магнітного поля використовується двоколекторний біполярний магнітотранзистор $VT1$. Частотний перетворювач утворено напівпровідниковими активними елементами: біполярні транзистори $VT4$ (BC847), $VT3$ (BC857), та польовий транзистор $VT2$ (BF998).

Основою пристрою є автогенератор, утворений транзисторною структурою $VT2$, $VT3$ та активною індуктивністю ($VT4$, $C1$, $R9$). Резистори $R5-R8$ забезпечують розміщення робочої точки транзисторів на спадній ділянці ВАХ транзисторної структури. Двоколекторний магнітотранзистор ввімкнено в схему перетворювача паралельно до транзисторної структури з від'ємним опором.

Оскільки, вольт-амперна характеристика на вихідних клеммах частотного перетворювача має спадну ділянку, то частотний перетворювач виступає генератором електричних коливань з певною частотою. При впливі зовнішнього магнітного поля на двоколекторний магнітотранзистор, змінюється напруга вихідного сигналу на його колекторів $K1$. При зміні напруги керування, яка є одночасно і напругою вихідного сигналу двоколекторного магнітотранзистора, змінюється частота генерації час-

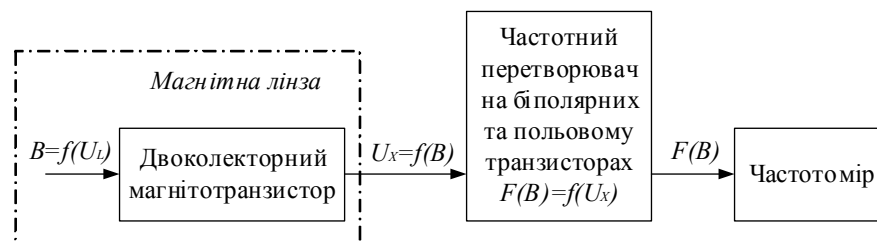


Рисунок 2 – Структурна схема радіовимірювального частотного перетворювача магнітного поля на транзисторній структурі

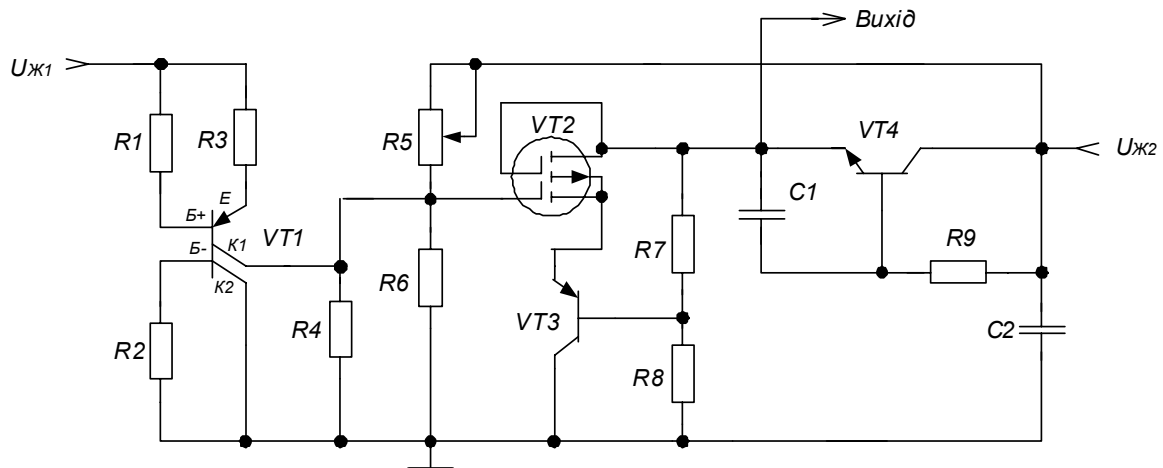


Рисунок 3 – Електрична схема радіовимірювального частотного перетворювача магнітного поля на транзисторній структурі

тотного перетворювача. Таким чином, чим вища напруженість магнітного поля, тим більша частота генерації вихідного коливання у радіовимірювального частотного перетворювача магнітного поля з частотним виходом.

ЕКСПЕРИМЕНТАЛЬНЕ ДОСЛІДЖЕННЯ РАДІОВИМІРЮВАЛЬНОГО МАГНІТНОГО ПЕРЕТВОРЮВАЧА

За результатами експериментальних досліджень було побудовано графіки залежності частоти генерації від напруженості магнітного поля (функції перетворення) рис. 4 при різних напругах керування.

Як показано на рис. 4, коли на магнітний сенсор діє лише природне магнітне поле Землі та відсутній вплив зовнішнього штучного магнітного поля, то при різних напругах керування частотний перетворювач має різні початкові частоти генерації. При напрузі керування $U_K = 1,6\text{ В}$ чутливість приладу максимальна, оскільки діапазон зміни частоти на виході досліджуваного пристрою становить від 900 кГц до 1065 кГц.

При подальшому збільшенні напруги керування діапазон зміни частоти на виході становить відповідно для $U_K = 1,8\text{ В}$ від 1050 кГц до 1198 кГц, та для $U_K = 2\text{ В}$ – від 1150 кГц до 1261 кГц. Таким чином, чутливість пристрою зменшується при збільшенні напруги керування.

Використовуючи експериментальні табличні дані та програму TableCurve2D, визначено аналітичний опис функції перетворення радіовимірювального перетворювача на основі біполярних, польового та магніточутливого двоколекторного транзисторів.

Найбільш оптимальною є функція перетворення

$$F(B) = a + \frac{b}{1 + \exp\left(\frac{-B + c}{d}\right)}, \quad (6)$$

де B – напруженість магнітного поля, а невідомі коефі-

цієнти мають наступні величини: $a = -548,14$; $b = 2477,5427$; $c = 1189,9973$; $d = 3405,9786$.

Використовуючи програму TableCurve2D, було встановлено, що відносна похибка функції перетворення становить $\pm 0,5\%$.

За результатами експериментальних досліджень побудовано графік залежності чутливості радіовимірювального

перетворювача $S(B) = \left| \frac{d(F(B))}{dB} \right|$ від напруженості магнітного поля при напрузі керування $U_K = 1,6\text{ В}$ (рис. 5).

Як показано на рис. 5, чутливість пристрою змінюється в залежності від рівня зовнішнього магнітного поля. Наприклад, при напруженості магнітного поля 100 мТл, чутливість пристрою становить 0,1765 кГц/мТл, а при рівні поля 1000 мТл – 0,1645 кГц/мТл. При порівнянні з попередніми результатами, досліджуваний радіовимірювальний перетворювач магнітного поля на біполярних та польовому транзисторах має більш лінійну залежність чутливості від магнітного поля ніж радіовимірювальний перетворювач, який був досліджений в [5].

ВИСНОВКИ

У статті досліджено виникнення інформативних сигналів у двоколекторному магнітотранзисторі при впливі на нього магнітного поля. Розроблено структурну та електричну схему радіовимірювального частотного перетворювача магнітного поля з частотним виходом.

При використанні розробленого радіовимірювального частотного перетворювача для вимірювання магнітного поля, ми в 4 рази підвищили точність вимірювання напруженості магнітного поля ніж у випадку застосування для вимірювання двоколекторного магнітотранзистора.

Також експериментально встановлено, що максимальна чутливість даного варіанта радіовимірювального перетворювача магнітного поля становить 0,1765 кГц/мТл. При

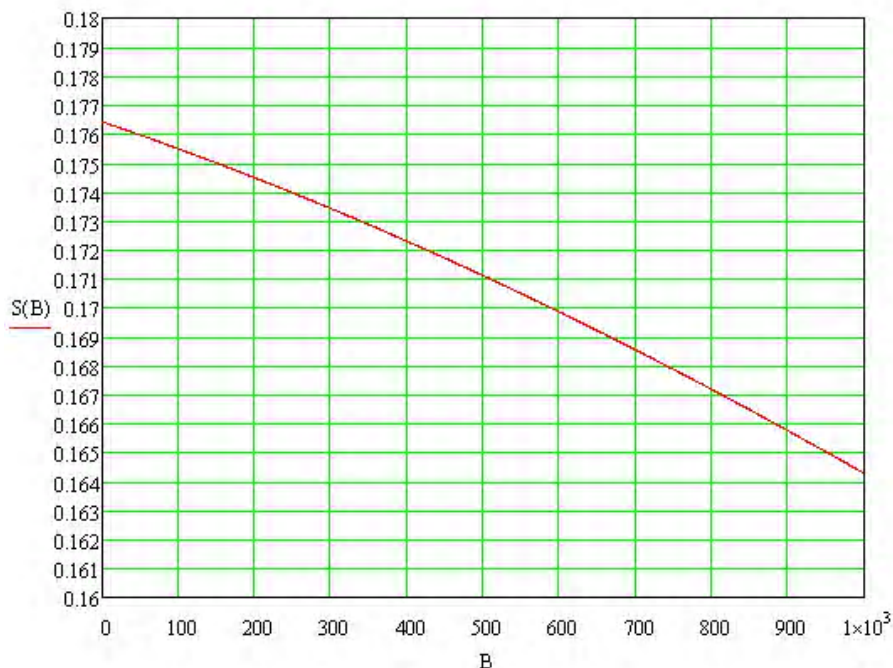


Рисунок 5 – Залежність чутливості радіовимірювального перетворювача від напруженості магнітного поля

цьому, відносна похибка перетворювача не перевищує $\pm 0,5\%$. При зміні напруженості магнітного поля в діапазоні 0 мТл ... 1000 мТл, досягнуто зміни частоти вихідного сигналу в діапазоні 900 кГц ... 1065 кГц. Перевагою розробленого перетворювача є більша лінійність характеристики чутливості, яка змінюється в межах від 0,1765 кГц/мТл до 0,1645 кГц/мТл.

СПИСОК ЛІТЕРАТУРИ

1. Готра З. Ю. Мікроелектронні сенсори фізичних величин : в 3 т. Т. 3. / З. Ю. Готра – Львів : Ліга-Прес, 2002. – 475 с.
2. Мікроелектронні сенсори магнітного поля з частотним виходом: монографія / [В. С. Осадчук, О. В. Осадчук]; під ред.

В. С. Осадчука, О. В. Осадчука. – Вінниця : ВНТУ, 2013. – 254 с.

3. Титце У. Полупроводниковая схемотехника. 12-ое издание. В 2-х томах / У. Титце, К. Шенк. – М. : ДМК Пресс, 2010. – 828 и 942 с.
4. Minin I. V. Microsensors / I. V. Minin, O. V. Minin. – India, Croatia : InTech, 2011. – 306 p.
5. Осадчук О. В. Радіовимірювальний перетворювач магнітного поля з магнітотранзистором та частотним вихідним сигналом / О. В. Осадчук, М. О. Притула, К. О. Коваль // Вісник Хмельницького національного університету. – 2015. – № 1. – С. 102–106.

Стаття надійшла до редакції 21.10.2015.
Після доробки 15.02.2016.

Осадчук А. В.¹, Коваль К. О.², Притула М. А.³

¹Д-р. техн. наук, професор, заведуючий кафедрою радіотехніки Вінницького національного технічного університету, Вінниця, Україна

²Канд. техн. наук, доцент, доцент кафедри радіотехніки Вінницького національного технічного університету, Вінниця, Україна

³Аспірант кафедри радіотехніки Вінницького національного технічного університету, Вінниця, Україна

РАДИОИЗМЕРИТЕЛЬНЫЙ ПРЕОБРАЗОВАТЕЛЬ МАГНИТНОГО ПОЛЯ НА ТРАНЗИСТОРНОЙ СТРУКТУРЕ

Решена задача измерения напряженности магнитного поля радиоизмерительным преобразователем магнитного поля на биполярных и полевому транзисторах. Предложено схемотехническое решение радиоизмерительного преобразователя, который обеспечивает преобразование напряженности магнитного поля в частоту электрического колебания, что позволяет повысить чувствительность устройства к напряженности магнитного поля. Проведен эксперимент по исследованию разработанного радиоизмерительного преобразователя магнитного поля. В результате проведенных экспериментов исследовано зависимости выходной частоты и чувствительности устройства от напряженности магнитного поля.

Ключевые слова: радиоизмерительный преобразователь, магнитотранзистор, напряженность магнитного поля, чувствительность.

Osadchuk O. V.¹, Koval K. O.², Prytula M. O.³

¹Dr.Sc., Professor, Head of the department of radio engineering, Vinnytsia National Technical University, Vinnytsia, Ukraine

²PhD, Associate professor, Associate professor of department of radio engineering, Vinnytsia National Technical University, Vinnytsia, Ukraine

³Graduate student of department of radio engineering, Vinnytsia National Technical University, Vinnytsia, Ukraine

THE RADIO MEASURING CONVERTER OF THE MAGNETIC FIELD BASED ON TRANSISTORS STRUCTURE

The problem of measuring the magnetic field strength of radio transmitter magnetic field on bipolar and field-effect transistors is solved. The circuit solution of radio measuring transducer is offered. It converts the magnetic field in the electrical oscillations frequency, that improves the sensitivity of the device to the intensity of the magnetic field. An experiment on proposed radio measuring transducer of the magnetic field is conducted. As a result of the experiment, we had the output frequency depending on the sensitivity of the device and the strength of the magnetic field.

Keywords: radiomeasuring converter, magnetic transistor, magnetic field strength, the sensitivity.

REFERENCES

1. Gotra Z. U. Microelectronni sensory fizychnykh velychyn: v 3 t. T. 3. Lviv, Liga-Pres, 2002, 475 p.
2. Osadchuk V. S., Osadchuk O. V.; pid red. V. S. Osadchuk, O. V. Osadchuk. Microelectronni sensory magnitnogo polya z chastotnym vihodom: monographiya. Vinnytsia, VNTU, 2013, 254 p.

3. Tyttse U., Shenk K. Poluprovodnikova shemotichnika. 12-e izdanie. V 2 tomah. Moscow, DMK Press, 2010, 828 and 942 p.
4. Minin Igor V., Oleg V. Minin Microsensors. India, Croatia, InTech, 2011, 306 p.
5. Osadchuk O. V., Prytula M. O., Koval K. O. The radiomeasuring converter of the magnetic field with magnetic transistor and frequency output signal, *Herald of Khmelnytskyi national university*, 2015, No. 1, pp. 102–106.

МАТЕМАТИЧНЕ ТА КОМП'ЮТЕРНЕ МОДЕЛЮВАННЯ

МАТЕМАТИЧЕСКОЕ И КОМПЬЮТЕРНОЕ МОДЕЛИРОВАНИЕ

MATHEMATICAL AND COMPUTER MODELLING

УДК 004.942

Прохоров А. В.¹, Кузнецова Ю. А.²¹Канд. техн. наук, доцент, доцент кафедры информационных управляющих систем Национального аэрокосмического университета им. Н. Е. Жуковского «ХАИ», Харьков, Украина²Канд. техн. наук, ст. преподаватель кафедры инженерии программного обеспечения Национального аэрокосмического университета им. Н. Е. Жуковского «ХАИ», Харьков, Украина

АГЕНТНОЕ МОДЕЛИРОВАНИЕ МУЛЬТИКЛАСТЕРНЫХ ТЕХНОЛОГИЧЕСКИХ КОМПЛЕКСОВ НАНОЭЛЕКТРОННЫХ ПРОИЗВОДСТВ

Проведен анализ особенностей автоматизированных производственных процессов и кластерной компоновки технологических комплексов нанoeлектронных производств, предложен подход к повышению производительности автоматизированного кластерного оборудования в полупроводниковом производстве на основе управления потоками полуфабрикатов. Предложена имитационная модель анализа процессов управления потоками полуфабрикатов и расчета характеристик нанoeлектронных производств, которая позволяет производить поиск рациональных решений, связанных с управлением потоков полуфабрикатов в мультикластерных технологических комплексах любой архитектуры с учетом ресурсных ограничений и факторов риска. Модель основывается на мультиагентном подходе, что дает следующие преимущества: автономность и индивидуальное поведение элементов модели (агентов), агенты имеют возможность адаптироваться и менять свое поведение, иметь динамические связи с другими агентами, которые могут формироваться и исчезать в процессе моделирования и др. Разработанная агентная имитационная модель учитывает сложную динамику нанoeлектронного производства и обеспечивает моделирование мультикластерных технологических комплексов различной топологии и реализует различные стратегии и алгоритмы управления потоками полуфабрикатов. Сформированы постановки задач планирования перемещений полуфабрикатов в мультикластерных технологических комплексах, связанные с различными стратегиями и критериями эффективности для логистики нанoeлектронного производства.

Ключевые слова: нанoeлектронное производство, мультикластерный технологический комплекс, кластер, пластина, поток полуфабрикатов, агентная модель.

НОМЕНКЛАТУРА

МТК – мультикластерный технологический комплекс;

КМДП – комплементарная структура «металл-диэлектрик-полупроводник»;

 S_j^T – стоимость перемещений транспортного модуля; c_{ij} – стоимость потребляемой электроэнергии на один оборот/перемещение транспортного модуля; f_{ij} – стоимость усредненного простоя транспортно-го модуля; r_{ij} – усредненная стоимость ремонта транспортного модуля, отнесенная на один оборот/перемещение; T_{o_j} – директивный срок прибытия пластины в заданный процессный модуль, который соответствует плано-

вому сроку начала работы основного технологического оборудования кластера;

 T_{ϕ_j} – фактический (рассчитываемый) срок прибытия пластины, полученный путем имитационного моделирования; M – количество полуфабрикатов, участвующих в технологическом процессе на планируемый период работы МТК; S_{ϕ} – фактическая стоимость перемещений; S_3 – задаваемая (допустимая) стоимость перемещений; T_{o_j} – допустимое запаздывание пластины, которое не приводит к срыву план-графика работы основного технологического оборудования МТК; α, β – значимость критериев стоимости и времени перемещений соответственно.

ВВЕДЕНИЕ

Современные полупроводниковые приборы и интегральные микросхемы представляют собой чрезвычайно сложные устройства, отдельные компоненты которых имеют размеры не более доли микрометра. Изготовление таких устройств осуществляется на монокристаллических пластинах из полупроводникового материала, которые. Таким образом, полупроводниковая пластина – полуфабрикат в технологическом процессе производства полупроводниковых приборов и микросхем. Производство полупроводниковых приборов осуществляется по многослойной технологии с одновременным изготовлением большого количества однотипных элементов. Производственный процесс состоит из нескольких этапов, число которых может варьироваться. Он включает комплекс подготовительных процессов, процессов групповой и индивидуальной обработки. Формирование самой структуры прибора происходит при групповой обработке, которая состоит из процессов окисления, диффузии примесей, эпитаксии, вакуумного напыления, фотолитографии и химической обработки.

Технологии нанoeлектронного производства сегодня используют кластерное оборудование, в котором реализуется замкнутый технологический цикл без выгрузки изделий в атмосферную среду рабочего помещения, что позволяет проводить высокоточные многоступенчатые технологические обработки при обеспечении автоматического контроля техпроцессов, межоперационных испытаний и оперативного структурного анализа. Кластерное оборудование состоит из центрального модуля перемещения пластин (транспортной камеры), на которой смонтированы универсальные камеры загрузки-выгрузки пластин, набор процессных модулей, реализующих процессы ионно-плазменной обработки на пластинах, (последовательно в каждом из них, без развакуумирования и, следовательно, исключая контакт изделий с атмосферой), набор аналитических модулей, выполняющих метрологический контроль поточного производства полупроводниковой продукции, а также переходных модулей или шлюзов в случае, когда имеет место мультикластерный технологический комплекс (МТК). Каждый кластер электрически и механически независим и имеет отдельные вакуумные и управляющие системы. Однако кластеры объединяются в единую технологическую систему через модули передачи, переворота и складирования пластин. Важным является то, что оборудование должно быть согласовано друг с другом по производительности.

Таким образом, МТК характеризуются: сложной структурой объектов и процессов многоцелевого функционирования; многоуровневостью; множеством структурно-компоновочных решений; многообразием технологических процессов и инвариантностью технологических маршрутов обработки; вероятностным характером функционирования и поведения, связанным со сложным взаимодействием с изменяющейся внешней средой.

Многовариантность структур МТК приводит к проблеме выбора рациональной структуры путем проведения подробного анализа динамики функционирования

каждого варианта с целью оценки его гибкости и эффективности. Даже незначительные изменения в конфигурации МТК (например, времени обработки или стратегий планирования) могут иметь значительное влияние на его производительность. В этой ситуации наибольшие трудности вызывают: отсутствие полного детерминизма в ходе производства и сложность технологических процессов.

1 ПОСТАНОВКА ЗАДАЧИ

Для создания такого сложного единого комплекса технологического оборудования требуется решение задач, связанных с эффективным взаимодействием технологических модулей между собой с целью обеспечения необходимых качественных характеристик технологического процесса, что в свою очередь обуславливает необходимость разработки методов планирования и оптимизации структурно-компоновочных решений и графика запуска полуфабрикатов. Поэтому актуальным является разработка математических методов и моделей, которые ориентированы на анализ новых схем организации и управления автоматизированным производством в нанoeлектронике. Целью данного исследования является повышение производительности автоматизированного кластерного оборудования в полупроводниковом производстве на основе управления потоками полуфабрикатов.

2 ОБЗОР ЛИТЕРАТУРЫ

Для решения системных задач планирования и управления сложными производственными системами [1] разработаны различные подходы, методы и модели. Среди аналитических методов моделирования МТК можно выделить: методы сетевого планирования и управления; методы линейного и динамического программирования; модели управления запасами; методы теории Марковских процессов; сети Петри и др. [2–4].

Совместное использование событийных графов и сетевых моделей для снижения размерности поиска при решении задач планирования в многокластерных технологических комплексах предложено в [5]. Авторы проводят анализ влияния процессов изменения потока пластин на общее время цикла, а также эффективность работы роботов.

В работе [6] продемонстрированы возможности, которые дает подход на основе имитации отжига для оптимизационного моделирования групп кластерного оборудования в полупроводниковом производстве, при этом авторы делают акцент на том, что требуется дальнейшее расширение алгоритмов и решение задач автоматизации их параметрической настройки для получения необходимых свойств адаптивности к динамике производственного процесса.

Генетический алгоритм для оптимизации времени производственного цикла при формировании последовательностей обработки пластин на кластерном оборудовании представлен в работе [7]. Развитие предложенного метода автор видит в двухэтапном применении генетического алгоритма: на первом – пластины распределяются между имеющимися кластерами, а на втором – формируется оптимальная последовательность их обработки.

В работе [8] разработана модель на основе цветных сетей Петри для анализа и повышения производительности кластерного оборудования для сухого травления. Результаты исследований позволили определить, что в условиях частой смены рецептов процессные камеры и шлюзы становятся узкими местами, а время производственного цикла может быть уменьшено за счет оптимизации рецептов.

Неотъемлемым элементом анализа сложных производственных систем, который позволяет наиболее гибко, полно и наглядно отражать динамические процессы, являются имитационные модели. На сегодняшний день сформировались и наиболее широко применяются три основных подхода: дискретно-событийное моделирование, модели системной динамики и агентное моделирование.

Работа [9] посвящена описанию имитационной модели автоматизированной поточной линии для производства интегральных микросхем. Рассмотрена возможность применения этой модели для составления расписания работы линии и задачи ее синтеза.

В работе [10] представлена простая имитационная модель, которая предназначена для прогнозирования времени производственного цикла в полупроводниковом производстве. В модели использованы две стратегии в обработке пластин – FIFO и на основе приоритетности партий, однако не рассматриваются вопросы моделирования сложных маршрутных процессов с параллельной обработкой пластин и балансировки загрузки технологического оборудования.

Дискретно-событийная модель технологического процесса обработки полупроводниковых пластин на кластерном оборудовании, разработанная в системе Anylogic, рассматривается в работе [11], однако в предложенной авторами реализации отсутствует учет стратегий планирования и диспетчеризации, а также не рассматриваются вопросы совместного моделирования технологических и контрольных операций.

Проведенный анализ показал, что существующие методы не в полной мере описывают различные компоновочные решения МТК, не учитывают сложную динамику технологических процессов и алгоритмов управления потоком полуфабрикатов.

3 МАТЕРИАЛЫ И МЕТОДЫ

Для построения агентной имитационной модели анализа процессов управления потоками полуфабрикатов

была выбрана среда моделирования AnyLogic, обладающая рядом преимуществ и поддерживающая различные парадигмы имитационного моделирования систем. Имитационная модель автоматизированного полупроводникового производства (рис. 1) состоит из компонент, каждый из которых отражает отдельный элемент динамики поведения МТК. В разработанной имитационной модели с помощью класса активных объектов (агентов) реализованы входной буфер, выходной буфер, кластер, линейный транспорт, что позволяет использовать в модели любое число экземпляров этих классов по числу элементов, участвующих в МТК. Кроме того в модели созданы простые классы: модуль, пластина, транспорт, робот, операция.

Агент входного буфера имеет параметры, которые отвечают за накопление информации о количестве пластин каждого типа. Модель, описывающая поведение этого агента, включает объекты, которые используются для формирования потоков пластин каждого типа. Агент выходного буфера имеет параметры, которые отвечают за накопление информации о количестве обработанных пластин каждого типа и числа необработанных пластин.

Агент транспортного модуля имеет следующие параметры: длина, скорость перемещения, наличие пластины, место назначения и другие. Модель линейного транспортного модуля реализует алгоритмы управления потоками полуфабрикатов, связанные с поиском подходящих и свободных процессных модулей, анализом загрузки процессных модулей кластеров, обеспечивая перемещение пластин между имеющимися в МТК кластерами.

Агент кластера (рис. 2) служит для моделирования кластерного оборудования радиальной топологии и имеет следующие параметры: номер кластера, число, наличие и тип рабочих модулей, число лап робота, его скорость, наличие пластины на входе и выходе кластера, состояние модулей и робота и другие. Кроме того здесь присутствуют элементы сбора статистики по времени ожидания и времени пребывания пластин в кластере.

Имитационная модель агента кластера отражает все основные технологические операции: занятие модуля пластиной; освобождение модуля; перемещение транспортного модуля к процессному модулю для последующего переноса пластины (холостой пробег); перенос пластины транспортным модулем из одного процессного модуля в другой модуль и др.

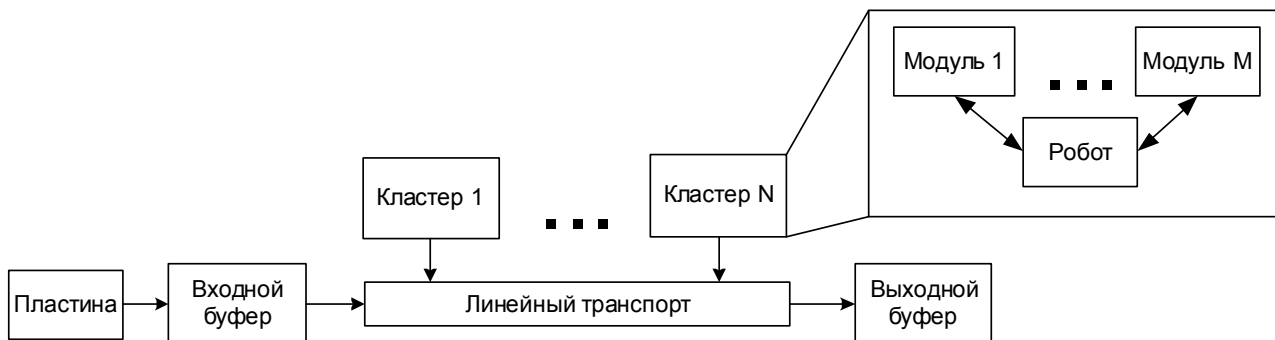


Рисунок 1 – Структура имитационной модели анализа процессов управления потоками полуфабрикатов в МТК

В модели реализуются алгоритмы поиска свободного оборудования (транспортных и процессных модулей) для выполнения технологических операций, в том числе анализируется возможность использования робота-манипулятора (двухлапного) в качестве буферного. Для этого в модель включены блоки принятия решения, которые в свою очередь основываются на целом ряде функций входящих в состав класса активного объекта агента кластера.

Необходимо отметить, что модель для задания различных стратегий управления потоками полуфабрикатов, обеспечивающих выполнение поставленных задач, может использовать специальный механизм, который предполагает организацию поиска среди множества альтернатив (модулей, очередей, пластин и др.). Так, например, альтернативными модулями являются такие, которые имеют условия, обеспечивающие возможность приема на обработку поступающих пластин данного типа. В качестве сценариев поиска могут быть заданы различные критерии: наименьшей стоимости, наименьшего коэффициента использования, наименьшего или наибольшего приоритета, кратчайшей очереди и др.

Основными стратегиями запуска пластин в производство являются следующие: больший приоритет имеет пластина, номер которой в партии меньше (FIFO – First In First Out); больший приоритет имеет пластина, номер которой в партии больше (LIFO – Last In First Out); больший приоритет имеет пластина, желаемый срок изготовления которой меньше (SD – Smaller Date); больший приоритет имеет пластина, время изготовления которой меньше (SPT – Short Processing Time); больший приоритет имеет пластина, количество партий запуска которой больше (LPQ – Large Parts Quantity); приоритеты выставляются в случайном порядке (R – Random). Решение о дальнейшем перемещении пластины может приниматься на основе вероятностных распределений (R – Random), значений параметров пластины, к первому свободному модулю (FF – First Free), к модулю с самой короткой очередью (SQ – Short Queue), последовательно (C – Cycle).

В качестве критериев эффективности работы транспортной системы обслуживания, используются два наиболее часто используемых в практике критерия: время доставки и условной стоимости перемещений. Критерий стоимости перемещений с учетом j -го, используемого в i -м маршруте транспортного модуля можно представить в виде:

$$S_j^T = \sum_i (c_{ij} + f_{ij} + r_{ij}). \quad (1)$$

Возможны следующие модели транспортного обслуживания МТК:

1. Обеспечить своевременное, без задержек, транспортное обслуживание МТК, с учетом ограничений на стоимость перемещений:

$$\Delta T_j = T_{\partial_j} - T_{\phi_j} \geq 0, \quad j = \overline{1, M}, \quad (2)$$

$$S_{\phi} \leq S_3. \quad (3)$$

2. Минимизировать транспортные расходы в планируемый период работы МТК с учетом ограничений на сроки доставки пластин:

$$\min S_{\phi}, \quad S_{\phi} \leq S_3, \quad (4)$$

$$T_{\phi_j} - T_{\partial_j} \leq T_{o_j}, \quad j = \overline{1, M}. \quad (5)$$

3. Обеспечить решение компромиссной задачи, минимизирующей стоимость перемещений и время доставки пластин с учетом ограничений на стоимость и время транспортного обслуживания:

$$\min(T_{\phi_j}, S_{\phi_j}), \quad (6)$$

$$T_{\phi_j} - T_{\partial_j} \leq T_{o_j}, \quad (7)$$

$$S_{\phi} \leq S_3. \quad (8)$$

Для решения задачи 3 использован следующий компромиссный критерий в виде аддитивной свертки:

$$K_j = \alpha T_j + \beta S_j, \quad 0 \leq \alpha \leq 1, \quad 0 \leq \beta \leq 1, \quad \alpha + \beta = 1, \quad (9)$$

$$T_j = \frac{T_j}{T_{\partial_j}}; \quad S_j = \frac{S_j}{S_{3j}}. \quad (10)$$

Значения α и β задаются с учетом предпочтений исследователя.

Благодаря многовариантному расчету на имитационной модели транспортного обслуживания МТК выбираются рациональные маршруты движения полуфабрикатов в зависимости от постановок транспортных задач.

Итоговые результаты по модели в целом: общее число поступивших в систему пластин; число обслуженных пластин; среднее время нахождения пластины в системе; максимальное время нахождения пластины в системе; диаграмма Ганта (с указанием ранних и поздних сроков начала и окончания операций).

Результат по работе каждого модуля: число обслуженных пластин, среднее время обслуживания одной пластины, коэффициент загрузки устройства, дефицит ресурса времени модуля. Положительное значение дефицита ресурса говорит о несбалансированном характере внешней нагрузки и пропускной способности устройства обработки. В этом случае запускается сценарий перераспределения нагрузки между аналогичными модулями.

Результат по каждой из очередей: максимальная длина очереди, средняя длина очереди, длина очереди в конце моделирования, среднее время нахождения пластины в очереди, максимальное время ожидания пластины в очереди.

Испытательный стенд, где производится настройка параметров моделирования, представлен на рис. 3.

В качестве входных данных модели предусматриваются: число радиальных кластеров, число модулей у кластера (остальные считаются не задействованными), тип каждого процессного модуля, число лап у роботов, чис-

которой цветом отображается выполнение операций, периоды ожидания, при отсутствии условий выполнения операций, простои и периоды блокирования работы как следствие имитации отказов оборудования; гистограммы занятости процессных модулей кластера и центрального транспортного робота; гистограмма распределения времени обработки пластин в кластере.

На рис. 5 представлен эксперимент при моделировании с одним типом пластин на двух кластерах.

На рис. 6 представлен эксперимент при моделировании с одним типом пластин на двух кластерах.

В ходе выполнения транспортных перемещений полуфабрикатов могут возникнуть непредвиденные обстоятельства, которые приводят к появлению рисков. Это, например, отсутствие требуемых материалов на модуле, закрытие участков из-за сбоев (робот уронил пластину), поломка транспортных модулей и т. д.

В модели предусмотрена имитация отказов оборудования (транспортных и процессных модулей), которая производилась путем задания вероятностного закона распределения интенсивности и времени восстановления. Имитация производилась блокированием работы модуля (рис. 7).

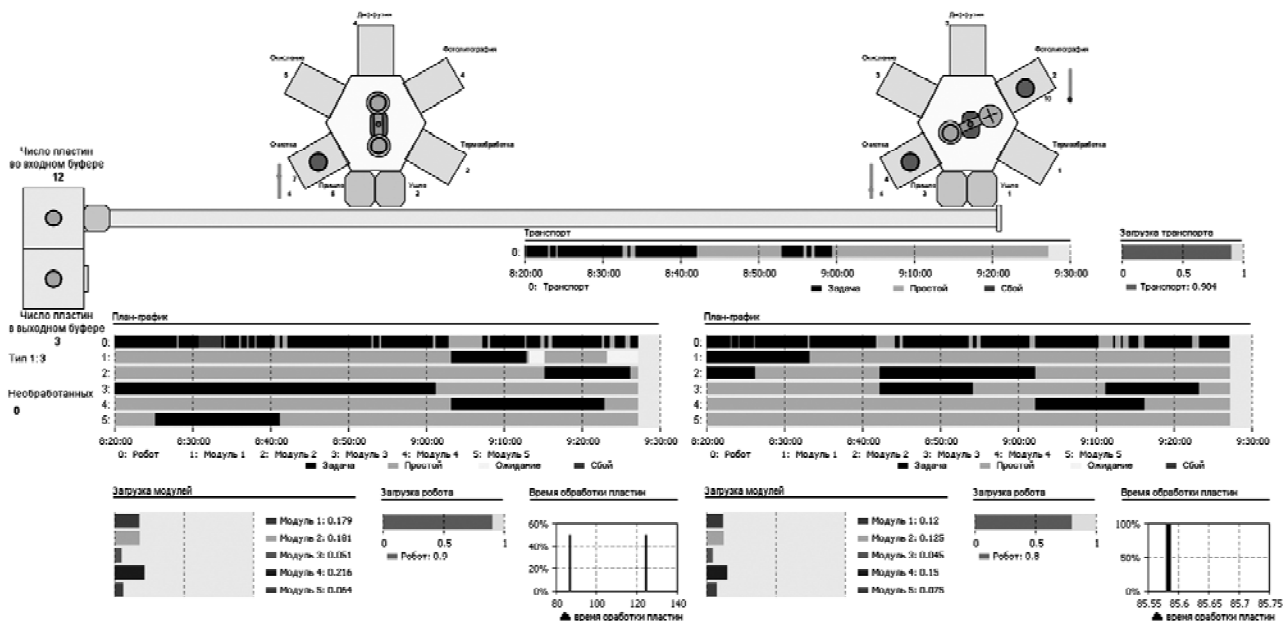


Рисунок 5 – Моделирование при одном типе пластин на двух кластерах

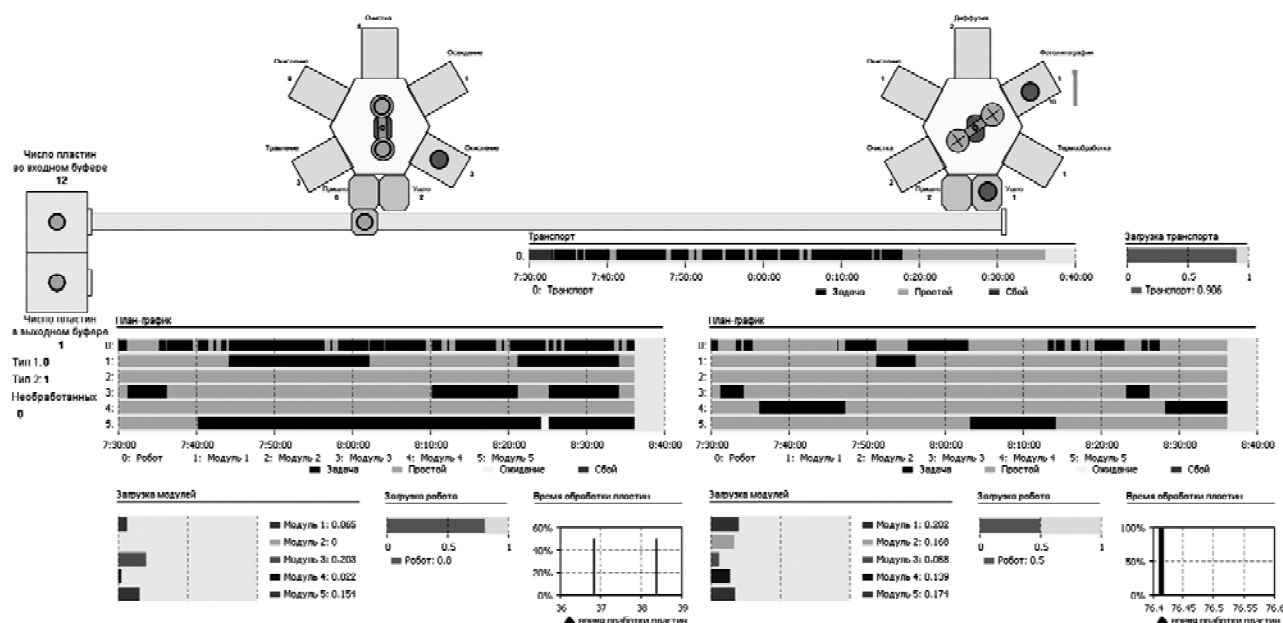


Рисунок 6 – Моделирование при двух типах пластин

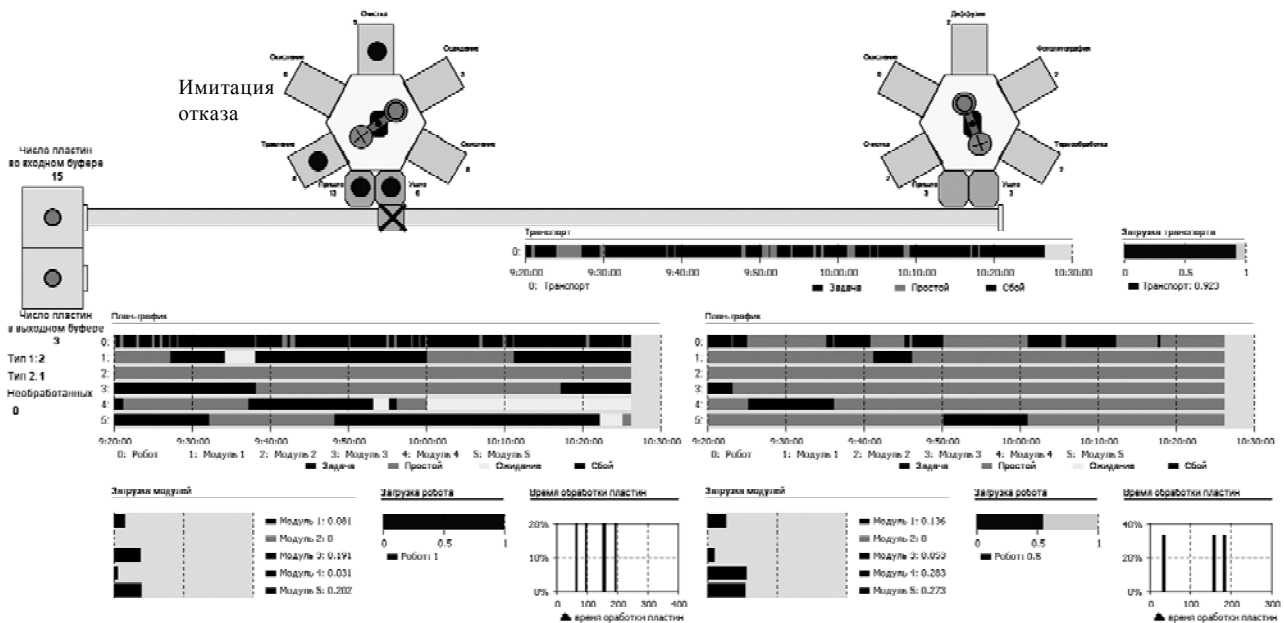


Рисунок 7 – Моделирование с отказами

5 РЕЗУЛЬТАТЫ

Таким образом, в процессе моделирования определяются:

- наиболее загруженные элементы производственной системы (узкие места – высокий процент загрузки, возникновение очередей);
- наиболее надежные элементы производственной системы;
- производительность системы (количество удовлетворенных заявок на производство продукции за заданный период времени);
- номенклатура и количество операций выполняемых на каждом модуле МТК и выявление несоответствий по их оснащенности и обеспеченности ресурсами;
- влияние транспортных операций на показатели работы системы полупроводникового производства;
- влияние «узких мест» на длительности технологических циклов.

Эта информация является исходной для решения задач автоматизированного управления мультикластерным технологическим комплексом нанoeлектронного производства.

6 ОБСУЖДЕНИЕ

Моделирование с одним кластером показало: большое среднее время пребывания пластин в системе; увеличение интенсивности поступления пластин всего лишь на 20 %, приводит к ухудшению значений характеристик эффективности МТК (среднее время пребывания пластин в МТК выросло в 1,5 раза, а число пластин, находящихся в очередях – почти в 2,5 раза (возникли простои в ожидании пластин в модулях, вследствие занятости транспортного робота). Добавление еще одного кластера в систему привело к значительному снижению времени пребывания пластин в системе, но при этом несколько возросла нагрузка на линейный транспорт. Был применен еще один способ разгрузки МТК, заключающийся в уменьшении вероятности передачи пластин к кластеру, являющемуся узким местом.

Добавление в обработку второго типа пластин потребовало включения в систему второго кластера. В этом случае увеличилась загрузка транспортных модулей системы, а с учетом того, что часть процессных модулей для выполнения технологических операций пластин обоих типов совпадает, то и загрузка процессных модулей. Также увеличилось число простоев из-за невозможности роботов принять уже обработанные пластины из процессных модулей (в случае, когда обе лапы робота заняты). В систему были заложены алгоритмы, которые были направлены на поиск свободного процессного модуля в имеющихся кластерах. Результаты моделирования показали, что удалось снизить время простоев, уменьшить время нахождения пластин в системе, однако, с сохранением высокой загруженности линейного транспорта.

Для улучшения характеристик обслуживания пластин в МТК необходимо разгрузить узкое место, которым является линейный транспорт. Для этого была увеличена скорость работы линейного транспорта в 1,5 раза, что согласуется с техническими возможностями оборудования. Анализ полученных результатов показал, что разгрузка узкого места позволила существенно уменьшить значения характеристик МТК: среднее время пребывания пластин в системе уменьшилось более чем в 2 раза, а число заявок, находящихся в очередях – почти в 3 раза.

Анализ пропускной способности основного технологического алгоритма показал, что наиболее существенное влияние на производительность МТК оказывает операция окисления, под которую выделены два процессных модуля. В соответствии с технологическим процессом, высокотемпературное окисление необходимо проводить на 5–10% пластин, обрабатываемых на МТК. Поэтому в модели было принято ограничение, что по этим техпроцессам возможно запускать пластины не чаще, чем каждую третью. Для сравнения была выполнена разгрузка этого узкого места путем добавления еще одного процессного модуля. За счет этого среднее вре-

мя ожидания пластин в МТК уменьшилось в 1,5 раза. В то же время, среднее время пребывания пластин в МТК увеличились более чем на 10%, что обусловлено большей длительностью обслуживания пластин.

При моделировании с отказами, как и следовало ожидать, результаты показали, что критичным по надежности элементом производственной системы является линейный транспорт.

В целом, реализация рассмотренных ранее алгоритмов управления потоком полуфабрикатов позволила снизить время простоев, уменьшить время нахождения пластин в системе, получить рациональные маршруты движения полуфабрикатов, а также сформировать рекомендации о структурно-компоновочных вариантах построения МТК.

ВЫВОДЫ

Таким образом, усовершенствована имитационная модель анализа процессов управления потоками полуфабрикатов и расчета характеристик нанoeлектронных производств, отличающаяся составом типовых блоков, необходимых и достаточных для построения имитационной модели мультикластерного технологического комплекса любой архитектуры, позволяющая производить поиск рациональных решений, связанных с управлением потоков полуфабрикатов с учетом ресурсных ограничений и факторов риска. Агентная имитационная модель обеспечивает моделирование МТК различной топологии и реализует разработанные стратегии и алгоритмы управления потоками полуфабрикатов.

Полученные результаты позволяют говорить о достижении поставленной цели исследования – повышение производительности автоматизированного кластерного оборудования в полупроводниковом производстве на основе управления потоками полуфабрикатов.

Практическая значимость исследования заключается в том, что разработанная модель позволяет производить поиск рациональных решений, связанных с управлением потоков полуфабрикатов с учетом ресурсных ограничений и отказов и может быть использована при проектировании и эксплуатации современных кластерных многофункциональных, высокопроизводительных установок, обладающих повышенной надежностью функционирования.

БЛАГОДАРНОСТИ

Данная работа выполнялась в рамках госбюджетных научно-исследовательских работ Министерства образования и науки Украины «Разработка методологических принципов, методов и математических моделей логистического анализа и управления сложными технологическими комплексами» и «Разработка методологических основ, системных моделей и интеллектуальных информационных технологий логистического управления распределенными производствами и проектами по их созданию».

СПИСОК ЛИТЕРАТУРЫ

1. Шишмарев В. Ю. Организация и планирование автоматизированных производств : учебник / В. Ю. Шишмарев. – М. : Академия, 2013. – 304 с.
2. Mohammad I. Handbook of sensor networks : compact wireless and wired sensing systems / edited by Mohammad Ilyas Mohammad and Imad Mahgoub. – Boca Raton, Florida : CRC Press LLC, 2005. – 776 p.
3. Lopez M. J. Systems of multiple cluster tools: configuration, reliability, and performance / M. J. Lopez, S. C. Wood // IEEE Trans. on Semiconductor Manufacturing. – 2003. – Vol. 16, № 2. – P. 170–178.
4. Акопов, А. С. Имитационное моделирование : учебник и практикум для академического бакалавриата / А. С. Акопов ; Высшая школа экономики (ВШЭ), Национальный исследовательский университет (НИУ). – Москва : Юрайт, 2015. – 389 с.
5. Challenges and Solution Approaches for the Online Simulation of Semiconductor Wafer Fabs / [D. Noack, M. Mosinski, O. Rose, P. Lendermann, and B. P. Gan] // Winter Simulation Conference : The premier international conference for simulation professionals in discrete and combined discrete-continuous simulation, AZ, USA, 11–14 December 2011 : proceedings. – Phoenix : WSC, 2011. – P. 1845–1856.
6. Uhlig T. Simulation-based optimization for groups of cluster tools in semiconductor manufacturing using simulated annealing / T. Uhlig, O. Rose // Winter Simulation Conference : The premier international conference for simulation professionals in discrete and combined discrete-continuous simulation, AZ, USA, 11–14 December 2011 : proceedings. – Phoenix : WSC, 2011. – P. 1857–1868.
7. Towards Realization of a High-Fidelity Simulation Model for Short-Term Horizon Forecasting in Wafer Fabrication Facilities / [W. Scholl, B. P. Gan, M. L. Peh and other] // Winter Simulation Conference : The premier international forum for disseminating recent advances in the field of system simulation, MD, USA, 05–08 December 2010 : proceedings. – Baltimore : WSC, 2010. – P. 2563–2574.
8. Zuberek W. M. Cluster tools with chamber revisiting – Modeling and analysis using timed Petri nets / W. M. Zuberek // IEEE Transactions on semiconductor manufacturing. – 2004. – Vol. 17, № 3. – P. 333–344.
9. Throughput Analysis of Linear Cluster Tools / [J. Yi, S. Ding, M. T. Zhang, V. D. Meulen] // IEEE International Conference on Automation Science and Engineering: International Conference, AZ, USA, 22–25 September 2007 : proceedings. – Scottsdale : IEEE, 2007. – P. 1063–1068.
10. Varadarajan A. A survey of dispatching rules for operational control in wafer Fabrication / A. Varadarajan, S. C. Sarin // Information Control Problems in Manufacturing: The 12th IFAC International Symposium, France, 17–19 May 2006 : proceedings. – Saint Etienne : ELSEVIER, 2006. – P. 709–720.
11. Wang P. Hybrid Model applied in the Semiconductor Production Planning / P. Wang, T. Murata // International MultiConference of Engineers and Computer Scientists: International MultiConference, China, 13–15 March 2013 : proceedings. – Hong Kong : Newswood Limited, 2013. – P. 964–968.

Статья поступила в редакцию 30.11.2015.

После доработки 14.12.2015.

Прохоров А. В.¹, Кузнецова Ю. А.²

¹Канд. техн. наук, доцент, доцент кафедри інформаційних управляючих систем Національного аерокосмічного університету ім. М. С. Жуковського «ХАІ», Харків, Україна

²Канд. техн. наук, ст. викладач кафедри інженерії програмного забезпечення Національного аерокосмічного університету ім. М. С. Жуковського «ХАІ», Харків, Україна

АГЕНТНЕ МОДЕЛЮВАННЯ МУЛЬТИКЛАСТЕРНИХ ТЕХНОЛОГІЧНИХ КОМПЛЕКСІВ НАНОЕЛЕКТРОННИХ ВИРОБНИЦТВ

Проведено аналіз особливостей автоматизованих виробничих процесів і кластерної компоновки технологічних комплексів напівпровідникових виробництв, запропоновано підхід до підвищення продуктивності автоматизованого кластерного обладнання в напівпровідниковому виробництві на основі управління потоками напівфабрикатів. Запропоновано імітаційну модель аналізу процесів управління потоками напівфабрикатів і розрахунку характеристик напівпровідникових виробництв, яка дозволяє проводити пошук раціональних рішень, пов'язаних з управлінням потоків напівфабрикатів у мультикластерних технологічних комплексах будь-якої архітектури з урахуванням ресурсних обмежень і факторів ризику. Модель ґрунтується на мультиагентному підході, що дає наступні переваги: автономність й індивідуальну поведінку елементів моделі (агентів), агенти мають можливість адаптуватися й змінювати свою поведінку, мати динамічні зв'язки з іншими агентами, які можуть формуватися і зникати у процесі моделювання та ін. Розроблена агентна імітаційна модель враховує складну динаміку напівпровідникового виробництва і забезпечує моделювання мультикластерних технологічних комплексів різної топології й реалізує різні стратегії та алгоритми управління потоками напівфабрикатів. Сформовано постановку завдань планування переміщень напівфабрикатів у мультикластерних технологічних комплексах, що пов'язані з різними стратегіями і критеріями ефективності для логістики напівпровідникових виробництв.

Ключові слова: напівпровідникове виробництво, мультикластерний технологічний комплекс, кластер, пластина, потік напівфабрикатів, агентна модель.

Prokhorov A. V.¹, Kuznetsova Yu. A.²

¹PhD of Engineering Science, Associate Professor, Associate Professor, Department of Information Control Systems of the National Aerospace University named after N. E. Zhukovsky «KhAI», Kharkiv, Ukraine

²PhD of Engineering Science, Senior Lecturer, Software Engineering Department of the National Aerospace University named after N. E. Zhukovsky «KhAI», Kharkiv, Ukraine

AGENT-BASED MODELING OF MULTICLUSTER TECHNOLOGICAL COMPLEXES OF NANOELECTRONIC MANUFACTURING

The analysis of the characteristics of automated production processes and the cluster layout of technological complexes of nanoelectronic manufacturing is carried out, an approach to improve the performance of automated cluster-based equipment in semiconductor manufacturing based on control flow of semi-finished products is proposed. A simulation model for the analysis of control flow processes of semi-finished products and calculation of nanoelectronic manufacturing characteristics, which allows you to search for rational decisions related to the control flows of semi-finished products in multicluster technological complexes of any architecture, taking into account resource constraints and risk factors also proposed in this article.

The model is based on multi-agent approach, which provides the following benefits: autonomy and individual behavior of model elements (agents), agents are able to adapt and change their behavior, have a dynamic relationship with other agents that can be configured to disappear in the process of modeling and other. The developed agent-based simulation model takes into account the complex dynamics of nanoelectronic manufacturing and provides a simulation multicluster technological complexes of different topologies and implements a variety of strategies and control algorithms by flows of semi-finished products. Formulations of the problems of movements planning of semi-finished products in multicluster technological complexes associated with the various strategies and performance criteria for nanoelectronic manufacturing logistics are formulated within the paper.

Keywords: nanoelectronic manufacturing, multi-cluster technological complex, cluster, plate, the flow of semi-finished products, agent model.

REFERENCES

- Shishmarev V. Y. Organization and planning of automated production : the textbook. Moscow, Academy 2013, 304.
- Mohammad I., Mahgoub I. Handbook of sensor networks : compact wireless and wired sensing systems. Boca Raton, Florida, CRC Press LLC, 2005, 776 p.
- Lopez M. J., Wood S. C. Systems of multiple cluster tools: configuration, reliability, and performance, *IEEE Trans. on Semiconductor Manufacturing*, 2003, Vol. 16, No. 2, pp. 170–178.
- Akopov A., Akopov S. Simulation : a tutorial and a workshop for the academic Bachelor, Higher School of Economics (HSE), National Research University (NRU). Moscow, Yuray, 2015, 389 p.
- Noack D., Mosinski M., O. Rose, Lendermann P., and Gan B. P. Challenges and Solution Approaches for the Online Simulation of Semiconductor Wafer Fabs, *Winter Simulation Conference : The premier international conference for simulation professionals in discrete and combined discrete-continuous simulation*, AZ, USA, 11–14 December 2011 : proceedings, Phoenix : WSC, 2011, pp. 1845–1856.
- Uhlig T., Rose O. Simulation-based optimization for groups of cluster tools in semiconductor manufacturing using simulated annealing, *Winter Simulation Conference : The premier international conference for simulation professionals in discrete and combined discrete-continuous simulation*, AZ, USA, 11–14 December 2011 : proceedings. Phoenix, WSC, 2011, pp. 1857–1868.
- Scholl W., Gan B. P., Peh M. L., Lendermann P., Noack D., Rose O., and Preuss P. Towards Realization of a High-Fidelity Simulation Model for Short-Term Horizon Forecasting in Wafer Fabrication Facilities, *Winter Simulation Conference : The premier international forum for disseminating recent advances in the field of system simulation*, MD, USA, 05 – 08 December 2010 : proceedings. Baltimore, WSC, 2010, pp. 2563–2574.
- Zuberek, W. M. Cluster tools with chamber revisiting – Modeling and analysis using timed Petri nets, *IEEE Transactions on semiconductor manufacturing*, 2004, Vol. 17, No. 3, pp. 333–344.
- Yi J. Ding S., Zhang M. T., Meulen V. D. Throughput Analysis of Linear Cluster Tools, *IEEE International Conference on Automation Science and Engineering: International Conference*, AZ, USA, 22 – 25 September 2007 : proceedings. Scottsdale, IEEE, 2007, pp. 1063–1068.
- Varadarajan A., Sarin S. C. A survey of dispatching rules for operational control in wafer Fabrication, *Information Control Problems in Manufacturing: The 12th IFAC International Symposium, France, 17–19 May 2006 : proceedings*. Saint Etienne: ELSEVIER, 2006, pp. 709–720.
- Wang P., Murata T. Hybrid Model applied in the Semiconductor Production Planning, *International MultiConference of Engineers and Computer Scientists: International MultiConference, China, 13–15 March 2013 : proceedings*. Hong Kong, Newswood Limited, 2013, pp. 964–968.

НЕЙРОИНФОРМАТИКА ТА ІНТЕЛЕКТУАЛЬНІ СИСТЕМИ

НЕЙРОИНФОРМАТИКА И ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ

NEUROINFORMATICS AND INTELLIGENT SYSTEMS

УДК 004.9

Бісікало О. В.¹, Васілевський О. М.²

¹Д-р техн. наук, професор, декан факультету комп'ютерних систем і автоматики Вінницького національного технічного університету, Вінниця, Україна

²Канд. техн. наук, доцент, професор кафедри метрології та промислової автоматики Вінницького національного технічного університету, Вінниця, Україна

ОЦІНКА НЕВИЗНАЧЕНОСТІ ВИМІРЮВАННЯ СЕНСУ ПРИРОДНО-МОВНИХ КОНСТРУКЦІЙ

Вирішено завдання оцінки невизначеності вимірювання сенсу природно-мовних конструкцій (ПМК) на основі формалізації понять мовного образу, штучної когнітивної системи та одиниці сенсу. В основу моделі бази знань штучної когнітивної системи закладено статистичну інформацію про асоціативну сполучуваність мовних образів, що надає підстави для уніфікованої оцінки одиниці та кількості сенсу ПМК. Запропоновано метод вимірювання сенсу ПМК на основі нечіткого відношення сенсу, який забезпечує врахування інформації про зв'язки між лемами тексту, що дозволяє отримати оцінку двох типів невизначеності вимірювання ознак сенсу. Отримано та інтерпретовано формальні оцінки невизначеності результатів вимірювання сенсу ПМК, що дозволяє врахувати інформацію про зв'язки між лемами для розв'язання прикладних задач комп'ютерної лінгвістики.

За допомогою розробленого на основі пакету DKPro Core програмного забезпечення проведено експерименти з дослідження запропонованого методу в задачі виявлення інформативних ознак тексту. В результаті проведених експериментів отримано залежності параметрів виявленого Парето-подібного закону розподілу зв'язків між лемами, аналіз яких дозволяє вважати показник середньої кількості зв'язків мовного образу найбільш інформативною чисельною ознакою тексту.

Ключові слова: сенс, невизначеність, текст, природно-мовна конструкція, штучна когнітивна системи, мовний образ, лема.

НОМЕНКЛАТУРА

$[\alpha_-; \alpha_+]$ – границі апіорно визначеного закону розподілу;

γ , ПМК – природно-мовна конструкція (образна конструкція);

λ , $\bar{q}$ – математичне сподівання;

μ – цілочисельний показник емоційного стану системи;

$\mu_Q(< i_l, i_j >)$ – функція належності нечіткого відношення Q ;

σ , σ_i, U_j – середньоквадратичне відхилення (СКВ);

$\Omega \subseteq I \times I$ – простір упорядкованих пар образів
 $I = \{i_1, i_2, \dots, i_L, \dots\}$;

$\omega \in \Omega$ – асоціативний зв'язок між парами образів (словосполучення);

A_Q – матриця суміжності розмірністю $L \times L$ на момент часу t_L ;

$G_Q(V, E)$ – граф Бержа;

k – параметр розподілу Парето;

k_{lj} – значення (частота зв'язку) ненульового lj -го елемента матриці A_Q ;

k_Σ – сумарне значення всіх зв'язків системи;

m – кількість ненульових елементів матриці суміжності A_Q ;

N – кількість відомих ШКС;

n – кількість спостережень;

nt – кількість образів, що розрізняє ШКС;

p – довірчий рівень;

S, S_i , ШКС – штучна когнітивна система;

$u_A(X)$ – оцінка невизначеності за типом А;

$u_B(X)$ – оцінка невизначеності за типом В;

x_i – спостереження стану бінарного нечіткого відношення образного сенсу Q ;

Cav – (синтагматичної асоціації вага) – одиниця вимірювання образного сенсу.

ВСТУП

Складність задач семантичного аналізу текстової інформації вважається однією з головних перешкод на шляху побудови штучного інтелекту в цілому та розв'язання з належною якістю значної частини задач комп'ютерної лінгвістики зокрема. В процесі онтогенезу людини вчиться та набуває нових знань все своє життя, внаслідок цього кожний природний інтелект є унікальним та динамічним явищем, здатним самовдосконалюватися та добре розуміти собі подібних. Тому конструювання лінгвістичних баз знань має базуватися саме на таких принципах, а проблема отримання нових формальних методів семантичного аналізу природно-мовних конструкцій на основі баз знань є актуальною. Потребують обґрунтування формальні підходи до створення штучних когнітивних систем, здатних імітувати діяльність людини в процесах оброблення, розуміння смислу та ефективного застосування вхідної текстової інформації.

В роботах [1, 2] було запропоновано та обґрунтовано введення одиниці вимірювання образного сенсу 1 Сав з метою розв'язання задач комп'ютерної лінгвістики, пов'язаних з моделюванням образного мислення людини. Але в процесі такого моделювання обов'язково потрібно врахувати суб'єктивний та динамічний характер онтогенезу пізнавальної, у тому числі мовленнєвої діяльності людини. Формально це можна зробити різними шляхами, одним з яких є оцінка невизначеності результату вимірювання сенсу як окремих природно-мовних конструкцій (ПМК), так і текстів та штучної когнітивної системи (ШКС) у цілому на певний момент часу. Відомо [3], що невизначеність вимірювання – це параметр, пов'язаний з результатом вимірювання, який характеризує дисперсію значень, що можуть бути достатньо обґрунтовано приписані вимірюваній величині. Але важливо, щоб величина, яка безпосередньо використовується для вираження невизначеності, має бути внутрішньо узгоджена: безпосередньо виведена з компонентів, які її утворюють, а також не повинна залежати від групування цих компонентів і від їх розкладу на субкомпоненти [4]. У відомих літературних джерелах, де розглянуті стандартні невизначеності вимірювання типів А та В, не було застосовано поняття невизначеності та основні вимоги до нього для розв'язання задач семантичного аналізу тексту.

Об'єктом дослідження обрано процес побудови лінгвістичних баз знань когнітивної системи, предмет дослідження – оцінка невизначеності формальних ознак сенсу ПМК. Мета роботи полягає в отриманні оцінки невизначеності вимірювання сенсу ПМК як компонентів ШКС. Для досягнення поставленої мети необхідно ввести формальне поняття ШКС, обґрунтувати метод вимірювання сенсу ПМК на основі нечіткого відношення, отримати та інтерпретувати формальні оцінки невизначеності результатів вимірювання сенсу ПМК.

1 ПОСТАНОВКА ЗАДАЧІ

На вхід будь-якої системи S_i з N відомих подається деякий потік $X = \{x_1, x_2, \dots\}$, що на момент часу t_L може бути визначений графом $G_Q(V, E)$ та відповідною матрицею суміжності A_Q розмірністю $L \times L$. Відомо також,

що в розрідженій матриці A_Q кількість ненульових lj -х елементів дорівнює m , а кожний з них набуває значення k_{lj} . Потрібно отримати оцінки невизначеності σ результатів спостережень k_{lj} кожної системи S_i , а також обчислити стандартні невизначеності типу А – $u_A(X)$ та типу В – $u_B(X)$ для всіх систем. З огляду на мету дослідження необхідно інтерпретувати та проаналізувати формальні результати у термінах предметної області комп'ютерної лінгвістики.

2 ОГЛЯД ЛІТЕРАТУРИ

Розглянемо основні вимоги до поняття невизначеності вимірювання, викладені у [4, 5]. Ідеальний метод оцінювання невизначеності результату вимірювання повинен бути універсальним: придатним для всіх видів вимірювань і для всіх типів вхідних даних, що використовуються у вимірюваннях. Внутрішня узгодженість величини, що безпосередньо використовується для вираження невизначеності, передбачає можливість прямого використання невизначеності одного результату як компонента оцінювання невизначеності іншого, в якому використовується перший результат.

Невизначеність результату вимірювання у загальному випадку складається з кількох компонентів, які можна згрупувати у дві категорії, залежно від способу оцінювання їх числового значення: тип А – компоненти, оцінені статистичними методами; тип В – компоненти, оцінені іншими способами. Кожний детальний звіт про невизначеності повинен містити повний перелік компонентів і для кожного з них – метод, який використовувався при одержанні його числового значення.

Компоненти категорії А зазвичай характеризуються оціненими дисперсіями σ_i^2 (або оціненими «стандартними відхиленнями» σ_i) і числом степенів вільності. У випадку необхідності слід зазначати коваріації. Компоненти категорії В повинні характеризуватися величинами U_j^2 , які можна розглядати як наближення до відповідних дисперсій, існування яких допускається. Величини U_j^2 можна розглядати як дисперсії, а U_j – як стандартні відхилення. При необхідності, коваріації повинні розглядатися аналогічно.

Комбінована невизначеність повинна характеризуватися числовим значенням, одержаним при застосуванні звичайного методу для складання дисперсій. Комбінована невизначеність і її компоненти повинні виражатися у формі «стандартних відхилень». Якщо в окремих випадках для одержання загальної невизначеності комбіновану невизначеність необхідно множити на коефіцієнт, то коефіцієнт множення повинен бути завжди зазначений. Загалом слово невизначеність (*uncertainty*) означає сумнів, і, таким чином, у широкому сенсі «невизначеність вимірювання» означає сумнів щодо вірогідності результату вимірювання (*uncertainty measuring*).

Отже, невизначеність результату вимірювання обов'язково є свідченням правдоподібності того, що результат вимірювання близький до значення вимірюваної величини; це просто оцінювання близькості результату вимірювання до найкращого значення, що відповідає наявним на цей час знанням. Введення поняття «невиз-

наченість вимірювання» є вимушеною мірою, необхідною для одноманітного і спрощеного оцінювання достовірності вимірювання (*evaluation of measuring authenticity*), оскільки її визначення здійснюється на основі одержуваних результатів вимірювання, відомих умов вимірювань і характеристик використовуваної апаратури, а не на невідомому дійсному значенні вимірюваної величини [6].

Для оцінювання x_i вхідної величини X_p , яка не була отримана в результаті повторних спостережень, пов'язані з ними оцінені дисперсія $u^2(x_i)$ або стандартна невизначеність $u(x_i)$ визначаються на базі наукового судження, що базується на всій доступній інформації про можливу змінність X_p . Тобто, стандартну невизначеність типу В одержують із передбачуваної функції щільності ймовірності, заснованої на мірі впевненості в тому, що подія обов'язково відбудеться (ця ймовірність часто називається суб'єктивною ймовірністю).

Оскільки інформацію для оцінки невизначеності вимірювання можуть складати дані попередніх вимірювань, розглянутий у [2] підхід дозволяє забезпечити процес вимірювання сенсу ПМК на основі нечіткої міри. Так, в [1] бінарне нечітке відношення, що задане на одній базисній множині (універсумі) мовних образів I , визначено як нечітке відношення

$$Q = \{ \langle i_l, i_j \rangle, \mu_Q(\langle i_l, i_j \rangle) \}, \quad (1)$$

де $\mu_Q(\langle i_l, i_j \rangle)$ – функція належності бінарного нечіткого відношення, що задається як відображення $\mu_Q: I \times I \rightarrow [0, 1]$. У виразі (1) через $\langle i_l, i_j \rangle$ позначено кортеж з двох елементів, причому $i_l \in I$, $i_j \in I$. Якщо носій Q_s нечіткого відношення Q є скінченним, то потужність цього нечіткого відношення чисельно дорівнює кількості кортежів його носія і позначається як $card(Q_s)$.

Якщо бінарне нечітке відношення (1) є базовою когнітивною характеристикою ШКС, тоді функцію належності $\mu_Q(\langle i_l, i_j \rangle)$ варто вважати природною чисельною мірою сенсу. Значення $\mu_Q(\langle i_l, i_j \rangle) = 1$, згідно з [1], отримало назву одиниці сенсу розміром один *Сав*. В загальному вигляді функція належності нечіткого відношення сенсу для пари мовних образів (на базовому рівні) задається як

$$\mu_Q(\langle i_l, i_j \rangle) = f(k_{lj}, t_L), \quad (2)$$

де k_{lj} – кількість зафіксованих ШКС зв'язків між l -м та j -м образами на момент часу t_L . Значення k_{lj} неважко отримати шляхом підрахунку кількості зафіксованих ШКС кортежів $\langle i_l, i_j \rangle$ на основі технологічних можливостей сучасних лінгвістичних пакетів, що дозволяє вперше застосувати та обґрунтувати поняття невизначеності вимірювання сенсу ПМК.

3 МАТЕРІАЛИ І МЕТОДИ

3.1 Поняття штучної когнітивної системи: формалізація та інтерпретація

Розглянемо систему S , яку в подальшому будемо називати штучною когнітивною системою, з точки зору процесів накопичення її бази знань. Нехай S здатна розпізнавати образи з нескінченної множини $I = \{i_1, i_2, \dots, i_L, \dots\}$ та

сприймати асоціативні зв'язки між парами образів як елементи множини $\omega \in \Omega$, де $\Omega \subseteq I \times I$ – простір упорядкованих пар. Для визначення образної конструкції застосуємо поняття F – сигма-алгебри (σ -алгебри) підмножин з Ω . Далі будемо вважати образною конструкцією будь-яку підмножину $\gamma \subseteq \Omega$, що має властивість $\gamma \in F$. Якщо, згідно з властивостями γ -алгебри [7], множини $A, B \in F$, то об'єднання, перетин і різниця A та B у теоретико-множинному сенсі також належать F .

Припустимо, що система S обмінюється інформацією із зовнішнім світом як чорним ящиком виключно у вигляді образних конструкцій, з яких розрізняють послідовність вхідних подій $X = \{x_1, x_2, \dots\}$ та множини образних реакцій системи $Y = \{y_1, y_2, \dots\}$, причому $x_i \in F$, $y_i \in F$. На рис. 1 зображено схему абстрактної моделі когнітивної діяльності, що включає у свій склад зовнішній «чорний ящик» та внутрішню ШКС, на вхід якої неперервно подається множина образів подій у вигляді потоку X . На виході ШКС з'являються образи Y , які є реакцією цієї системи на зовнішню ситуацію X згідно з підходом до моделювання образного мислення людини [2].

Закладемо як базовий *онтогенетичний принцип* побудови ШКС – когнітивний ресурс Ω системи S , що визначає сенс її функціонування, отримується виключного шляхом послідовного накопичення параметрів чергових ω з зовнішнього «чорного ящика» та подальшого самовдосконалення множини Ω . Формально онтогенетичний принцип відображається в тому, що базу знань системи S

будуємо як $C = \bigcup_{i=1}^{m'} x_i$, де m' – загальна кількість сприйня-

тих системою на даний час вхідних образних конструкцій.

З метою розв'язання прикладних задач комп'ютерної лінгвістики інтерпретуємо складові отриманої абстрактної моделі когнітивної діяльності. Для ШКС лінгвістичного типу образом i пропонується вважати мовний образ, що наближено задається лексемою або словоформою [8]. Тоді аналогом асоціативного зв'язку між парами образів ω є словосполучення, а образної конструкції γ – речення, мовне висловлювання, загалом – ПМК. Накопичення ШКС когнітивного ресурсу Ω відбувається шляхом опрацювання множини текстів, а наслідком цього є побудова лінгвістичної бази знань C .

На відміну від відомих моделей знань комп'ютерної лінгвістики, в яких словник словоформ поєднується з

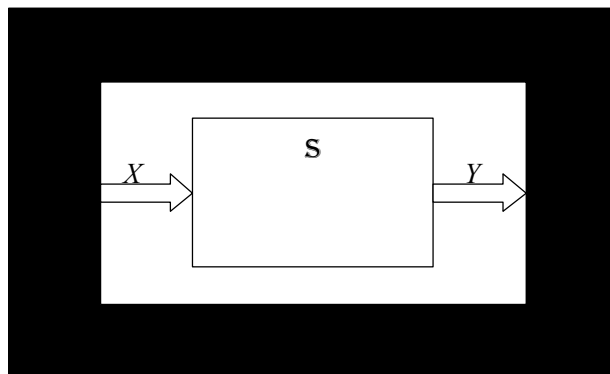


Рисунок 1 – Схема абстрактної моделі когнітивної діяльності

множинами морфологічних, синтаксичних та семантичних правил, основу бази знань S складають виключно асоціативні знання про сполучуваність мовних образів i . Це надає підстави для уніфікованої оцінки одиниці та кількості сенсу ПМК.

3.2 Метод вимірювання сенсу ПМК на основі нечіткого відношення

У відповідності до запропонованого підходу [9] деталізуємо функцію належності, що породжує бінарне нечітке відношення сенсу (1) на таких 3-х послідовних рівнях, побудованих на базовому (2):

1. Рівень імовірнісного прогнозування – з метою нормування функції належності у проміжку $[0, 1]$ передбачено розрахунок статистичної оцінки λ (математичного сподівання): якщо для nt відомих ШКС на момент часу

t_L образів $k_{\Sigma} = \sum_{l=1}^{nt} \sum_{j=1}^{nt} k_{lj}$, а m – кількість усіх ненульових

кортежів $< i_l, i_j >$, то $\lambda = k_{\Sigma} / m$ – в цьому випадку застосуємо відому сигмоїдальну функцію [10]

$$\mu_Q(< i_l, i_j >) = f_1(k_{lj}, \lambda) = 1 / (1 + e^{-k_{lj} + \lambda}), \quad (3)$$

Внаслідок нормування з'являється характерна властивість функції належності, отриманої за методом, що

пропонується – середнє значення $\overline{\mu_Q} = \frac{1}{m} \sum_{j=1}^m \mu_{Qj} = 0,5$.

2. Рівень врахування емоційного стану – введено можливість врахування бінарної моделі емоцій ШКС [9] за рахунок показника $\mu = \{..., -2, -1, 1, 2, ...\}$, тоді

$$\mu_Q(< i_l, i_j >) = f_2(k_{lj}, \lambda, \mu) = 1 / (1 + e^{\frac{k_{lj} - \lambda}{|\mu|}}). \quad (4)$$

При $\mu = -1 \vee 1$ емоції не впливають на сенс функціонування ШКС, а функція належності (4) вироджується у функцію (3). Збільшення показника μ симетрично згляд-

жує сигмоїдальну функцію f_2 , що продемонстровано на рис. 2.

3. Рівень врахування мотиваційної компоненти на основі образів-центрів потреб – запропоновано модель мотиву ШКС на момент часу t_L вважати досягнення образу-центру потреби j' , а також розрахувати дисперсію та середньоквадратичне відхилення результатів спостережень k_{lj} як

$$D = \frac{1}{m} \sum_{l=1}^{nt} \sum_{j=1}^{nt} (k_{lj} - \lambda)^2 \mid k_{lj} > 0 \text{ і } \sigma = \sqrt{D}. \quad (5)$$

Отримане значення σ будемо вважати невизначеністю, що обумовлена недосконалістю моделі мотиву ШКС. Характеризує цю невизначеність зокрема недосконалість базової залежності (3), на основі якої пропонується врахувати мотиваційну компоненту на основі образів-центрів потреб.

В залежності від ступеня наближення r пари образів $< i_l, i_j >$ до j' , функцію (4) можна зміщувати вліво за віссю абсцис шляхом зменшення математичного сподівання для цієї пари $\lambda_{lj} = \lambda - r \cdot \sigma$, де $r = \{0, 1, 2, 3\}$, зрештою маємо

$$\mu_Q(< i_l, i_j >) = f_3(k_{lj}, \lambda_{lj}, \sigma, \mu, i') = 1 / (1 + e^{\frac{k_{lj} - \lambda_{lj}}{|\mu|}}). \quad (6)$$

Питання побудови окремого алгоритму для визначення ступеня наближеності r пари $< i_l, i_j >$ до образу-потреби j' та введення додаткового рівня врахування рефлексів та результатів зовнішнього навчання розглянуто у [9]. Зауважимо, що, на відміну від (3) та (4), у функції належності відношення сенсу (6) внаслідок локальних зсувів математичного сподівання зникає властивість $\overline{\mu_Q} = 0,5$, що, на думку авторів, свідчить про належну формальну інтерпретацію відомих фактів з психології та фізіології щодо протиріч між загальноприйнятим

Сигмоїдальна функція належності відношення сенсу

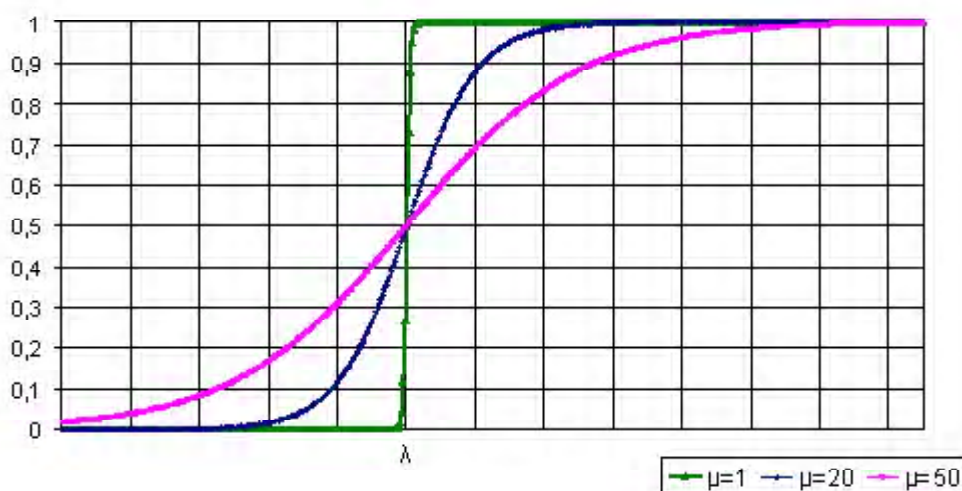


Рисунок 2 – Вплив показника μ на функцію належності (4)

(середньостатистичним) сенсом і діями під впливом сильних мотивів.

3.3 Невизначеність результатів вимірювання сенсу ПМК

Розглянутий підхід до вимірювання сенсу відповідає лінгвістичній базі знань однієї ШКС, вихідними даними якої може бути як окремих текст, так і деяка унікальна множина текстів. При цьому потрібно розуміти, що за кожним текстом стоїть так само унікальний світогляд автора, втілений у його мові. Для розв'язання задачі виявлення інформативних ознак тексту важливим є визначення достовірності бази знань у цілому та сенсу однієї пари мовних образів у вигляді $\mu_Q(<i_l, i_j>)$, як базової складової цих знань зокрема. Оскільки фактично йдеться про вимірювання сенсу, то для оцінки достовірності пропонується застосувати поняття невизначеності результатів множинного вимірювання сенсу ПМК.

У першому наближенні будемо вважати, що суб'єктивна оцінка кількості сенсу однієї пари мовних образів втілюється у статистичний ряд чисельних значень для N різних ШКС. Отже, для довільного кортежу $<i_l, i_j>$ вимірювана згідно (3) величина $Y = \mu_Q(<i_l, i_j>)$ функціонально залежить від результатів її багаторазових вимірювань $X_1, X_2, \dots, X_N$ для різних ШКЛ та, в загальному випадку, має такий вигляд

$$Y = f(X_1, X_2, \dots, X_N). \quad (7)$$

Оцінку вимірюваної величини Y , позначену y , одержимо із загального рівняння (7), використовуючи вхідні оцінки $x_1, x_2, \dots, x_N$ для N значень величин $X_1, X_2, \dots, X_N$. Отже, вихідна оцінка y , яка є результатом вимірювання, виражається таким чином

$$y = f(x_1, x_2, \dots, x_N).$$

Базовою оцінкою математичного сподівання або очікуваного значення μ_Q величини q , що змінюється випадковим чином, є середнє арифметичне або середнє значення $\bar{q}$ із n спостережень

$$\bar{q} = \frac{1}{n} \sum_{k=1}^n q_k. \quad (8)$$

Експериментальне стандартне відхилення, що характеризує змінність значень q_k , або, точніше, їхню дисперсію σ^2 щодо середнього значення $\bar{q}$, розраховують за формулою [6]

$$u_A(q_k) = \sqrt{\frac{\sum_{k=1}^n (q_k - \bar{q})^2}{n-1}}. \quad (9)$$

Оскільки за результат багаторазових вимірювань приймають середнє значення $\bar{q}$, то важливо оцінити його

дисперсію. Найкраща оцінка $\sigma^2(\bar{q}) = \sigma^2/n$ дисперсії середнього значення $u_A^2(\bar{q})$ виражається як

$$u_A^2(\bar{q}) = \frac{u_A^2(q_k)}{n}. \quad (10)$$

Експериментальна дисперсія середнього $u_A^2(\bar{q})$ і експериментальне стандартне відхилення середнього значення $u_A(\bar{q})$, що дорівнює позитивному квадратному кореню з оцінки дисперсії $u_A^2(\bar{q})$, кількісно визначають, наскільки добре $\bar{q}$ оцінює очікування μ_Q величини q . З урахуванням виразів (9) та (10) експериментальне стандартне відхилення середнього значення $u_A(\bar{q})$ розраховується за формулою [6]

$$u_A(\bar{q}) = \sqrt{\frac{\sum_{k=1}^n (q_k - \bar{q})^2}{n(n-1)}}. \quad (11)$$

Для більш глибокого врахування суб'єктивного характеру вимірюваного сенсу кортежів у функції (7) застосуємо складові стандартної невизначеності типу В, які, як правило визначають|обчислюють, вивчають| на основі інформації про верхні і нижні границі $[\alpha_-; \alpha_+]$ передбачуваного закону розподілу чи через інтервал U , що має заданий довірчий рівень довіри p .

Для визначення стандартної невизначеності типу В потрібно взяти позитивний квадратний корінь з добутку довірчого рівня кожного значення та квадрата відхилення цього значення і всі добутки такого виду додати. В результаті загальний вигляд формули для обчислення стандартної невизначеності типу В при дискретних даних має вигляд

$$u_B(X) = \sqrt{\sum_{i=1}^n \left(x_i - \sum_{i=1}^n x_i p_i \right)^2 p_i} = \sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 p_i}. \quad (12)$$

Якщо для значення величини X_i можна оцінити верхню та нижню границю $[\alpha_-; \alpha_+]$, то стандартні невизначеності типу В, в припущенні про можливий вигляд закону розподілу, можна визначити за формулами [4, 5, 6]:

а) для трикутного закону розподілу

$$u_B(X_i) = \frac{\alpha_+ - \alpha_-}{\sqrt{24}}; \quad (13)$$

б) для експоненціального закону розподілу

$$u_B(X_i) = \sqrt{\frac{(\alpha_+ - x)(x - \alpha_-) - (\alpha_+ - 2x + \alpha_-)}{\lambda}}, \quad (14)$$

де x – очікуване значення, а λ – параметр розподілу;

в) для закону розподілу Парето

$$u_B(X_i) = \frac{x_m}{k-1} \sqrt{\frac{k}{k-2}}, \quad (15)$$

де x_m – початкове значення x , а k – параметр розподілу (значення щільності для x_m);

д) для рівномірного закону розподілу

$$u_B(X_i) = \frac{\alpha_+ - \alpha_-}{\sqrt{12}}. \quad (16)$$

Для заданих інтервалів U_p із відомим рівнем довіри p , в припущенні нормального закону розподілу, невизначеність типу В визначається за формулою

$$u_B(X_i) = \frac{U_p}{k_p},$$

де k_p – коефіцієнт охоплення, який для нормального закону розподілу, дорівнює 1,64; 1,96; 2,58 і 3 для довірчих рівнів 0,9; 0,95; 0,99 і 0,9973. За відсутності інформації про наявність законів (13)–(16) розподілу вхідної величини X_i для симетричних границь $\pm \alpha_i$ стандартну невизначеність типу В визначають за формулою

$$u_B(X_i) = \frac{2\alpha_i}{\sqrt{12}} = \frac{\alpha_i}{\sqrt{3}}, \quad (17)$$

яка може бути застосована на початковому етапі експериментального дослідження ШКС.

4 ЕКСПЕРИМЕНТИ

З метою експериментальної перевірки результатів оцінки невизначеності вимірювання сенсу ПМК як компонентів ШКС за допомогою запропонованого методу було застосовано відомий лінгвістичний пакет DKPro Core, який базується на платформі Apache UIMA framework [12]. Для реалізації серії експериментів було розроблено додаткову Java-програму (додаток 1), що використовує та удосконалює колекцію програмних компонентів для обробки природної мови DKPro Core [13]. Особливість розробленої програми, що орієнтована на технологію Java/Maven/Eclipse, полягає у визначенні списку лем тексту та складних залежностей згідно [14] між цими лемами у вигляді списку з m зв'язків.

Експериментальною базою було обрано три відомі літературні твори з відкритого джерела Project Gutenberg [15], а саме англomовні (авторські) варіанти 4-х текстів різного обсягу: «Аліса в країні див» (Л. Керол, 1 – уривок з 4204 слів та 2 – повна версія з 26690 слів), 3 – «Біле ікло» (Дж. Лондон,

48907 слів) та 4 – «Троє у човні без врахування собаки» (Дж. К. Джером, 67328 слів). Мета серії експериментів полягала у дослідженні базових характеристик невизначеності кожного з 4-х текстів, а також у отриманні оцінки невизначеності множини спільних для всіх текстів пар мовних образів $\langle i_l, i_j \rangle$ згідно з запропонованим методом.

5 РЕЗУЛЬТАТИ

У результаті дослідження формалізовано та інтерпретовано для предметної галузі комп'ютерної лінгвістики поняття штучної когнітивної системи, закладено базовий онтогенетичний принцип побудови ШКС. Отримано формальні характеристики методу створення бінарного нечіткого відношення образного сенсу Q ШКС S_Q шляхом моделювання понять мотиваційної мети та емоційного стану. Запропоновано принципи послідовної багаторівневої побудови функції належності $\mu_Q(\langle i_l, i_j \rangle)$, що породжує нечітке відношення Q , визначено характерну властивість $\overline{\mu_Q} = 0,5$ методу вимірювання сенсу ПМК. Згідно з ним для задачі виявлення інформативних ознак тексту отримано формальні теоретичні оцінки невизначеності σ результатів спостережень k_{ij} кожної ШКС S_i , а також розраховані стандартні невизначеності типу А – $u_A(X)$ та типу В – $u_B(X)$ для всіх ШКС.

За допомогою розробленого в [13] програмного забезпечення на основі пакету DKPro Core було отримано результати обробки 4-х обраних англomовних текстів, що можуть інтерпретуватися як 4 різні ШКС. Основні результати обробки у відповідності до (5) представлено в табл. 1, де 3 останні стовпці вміщують такі дані:

- відсоток СКВ σ від оцінки математичного сподівання λ ;
- кількість визначених засобами DKPro Core лем тексту;
- середня кількість різних зв'язків для однієї лем тексту.

Отримані гістограми експериментальних законів щільності розподілу показали значну зовнішню схожість до розподілу за законом Парето, що демонструє приклад порівняння експериментального результату для тексту 1 (Carrol_part) з теоретичною щільністю розподілу Парето зі значенням параметру $k = 2108$.

Аналіз відсортованих за спаданням k_{ij} списків пар мовних образів $\langle i_l, i_j \rangle$ дозволив виявити 4 спільні пари у верхній частині списків, вихідні дані та результати оцінки $\bar{q}$ за (8) та невизначеності за типами А та В згідно з (11) і (12) яких представлено в табл. 2.

Таблиця 1 – основні результати обробки 4-х англomовних текстів

Текст	m	k_Σ	λ	σ	%	Кільк. лем	Сер. кільк. зв'язків
1 Carrol part	2360	2812	1,191525424	0,778805721	65,36%	762	3,0971
2 Carrol full	12156	17786	1,463145772	2,245695112	153,48%	2121	5,7313
3 London	25244	31234	1,237284107	1,259517221	101,80%	5702	4,4272
4 Jerom	33316	47091	1,413465002	2,044626970	144,65%	6048	5,5086

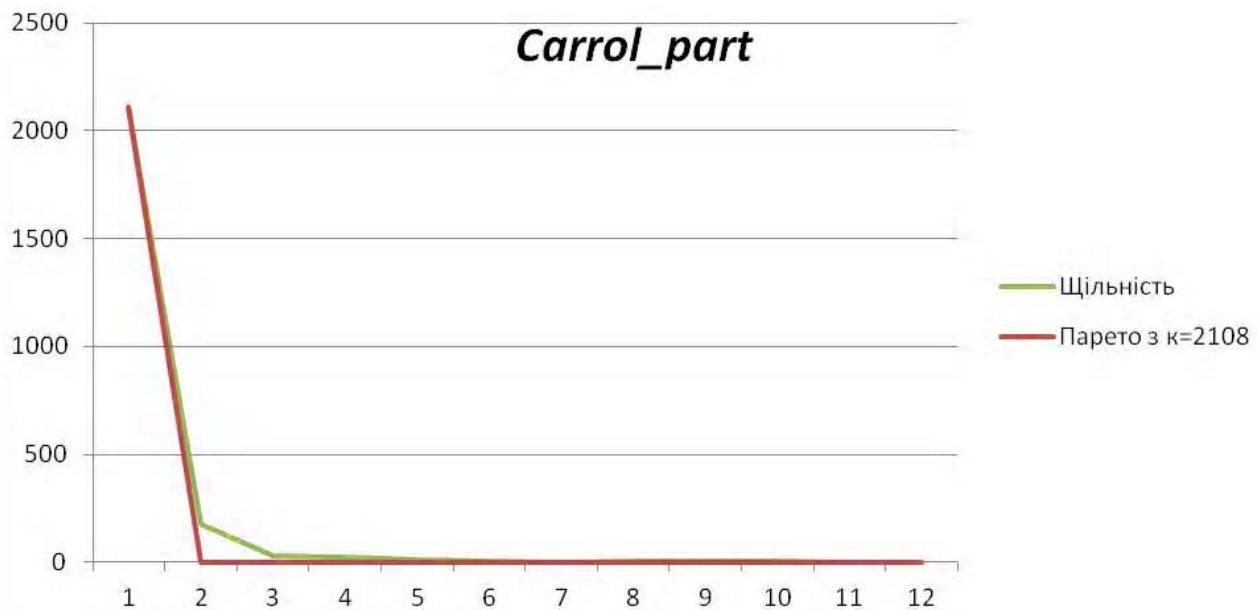


Рисунок 3 – Аналіз експериментального закону щільності розподілу для тексту 1

Таблиця 2 – Результати оцінки невизначеності 4-х обраних пар мовних образів

Текст		<i>go-back</i>	<i>say-I</i>	<i>know-I</i>	<i>see-I</i>
1	<i>Carrol part</i>	0,859177412	0,943132060	0,943132060	0,943132060
2	<i>Carrol full</i>	0,998553057	0,999999514	0,386239825	0,998553057
3	<i>London</i>	0,999999948	1,000000000	1,000000000	0,999999993
4	<i>Jerom</i>	0,999999537	1,000000000	1,000000000	1,000000000
$\bar{q}$		0,928865234	0,971565787	0,664685943	0,970842558
$u_A(X)$		0,028449934	0,01160802	0,113675151	0,011312764
		0,036641087	0,01495014	0,146403893	0,014569874

6 ОБГОВОРЕННЯ

Отримані в результаті експерименту чисельні оцінки невизначеності результатів вимірювання сенсу пар мовних образів дозволяють отримати нову інформацію щодо текстів, що аналізуються. Представлення кожного тексту, як окремої ШКС демонструє, що експериментальний закон щільності розподілу для характеристики k_{ij} пар мовних образів дуже подібний до розподілу Парето. Однак такому висновку не відповідають оцінки математичного сподівання λ , яке б мало зменшуватися та наближатися до 1 ($\lambda_{Pareto} = \frac{k \cdot x_m}{k-1}$) зі збільшенням числа пар [16], а також СКВ σ , яке занадто велике для розподілу Парето. Наприклад, для тексту 1 згідно з (5) $\sigma = 0,7788$, що складає 65,36% від λ . Аналогічні оцінки відповідно до залежностей (15) для розподілу Парето та (17) для загального випадку з малим значенням $\alpha_i = \pm 0,01$: $\sigma_1 = 0,0004748$ (0,04%) та $\sigma_2 = 0,58$ (0,48%).

Проте, аналіз даних таблиці 1 надає формальні підстави для висунення гіпотези – найбільш інформативною характеристикою ШКС є середня кількість зв'язків для однієї леми (мовного образу). Обґрунтування – коефіцієнт кореляції Пірсона для стовпчиків з λ та «Кількість

лем» для всіх 4-х ШКС дорівнює 0,198, але для пар стовпчиків λ та «Середня кількість зв'язків» – 0,945. Одночасно для пар стовпчиків σ та «Середня кількість зв'язків» коефіцієнт кореляції дорівнює 0,984, а для пар стовпчиків «%» та «Середня кількість зв'язків» – 0,996. Це дозволяє вважати, що закон розподілу є лише Парето-подібним, проте невизначеність сенсу ШКС (параметр σ) прямо пропорційна середній кількості зв'язків. Висунута гіпотеза потребує масштабнішої експериментальної перевірки та уточнення.

Дані таблиці 2 демонструють високу ступінь сенсоподібності згідно з висунутим підходом для 4-х обраних пар мовних образів, що використовувалися 3-ма різними авторами. Загальна тенденція полягає у тому, що оцінки невизначеності $u_A(X)$ за типом А менші відповідних оцінок $u_B(X)$ за типом В для всіх ШКС приблизно у 1,5 рази. При цьому відсоток невизначеності не перевищує 4% від оцінки математичного сподівання $\bar{q}$ для всіх пар $\mu_Q(< i_l, i_j >)$, окрім пари «know-I» (до 22,03%), що має зрозуміле пояснення – в обраному уривку 1 тексту Л. Керола ця пара зустрічається відносно набагато частіше, ніж у цьому ж творі 2 («Аліса у країні казок») у цілому. Такі результати дозволяють сподіватися, що зап-

ропонований підхід дозволить підвищити якість розв'язання задач автоматичного семантичного аналізу текстів, зокрема визначення авторства. Однак цілком ймовірно, що аналогічне порівняння пар, що зустрічаються рідко (знаходяться у нижній частині відсортованих списків) продемонструє високу невизначеність.

Подальших досліджень потребує також визначення законів розподілу експериментальних значень $\mu_Q(<i_l, i_j>)$ та отримання суб'єктивних характеристик бази знань ШКС у представленні динамічної невизначеності вимірювань.

ВИСНОВКИ

Внаслідок проведених досліджень розв'язано актуальну задачу оцінки невизначеності вимірювання сенсу ПМК як компонентів ШКС, яка безпосередньо пов'язана з проблемою розуміння смислу текстової інформації. Набув подальшого розвитку метод вимірювання сенсу ПМК на основі нечіткого відношення, який, на відміну від існуючих, базується на введених формальних поняттях штучної когнітивної системи та мовного образу, що дозволяє отримати вихідні статистичні дані для оцінки результатів невизначеності вимірювання типів А та В. Уперше отримано та інтерпретовано формальні оцінки невизначеності результатів вимірювання сенсу ПМК, що дозволяє врахувати інформацію про зв'язки між лемами тексту для розв'язання задачі виявлення інформативних ознак тексту.

Практичне значення отриманих результатів полягає в отриманні програмного технологічного інструментарію на основі лінгвістичного пакету DKPro Core, що дозволяє реалізувати запропонований метод для семантичного аналізу англійських текстів. За результатами проведеної серії експериментів виявлено, що закон розподілу зв'язків між лемами тексту є Парето-подібним, проте має суттєві формальні відмінності від класичного розподілу Парето, зокрема суттєво більші оцінки математичного сподівання λ (до 46,3 %) та СКВ σ (на кілька порядків).

З точки зору запропонованого підходу до визначення сенсу ПМК збільшення розмірів тексту за кількістю слів та, відповідно, його словникового складу за кількістю лем не впливає на параметри закону розподілу та невизначеність сенсу окремої ШКС. Аналіз отриманих результатів дозволяє вважати показник середньої кількості зв'язків мовного образу найбільш інформативною ознакою тексту, оскільки коефіцієнт кореляції Пірсона між ним та параметрами, пов'язаними з невизначеністю сенсу більший за 0,945.

Порівняння оцінок невизначеності 4-х пар мовних образів, що використовувалися 3-ма різними авторами, показало високу ступінь сенсоподібності таких пар згідно з висунутим підходом. При цьому оцінки невизначеності $u_A(X)$ за типом А пропорційно менші відповідних оцінок $u_B(X)$ за типом В для всіх ШКС приблизно у 1,5 рази, що дозволяє обмежитися знаходженням тільки однієї оцінки невизначеності $u_A(X)$.

Отримані результати досліджень, а саме формальні показники невизначеності сенсу та середня кількість

зв'язків мовного образу, мають перспективи використання в задачах семантичного аналізу ПМК, зокрема класифікації, класифікації та визначення авторства текстів.

ПОДЯКИ

Проведені дослідження здійснювались у межах держбюджетної науково-дослідної роботи Вінницького національного технічного університету за темою «Інтелектуальна інформаційна технологія образного аналізу тексту та синтезу інтегрованої бази знань природно-мовного контенту» (№ держреєстрації 0114U003462), а також у відповідності до плану досліджень науково-дослідного центру прикладної та комп'ютерної лінгвістики ВНТУ.

ДОДАТОК 1

Текст Java-програми *MyBasePipeline3*:

```
package de.tudarmstadt.ukp.tutorial.gscl2013.dkpro;
import static org.apache.uima.fit.factory.AnalysisEngineFactory.createEngineDescription;
import static org.apache.uima.fit.factory.CollectionReaderFactory.createReaderDescription;
import static org.apache.uima.fit.util.JCasUtil.select;
import java.io.FileNotFoundException;
import java.io.PrintWriter;
import java.io.UnsupportedEncodingException;
import java.util.ArrayList;
import java.util.Arrays;
import java.util.Collection;
import java.util.HashMap;
import java.util.HashSet;
import java.util.List;
import java.util.Map;
import java.util.Set;
import javax.xml.transform.TransformerConfigurationException;
import org.apache.uima.fit.pipeline.JCasIterable;
import org.apache.uima.fit.util.JCasUtil;
import org.apache.uima.jcas.JCas;
import org.jgrapht.ext.GraphMLExporter;
import org.jgrapht.ext.IntegerEdgeNameProvider;
import org.jgrapht.ext.IntegerNameProvider;
import org.jgrapht.ext.StringEdgeNameProvider;
import org.jgrapht.ext.StringNameProvider;
import org.jgrapht.graph.ClassBasedEdgeFactory;
import org.jgrapht.graph.DefaultDirectedWeightedGraph;
import org.xml.sax.SAXException;
import de.tudarmstadt.ukp.dkpro.core.api.coref.type.CoreferenceChain;
import de.tudarmstadt.ukp.dkpro.core.api.coref.type.CoreferenceLink;
import de.tudarmstadt.ukp.dkpro.core.api.segmentation.type.Lemma;
import de.tudarmstadt.ukp.dkpro.core.api.segmentation.type.Sentence;
import de.tudarmstadt.ukp.dkpro.core.api.segmentation.type.Token;
import de.tudarmstadt.ukp.dkpro.core.api.syntax.type.dependency.Dependency;
import de.tudarmstadt.ukp.dkpro.core.io.text.TextReader;
import de.tudarmstadt.ukp.dkpro.core.opennlp.OpenNlpPosTagger;
import de.tudarmstadt.ukp.dkpro.core.stanfordnlp.StanfordCoreferenceResolver;
import de.tudarmstadt.ukp.dkpro.core.stanfordnlp.StanfordParser;
import de.tudarmstadt.ukp.dkpro.core.tokit.BreakIteratorSegmenter;
public class MyBasePipeline3 {
    private static class LinkCounter { // класс для связи Map с
        графом DefaultDirectedWeightedGraph
        private Map<String, Map<String, Integer>> links;
        private DefaultDirectedWeightedGraph<String,
            RelationshipEdge<String>> g;
        LinkCounter() { //
            links = new HashMap<String, Map<String, Integer>>();
            g = new DefaultDirectedWeightedGraph<String,
                RelationshipEdge<String>>() {
                    new ClassBasedEdgeFactory(RelationshipEdge.class));
                }
        }
        void addLink(String from, String to, String type) { // метод для
            добавления новой связи
            from = from.toLowerCase();
            to = to.toLowerCase();
            Map<String, Integer> mapFrom = links.get(from);
            if (mapFrom == null) {
                mapFrom = new HashMap<String, Integer>();
                links.put(from, mapFrom);
            }
        }
    }
}
```

```

Integer countTo = mapFrom.get(to);
mapFrom.put(to, (countTo == null) ? 1 : countTo + 1);
if (!g.containsVertex(from)) {
    g.addVertex(from);
}
if (!g.containsVertex(to)) {
    g.addVertex(to);
}
RelationshipEdge<String> edge = new RelationshipEdge<String>(from,
to, type);
g.addEdge(from, to, edge);
}
void saveLinks(String filename)
throws FileNotFoundException, UnsupportedEncodingException {
    // метод для запаминання статистики зв'язей в файлі links.csv
    PrintWriter writer = new PrintWriter(filename, "UTF-8");
    for (String keyFrom : links.keySet()) {
        Map<String, Integer> mapFrom = links.get(keyFrom);
        for (String keyTo : mapFrom.keySet()) {
            Integer count = mapFrom.get(keyTo);
            writer.println(keyFrom + " " + keyTo + " " + count);
        }
    }
    writer.close();
}
/* метод - в графі шукати v2 і замінювати програмно на v1, т.е.
найти все зв'язи для v2, запаминати, видалити v2, вставити вузол
v1, додати зв'язи */
void V1ChangeV2(String v1, String v2) {
    try {
        v1 = v1.toLowerCase();
        v2 = v2.toLowerCase();
        Set<RelationshipEdge<String>> targ1 = g.outgoingEdgesOf(v2);
        if (targ1 != null) {
            Set<RelationshipEdge<String>> targ1copy = new
HashSet<RelationshipEdge<String>>();
            targ1copy.addAll(targ1);
            for (RelationshipEdge v : targ1copy) {
                addLink(v1, g.getEdgeTarget(v), v.toString());
                System.out.printf("%n" + v1 + " " + g.getEdgeTarget(v) + " " + v);
                g.removeEdge(v);
                //delLink(v2, g.getEdgeTarget(v), v.toString());
            }
        }
        Set<RelationshipEdge<String>> targ2 = g.incomingEdgesOf(v2);
        if (targ2 != null) {
            for (RelationshipEdge v : targ2) {
                addLink(g.getEdgeSource(v), v1, v.toString());
                System.out.printf("%n" + g.getEdgeSource(v) + " " + v1 + " " + v);
                //g.removeEdge(v);
                //delLink(g.getEdgeSource(v), v2, v.toString());
            }
        }
        g.removeVertex(v2);
    } catch (Exception e) {
        System.out.println("\nНесподіванка спіткала українських науковців
під час зміни " + v1 + " на " + v2);
        //e.printStackTrace();
    }
}
private void saveGraph(String filename)
throws FileNotFoundException, UnsupportedEncodingException,
SAXException, TransformerConfigurationException {
    // метод для
запаминання графа в файлі graph.xml
    GraphMLExporter<String, RelationshipEdge<String>> me =
new GraphMLExporter<String, RelationshipEdge<String>>(
new IntegerNameProvider<String>(),
new StringNameProvider<String>(),
new IntegerEdgeNameProvider<RelationshipEdge<String>>(),
new StringEdgeNameProvider<RelationshipEdge<String>>());
    PrintWriter writer = new PrintWriter(filename, "UTF-8");
    me.export(writer, g);
}
Map<String, Map<String, Integer>> getLinks() {
    return links;
}
DefaultDirectedWeightedGraph<String, RelationshipEdge<String>>
getGraph() {
    return g;
}
}
public static void main(String[] args) throws Exception {

```

```

главный метод класса MyBasePipeline3
JCasIterable pipeline = new JCasIterable(// запуск программного
конвейера для последовательного аннотирования
// (создания многоуровневой разметки) текста
createReaderDescription(TextReader.class,
TextReader.PARAM_SOURCE_LOCATION, "input/Obama.txt",
TextReader.PARAM_LANGUAGE, "en"), // чтение текста
createEngineDescription(BreakIteratorSegmenter.class), // сегмен-
тирование текста
createEngineDescription(OpenNlpPosTagger.class), // морфологичес-
кая разметка
createEngineDescription(StanfordParser.class,
StanfordParser.PARAM_VARIANT, "rnn",
StanfordParser.PARAM_MODE,
StanfordParser.DependenciesMode.CC_PROPAGATED), // синтаксическая
разметка, учитывающая
// сложные зависимости между парами лемм
// Stem
//createEngineDescription(SnowballStemmer.class),
// Lemma
//createEngineDescription(MateLemmatizer.class),
// NamedEntity
//createEngineDescription(OpenNlpNameFinder.class,
// OpenNlpNameFinder.PARAM_VARIANT, "person"),
//createEngineDescription(OpenNlpNameFinder.class,
// OpenNlpNameFinder.PARAM_VARIANT, "organization"),
//CoreferenceChain, CoreferenceLink
createEngineDescription(StanfordCoreferenceResolver.class)//
поиск соответствия местоимений
// SemanticPredicate, SemanticArgument
//createEngineDescription(ClearNlpSemanticRoleLabeler.class)
);
PrintWriter writer = new PrintWriter("output/output.txt", "UTF-
8");// запись зависимостей по предложениям в файл output.txt
for (JCas jcas : pipeline) {
    LinkCounter linkCounter = new LinkCounter();
    for (Sentence sentence : select(jcas, Sentence.class)) {
        writer.println("sentence: " + sentence.getCoveredText());
        // запись исходного предложения
        Collection<Token> tokens = JCasUtil.selectCovered(jcas,
Token.class, sentence);
        List<String> ts = new ArrayList<String>(tokens.size());
        /* исключение неинформативных для анализа типов зависимостей */
        List<String> excludes = Arrays.asList(new String[]{"det",
"punct", "cop", "cc", "aux", "auxpass", "expl", "mark", "num",
"number", "quantmod", "ref"/**/});
        for (Dependency dep : JCasUtil.selectCovered(jcas,
Dependency.class, sentence)) {
            String type = dep.getDependencyType();
            if (!excludes.contains(type)) {
                String govLemma = safeVal(dep.getGovernor());
                String depLemma = safeVal(dep.getDependent());
                String depnl = dep.getDependencyType();
                writer.println(depnl + " (" + govLemma + " " + depLemma + ")");
                // запись очередной зависимости
                linkCounter.addLink(govLemma, depLemma, depnl);
            }
        }
        //System.out.printf("%n - Semantic structure -%n");
        //for (SemanticPredicate pred : selectCovered(
//SemanticPredicate.class, sentence)) {
        //System.out.printf(" %16s %10s", pred.getCoveredText(),
//pred.getCategory());
        //for (SemanticArgument arg : select(pred.getArguments(),
//SemanticArgument.class)) {
        //System.out.printf("\t%s:%s", arg.getRole(),
//arg.getCoveredText());
        //}
        //System.out.printf("%n");
        //}
        System.out.printf("%n== Coreference chains (for the whole
document) ==%n");
        for (CoreferenceChain chain : select(jcas,
CoreferenceChain.class)) {
            // цикл по все найденным кореференци-
ям для имен и местоимений
            CoreferenceLink link = chain.getFirst();
            //System.out.println(link);
            String v1 = "#";
            String v2 = "$";

```

```

while (link != null) {
    //String v = link.getCoveredText().trim();
    String v = link.getCoveredText();
    while (v.indexOf(" ") != -1) v = v.substring(1+v.indexOf(" "));
    System.out.printf("\n:: %s |%s|", link.getCoveredText(),
        link.getReferenceType());
    if (link.getReferenceType() == "PROPER" ||
        link.getReferenceType() == "NOMINAL") v1=v;
    if (link.getReferenceType() == "PRONOMINAL") v2=v;
    if (link.getReferenceRelation() != null) {
        //System.out.printf("-[%s]", link.getReferenceRelation());
    }
    link = link.getNext();
}
if (!v1.equals("#") & !v2.equals("$")) { // если кореференции
    найдены, то проводим замену местоимений на номиналы
    System.out.printf(v1+" "+v2);
    linkCounter.V1ChangeV2(v1, v2); // вызов метода V2ChangeV1
}
System.out.printf("\n");
}
linkCounter.saveLinks("output/links.csv"); // вызов метода
saveLinks
linkCounter.saveGraph("output/graph.xml"); // вызов метода
saveGraph
//GraphAnalyzer.showGraph(linkCounter.getGraph()); // визуа-
лизация графа через вызов метода showGraph класса GraphAnalyzer
}
writer.close();
}
private static String safeVal(Token t) { // функция для опреде-
ления леммы для слова (токена)
Lemma l = t.getLemma();
return l != null ? l.getValue() : "";
}
}

```

СПИСОК ЛІТЕРАТУРИ

1. Бисикало О. В. Субъективная единица смысла образных конструкций / О. В. Бисикало // *Nauka: teoria i praktyka* – 2009 : materialy V miedzynar. naukow-i-praktycznej konf., (Przemysl, 7–15 sierpnia 2009). – Przemysl : Nauka i studia, 2009. – Vol. 6. – P. 9–12.
2. Бісикало О. В. Інфологічний підхід до моделювання образного мислення людини [Електронний ресурс] / О. В. Бісикало // *Вісник СумДУ (Серія «Технічні науки»)*. – 2009. – № 2. – С. 15–20. – Режим доступу: http://visnyk.sumdu.edu.ua/arhiv/2009/Tech_2_09/09bovoml.pdf.
3. Vasilevskyi O. M. Calibration method to assess the accuracy of measurement devices using the theory of uncertainty. *International Journal of Metrology and Quality Engineering*, 2014, 5.04: 403. – № 3 (7). – 2006. – P. 147–151.
4. Руководство по выражению неопределенностей измерения = Guide to the Expression of Uncertainty in Measurement : [на-

учн. редактор Слаев В. А.]. – Санкт-Петербург : НПО ВНИИМ им. Д. М. Менделеева, 1999. – 134 с.

5. Васілевський О. М. Алгоритм оцінювання невизначеності у вимірюваннях при виконанні метрологічних робіт / О. М. Васілевський // *Інформаційні технології та комп'ютерна інженерія*. – № 3 (7). – 2006. – С. 147–151.
6. Применение «Руководства по выражению неопределенности измерений» : МИ 2552-99. – Офиц. изд. – Санкт-Петербург : ВНИИМ им. Д. И. Менделеева, 1999. – 27 с.
7. Колмогоров А. Н. Основные понятия теории вероятностей / А. Н. Колмогоров. – 2-е изд. – М. : Наука, 1974. – 120 с.
8. Бісикало О. В. Формалізація понять мовного образу та образного сенсу природно-мовних конструкцій / О. В. Бісикало // *Математичні машини і системи*. – 2012. – № 2. – С. 70–73.
9. Бісикало О. В. Формальні методи образного аналізу та синтезу природно-мовних конструкцій : монографія / О. В. Бісикало. – Вінниця : БНТУ, 2013. – 316 с.
10. Раскин Л. Г. Нечеткая математика. Основы теории. Приложение / Л. Г. Раскин, О. В. Серая. – Х. : Парус, 2008. – 352 с.
11. Загальні вимоги до компетентності випробувальних та калібрувальних лабораторій : ДСТУ ISO/IEC 17025-2001. – [Чинний від 2001 – 01 – 01]. – К. : Держстандарт України, 2001. – 31 с. – (Національний стандарт України).
12. Gurevych I. Darmstadt Knowledge Processing Repository Based on UIMA [Electronic resource] / I. Gurevych, M. Muhlhauser, Ch. Muller, J. Steimle, M. Weimer, T. Zesch. – February 9, 2007. – Available at: \www\URL: https://www.ukp.tu-darmstadt.de/fileadmin/user_upload/Group_UKP/publikationen/2007/gldv-uima-ukp.pdf.
13. Бісикало О. В. Метод вилучення образних знань з англomовного тексту на основі інструментальних засобів пакету DKPro Core / О. В. Бісикало, І. Гуревич // *Контроль і управління в складних системах: XII міжнар. конф., 14–16 жовтня 2014 р.: тези доповідей*. – Вінниця, 2014. – С. 51.
14. Stanford parser [Електронний ресурс] // *Stanford Dependencies*. – Назва з екрану. – Режим доступу: <http://nlp.stanford.edu/software/stanford-dependencies.shtml>.
15. Free ebooks – Project Gutenberg [Електронний ресурс] / Project Gutenberg Literary Archive Foundation. – Режим доступу: <https://www.gutenberg.org/>.
16. Бісикало О. В. Статистичний аналіз складних залежностей у тексті / О. В. Бісикало // *Вісник Нац. ун-ту «Львівська політехніка» : Інформаційні системи та мережі*. – 2015. – № 814. – С. 228–236.

Стаття надійшла до редакції 19.11.2015.

Після доробки 25.12.2016.

Бисикало О. В.¹, Василевский А. Н.²

¹Д-р техн. наук, професор, декан факультета компьютерных систем и автоматизации Винницкого национального технического университета, Винница, Украина

²Канд. техн. наук, доцент, профессор кафедры метрологии и промышленной автоматизации Винницкого национального технического университета, Винница, Украина

ОЦЕНКА НЕОПРЕДЕЛЕННОСТИ ИЗМЕРЕНИЯ СМЫСЛА ЕСТЕСТВЕННО-ЯЗЫКОВЫХ КОНСТРУКЦИЙ

Решена задача оценки неопределенности измерения смысла естественно-языковых конструкций (ЕЯК) на основе формализации понятий лингвистического образа, искусственной когнитивной системы и единицы смысла. В основу модели базы знаний искусственной когнитивной системы заложена статистическая информация относительно ассоциативной сочетаемости лингвистических образов, что обеспечивает возможность унифицированной оценки единицы и количества смысла ЕЯК. Предложен метод измерения смысла ЕЯК на основе нечеткого отношения смысла, обеспечивающий использование информации про связи между леммами текста, что позволяет получить оценку двух типов неопределенности измерения формальных признаков смысла. Получены и интерпретированы формальные оценки неопределенности результатов измерения смысла ЕЯК, что позволяет учитывать информацию про связи между леммами для решения прикладных задач компьютерной лингвистики.

С помощью разработанного на основе пакета DKPro Core программного обеспечения проведены эксперименты с целью исследования предложенного метода в задаче определения информативных признаков текста. В результате проведенных экспериментов полу-

чены зависимости параметров обнаруженного Парето-подобного закона распределения связей между леммами, анализ которых позволяет считать показатель среднего количества связей лингвистического образа наиболее информативным численным признаком текста.

Ключевые слова: смысл, неопределенность, текст, естественно-языковая конструкция, искусственная когнитивная система, лингвистический образ, лемма.

Bisikalo O. V.¹, Vasilevskyi O. M.²

¹Dr.Sc., Professor, Dean of faculty for computer systems and automation, Vinnytsia National Technical University, Vinnytsia, Ukraine

²PhD, Associate professor, Professor of department of metrology and industrial automatics, Vinnytsia National Technical University, Vinnytsia, Ukraine

EVALUATION OF UNCERTAINTY MEASURING OF SENSE OF THE NATURAL LANGUAGE CONSTRUCTS

The task of evaluation of measurement uncertainty meaning of natural language constructs (NLC) based on formalization of the concepts of linguistic image, artificial cognitive systems and unit of sense is resolved. The basis of model the knowledge base of artificial cognitive system laid down statistical information regarding the associative compatibility of linguistic images, which enables unified evaluation the unit and the quantity of sense NLC. The method for measuring the sense of NLC based on fuzzy relation of meaning is offered. It provides to use information about the links between lemmas of text that allows you to estimate the measurement uncertainty of two types of sense signs. The results of the formal evaluation of the uncertainty of measurement sense of NLC are received and interpreted what enables into account information about the relationship between lemmas for solve tasks of computational linguistics.

With developed on the basis of the package DKPro Core Software conducted experiments to study the proposed method in the problem of the definition of informative features of the text. The experiments obtained dependence of the parameters detected Pareto-like distribution law relations between lemmas, whose analysis suggests that average number of connections of linguistic image is the most informative numerical feature for the text.

Keywords: sense, uncertainty, text, natural language construct, artificial cognitive system, linguistic image, lemma.

REFERENCES

1. Bisikalo O.V. Sub'ektivnaya edinita smysla obraznykh konstruktsiy, Nauka: teoria i praktika – 2009: materialy V miedzynar. naukowi-praktycznej konf., (Przemysl, 7–15 sierpnia 2009). Przemysl, Nauka i studia, 2009, Vol. 6, pp. 9–12
2. Bisikalo O. V. Infologichnyi pidhid do modelyuvannya obraznogo mislennya lyudini [Elektronnyi resurs], Visnik SumDU (Seriya "Tehnichni nauki"), 2009, No. 2, pp. 15–20. Rezhim dostupu: http://visnyk.sumdu.edu.ua/arhiv/2009/Tech_2_09/09bovoml.pdf.
3. Vasilevskyi O. M. Calibration method to assess the accuracy of measurement devices using the theory of uncertainty, International Journal of Metrology and Quality Engineering, 2014, 5.04: 403, No. 3 (7), 2006, pp. 147–151.
4. Rukovodstvo po vyrazheniyu neopredelennostey izmereniya = Guide to the Expression of Uncertainty in Measurement : [nauchn. redaktor Slaev V. A.]. Sankt-Peterburg, NPO VNIIM im. D. M. Mendeleeva, 1999, 134 p.
5. Vasilevskyi O. M. Algoritm otsinyuvannya neviznachenosti u vimiryuvannyakh pri vikonanni metrologichnih robit, Informatsiyni tehnologiyi ta komp'yuterna inzheneriya, 2006, No. 3 (7), pp. 147–151.
6. Primenenie «Rukovodstva po vyrazheniyu neopredelennosti izmereniy», MI 2552-99. Ofits. izd. Sankt-Peterburg, VNIIM im. D. I. Mendeleeva, 1999, 27 p.
7. Kolmogorov A. N. Osnovnyie ponyatiya teorii veroyatnostey 2-e izd. Moscow, Nauka, 1974, 120 p.
8. Bisikalo O.V. Formalizatsiya ponyat movnogo obrazu ta obraznogo sensu prirodno-movnih konstruktsiy, Matematichni mashini i sistemi, 2012, No. 2, pp. 70–73.
9. Bisikalo O.V. Formalni metodi obraznogo analizu ta sintezu prirodno-movnih konstruktsiy : monografiya. Vinnitsya, VNTU, 2013, 316 p.
10. Raskin L. G., Seraya O. V. Nechetkaya matematika. Osnovy teorii. Prilozheniya. Har'kov, Parus, 2008, 352 p.
11. Zagalni vimogi do kompetentnosti viprobuvalni kh ta kalibruvalnih laboratoriy : DSTU ISO/IEC 17025-2001. [Chinniy vid 2001 – 01 – 01]. Kiev, Derzhstandart UkraYini, 2001, 31 p. (Natsionalniy standart UkraYini).
12. Gurevych I., Muhlhauser M., Muller Ch., Steimle J., Weimer M., Zesch T. Darmstadt Knowledge Processing Repository Based on UIMA [Electronic resource]. February 9, 2007, Available at: www.ukp.tu-darmstadt.de/fileadmin/user_upload/Group_UKP/publikationen/2007/gldv-uima-ukp.pdf.
13. Bisikalo O. V., Gurevich I. Metod viluchennya obraznih znan z anglo-movnogo tekstu na osnovi instrumentalnih zasobiv paketu DKPro Core, Kontrol i upravlinnya v skladnih sistemah: XII mizhnar. konf., 14–16 zhovtnya 2014 r.: tezi dopovidey. Vinnitsya, 2014, pp. 51.
14. Stanford parser [Elektronnyi resurs]. Stanford Dependencies. Nazva z ekranu. Rezhim dostupu: <http://nlp.stanford.edu/software/stanford-dependencies.shtml>.
15. Free ebooks – Project Gutenberg [Elektronnyi resurs]. Project Gutenberg Literary Archive Foundation. Rezhim dostupu: <https://www.gutenberg.org/>.
16. Bisikalo O. V. Statistichnyi analiz skladnih zalezhnostey u teksti, Visnik Nats. un-tu «Lvivska politehnika», Informatsiyni sistemi ta merezhi, 2015, No. 814, pp. 228–236.

ЕВОЛЮЦІЙНА КАСКАДНА СИСТЕМА НА ОСНОВІ НЕЙРО-ФАЗЗИ ВУЗЛІВ

У статті запропоновано еволюційну каскадну систему обчислювального інтелекту на основі нейро-фаззи вузлів, а також алгоритми її навчання. Запропонована система в процесі навчання налаштовує не тільки свої параметри, але й архітектуру в online режимі. У якості вузлів еволюційної каскадної системи запропоновано використовувати нейро-фаззи системи. Для налаштування параметрів функцій належності нейро-фаззи вузлів запропоновано алгоритм, що ґрунтується на використанні градієнтної процедури мінімізації критерію навчання. У процесі навчання нейро-фаззи вузлів налаштовуються синаптичні ваги, центри і параметри ширини функцій належності нейро-фаззи вузлів. Розроблено програмне забезпечення, що реалізує запропоновану архітектуру еволюційної каскадної нейро-фаззи системи. Проведено ряд експериментів з дослідження властивостей запропонованої системи. Результати експериментів підтвердили те, що запропоновану систему можна використовувати для розв'язання широкого кола задач інтелектуального аналізу даних, при цьому обробка даних відбувається в online режимі. Система характеризується простотою чисельної реалізації, а процес обробки даних може бути пришвидшено завдяки можливості паралельного налаштування вузлів еволюційної каскадної системи. Характерною особливістю запропонованої системи є те, що для її налаштування не потрібний великий об'єм навчальної вибірки.

Ключові слова: гібридна система, обчислювальний інтелект, каскадна система, нейро-фаззи система, функція належності, еволюційна система.

НОМЕНКЛАТУРА

TSK – Takagi–Sugeno–Kang system;

ANFIS – adaptive neuro-fuzzy inference system;

МГУА – Метод Групового Урахування Аргументів;

МНК – метод найменших квадратів;

$x(k)$ – вектор вхідних сигналів;

$x_i(k)$ – i -та компонента вектора вхідних сигналів;

h – кількість функцій належності;

k – номер спостереження у навчальній вибірці;

n – кількість входів системи;

N – кількість спостережень у навчальній вибірці;

S – кількість вузлів у першому прихованому шарі системи;

SB – блок селекції;

$N^{[l]}$ – i -й прихований шар системи;

$\hat{y}_j^{[1]}(k)$ – вихідний сигнал j -го вузла першого прихованого шару для k -го вектора вхідних сигналів;

$\hat{y}_j^{[1]*}(k)$ – вихідний сигнал блоку селекції для j -го вузла першого прихованого шару для k -го вектора вхідних сигналів;

$\hat{y}^{[i]}(k)$ – вихідний сигнал i -го прихованого шару системи;

$\hat{y}(k)$ – вихідний сигнал вузла для k -го вектора вхідних сигналів;

$u(k)$ – навчальний сигнал для k -го вектора вхідних сигналів;

μ_{li} – функція належності i -ї компоненти вектора вхідних сигналів до l -го центроїда;

C_{li} – параметр, що визначає центр l -ї функції належності i -ї компоненти;

σ_{li} – параметр ширини l -ї функції належності i -ї компоненти;

$\tilde{x}_l(k)$ – добуток значень l -ї функції належності для k -го вектора вхідних сигналів;

w_l^{ij} – синаптична вага добутку значень l -ї функції належності для вузла з i -ю та j -ю компонентами;

w^{ij} – вектор синаптичних ваг для вузла з i -ю та j -ю компонентами;

$E(k)$ – значення функції критерію навчання для k -го вектора вхідних сигналів;

η_c – параметр кроку навчання для параметра центра функції належності;

η_σ – параметр кроку навчання для параметра ширини функції належності;

$\exp(x)$ – експонента;

α – параметр забування;

δ_{lr} – дельта-функція.

ВСТУП

У наш час гібридні системи обчислювального інтелекту і, перш за все, штучні нейронні мережі та нейро-фаззи системи отримали широке розповсюдження для вирішення різних задач інтелектуального аналізу даних: розпізнавання образів, ідентифікації, емуляції, інтелектуального керування тощо. При цьому на практиці достатньо часто виникають ситуації, коли дані, що підлягають обробці, надходять до входу системи послідовно в online режимі. Така ситуація є предметом розгляду напрямків, що інтенсивно розвиваються в наш час і відомі як Dynamic Data Mining та Data Stream Mining [1], а для роботи в цих умовах найкраще пристосовані так звані «еволюційні системи обчислювального інтелекту» [2–6], які

у процесі навчання налаштовують не тільки свої параметри, але й архітектуру.

Треба відзначити, що процес навчання суттєво ускладнюється, коли навчальні вибірки містять невелику кількість спостережень, що призводить до того, що система або не встигає налаштувати свої параметри (не кажучи про архітектуру), або виникає ефект «перенавчання» (overfitting).

У зв'язку з цим синтез гібридних систем обчислювального інтелекту, що навчаються в online режимі та здатні змінювати свою архітектуру у процесі навчання, є досить цікавою та актуальною задачею.

1 ПОСТАНОВКА ЗАДАЧІ

Нехай задана вибірка даних, кожне спостереження з якої має вигляд $(n \times 1)$ -вимірний вектора сигналів $x(k) = (x_1(k), x_2(k), \dots, x_n(k))^T$ ($k = 1, 2, \dots, N$ – номер спостереження у навчальній вибірці).

Тоді задача дослідження полягає в навчанні еволюційної каскадної нейро-фаззі системи для обробки вхідних даних і отриманні вихідних значень для розв'язку певної задачі інтелектуального аналізу даних.

Результатом роботи запропонованої системи є вихідний сигнал $\hat{y}(k)$ для кожного спостереження з вибірки даних.

У якості цільової функції використовується середньоквадратична помилка $E(k) = \frac{1}{N} \sum_{k=1}^N (y(k) - \hat{y}(k))^2$.

2 ОГЛЯД ЛІТЕРАТУРИ

На сьогоднішній день існує багато еволюційних систем обчислювального інтелекту, що здатні обробляти дані в online режимі. В основі більшості з них лежать багатошарові нейро-фаззі системи типу TSK або ANFIS. Але у випадках, коли навчальні вибірки містять невелику кількість спостережень, ці системи можуть не встигнути налаштувати свої параметри та архітектуру. Перебороти вказані труднощі можна, декомпозувавши вихідну проблему на ряд задач меншої розмірності, для вирішення яких кількість наявних даних є достатньою.

Одним з найбільш ефективних підходів, що ґрунтуються на цьому принципі, є Метод Групового Урахування Аргументів (МГУА) [7–10], який, на жаль, погано пристосований до роботи в online режимі. Набагато простіше адаптувати до цього режиму так звані «каскадні нейронні мережі» [11–14], хоча до суттєвих недоліків цих систем можна віднести те, що в них використовується «заморожування» (freezing) параметрів.

Спроба створення «гібрида» МГУА і каскадних нейронних мереж, що адаптований до роботи в online режимі, була здійснена в [15], де в якості вузлів були використані складені R нейрони. Ця мережа, що володіє високою швидкістю та апроксимуючими можливостями, налаштовує всі свої параметри і архітектуру в online режимі, однак при цьому потребує достатньо великої за об'ємом навчальні вибірки.

У цій роботі зроблено спробу синтезу еволюційної каскадної системи обчислювального інтелекту, яка призначена для обробки потоку даних, що надходять в online режимі, і яка має суттєво меншу в порівнянні з відомими аналогами кількість параметрів, що налаштовуються.

3 МАТЕРІАЛИ І МЕТОДИ

На рис. 1 наведено архітектуру еволюційної каскадної системи.

До нульового шару системи надходить $(n \times 1)$ -вимірний вектор вхідних сигналів $x(k) = (x_1(k), x_2(k), \dots, x_n(k))^T$ (тут $k = 1, 2, \dots, N$ – номер спостереження у навчальній вибірці або індекс поточного дискретного часу), який далі передається на перший прихований шар, що містить c_n^2 вузлів-нейронів, кожен з яких має два входи. На виходах вузлів $N^{[1]}$ першого прихованого шару формуються вихідні сигнали $\hat{y}_s^{[1]}(k)$, $s = 1, 2, \dots, 0,5n(n-1) = c_n^2$. Далі ці сигнали надходять до блоку селекції SB , який виконує сортування вузлів першого прихованого шару за прийнятим критерієм (найчастіше за зростанням значення середнього квадрата помилки). Виходи блоку селекції $\hat{y}_1^{[1]*}(k)$ і

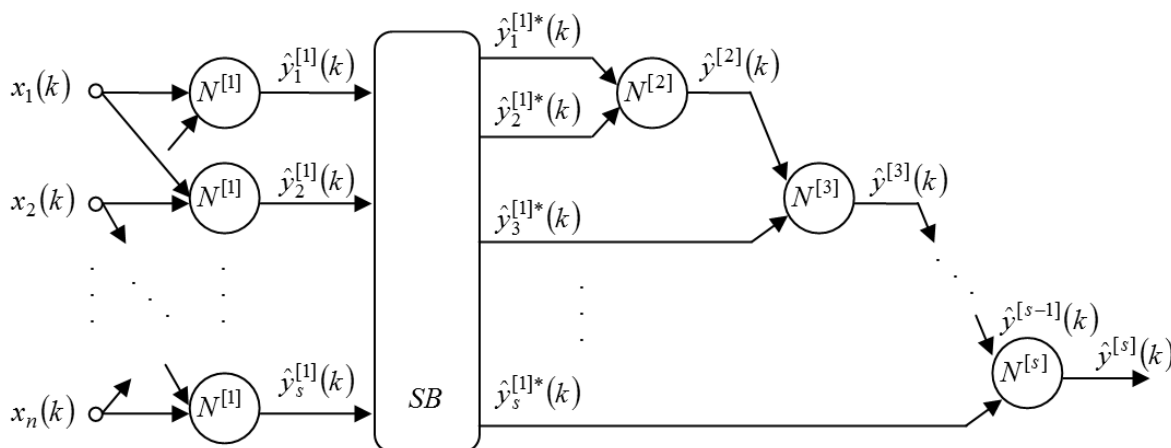


Рисунок 1 – Архітектура еволюційної каскадної системи

$\hat{y}_2^{[1]*}(k)$ надходять до входу єдиного вузла-нейрона другого шару, на виході якого формується вихідний сигнал $\hat{y}^{[2]}(k)$. Цей вихідний сигнал разом з вихідним сигналом блока селекції $\hat{y}_3^{[1]*}(k)$ надходить до входу вузла-нейрона наступного шару. Процес нарощування каскадів триває до досягнення необхідної точності. Варто зауважити, що всі нейрони системи налаштовуються незалежно один від одного. Завдяки цьому об'єм необхідної навчальної вибірки можна скоротити.

У якості вузла каскадної мережі запропоновано використовувати систему, архітектура якої наведена на рис. 2.

Цей вузол є по суті нейро-фаззі системою Ванга-Менделя [16, 17]. На вхід вузла подається двовимірний вектор вхідних сигналів $x(k) = (x_i(k), x_j(k))^T$, що підлягають обробці. Перший шар вузла реалізує фаззифікацію вхідних змінних. Він містить $2h$ функцій належності $\mu_{li}(x_i(k))$, $\mu_{lj}(x_j(k))$, $l = 1, 2, \dots, h$. Елементи першого шару обчислюють рівні належності $0 < \mu_{li}(x_i(k)) \leq 1$, $0 < \mu_{lj}(x_j(k)) \leq 1$. У якості функцій належності найчастіше використовуються гаусіани

$$\mu_{li}(x_i(k)) = \exp\left(-\frac{(x_i(k) - c_{li}(k))^2}{2\sigma_{li}^2(k)}\right), \quad (1)$$

де $c_{li}(k)$ – параметр, що визначає центр функції належності, $\sigma_{li}(k)$ – параметр ширини функції належності.

Другий шар забезпечує агрегування рівнів належності, що обчислюються в першому шарі. На виходах другого шару з'являється h агрегованих сигналів $\tilde{x}_l(k) = \mu_{li}(x_i(k))\mu_{lj}(x_j(k))$.

Третій шар – це шар синаптичних ваг, що підлягають налаштуванню у процесі навчання. Виходами третього шару є значення

$$w_l^{ij}\mu_{li}(x_i(k))\mu_{lj}(x_j(k)) = w_l^{ij}\tilde{x}_l(k).$$

Четвертий шар утворений двома суматорами і обчислює суми вихідних сигналів другого та третього шарів. Виходами четвертого шару є значення

$$\begin{cases} \sum_{l=1}^h w_l^{ij}\mu_{li}(x_i(k))\mu_{lj}(x_j(k)) = \sum_{l=1}^h w_l^{ij}\tilde{x}_l(k), \\ \sum_{l=1}^h \mu_{li}(x_i(k))\mu_{lj}(x_j(k)) = \sum_{l=1}^h \tilde{x}_l(k). \end{cases}$$

У п'ятому (вихідному) шарі нейрона відбувається нормалізація, в результаті якої обчислюється вихідний сигнал вузла $\hat{y}_s^{[1]}(k)$:

$$\begin{aligned} \hat{y}_s^{[1]}(k) &= \frac{\sum_{l=1}^h w_l^{ij}\mu_{li}(x_i(k))\mu_{lj}(x_j(k))}{\sum_{p=1}^h \mu_{pi}(x_i(k))\mu_{pj}(x_j(k))} = \frac{\sum_{l=1}^h w_l^{ij}\tilde{x}_l(k)}{\sum_{p=1}^h \tilde{x}_p(k)} = \\ &= \sum_{l=1}^h w_l^{ij} \frac{\tilde{x}_l(k)}{\sum_{p=1}^h \tilde{x}_p(k)} = \sum_{l=1}^h w_l^{ij} \phi_l^{ij}(x(k)) = (w^{ij})^T \phi^{ij}(x(k)), \end{aligned}$$

$$\text{де } \phi_l^{ij}(x(k)) = \frac{\tilde{x}_l(k)}{\sum_{p=1}^h \tilde{x}_p(k)} =$$

$$= \mu_{li}(x_i(k))\mu_{lj}(x_j(k)) \left(\sum_{p=1}^h \mu_{pi}(x_i(k))\mu_{pj}(x_j(k)) \right)^{-1},$$

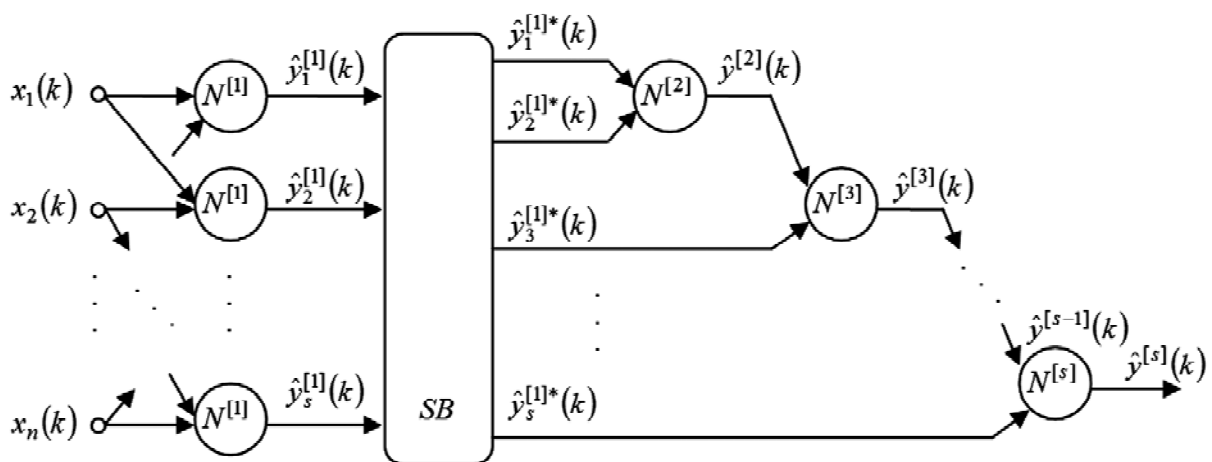


Рисунок 2 – Вузол каскадної мережі

$$w^{ij} = (w_1^{ij}, \dots, w_h^{ij})^T, \quad \varphi^{ij}(x(k)) = (\varphi_1^{ij}(x(k)), \dots, \varphi_h^{ij}(x(k)))^T.$$

Аналогічним чином виконуються обчислення в інших вузлах прихованих шарів. Такий вузол системи подібно до нормалізованої радіально-базисної нейронної мережі реалізує нелінійне відображення вхідних сигналів у вихідні, але містить меншу кількість параметрів, що налаштовуються.

Оскільки вихідний сигнал кожного з вузлів системи $\hat{y}_i^{[1]*}(k)$ лінійно залежить від синаптичних ваг $w_l^{ij}(k)$, для їх налаштування може бути використаний стандартний метод найменших квадратів (МНК). Якщо дані надходять на обробку послідовно в online режимі, то використовується рекурентна форма МНК

$$\begin{cases} w^{ij}(k+1) = w^{ij}(k) + \frac{P^{ij}(k)(y(k) - (w^{ij}(k))^T \varphi^{ij}(x(k))) \varphi^{ij}(x(k))}{\alpha + (\varphi^{ij}(x(k)))^T P^{ij}(k) \varphi^{ij}(x(k))}, \\ P(k+1) = \frac{1}{\alpha} \left(P^{ij}(k) - \frac{P^{ij}(k) \varphi^{ij}(x(k)) (\varphi^{ij}(x(k)))^T P^{ij}(k)}{\alpha + (\varphi^{ij}(x(k)))^T P^{ij}(k) \varphi^{ij}(x(k))} \right), \quad 0 < \alpha \leq 1. \end{cases}$$

Також для налаштування синаптичних ваг може бути використаний оптимальний за швидкістю градієнтний однокроковий алгоритм Качмажа-Уїдрю-Гоффа у вигляді

$$w^{ij}(k+1) = w^{ij}(k) + \frac{y(k) - (w^{ij}(k))^T \varphi^{ij}(x(k))}{\|\varphi^{ij}(x(k))\|^2} \varphi^{ij}(x(k)).$$

Для налаштування параметрів центрів і ширини функцій належності можна використати градієнтні процедури мінімізації критерію навчання

$$E_s^{[1]}(k) = \frac{1}{2} (y(k) - \hat{y}_s^{[1]}(k))^2 = \frac{1}{2} (y(k) - (w^{ij}(k))^T \varphi^{ij}(x(k)))^2. \quad (2)$$

Для спрощення розрахунку похідних і прискорення обчислення значення функції належності можна ввести позначення

$$\tilde{\sigma}_{ri}^2(k) = -0,5 \sigma_{ri}^{-2}(k).$$

При застосуванні методу найшвидшого спуску відповідні формули адаптації у загальному випадку для $(n \times 1)$ -вимірного вектора вхідних сигналів набувають вигляду

$$\begin{cases} c_{ri}(k+1) = c_{ri}(k) - \eta_c \frac{\partial E_s^{[1]}(k)}{\partial c_{ri}}, \\ \tilde{\sigma}_{ri}^2(k+1) = \tilde{\sigma}_{ri}^2(k) - \eta_\sigma \frac{\partial E_s^{[1]}(k)}{\partial \tilde{\sigma}_{ri}^2}, \end{cases}$$

де η_c , η_σ – параметри кроку навчання для параметрів центра і ширини відповідно; $r = 1, 2, \dots, h$; $i = 1, 2, \dots, n$.

При використанні в якості функцій належності традиційних гаусіанів (1) відповідні формули градієнта цільової функції (2) для однієї пари $(x(k), y(k))$ навчальних даних набувають вигляду

$$\begin{cases} \frac{\partial E_s^{[1]}(k)}{\partial c_{ri}} = \left((w^{ij}(k))^T \varphi^{ij}(x(k)) - y(k) \right) \sum_{l=1}^h w_l^{ij}(k) \frac{\partial \varphi_l^{ij}(x(k))}{\partial c_{ri}}, \\ \frac{\partial E_s^{[1]}(k)}{\partial \tilde{\sigma}_{ri}^2} = \left((w^{ij}(k))^T \varphi^{ij}(x(k)) - y(k) \right) \sum_{l=1}^h w_l^{ij}(k) \frac{\partial \varphi_l^{ij}(x(k))}{\partial \tilde{\sigma}_{ri}^2}, \end{cases}$$

де

$$\varphi_l^{ij}(x(k)) = \frac{\mu_{li}(x_i(k)) \mu_{lj}(x_j(k))}{\sum_{p=1}^h \mu_{pi}(x_i(k)) \mu_{pj}(x_j(k))}. \quad (3)$$

Похідні $\frac{\partial \varphi_l^{ij}(x(k))}{\partial c_{ri}}$ і $\frac{\partial \varphi_l^{ij}(x(k))}{\partial \tilde{\sigma}_{ri}^2}$, визначені на ос-

нові (3), можна записати як

$$\begin{cases} \frac{\partial \varphi_l^{ij}(x(k))}{\partial c_{ri}} = \frac{\delta_{lr} \sum_{p=1}^h \mu_{pi}(x_i(k)) \mu_{pj}(x_j(k)) - \mu_{li}(x_i(k)) \mu_{lj}(x_j(k))}{\left(\sum_{p=1}^h \mu_{pi}(x_i(k)) \mu_{pj}(x_j(k)) \right)^2} \times \\ \times \mu_{rj}(x_j(k)) \frac{\partial \mu_{ri}(x_i(k))}{\partial c_{ri}}, \\ \frac{\partial \varphi_l^{ij}(x(k))}{\partial \tilde{\sigma}_{ri}^2} = \frac{\delta_{lr} \sum_{p=1}^h \mu_{pi}(x_i(k)) \mu_{pj}(x_j(k)) - \mu_{li}(x_i(k)) \mu_{lj}(x_j(k))}{\left(\sum_{p=1}^h \mu_{pi}(x_i(k)) \mu_{pj}(x_j(k)) \right)^2} \times \\ \times \mu_{rj}(x_j(k)) \frac{\partial \mu_{ri}(x_i(k))}{\partial \tilde{\sigma}_{ri}^2}, \end{cases}$$

де δ_{lr} – дельта-функція. Похідні $\frac{\partial \mu_{ri}(x_i(k))}{\partial c_{ri}}$ і

$\frac{\partial \mu_{ri}(x_i(k))}{\partial \tilde{\sigma}_{ri}^2}$, визначені на основі (1), можна записати як

$$\begin{cases} \frac{\partial \mu_{ri}(x_i(k))}{\partial c_{ri}} = \frac{x_i(k) - c_{ri}(k)}{\sigma_{ri}^2(k)} \exp \left(-\frac{(x_i(k) - c_{ri}(k))^2}{2\sigma_{ri}^2(k)} \right), \\ \frac{\partial \mu_{ri}(x_i(k))}{\partial \tilde{\sigma}_{ri}^2} = (x_i(k) - c_{ri}(k))^2 \exp \left(-\frac{(x_i(k) - c_{ri}(k))^2}{2\sigma_{ri}^2(k)} \right). \end{cases}$$

Таким чином, виконується паралельне налаштування всіх синаптичних ваг, параметрів центрів та ширини функцій належності вузлів першого прихованого шару системи. Вузли наступних шарів, що утворюють каскадну структуру, налаштовуються аналогічно нейронам першого прихованого шару з тією лише різницею, що їхніми входами є вихідні сигнали $\hat{y}_i^{[1]*}(k)$ блоку селекції SB та попередніх каскадів, а навчальний сигнал $y(k)$ є спільним для всіх елементів еволюційної каскадної системи.

4 ЕКСПЕРИМЕНТИ

Ефективність запропонованої еволюційної каскадної нейро-фаззі мережі та її процедур навчання була продемонстрована при розв'язанні задачі прогнозування нестационарних сигналів.

У якості тестової вибірки було обрано часовий ряд, що описує щомісячний тиск над рівнем моря за період з 1882 по 1998 рік (Darwin sea level pressure, вибірка отримана зі сховища даних DataMarket). Загальний об'єм вибірки – 1400 спостережень. Навчальна вибірка містила 1100 спостережень, перевірна – 300 спостережень. Для порівняння ефективності запропонованої системи з ефективністю інших існуючих систем було обрано багатошаровий перцептрон, радіально-базисну нейронну мережу та ANFIS.

У якості критерію якості прогнозування було обрано середньоквадратичну помилку (MSE).

На рис. 3 наведено фазовий портрет ряду, що прогнозується.

5 РЕЗУЛЬТАТИ

Результати проведених експериментів наведено в табл. 1.

6 ОБГОВОРЕННЯ

Як можна побачити з табл. 1, запропонована еволюційна каскадна система досить непогано показала себе при розв'язанні задачі прогнозування. Варто відзначити, що час навчання запропонованої системи був одним з найкращих. Якщо порівнювати якість прогнозів, еволюційна каскадна система продемонструвала найкращий результат.

Результат прогнозування для перевірконої вибірки наведений на рис. 4. Суцільною чорною лінією показаний ряд, що прогнозується, пунктирною – прогноз системи, суцільною сірою – значення помилки.

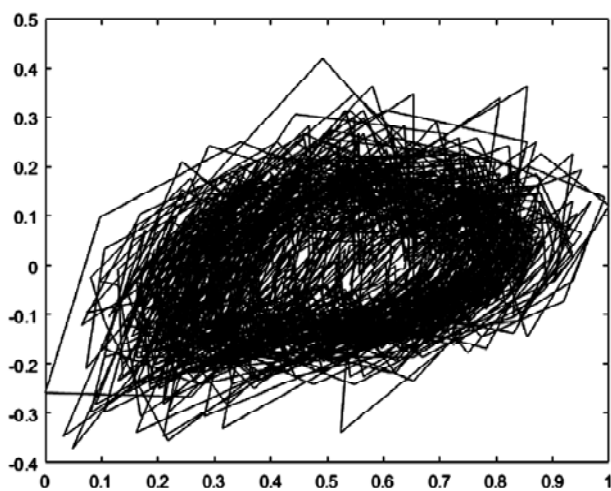


Рисунок 3 – Фазовий портрет ряду

Таблиця 1 – Результати експериментів

Тип системи	Помилка навчання	Помилка тестування	Час навчання, с
Багатошаровий перцептрон	0,0150	0,0168	0,2500
Радіально-базисна нейронна мережа	0,0172	0,0224	0,2391
ANFIS	0,0157	0,0165	0,2031
Еволюційна каскадна система	0,0146	0,0156	0,2067

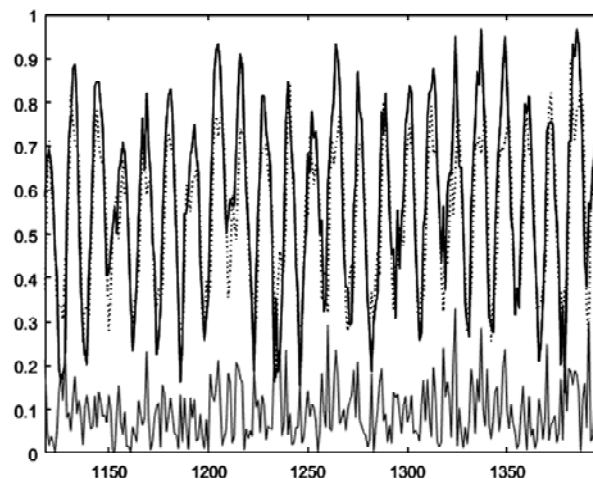


Рисунок 4 – Результати прогнозування ряду

ВИСНОВКИ

Наукова новизна роботи полягає у тому, що запропоновано еволюційну каскадну нейро-фаззі систему та online алгоритм навчання всіх її параметрів, характерною особливістю якої є те, що для свого налаштування вона не потребує великих об'ємів навчальної вибірки. В основі системи лежать ідеї Метода Групового Урахування Аргументів та еволюційних нейро-фаззі систем, що дозволяє знайти не тільки оптимальні значення параметрів, але й отримати найкращу архітектуру для кожної конкретної задачі. Система характеризується простотою чисельної реалізації, а можливість паралельного налаштування її вузлів збільшує швидкість процесу обробки даних.

До практичної цінності отриманих результатів можна віднести те, що запропоновану систему можна використовувати для розв'язання широкого кола задач інтелектуального аналізу даних.

ПОДЯКИ

Роботу виконано в рамках держбюджетної науково-дослідної теми Харківського національного університету радіоелектроніки № 307 «Динамічний інтелектуальний аналіз послідовностей нечіткої інформації за умов суттєвої невизначеності на основі гібридних систем обчислювального інтелекту».

СПИСОК ЛІТЕРАТУРИ

1. Bifet A. Adaptive Stream Mining: Pattern Learning and Mining from Evolving Data Streams / A. Bifet. – IOS Press, 2010. – 224 p.
2. Kasabov N. Evolving fuzzy neural networks: theory and applications for on-line adaptive prediction, decision making and control / N. Kasabov // Australian Journal of Intelligent Information Processing Systems. – 1998. – Vol. 5, Issue 3. – P. 154–160.
3. Kasabov N. Evolving fuzzy neural networks for on-line supervised/unsupervised, knowledge-based learning / N. Kasabov // IEEE Transactions on Man, Machine, and Cybernetics. – 2001. – Vol. 31, Issue 6. – P. 902–918.
4. Kasabov N. Evolving Connectionist Systems / N. Kasabov. – London: Springer-Verlag, 2003. – 307 p.
5. Lughofer E. Evolving Fuzzy Systems – Methodologies, Advanced Concepts and Applications / E. Lughofer. – Berlin: Springer, 2011. – 410 p.
6. Ivakhnenko A. G. Polynomial theory of complex systems / A. G. Ivakhnenko // IEEE Transactions on Systems, Man, and Cybernetics. – 1971. – Vol. 1, Issue 4. – P. 364–378.
7. Ивахненко А. Г. Самообучающиеся системы распознавания и автоматического управления / А. Г. Ивахненко. – Киев: Техніка, 1969. – 392 с.
8. Ивахненко А. Г. Долгосрочное прогнозирование и управление сложными системами / А. Г. Ивахненко. – Киев: Техніка, 1975. – 311 с.
9. Ивахненко А. Г. Помехоустойчивость моделирования / А. Г. Ивахненко, В. С. Степанюк. – Киев: Наукова думка, 1985. – 216 с.

10. Fahlman S. E. The cascade-correlation learning architecture / S. E. Fahlman, C. Lebiere // *NIPS 1989 : Neural Information Processing Systems Conference*, Denver, 27-30 November 1989 : proceedings. – San Mateo, CA : Morgan Kaufman, 1990. – P. 524–532.
11. Prechelt L. Investigation of the CasCor family of learning algorithms / L. Prechelt // *Neural Networks*. – 1997. – Vol. 10, Issue 5. – P. 885–896.
12. Bodyanskiy Ye. Hybrid cascade neural network based on wavelet-neuron / Ye. Bodyanskiy, O. Kharchenko, O. Vynokurova // *Information Theories and Applications*. – 2011. – Vol. 18, Issue 4. – P. 335–343.
13. Bodyanskiy Ye. Evolving cascade neural network based on multidimensional Epanechnikov's kernels and its learning algorithm / Ye. Bodyanskiy, P. Grimm, N. Teslenko // *Information Technologies and Knowledge*. – 2011. – Vol. 5, Issue 1. – P. 25–30.
14. Wang L.-X. Fuzzy basis functions, universal approximation, and orthogonal least-squares learning / L.-X. Wang, J. M. Mendel // *IEEE Transactions on Neural Networks*. – 1992. – Vol. 3, Issue 5. – P. 807–814.
15. Wang L.-X. Adaptive Fuzzy Systems and Control. Design and Stability Analysis / L. X. Wang. – Upper Saddle River : Prentice Hall, 1994. – 256 p.

Стаття надійшла до редакції 29.03.2016.

Після доробки 11.04.2016.

Бодянский Е. В.¹, Тищенко А. К.², Бойко Е. А.³

¹Д-р техн. наук, научный руководитель Проблемной НИЛ АСУ, Харьковский национальный университет радиоэлектроники, Харьков, Украина

²Канд. техн. наук, старший научный сотрудник Проблемной НИЛ АСУ, Харьковский национальный университет радиоэлектроники, Харьков, Украина

³Аспирантка кафедры искусственного интеллекта Харьковского национального университета радиоэлектроники, Харьков, Украина

ЭВОЛЮЦИОНИРУЮЩАЯ КАСКАДНАЯ СИСТЕМА НА ОСНОВЕ НЕЙРО-ФАЗЗИ УЗЛОВ

В статье предложена эволюционирующая каскадная система вычислительного интеллекта на основе нейро-фаззи узлов, а также алгоритмы её обучения. Предложенная система в процессе обучения настраивает не только свои параметры, но и архитектуру в online режиме. В качестве узлов эволюционирующей каскадной системы предлагается использовать нейро-фаззи системы. Для настройки параметров функций принадлежности нейро-фаззи узлов предложен алгоритм, основанный на использовании градиентной процедуры минимизации критерия обучения. В процессе обучения нейро-фаззи узлов настраиваются синаптические веса, центры и параметры ширины функций принадлежности нейро-фаззи узлов. Разработано программное обеспечение, которое реализует предложенную архитектуру эволюционирующей каскадной нейро-фаззи системы. Проведен ряд экспериментов по исследованию свойств предложенной системы. Результаты экспериментов подтвердили то, что предложенную систему можно использовать для решения широкого круга задач интеллектуального анализа данных, при этом обработка данных происходит в online режиме. Система характеризуется простотой численной реализации, а процесс обработки данных ускоряется благодаря возможности параллельной настройки узлов эволюционирующей каскадной системы. Отличительной особенностью предложенной системы является то, что для её настройки не требуется большой объем обучающей выборки.

Ключевые слова: гибридная система, вычислительный интеллект, каскадная система, нейро-фаззи система, функция принадлежности, эволюционирующая система.

Bodyanskiy Ye. V.¹, Tyshchenko O. K.², Boiko O. O.³

¹Dr.Sc., Scientific Head of the Control Systems Research Laboratory, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine

²Ph.D., Senior Researcher at the Control Systems Research Laboratory, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine

³Ph.D. student at the Artificial Intelligence Department, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine

AN EVOLVING CASCADE SYSTEM BASED ON NEURO-FUZZY NODES

An evolving cascade system based on fuzzy-neurons and its learning procedures are proposed in the paper. During a learning procedure in an online mode, the proposed system tunes both its parameters and its architecture. Neuro-fuzzy systems are proposed as nodes of the evolving cascade system. A method based on the gradient procedure of a learning criterion minimization is proposed for membership functions' tuning in the neuro-fuzzy nodes. Synaptic weights, centers and width parameters of the membership functions are tuned during the learning procedure. Software that implements the proposed evolving cascade neuro-fuzzy system's architecture has been developed. A number of experiments has been held in order to research the proposed system's properties. Experimental results have proven the fact that the proposed system could be used to solve a wide range of Data Mining tasks. Data sets are processed in an online mode. The proposed system provides computational simplicity, and data sets are processed faster due to the possibility of parallel tuning for the evolving cascade system. A distinguishing feature of the proposed system is that there is no need of a large training set for the system to be tuned.

Keywords: hybrid system, Computational Intelligence, cascade system, neuro-fuzzy system, membership function, evolving system.

REFERENCES

1. Bifet A. Adaptive Stream Mining: Pattern Learning and Mining from Evolving Data Streams. IOS Press, 2010, 224 p.
2. Kasabov N. Evolving fuzzy neural networks : theory and applications for on-line adaptive prediction, decision making and control, *Australian Journal of Intelligent Information Processing Systems*, 1998, Vol. 5, Issue 3, pp. 154–160.
3. Kasabov N. Evolving fuzzy neural networks for on-line supervised/unsupervised, knowledge-based learning, *IEEE Transactions on Man, Machine, and Cybernetics*, 2001, Vol. 31, Issue 6, pp. 902–918.
4. Kasabov N. Evolving Connectionist Systems. London, Springer-Verlag, 2003, 307 p.
5. Lughofer E. Evolving Fuzzy Systems – Methodologies, Advanced Concepts and Applications. Berlin, Springer, 2011, 410 p.
6. Ivakhnenko A. G. Polynomial theory of complex systems, *IEEE Transactions on Systems, Man, and Cybernetics*, 1971, Vol. 1, Issue 4, pp. 364–378.
7. Ivakhnenko A. G. Samoobuchayushiesya sistemy raspoznavaniya i avtomaticheskogo upravleniya. Kiev, Technika, 1969, 392 p.
8. Ivakhnenko A. G. Dolgosrochnoe prognozirovanie i upravlenie slozhnyimi sistemami. Kiev, Technika, 1975, 311 p.
9. Ivakhnenko A. G., Stepashko V. S. Pomexoustojchivost' modelirovaniya. Kiev, Naukova dumka, 1985, 216 p.
10. Fahlman S. E., Lebiere C. The cascade-correlation learning architecture, *NIPS 1989 : Neural Information Processing Systems Conference, Denver, 27–30 November 1989 : proceedings*. San Mateo, CA, Morgan Kaufman, 1990, pp. 524–532.
11. Prechelt L. Investigation of the CasCor family of learning algorithms, *Neural Networks*, 1997, Vol. 10, Issue 5, pp. 885–896.
12. Bodyanskiy Ye., Kharchenko O., Vynokurova O. Hybrid cascade neural network based on wavelet-neuron, *Information Theories and Applications*, 2011, Vol. 18, Issue 4, pp. 335–343.
13. Bodyanskiy Ye., Grimm P., Teslenko N. Evolving cascade neural network based on multidimensional Epanechnikov's kernels and its learning algorithm, *Information Technologies and Knowledge*, 2011, Vol. 5, Issue 1, pp. 25–30.
14. Wang L.-X., Mendel J. M. Fuzzy basis functions, universal approximation, and orthogonal least-squares learning, *IEEE Transactions on Neural Networks*, 1992, Vol. 3, Issue 5, pp. 807–814.
15. Wang L.-X. Adaptive Fuzzy Systems and Control. Design and Stability Analysis. Upper Saddle River, Prentice Hall, 1994, 256 p.

РАСЧЕТ И АНАЛИЗ ПОВЕДЕНИЯ НЕПОЛНОСТЬЮ ОПРЕДЕЛЕННЫХ ФУНКЦИЙ МЕТОДОМ ДЕТЕРМИНИЗАЦИИ

Рассмотрены существующие подходы к расчету, анализу, синтезу и оптимизации систем в условиях неопределенности. Исследование неопределенных систем формулируется в виде задач расчета, анализа и синтеза различных функций с недетерминированными параметрами, служащих соответствующими характеристиками данных систем. Все эти задачи значительно сложнее их детерминированных аналогов, которые приходится решать при исследовании систем с детерминированными (точно известными) параметрами. Такое усложнение связано с тем, что алгебра недетерминированных чисел сложнее алгебры детерминированных чисел. В статье сформулирована и подробно описана задача вычисления и анализа поведения неполностью определенной функции, заданной с точностью до интервала значений. Для решения указанной задачи предложен алгоритм детерминизации, который позволяет свести задачу к двум аналогичным – для верхней и нижней граничных функций исходной неполностью определенной функции. В этом алгоритме автором использован аппарат интервальной математики и интервально-дифференциального исчисления. Далее выделены различные типы возможного поведения интервальных функций (постоянство, возрастание, убывание, расширение, сужение) и различные типы экстремальных точек таких функций (например, точка максимума, точка минимума, точка максимального расширения, точка минимального расширения). Доказаны теоремы, которые позволяют определять участки различного поведения интервальных функций и точки с различными видами экстремума. Подробно рассмотрена работа предложенного алгоритма детерминизации, позволяющего анализировать поведение интервальных функций. Работа проиллюстрирована на конкретном примере.

Ключевые слова: оптимизация, детерминированная функция, интервальная функция, анализ поведения функций, неопределенность.

НОМЕНКЛАТУРА

$x_1, x_2, \dots$ – вещественные независимые переменные;

y_1, y_2 – вещественные зависимые переменные;

X – множество независимых переменных;

Y – множество зависимых переменных;

a_i, b_i – вещественные числа;

$f_1, f_2, \dots$ – вещественные детерминированные функции;

$\tilde{f} = [f_1, f_2]$ – интервальная функция;

$\tilde{a} = [a_1, a_2], \tilde{b} = [b_1, b_2]$ – интервальные числа;

$\vee$ – символ дизъюнкции (взятия максимума);

$\wedge$ – символ конъюнкции (взятия минимума).

ВВЕДЕНИЕ

Современная наука и практика обработки информации успешно справляется с задачами исследования различных систем с полностью определенными (детерминированными) параметрами. Эти задачи обычно формулируются как задачи расчета, анализа и синтеза тех или иных функций с детерминированными параметрами, служащих соответствующими характеристиками изучаемых систем. Но на практике часто встречаются другие системы – системы с неточно известными, т.е. неполностью определенными (недетерминированными) параметрами. Причины появления таких систем заключаются в естественной неопределенности, свойственной многим реальным процессам, происходящим в системах; в неточном задании параметров большинства систем из-за неизбежных погрешностей при их вычислении или измерении; в изменении во времени параметров систем; в необходимости или целесообразности совместного исследования целых семейств однотипных сис-

тем, имеющих одинаковые функции-характеристики и различающихся лишь значениями параметров этих функций. Учет неопределенности систем особенно важен при их проектировании, поскольку полная определенность в работе системы появляется лишь на последних этапах ее создания.

Исследование введенных неопределенных систем формулируется в виде задач расчета, анализа и синтеза различных функций с недетерминированными параметрами, служащих соответствующими характеристиками данных систем. Все эти задачи значительно сложнее их вышеупомянутых детерминированных аналогов, которые приходится решать при исследовании систем с детерминированными параметрами. Усложнение связано с тем, что алгебра недетерминированных чисел сложнее алгебры детерминированных чисел.

В настоящей статье рассматриваются задачи расчета и анализа неточно заданных (недетерминированных) функций интервального типа. В качестве математического аппарата используется интервальная алгебра и интервально-дифференциальное исчисление.

1 ПОСТАНОВКА ЗАДАЧИ

Рассмотрим обычную (детерминированную) функцию одной независимой переменной

$$y = f(x), \quad (1)$$

однозначно отображающую заданное множество $X = \{x\}$ независимых переменных x в заданное множество $Y = \{y\}$ зависимых переменных y в соответствии с некоторым законом f , который и называется функцией. Хорошо известно, что задача расчета (вычисления значений) функции (1) решается с помощью адекватного

этой задаче математического аппарата алгебры вещественных чисел, при использовании подходящих методов вычисления, а задача анализа поведения функции (1) – с помощью адекватного ей аппарата классического дифференциального исчисления [1].

Далее, рассмотрим недетерминированную (конкретно, интервальную) функцию одной независимой переменной [2]

$$\tilde{y} = \tilde{f}(x), \quad (2)$$

однозначно отображающую заданное множество $X = \{x\}$ независимых вещественных (как и в (1)) переменных x в заданное множество $\tilde{Y} = \{\tilde{y}\}$ зависимых переменных-интервалов $\tilde{y} = [y_1, y_2]$, в соответствии с законом $\tilde{f}$, который и называется интервальной функцией. По определению (2), любую интервальную функцию можно представить в виде пары обычных функций

$$\tilde{f} = \{f_1, f_2\}, \quad (3)$$

которые имеют вид

$$y_1 = f_1(x), y_2 = f_2(x). \quad (4)$$

Из соотношений (3), (4) видно, что интервальная функция $\tilde{f}$ эквивалентна паре обычных функций f_1, f_2 , из которых первая однозначно отображает заданное множество $X = \{x\}$ независимых переменных x функции $\tilde{f}$ в множество $Y_1 = \{y_1\}$ нижних границ интервалов $\tilde{y} = [y_1, y_2]$ – зависимых переменных этой функции, а вторая однозначно отображает то же множество $X = \{x\}$ в множество $Y_2 = \{y_2\}$ верхних границ интервалов $\tilde{y} = [y_1, y_2]$ – зависимых переменных функции.

Задача настоящей работы заключается в построении двух систематических процедур (алгоритмов), связанных с изучением поведения интервальных функций (2). А именно: 1) процедура расчета (т.е. вычисления значений) интервальной функции; 2) процедура анализа поведения интервальной функции.

2 ОБЗОР ЛИТЕРАТУРЫ

Проблема изучения объектов, характеризующихся той или иной неопределенностью, возникла впервые в начале Второй мировой войны, в связи с необходимостью управления огнем зенитной артиллерии, в условиях случайного движения воздушных целей. Соответствующими задачами занимались выдающиеся математики-вероятностники Н. Винер [3] и А. Н. Колмогоров [4] и их многочисленные последователи. Однако широкое развитие исследований по изучению гражданских объектов, работающих в условиях неопределенности, началось только в конце 1950-х – начале 1960-х гг., в рамках математической статистики и ее новых направлений – обработки данных и планирование экспериментов [5, 6].

1970-е – 1980-е годы привели к более широкому пониманию неопределенности, включившей в себя теперь не только случайность, но и незнание, неединственность возможных исходов, неопределенность целей, многокри-

териальность при решении задач оптимизации. В связи с этим появились новые подходы к описанию неопределенности: теория нечетких множеств, принцип недоопределенной модели, принятие решений в многокритериальных задачах [7, 8, 9].

Наконец, с 1980-х годов начал интенсивно применяться подход к описанию неопределенности, базирующийся на интервальной математике, позволяющей получать оценки характеристик неопределенных систем с гарантированной точностью [10–16]. При этом указанный подход применялся сначала в метрологии для определения интервального значения известной функции при интервальных значениях аргументов. Затем его развитие пошло по двум направлениям. За рубежом этот подход развивался как средство автоматического учета ошибок округления при численном решении задач на компьютерах, в то время как в СССР и России ученые развивали его с целью нахождения области возможных значений результата вычислений с учетом структуры данных и функций, заданных в символьном виде. Но во всех упомянутых выше подходах к изучению систем с неопределенностями не фигурирует элементарная по форме, но очень важная в практических приложениях базовая задача изучения поведения непостоянно определенной функции, служащей характеристикой такой системы, которая аналогична базовой для математического анализа задаче анализа поведения полностью определенной функции. Настоящая статья направлена на решение именно этой задачи.

3 МАТЕРИАЛЫ И МЕТОДЫ

Начнем с решения задачи расчета (вычисления значений) интервальной функции. При этом возможны два случая.

Случай 1. Интервальная функция задана в разделенном виде, в котором верхняя и нижняя границы интервального значения функции выражены каждая по отдельности. Этот вид представления интервальной функции вытекает из соотношений (2)–(4). Именно, из (2), (3) следует явное представление интервальной функции в виде интервала

$$[y_1, y_2] = [f_1(x), f_2(x)], \quad (5)$$

границы которого согласно (5) выражаются формулами

$$y_1 = f_1(x), y_2 = f_2(x). \quad (6)$$

Итак, вычисление интервального значения $\tilde{y} = [y_1, y_2]$ интервальной функции (2), соответствующего значению x независимой переменной этой функции, осуществляется по алгоритму:

Шаг 1. Записываем вычисляемую интервальную функцию (2) в разделенном виде (5), (6) с помощью нижней f_1 и верхней f_2 граничных функций функции (2).

Шаг 2. Вычисляем нижнюю граничную функцию f_1 , используя для этого какой-либо подходящий известный метод вычисления обычных (детерминированных, т.е. полностью определенных) функций [17].

Шаг 3. Вычисляем соответствующую верхнюю граничную функцию f_2 , используя ту же методику, что и на шаге 2.

Шаг 4. Соединяя вычисленные значения нижней f_1 и верхней f_2 граничных функций, получим явное представление (5) вычисленной интервальной функции (2) в виде интервала.

Случай 2. Интервальная функция задана в неразделенном виде, т.е. в виде суперпозиции элементарных интервальных функций: интервального сложения и вычитания, умножения интервала на вещественное число, умножения и деления интервалов [11]. В указанном случае перед собственно вычислением интервальной функции приводится к разделенному виду, после чего к функции применяется 4-шаговый алгоритм случая 1. Приведение любой интервальной функции к разделенному виду можно осуществить с помощью основных формул интервальной математики, выражающих результаты элементарных преобразований интервалов [11]

$$\begin{aligned} [a_1, a_2] + [b_1, b_2] &= [a_1 + b_1, a_2 + b_2]; \\ [a_1, a_2] - [b_1, b_2] &= [a_1 - b_2, a_2 - b_1]; \\ k \cdot [a_1, a_2] &= \begin{cases} [ka_1, ka_2], & k > 0, \\ [ka_2, ka_1], & k < 0; \end{cases} \\ [a_1, a_2] \cdot [b_1, b_2] &= [\min_{i,j}(a_i \cdot b_j), \max_{i,j}(a_i \cdot b_j)]; \\ [a_1, a_2] / [b_1, b_2] &= [a_1, a_2] \cdot [1/b_2, 1/b_1]. \end{aligned} \quad (7)$$

Пример 1. Привести к разделенному виду интервальную функцию

$$\tilde{y} = ([2, 3]x + [1, 2]x) \cdot ([5, 7]x + [4, 6]x) \text{ в области } x \geq 0.$$

Решение. Применяя к заданной интервальной функции последовательно третью, первую и четвертую формулы (7), получим нужный вид функции

$$\begin{aligned} \tilde{y} &= [y_1, y_2] = ([2x, 3x] + [x, 2x]) \times \\ &\times ([5x, 7x] + [4x, 6x]) = [3x, 5x] \cdot [9x, 13x] = \\ &= [\min(3 \cdot 9x^2, 3 \cdot 13x^2, 5 \cdot 9x^2, 5 \cdot 13x^2), \\ &\max(3 \cdot 9x^2, 3 \cdot 13x^2, 5 \cdot 9x^2, 5 \cdot 13x^2)] = [27x^2, 65x^2]. \end{aligned}$$

Таким образом, $y_1 = 27x^2$, $y_2 = 65x^2$, и наконец, разделенная форма заданной интервальной функции $\tilde{y} = [y_1, y_2] = [27x^2, 65x^2]$.

Перейдем к описанию уже упомянутой ранее (п. 1) задачи анализа поведения интервальной функции. Постановка этой задачи аналогична постановке задачи анализа поведения обычной детерминированной функции и включает, в первую очередь, отыскание

- 1) интервалов возрастания функции;
- 2) интервалов убывания функции;
- 3) интервалов постоянства значений функции;
- 4) точек максимума функции;
- 5) точек минимума функции.

Кроме того, постановка задачи анализа поведения интервальной функции может включать отыскание особых интервалов (особых точек) такой функции. Существование таких интервалов (таких точек) связано с интервальным характером этой функции. У обычных детерминированных функций такие интервалы (точки) отсутствуют.

Очевидно, что решение задач анализа поведения интервальной функции требует сравнения величин интервалов. В связи с этим ниже кратко изложены основные результаты теории сравнения интервалов [18, 19]. Эти результаты составляют теоретическую базу детерминистического подхода к принятию решений в условиях интервальной неопределенности. Теоретические основы других возможных подходов к принятию решений в условиях неопределенности можно найти в литературе [20–23]. Принятие решений в условиях неопределенности является важной составной частью онтологии проектирования [24].

Рассмотрим два интервала $\tilde{a} = [a_1, a_2]$ и $\tilde{b} = [b_1, b_2]$. Попытаемся сравнить величины этих интервалов, рассматривая их как интервальные числа. Прямое сравнение интервалов $\tilde{a}$ и $\tilde{b}$ на основе отношений отдельных пар вещественных чисел (a_i, b_j) , где $a_i \in \tilde{a}$, $b_j \in \tilde{b}$, не всегда возможно, так как в общем случае одни пары чисел будут находиться в отношении $(a_i > b_j)$, а другие – в противоположном отношении $(a_i < b_j)$. Поэтому остается только реализовать сравнение интервалов на теоретико-множественном уровне, рассматривая каждый из интервалов как единое целое, не делимое на части. При этом операции нахождения максимума $\vee$ и минимума $\wedge$ двух интервалов $\tilde{a} = [a_1, a_2]$ и $\tilde{b} = [b_1, b_2]$ можно ввести в виде следующих теоретико-множественных конструкций

$$\tilde{a} \vee \tilde{b} = \{a \vee b \mid a \in \tilde{a}, b \in \tilde{b}\}, \quad \tilde{a} \wedge \tilde{b} = \{a \wedge b \mid a \in \tilde{a}, b \in \tilde{b}\}. \quad (8)$$

Таким образом, взятие максимума (минимума) двух интервалов $\tilde{a}$ и $\tilde{b}$ определяется, согласно (8), как нахождение множества максимумов (минимумов) двух точных величин a и b , при условии, что эти величины пробегает все возможные значения соответственно из интервалов $\tilde{a}$ и $\tilde{b}$. Теперь для того, чтобы интервалы $\tilde{a}$ и $\tilde{b}$ можно было сравнить по величине, при этом установив их отношение $\tilde{a} \geq \tilde{b}$ или $\tilde{a} \leq \tilde{b}$, нужно, чтобы 1) введенные операции $\vee, \wedge$ над этими интервалами существовали; 2) эти операции давали в результате один из операндов: $\tilde{a}$ или $\tilde{b}$; 3) указанные две операции были согласованы, т.е. если большим (меньшим) оказывается один из интервалов, то меньшим (большим) является другой из них. Сформулированное условие сравнимости величин интервалов является, очевидно, необходимым и достаточным.

Нетрудно доказать, что условие согласованности операций $\vee$ и $\wedge$ над интервалами всегда выполняется. Очевидно также, что эти операции существуют для любой пары $\tilde{a}, \tilde{b}$, причем результатом операции в общем случае оказывается некоторый новый интервал, отличный как от $\tilde{a}$, так и от $\tilde{b}$. Таким образом, необходимым и достаточным условием сравнимости интервалов $\tilde{a}$ и $\tilde{b}$ оказывается условие, по которому операции $\tilde{a} \vee \tilde{b}$ и $\tilde{a} \wedge \tilde{b}$ должны иметь своим результатом один из интервалов – $\tilde{a}$ или $\tilde{b}$. Из этой формулировки условия сравнимости интервалов выводятся различные его конструктивные формы, удобные для практического применения. Эти формы содержатся в следующих теоремах 1–4.

Теорема 1. Для того чтобы два интервала $\tilde{a}=[a_1, a_2]$ и $\tilde{b}=[b_1, b_2]$ были сравнимы по величине и находились в отношении $\tilde{a} \geq \tilde{b}$, необходимо и достаточно выполнения условий

$$a_1 \geq b_1, a_2 \geq b_2, \quad (9)$$

а для того чтобы эти интервалы были сравнимы по величине и находились в отношении $\tilde{a} \leq \tilde{b}$, необходимо и достаточно выполнения условий

$$a_1 \leq b_1, a_2 \leq b_2. \quad (10)$$

Из этой теоремы вытекает, что любые интервалы $\tilde{a}, \tilde{b}$ сравнимы по величине (по отношению $\geq$ или $\leq$) и находятся в этом отношении только тогда, когда в таком же отношении находятся их одноименные границы a_1, b_1 и a_2, b_2 .

Значение теоремы 1 в том, что она сводит сравнение интервалов и выбор большего (меньшего) из них к очевидной операции сравнения границ указанных интервалов, являющихся вещественными числами.

Теорема 2. Для того чтобы два интервала $\tilde{a}=[a_1, a_2]$ и $\tilde{b}=[b_1, b_2]$ были несравнимы по величине (по отношению $\geq$ или $\leq$), т. е. не находились в отношении $\tilde{a} \geq \tilde{b}$ или $\tilde{a} \leq \tilde{b}$, необходимо и достаточно выполнения условий

$$(a_1 < b_1, a_2 > b_2) \text{ или } (b_1 < a_1, b_2 > a_2). \quad (11)$$

Из приведенной выше теоремы 2 вытекает, что интервалы $\tilde{a}$ и $\tilde{b}$ не сравнимы по отношениям $\geq$ и $\leq$ только тогда, когда один из них полностью накрывает другой. Значение теоремы состоит в том, что она выявляет существование определенных случаев несравнимости интервалов по отношениям $\geq$ и $\leq$, в отличие от обычных вещественных чисел, которые всегда сравнимы по этим отношениям. Несравнимость некоторых интервалов – естественный результат того, что интервальные числа, в отличие от обычных вещественных чисел, задаются не точно, а с неопределенностью (т. е. число принимает некоторое значение в заданном интервале, но при этом не уточняется, какое именно это значение).

Будем рассматривать теперь систему нескольких интервалов

$$\tilde{a}_1=[a_{11}, a_{12}], \tilde{a}_2=[a_{21}, a_{22}], \tilde{a}_3=[a_{31}, a_{32}], \dots \quad (12)$$

Сравнение по отношениям $\geq, \leq$ величин интервалов выписанной системы (12), рассматриваемых как интервальные числа, реализуется в результате попарного сравнения указанных интервалов, выполняемого в соответствии с теоремами 1, 2. Основные результаты, получаемые этим путем, содержатся в теоремах 3 и 4.

Теорема 3. Для того чтобы в системе нескольких интервалов (12) существовал максимальный интервал, который находится со всеми остальными интервалами в отношении $\geq$, и этим интервалом являлся $\tilde{a}_1$, необходимо и достаточно, чтобы границы этого интервала были расположены относительно одноименных границ всех остальных интервалов согласно условиям

$$\begin{aligned} a_{11} &\geq a_{21}, a_{11} \geq a_{31}, a_{11} \geq a_{41}, \dots \\ a_{12} &\geq a_{22}, a_{12} \geq a_{32}, a_{12} \geq a_{42}, \dots \end{aligned} \quad (13)$$

Теорема 4. Для того чтобы в системе нескольких интервалов (12) существовал минимальный интервал, который находится со всеми остальными интервалами в отношении $\leq$, и этим интервалом являлся $\tilde{a}_1$, необходимо и достаточно, чтобы границы данного интервала были расположены относительно одноименных границ всех остальных интервалов согласно условиям

$$\begin{aligned} a_{11} &\leq a_{21}, a_{11} \leq a_{31}, a_{11} \leq a_{41}, \dots \\ a_{12} &\leq a_{22}, a_{12} \leq a_{32}, a_{12} \leq a_{42}, \dots \end{aligned} \quad (14)$$

Как показывают теоремы 3, 4, интервал является максимальным (минимальным) в системе интервалов только тогда, когда максимальны (минимальны) его нижняя граница – среди нижних границ всех интервалов и верхняя граница – среди верхних границ всех интервалов. Подобно случаю сравнения двух интервалов, сравнение произвольного числа интервалов не выявит экстремального интервала, если интервалы, входящие в систему, попарно не сравнимы.

До сих пор мы рассматривали процедуры выделения, вообще говоря, нестрого максимального (нестрого минимального) интервала, основанные на теоретико-множественных операциях (8) вычисления нестрогого максимума (нестрогого минимума) двух интервалов. Аналогично этому вводятся процедуры выделения строго максимального (строго минимального) интервала, т. е. единственного интервала, являющегося максимальным (минимальным).

Будем считать равными по определению совпадающие интервалы, т. е. для произвольных интервалов $\tilde{a}=[a_1, a_2]$ и $\tilde{b}=[b_1, b_2]$, условие их равенства вводится таким образом:

$$(\tilde{a} = \tilde{b}) \Leftrightarrow (a_1 = b_1, a_2 = b_2). \quad (15)$$

Из выражения (15) следует, что неравными являются интервалы, удовлетворяющие следующему условию:

$$(\tilde{a} \neq \tilde{b}) \Leftrightarrow (a_1 \neq b_1 \text{ или } a_2 \neq b_2). \quad (16)$$

Теперь определение того факта, что некоторый интервал $\tilde{a}$ является строго максимальным из двух интервалов $\tilde{a}, \tilde{b}$, можно записать в виде

$$(\tilde{a} > \tilde{b}) \Leftrightarrow (\tilde{a} \vee \tilde{b} = \tilde{a}, \tilde{a} \wedge \tilde{b} = \tilde{b}, \tilde{a} \neq \tilde{b}), \quad (17)$$

аналогичным образом, определение того, что некоторый интервал $\tilde{a}$ является строго минимальным из двух интервалов $\tilde{a}, \tilde{b}$, записывается в виде

$$(\tilde{a} < \tilde{b}) \Leftrightarrow (\tilde{a} \vee \tilde{b} = \tilde{b}, \tilde{a} \wedge \tilde{b} = \tilde{a}, \tilde{a} \neq \tilde{b}). \quad (18)$$

Здесь $\vee$ и $\wedge$ – теоретико-множественные операции (8) вычисления максимума и минимума двух интервалов.

Из формулировки условий (17), (18) сравнимости интервалов в виде строгих неравенств между ними можно вывести различные конструктивные формы этих условий, удобные для использования. Такие формы содержатся в теоремах 5–8, которые подобны теоремам 1–4, дающим удобные конструктивные формы сравнимости интервалов в виде нестрогих неравенств.

Теорема 5. Для того чтобы два интервала $\tilde{a}=[a_1, a_2], \tilde{b}=[b_1, b_2]$ были сравнимы по величине и находились в отношении $\tilde{a} > \tilde{b}$, необходимо и достаточно выполнения условий

$$(a_1 > b_1, a_2 \geq b_2) \text{ или } (a_1 \geq b_1, a_2 > b_2), \quad (19)$$

а для того чтобы эти интервалы были сравнимы по величине и находились в отношении $\tilde{a} < \tilde{b}$, необходимо и достаточно выполнения условий

$$(a_1 < b_1, a_2 \leq b_2) \text{ или } (a_1 \leq b_1, a_2 < b_2). \quad (20)$$

Из теоремы 5 ясно, что интервалы $\tilde{a}$ и $\tilde{b}$ сравнимы по величине (отношению $>$ или $<$) и находятся в указанном отношении лишь тогда, когда в том же отношении находятся их нижние границы a_1, b_1 (при этом верхние границы a_2, b_2 находятся в соответствующем нестрогом отношении – для $>$ это $\geq$, а для $<$ это $\leq$) или же, наоборот, их верхние границы (при этом нижние границы a_1, b_1 находятся в соответствующем нестрогом отношении).

Теорема 6. Для того чтобы два интервала $\tilde{a}=[a_1, a_2]$ и $\tilde{b}=[b_1, b_2]$ были не сравнимы по величине (по отношению $>$ или $<$), т.е. не находились в отношении $\tilde{a} > \tilde{b}$ или $\tilde{a} < \tilde{b}$, необходимо и достаточно выполнения условий

$$(a_1 < b_1, a_2 > b_2) \text{ или } (b_1 < a_1, b_2 > a_2) \text{ или } (a_1 = b_1, a_2 = b_2). \quad (21)$$

Итак, произвольные интервалы $\tilde{a}$ и $\tilde{b}$ не сравнимы по отношениям $>$ и $<$ только в тех случаях, когда один из них полностью накрывает другой либо когда интервалы равны между собой.

Значение теоремы 6 в том, что она показывает существование случаев несравнимости интервалов по отношениям $>$ и $<$ даже тогда, когда сравниваемые интервалы не равны между собой, а только накрывают один другой. Эта ситуация отличается от ситуации со сравнением вещественных чисел, где неравные числа всегда сравнимы по отношениям $>$ и $<$.

Рассмотрим теперь систему нескольких интервалов (12). Сравнение по отношениям $>$ и $<$ величин интервалов системы (12), рассматриваемых как интервальные числа, реализуется путем попарного сравнения этих интервалов, в соответствии с теоремами 5 и 6. Основные результаты, получаемые этим путем, изложены в теоремах 7, 8.

Теорема 7. Для того чтобы в системе нескольких интервалов (12) существовал максимальный интервал, который находится со всеми остальными интервалами в отношении $>$, и этим интервалом был $\tilde{a}_1$, необходимо и достаточно, чтобы границы этого интервала были расположены относительно одноименных границ всех остальных интервалов согласно условиям

$$\left. \begin{aligned} (a_{11} > a_{21}, a_{12} \geq a_{22}) \text{ или } (a_{11} \geq a_{21}, a_{12} > a_{22}) \\ (a_{11} > a_{31}, a_{12} \geq a_{32}) \text{ или } (a_{11} \geq a_{31}, a_{12} > a_{32}) \\ (a_{11} > a_{41}, a_{12} \geq a_{42}) \text{ или } (a_{11} \geq a_{41}, a_{12} > a_{42}) \\ \dots\dots\dots \end{aligned} \right\}. \quad (22)$$

Теорема 8. Для того чтобы в системе нескольких интервалов (12) существовал минимальный интервал, который находится со всеми остальными интервалами в

отношении $<$, и этим интервалом являлся $\tilde{a}_1$, необходимо и достаточно, чтобы границы данного интервала были расположены относительно одноименных границ всех остальных интервалов согласно условиям

$$\left. \begin{aligned} (a_{11} < a_{21}, a_{12} \leq a_{22}) \text{ или } (a_{11} \leq a_{21}, a_{12} < a_{22}) \\ (a_{11} < a_{31}, a_{12} \leq a_{32}) \text{ или } (a_{11} \leq a_{31}, a_{12} < a_{32}) \\ (a_{11} < a_{41}, a_{12} \leq a_{42}) \text{ или } (a_{11} \leq a_{41}, a_{12} < a_{42}) \\ \dots\dots\dots \end{aligned} \right\}. \quad (23)$$

4 АНАЛИЗ ПОВЕДЕНИЯ ИНТЕРВАЛЬНЫХ ФУНКЦИЙ

В предыдущем разделе мы изложили вспомогательный для задачи анализа поведения интервальной функции материал, связанный со сравнением интервальных величин. Теперь мы займемся собственно анализом поведения интервальных функций.

Рассмотрим произвольную интервальную функцию (2). Будем считать, что эта функция задана в разделенном виде (5), (6). Это не ограничивает общности рассмотрения, так как функция, заданная в неразделенном виде, всегда может быть приведена к разделенному виду (п. 2). Будем также полагать, что нижняя и верхняя граничные функции интервальной функции непрерывны и дифференцируемы. Сформулируем условия, при которых функция возрастает, убывает, остается постоянной, достигает максимума (минимума), ведет себя иным способом.

По аналогии с обычными (детерминированными) функциями [1] мы введем понятия возрастания, убывания, постоянства, максимума и минимума интервальной функции.

Определение 1. Интервальная функция $\tilde{y} = \tilde{f}(x)$ называется возрастающей на интервале (a, b) , если для любых x_1 и x_2 из указанного интервала, для которых $x_1 < x_2$, выполняется неравенство $\tilde{f}(x_1) < \tilde{f}(x_2)$.

Определение 2. Интервальная функция $\tilde{y} = \tilde{f}(x)$ называется убывающей на интервале (a, b) , если для любых x_1 и x_2 из данного интервала, для которых $x_1 < x_2$, выполняется неравенство $\tilde{f}(x_1) > \tilde{f}(x_2)$.

Определение 3. Интервальная функция $\tilde{y} = \tilde{f}(x)$ называется постоянной на интервале (a, b) , если для любых x_1 и x_2 из этого интервала, для которых $x_1 < x_2$, выполняется равенство $\tilde{f}(x_1) = \tilde{f}(x_2)$.

Определение 4. Точка $x = x_0$ называется точкой максимума заданной интервальной функции $\tilde{y} = \tilde{f}(x)$, а само число $\tilde{f}(x_0)$ – максимумом этой функции, если для всех точек x из некоторой окрестности точки x_0 , не совпадающих с x_0 , истинно строгое неравенство $\tilde{f}(x_0) > \tilde{f}(x)$.

Определение 5. Точка $x = x_0$ называется точкой минимума интервальной функции $\tilde{y} = \tilde{f}(x)$, а число $\tilde{f}(x_0)$ – минимумом этой функции, если для всех точек x из некоторой окрестности точки x_0 , не совпадающих с x_0 , выполняется строгое неравенство $\tilde{f}(x_0) < \tilde{f}(x)$.

Определение 6. Интервальная функция $\tilde{y} = \tilde{f}(x)$ называется расширяющейся на интервале (a, b) , если для любых точек x_1 и x_2 из этого интервала, для которых $x_1 < x_2$, интервал $\tilde{f}(x_2)$ полностью «накрывает» интервал $\tilde{f}(x_1)$.

Определение 7. Интервальная функция $\tilde{y} = \tilde{f}(x)$ называется сужающейся на интервале (a, b) , если для любых точек x_1, x_2 из этого интервала, для которых $x_1 < x_2$, интервал $\tilde{f}(x_1)$ полностью покрывает интервал $\tilde{f}(x_2)$.

Определение 8. Точку $x = x_0$ будем называть точкой максимального расширения интервальной функции $\tilde{y} = \tilde{f}(x)$, а число $D(x_0) = f_2(x_0) - f_1(x_0)$ – максимальной шириной указанной функции, если для всех точек x из некоторой окрестности точки x_0 , не совпадающих с x_0 , интервал полностью покрывает интервал $\tilde{f}(x)$.

Определение 9. Точку $x = x_0$ будем называть точкой максимального сужения интервальной функции $\tilde{y} = \tilde{f}(x)$, а число $d(x_0) = f_2(x_0) - f_1(x_0)$ – минимальной шириной функции $\tilde{f}(x)$, если для всех точек x некоторой окрестности точки x_0 , не совпадающих с x_0 , интервал $\tilde{f}(x)$ полностью «накрывает» $\tilde{f}(x_0)$.

Сформулируем и докажем условия, которые определяют то или иное поведение интервальной функции.

Теорема 9. Для того чтобы интервальная функция $\tilde{y} = \tilde{f}(x)$ являлась возрастающей на интервале (a, b) , необходимо и достаточно, чтобы на указанном интервале ее нижняя граничная функция $f_1(x)$ была возрастающей, а верхняя граничная функция $f_2(x)$ – неубывающей либо, наоборот, функция $f_2(x)$ была возрастающей, а функция $f_1(x)$ – неубывающей.

Доказательство. Представим нашу интервальную функцию $\tilde{y} = \tilde{f}(x)$ в интервальной форме (5): $\tilde{y} = [f_1(x), f_2(x)]$. Возрастание данной функции на интервале (a, b) по определению (1) означает, что для любых x_1, x_2 из этого интервала, таких, что $x_1 < x_2$, выполняется неравенство $[f_1(x_1), f_2(x_1)] < [f_1(x_2), f_2(x_2)]$, которое, по теореме 5, приводит к двум возможным вариантам:

$$f_1(x_1) < f_1(x_2), f_2(x_1) \leq f_2(x_2) \text{ или } f_1(x_1) \leq f_1(x_2), f_2(x_1) < f_2(x_2).$$

Итак, либо на всем интервале (a, b) нижняя граничная функция $f_1(x)$ нашей интервальной функции $\tilde{y} = \tilde{f}(x)$ является возрастающей, а верхняя граничная функция $f_2(x)$ – неубывающей либо наоборот, $f_2(x)$ является возрастающей, $f_1(x)$ – неубывающей, что и требовалось доказать.

Теорема 10. Для того чтобы интервальная функция $\tilde{y} = \tilde{f}(x)$ была убывающей на интервале (a, b) , необходимо и достаточно, чтобы на указанном интервале ее нижняя граничная функция $f_1(x)$ была убывающей, а верхняя граничная функция $f_2(x)$ – невозрастающей либо, наоборот, функция $f_2(x)$ была убывающей, а функция $f_1(x)$ – невозрастающей.

Доказательство аналогично доказательству теоремы 9, однако использует определение 2 вместо определения 1.

Теорема 11. Для того чтобы интервальная функция $\tilde{y} = \tilde{f}(x)$ была постоянной на интервале (a, b) , необходимо и достаточно, чтобы на этом интервале ее нижняя $f_1(x)$ и верхняя $f_2(x)$ граничные функции были постоянными.

Доказательство следует прямо из определения интервала как множества всех вещественных чисел между заданными двумя числами – границами интервала, включая сами границы.

Теорема 12. Для того чтобы $x = x_0$ была точкой максимума интервальной функции $\tilde{y} = \tilde{f}(x)$, а число $\tilde{f}(x_0)$ – максимумом этой функции, необходимо и достаточно, чтобы в указанной точке достигала максимума ее нижняя граничная функция $f_1(x)$ и не достигала минимума ее верхняя граничная функция $f_2(x)$ либо, наоборот, достигала максимума ее верхняя граничная функция $f_2(x)$ и не достигала минимума ее нижняя граничная функция $f_1(x)$.

Доказательство. Представим нашу интервальную функцию $\tilde{y} = \tilde{f}(x)$ в интервальной форме (5): $\tilde{y} = [f_1(x), f_2(x)]$. Существование максимума этой функции в точке $x = x_0$ по определению 4 означает, что для всех точек x из некоторой окрестности точки x_0 , не совпадающих с x_0 , выполняется неравенство $\tilde{f}(x_0) > \tilde{f}(x)$ или, в интервальной форме $[f_1(x_0), f_2(x_0)] > [f_1(x), f_2(x)]$, причем последнее неравенство, согласно теореме 5, эквивалентно условию

$$(f_1(x_0) > f_1(x), f_2(x_0) \geq f_2(x)) \text{ или } (f_1(x_0) \geq f_1(x), f_2(x_0) > f_2(x)).$$

Условия левой скобки показывают, что в точке x_0 функция $f_1(x)$ обращается в максимум, функция $f_2(x)$ при этом не обращается в минимум. Условия правой скобки говорят, что в точке x_0 функция $f_2(x)$ обращается в максимум, а функция $f_1(x)$ не обращается в минимум, что и требовалось доказать.

Теорема 13. Для того чтобы $x = x_0$ была точкой минимума интервальной функции $\tilde{y} = \tilde{f}(x)$, а число $\tilde{f}(x_0)$ – минимумом этой функции, необходимо и достаточно, чтобы в указанной точке достигала минимума ее нижняя граничная функция $f_1(x)$ и не достигала максимума ее верхняя граничная функция $f_2(x)$ либо, наоборот, достигала минимума ее верхняя граничная функция $f_2(x)$ и не достигала максимума ее нижняя граничная функция $f_1(x)$.

Доказательство теоремы 13 аналогично доказательству теоремы 12, но использует определение 5.

Теорема 14. Для того чтобы интервальная функция $\tilde{y} = \tilde{f}(x)$ являлась расширяющейся на интервале (a, b) , необходимо и достаточно, чтобы на этом интервале была убывающей ее нижняя граничная функция $f_1(x)$ и была возрастающей ее верхняя граничная функция $f_2(x)$.

Доказательство. Для любых точек x_1, x_2 из интервала (a, b) , таких, что $x_1 < x_2$, в силу расширения функции f

на этом интервале должно выполняться условие: интервал $\tilde{f}(x_2)$ полностью покрывает интервал $\tilde{f}(x_1)$. Выражая интервалы $\tilde{f}(x_1)$ и $\tilde{f}(x_2)$ в явной интервальной форме $\tilde{f}(x_1)=[f_1(x_1), f_2(x_1)]$, $\tilde{f}(x_2)=[f_1(x_2), f_2(x_2)]$, имеем условие накрытия интервалов:

$$f_1(x_2) < f_1(x_1), f_2(x_2) > f_2(x_1).$$

Первое из неравенств показывает, что функция $f_1(x)$ убывающая, а второе – что функция $f_2(x)$ возрастающая. Что и требовалось доказать.

Теорема 15. Для того чтобы интервальная функция $\tilde{y}=\tilde{f}(x)$ была сужающейся на интервале (a,b) , необходимо и достаточно, чтобы на данном интервале была убывающей ее верхняя граничная функция $f_2(x)$ и была возрастающей ее нижняя граничная функция $f_1(x)$.

Доказательство аналогично доказательству теоремы 14.

Теорема 16. Для того чтобы точка $x = x_0$ являлась точкой максимального расширения интервальной функции $\tilde{y}=\tilde{f}(x)$, а число $D(x_0)=f_2(x_0)-f_1(x_0)$ – максимальной шириной этой функции, необходимо и достаточно, чтобы в некоторой окрестности точки x_0 слева функция $\tilde{y}=\tilde{f}(x)$ была расширяющейся, а в некоторой окрестности точки x_0 справа она была сужающейся.

Теорема 17. Для того чтобы точка $x = x_0$ была точкой максимального сужения интервальной функции $\tilde{y}=\tilde{f}(x)$, а число $d(x_0)=f_2(x_0)-f_1(x_0)$ минимальной шириной данной функции, необходимо и достаточно, чтобы в некоторой окрестности точки x_0 слева функция $\tilde{y}=\tilde{f}(x)$ была сужающейся, а в некоторой окрестности точки x_0 справа она была расширяющейся.

Доказательства теорем 16, 17 вытекают прямо из определений 8, 9 точек максимального расширения и сужения интервальной функции.

Анализ поведения интервальной функции $\tilde{y}=\tilde{f}(x)$, как показывает изложенный в этом параграфе материал, всегда сводится к анализу поведения двух обычных детерминированных функций: нижней $f_1(x)$ и верхней $f_2(x)$ граничных функций функции $\tilde{f}(x)$. Это позволяет нам использовать для анализа поведения интервальных функций хорошо известные и разработанные методы анализа поведения обычных (детерминированных) функций, основанные на применении классического дифференциального исчисления [1]. При этом алгоритм анализа поведения произвольной интервальной функции может быть описан следующим образом.

Шаг 1. Проверка формы, в которой представлена интервальная функция $\tilde{y}=\tilde{f}(x)$, подлежащая анализу. Если она неразделенная, т.е. не имеющая вид интервала $\tilde{y}=[y_1, y_2]=[f_1(x), f_2(x)]$, где $y_1=f_1(x)$ и $y_2=f_2(x)$ – соответственно нижняя и верхняя граничные функции заданной интервальной функции, то выполняем переход к шагу 2. Если она разделенная, т.е. имеющая указанный вид, то переход к шагу 3.

Шаг 2. Приведение функции $\tilde{y}=\tilde{f}(x)$ из неразделенного вида к разделенному с помощью основных формул интервальной математики (7) (см. пример 1).

Шаг 3. Анализ поведения нижней граничной функции $y_1=f_1(x)$ нашей интервальной функции $\tilde{y}=\tilde{f}(x)$ с помощью известных методов анализа поведения обычных (детерминированных) функций, на основе классического дифференциального исчисления. В ходе анализа устанавливаем интервалы возрастания и убывания функции f_1 и точки ее максимумов и минимумов.

Шаг 4. Анализ поведения верхней граничной функции $y_2=f_2(x)$ той же самой интервальной функции. Он выполняется теми же методами и по той же программе, что и предыдущий шаг.

Шаг 5. Составление сводной таблицы поведения граничных функций $y_1=f_1(x)$, $y_2=f_2(x)$ путем заполнения в ней первых трех строк (табл. 1), в соответствии с результатом анализа поведения функций $f_1(x)$ и $f_2(x)$ (шаги 3, 4).

Шаг 6. Анализ сводной таблицы с помощью теорем 9–15, позволяющих идентифицировать последовательные интервалы $(-\infty, x_1)$, (x_1, x_2) , ..., (x_{n-1}, x_n) , (x_n, ∞) в ней как интервалы возрастания, убывания, расширения или сужения анализируемой интервальной функции $\tilde{y}=\tilde{f}(x)$, а промежуточные точки $x_1, x_2, \dots, x_n$ между интервалами как точки максимума (минимума) или точки максимума (минимума) ее расширения или сужения. Например, если на некотором интервале (x_i, x_{i+1}) функция $f_2(x)$ возрастает, а на соседнем интервале (x_{i+1}, x_{i+2}) убывает, так что в точке x_{i+1} она максимальна, и при этом функция $f_1(x)$ на обоих интервалах постоянная, то согласно теоремам 9, 10, 12 интервальная функция $\tilde{y}=\tilde{f}(x)$ на интервале (x_i, x_{i+1}) возрастает, в точке x_{i+1} достигает максимума, затем на интервале (x_{i+1}, x_{i+2}) убывает.

После выполнения шага 6 заполняется четвертая строка сводной таблицы поведения и на этом анализ поведения заданной интервальной функции заканчивается.

5 РЕЗУЛЬТАТЫ

Характерный возможный вид четвертой строки сводной таблицы поведения интервальной функции проиллюстрирован в табл. 1.

По результатам анализа можно вычертить график функции $\tilde{y}=\tilde{f}(x)$ (рис. 1).

6 ОБСУЖДЕНИЕ

Проблема анализа поведения непостоянно определенных функций существенно отличается от аналогичной проблемы для постоянно определенных функций, что связано с известным отличием проблемы сравнения непостоянно определенных (в рассматриваемом случае – интервальных) чисел от проблемы сравнения постоянно определенных чисел.

Вследствие этих отличий для анализа поведения непостоянно определенных функций имеются простые и ясные методики, основанные на использовании дифференциального исчисления, позволяющие разбить область определения любой функции на подобласти, в которых

Таблица 1 – Фрагмент сводной таблицы поведения интервальной функции

y	$(-\infty, x_1)$	x_1	(x_1, x_2)	x_2	(x_2, x_3)	x_3	(x_3, x_4)	...	(x_{n-1}, x_n)	x_n	(x_n, ∞)
$y_1 = f_1(x)$	монотонная	возможный экстремум f_1 или f_2	монотонная	возможный экстремум f_1 или f_2	монотонная	возможный экстремум f_1 или f_2	монотонная	...	монотонная	возможный экстремум f_1 или f_2	монотонная
$y_2 = f_2(x)$	монотонная	возможный экстремум f_1 или f_2	монотонная	возможный экстремум f_1 или f_2	монотонная	возможный экстремум f_1 или f_2	монотонная	...	монотонная	возможный экстремум f_1 или f_2	монотонная
$\tilde{y} = \tilde{f}(x)$	возрастание	максимум	убывание	минимум	расширение	максимальное расширение	сужение	...	сужение	максимальное сужение	расширение

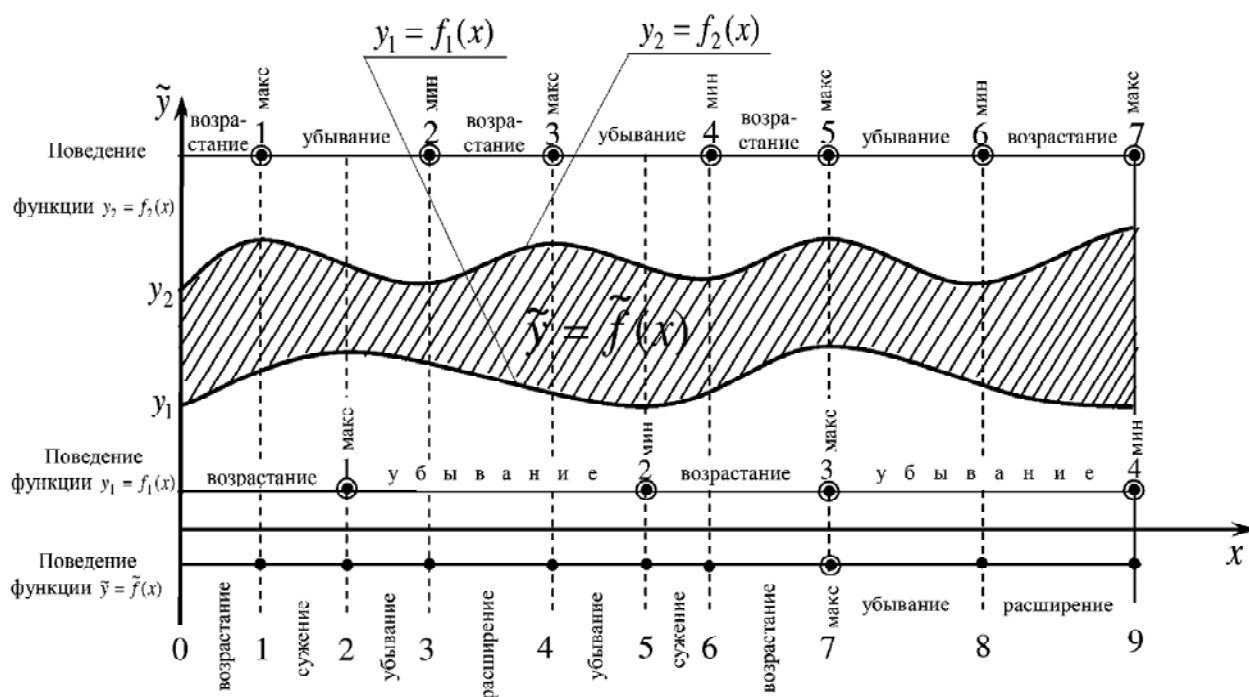


Рисунок 1 – Пример интервальной функции $\tilde{y} = \tilde{f}(x)$ ($y_1 = f_1(x)$ – нижняя граничная функция, $y_2 = f_2(x)$ – верхняя граничная функция)

функция изменяется монотонно: возрастает или убывает, в то время как анализ поведения неполностью определенной – интервальной – функции оказывается более сложным.

Этот анализ, как показано в статье, базируется на сравнении интервалов путем сравнения их границ и далее сводится к анализу поведения двух полностью определенных функций, являющихся нижней и верхней граничными функциями исходной интервальной функции.

Такой анализ позволяет разбить область определения произвольной неполностью определенной (интервальной) функции на подобласти, в которых интервал возможных значений функции изменяется одним из четырех способов: монотонно возрастает, монотонно убывает, монотонно расширяется, монотонно сужается.

При этом, если первые два способа поведения недетерминированной (интервальной) функции аналогичны соответствующим способам поведения полностью определенной (детерминированной) функции, то третий и четвертый способы поведения интервальной функции являются принципиально новыми и невозможны у полностью определенных функций. Последние два способа

поведения интервальной функции возникают вследствие появления у функции подобластей, в которых различные значения этой функции несравнимы и потому стандартное поведение функции – возрастание или убывание – оказывается невозможным. Конечно, два указанных нестандартных способа поведения интервальной функции являются следствием недостатка информации об этой функции, проявляющемся в ее неточном, интервальном задании.

ЗАКЛЮЧЕНИЕ

В статье сформулирована и рассмотрена задача вычисления и анализа поведения неполностью определенной функции, заданной с точностью до интервала значений.

В качестве математического аппарата использованы интервальная алгебра и дифференциальные характеристики верхней и нижней границ интервальных функций, которые можно рассматривать как специальное дифференциальное исчисление для неточно задаваемых функций интервального типа.

Разработаны систематические методы решения задач расчета и анализа поведения недетерминированных фун-

кий інтервального типу, в яких функції визначаються з точністю до інтервалу можливих значень.

Предложено алгоритм детермінізації, який дозволяє звести розв'язувану задачу до двох аналогічних – для верхньої і нижньої граничних функцій вихідної не повністю визначеної функції.

Виділені різні типи поведінки інтервальних функцій (постійність, зростання, зменшення, розширення, звуження) і різні типи екстремальних точок таких функцій (наприклад, точка максимуму, точка мінімуму, точка максимального розширення, точка мінімального звуження).

Доказано теореми, які дозволяють визначати ділянки різної поведінки інтервальних функцій і точки з різними видами екстремуму.

СПИСОК ЛІТЕРАТУРИ

1. Фихтенгольц Г. М. Курс диференціального і інтегрального числення. Т. 1 / Г. М. Фихтенгольц. – М.: Физматлит, 2001. – 616 с.
2. Левин В. И. Интервальная производная и начала недетерминистского дифференциального исчисления / В. И. Левин // Онтология проектирования. – 2013. – № 4. – С. 72–84.
3. Wiener N. Extrapolation, Interpolation and Smoothing of Stationary Time Series / N. Wiener. – N.Y.: Technology Press and Wiley, 1949. – 180 p. (расширенный отчет 1942 года).
4. Колмогоров А. Н. Интерполирование и экстраполирование стационарных случайных последовательностей / А. Н. Колмогоров // Известия АН СССР. Математика. – 1941. – № 5. – С. 3–14.
5. Налимов В. В. Статистические методы планирования экстремальных экспериментов / В. В. Налимов, Н. А. Чернова. – М.: Наука, 1965. – 340 с.
6. Налимов В. В. Теория эксперимента / В. В. Налимов, Н. А. Чернова. – М.: Наука, 1971. – 320 с.
7. Заде Л. А. Понятие лингвистической переменной и его применение к принятию приближенных решений / Л. А. Заде. – М.: Мир, 1976. – 176 с.
8. Нариньяни А. С. Недоопределенность в системе представления и обработки знаний / А. С. Нариньяни // Известия АН СССР. Техническая кибернетика. – 1986. – № 5. – С. 17–25.
9. Hyvonen E. Constraint Reasoning Based on Interval Arithmetic: the Tolerance Propagation Approach / E. Hyvonen // Artificial Intelligence. – 1992. Vol. 58. – P. 19.
10. Moore R. E. Interval Analysis / R. E. Moore. – N. Y.: Prentice-Hall, 1966. – 230 p.
11. Алефельд Г. Введение в интервальные вычисления / Г. Алефельд, Ю. Херцбергер. – М.: Мир, 1987. – 360 с.
12. Канторович Л. В. О некоторых новых подходах к вычислительным методам и обработке наблюдений / Л. В. Канторович // Сибирский математический журнал. – 1962. Т. 3, № 5.
13. Вошинин А. П. Оптимизация в условиях неопределенности / А. П. Вошинин, Г. Р. Сотиров. – М.: МЭИ-Техника, 1989. – 226 с.
14. Вошинин А. П. Интервальный анализ данных / А. П. Вошинин, А. Ф. Бочков, Г. Р. Сотиров // Заводская лаборатория. – 1990. – № 7. – С. 76–81.
15. Куржанский А. Б. Задача идентификации – теория гарантированных оценок / А. Б. Куржанский // Автоматика и телемеханика. – 1991. – № 4. – С. 75–89.
16. Вошинин А. П. Интервальный метод калибровки / А. П. Вошинин // Датчики и системы. – 2000. – № 7. – С. 63–69.
17. Милн В. Э. Численный анализ / В. Э. Милн. – М.: Издательство иностранной литературы, 1980. – 350 с.
18. Левин В. И. Интервальные методы оптимизации систем в условиях неопределенности / В. И. Левин. – Пенза: Изд-во Пензенского технологического ин-та, 1999. – 101 с.
19. Левин В. И. Оптимизация в условиях интервальной неопределенности. Метод детерминизации / В. И. Левин // Автоматика и вычислительная техника. – 2012. – № 4. – С. 157–163.
20. Tsoukias A. A Characterization of PQI Interval Orders / A. Tsoukias, P. Vincke // Discrete Applied Mathematics. – 2003. – № 127 (2). – P. 387–397.
21. Ozturk M. Positive and Negative Reasons in the Interval Comparisons: Valued PQI Interval Orders / M. Ozturk, A. Tsoukias // Proceedings of the 10th Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU-2004), Perugia, Italy, July 4–9, 2004. – P. 983–989.
22. Давыдов Д. В. Идентификация параметров линейных интервальных управляемых систем с интервальным наблюдением / Д. В. Давыдов // Известия РАН. Теория и системы управления. – 2008. – № 6. – С. 25–29.
23. Пиявский С. А. Простой и универсальный метод принятия решений в пространстве критериев «стоимость-эффективность» / С. А. Пиявский // Онтология проектирования. – 2014. – № 3. – С. 89–102.
24. Боргест Н. М. Ключевые термины онтологии проектирования / Н. М. Боргест // Онтология проектирования. – 2013. – № 3. – С. 9–31.

Статья поступила в редакцию 10.01.2016.

После доработки 25.01.2013.

Левин В. И.

Д-р техн. наук, профессор, ведущий научный сотрудник Пензенского государственного технологического университета, Пенза, Россия

РОЗРАХУНОК І АНАЛІЗ ПОВЕДІНКИ НЕЦІЛКОМ ВИЗНАЧЕНИХ ФУНКЦІЙ МЕТОДОМ ДЕТЕРМІНАЦІЇ

Розглянуто існуючі підходи до розрахунку, аналізу, синтезу й оптимізації систем в умовах невизначеності. Дослідження невизначених систем формулюється у вигляді задач розрахунку, аналізу і синтезу різних функцій з недетермінованими параметрами, що слугують відповідними характеристиками даних систем. Усі ці задачі є значно складнішими за їхніх детермінованих аналогів, що приходить вирішувати при дослідженні систем з детермінованими (точно відомими) параметрами. Таке ускладнення пов'язане з тим, що алгебра недетермінованих чисел є складнішою алгебри детермінованих чисел. У статті сформульована і докладно описана задача обчислення й аналізу поведінки нецілком визначеної функції, заданої з точністю до інтервалу значень. Для вирішення зазначеної задачі запропонований алгоритм детермінізації, що дозволяє звести задачу до двох аналогічних – для верхньої і нижньої граничних функцій вихідної нецілком визначеної функції. У цьому алгоритмі автором використаний апарат інтервальної математики та інтервально-диференційного числення. Далі виділені різні типи можливого поведінки інтервальних функцій (сталість, зростання, зменшення, розширення, звуження) і різні типи екстремальних точок таких функцій (наприклад, точка максимуму, точка мінімуму, точка максимального розширення, точка мінімального звуження). Доведено теореми, що дозволяють визначати ділянки різного поведінки інтервальних функцій і точки з різними видами екстремуму. Докладно розглянута робота запропонованого алгоритму детермінізації, що дозволяє аналізувати поведінку інтервальних функцій. Робота проілюстрована на конкретному прикладі.

Ключові слова: оптимізація, детермінована функція, інтервальна функція, аналіз поведінки функцій, невизначеність.

Levin V. I.

Dr Sc., Professor of Mathematical Department of Penza State Technological University, Penza, Russia

CALCULATION AND ANALYSIS OF INCOMPLETELY DEFINED FUNCTIONS BY DETERMINATION METHOD

This article reviews current approaches to the calculation, analysis, synthesis and optimization under uncertainty. Studying uncertain systems is formulated as problems of the calculation, analysis and synthesis of various non-deterministic functions with parameters that serve as the relevant characteristics of these systems. All these problems are much more difficult their deterministic counterparts which should be solved in the study of systems with deterministic (exactly known) parameters. Complexity is due to the fact that the non-deterministic algebra is more complicated than algebra of deterministic numbers. The article stated and described in detail the problem of calculating and analyzing the behavior of a function which is given up to a range of values. To solve this problem, the algorithm of determination is presented. This algorithm reduces the problem to the two same – for the lower and upper boundary functions of the original incompletely defined function. In this algorithm author uses interval mathematics and interval-differential calculus. The different types of possible behavior of interval functions are highlighted (consistency, increase, decrease, expansion, contraction) and various types of extreme points of such functions (for example, the maximum point, a minimum point, the point of maximum expansion, the point of minimum extension) are shown. Theorems that allow you to define areas of different behavior of interval functions and points with different types of extreme are proved. The work of the proposed algorithm of determination for analyzing the behavior of interval functions is considered in detail. Operation of algorithm is illustrated by concrete example.

Keywords: optimization, deterministic function, non-deterministic function, uncertainty, analysis of behavior of function, uncertainty.

REFERENCES

1. Fih tengolc G.M. Kurs differencialnogo i integralnogo ischisleniya. V. 1. Moscow, Fizmatlit, 2001, 616 p.
2. Levin V.I. Intervalnaya proizvodnaya i nachala nedeterministicheskogo differencialnogo ischisleniya, *Ontologiya proektirovaniya*, 2013, No. 4, pp. 72–84.
3. Wiener N. Extrapolation, Interpolation and Smoothing of Stationary Time Series. N.Y., Technology Press and Wiley, 1949, 180 p. (expired report of 1942 year).
4. Kolmogorov A. N. Interpolirovanie i extrapolirovanie stacionarnykh sluchaynykh posledovatel'nostey, *Izvestiya AN SSSR: Seriya Matematicheskaya*, 1941, No. 5, pp. 3–14.
5. Nalimov V. V., Chernova N. A. Statisticheskie metody planirovaniya ekstremal'nykh eksperimentov. Moscow, Nauka, 1965, 340 p.
6. Nalimov V. V., Chernova N. A. Teoriya experimenta. Moscow, Nauka, 1971, 320 p.
7. Zadeh L. A. The Concept of a Linguistic Variable and its Application to Approximate Reasoning, *Information Sciences*, 1975, No. 8; 9, pp. 199–249, 301–357; 43–80.
8. Narinyani A. S. Nedopredelennost v sisteme predstavleniya i obrabotki znaniy, *Journal of Computer and Systems Sciences International*, 1986. Vol. 24, No. 5, pp. 17–25.
9. Hyvonen E. Constraint Reasoning Based on Interval Arithmetic: the Tolerance Propagation Approach, *Artificial Intelligence*, 1992, Vol. 58, pp. 19.
10. Moore R.E. Interval Analysis. N. Y., Prentice-Hall, 1966, 230 p.
11. Alefeld G., Herzberger J. Introduction to Interval Computation. N.Y., Academic Press, 1983, 352 p.
12. Kantorovich L.V. O nekotorykh novykh podkhodakh k vychislitel'nykh metodam i obrabotke nablyudeniy, *Siberian Mathematical Journal*, 1962, Vol. 3, No. 5.
13. Voschinin A.P. G.R. Sotirov. Optimizatsiya v usloviyakh neopredelennosti. Moscow, MEI-Tehnika, 1989, 226 p.
14. Voschinin A. P., Bochkov A. F., Sotirov G. R. Intervalniy analiz dannykh. Industrial Laboratory, 1990, Vol. 56, No. 7, pp. 76–81.
15. Kurzhanskii A.B. Identification Problem – Theory of Guaranteed Estimates, *Automation and Remote Control*, 1991, Vol. 52, No. 4, Part 1, pp. 447–465.
16. Voschinin A.P. Intervalniy metod kalibrovki, *Datchiki i Sistemy*, 2000, No. 7, pp. 63–69.
17. Milne W. E. Numerical Calculus. Princeton, University Press, 1949, 393 p.
18. Levin V. I. Intervalnye metody optimizatsii system v usloviyakh neopredelennosti. Penza, Penza Technological Institute Publishing, 1999, 95 p.
19. Levin V.I. Optimization in Terms of Interval Uncertainty: The Determinization Method, *Automatic Control and Computer Sciences*, 2012, Vol. 46, No. 4, pp. 157–163.
20. Tsoukias A., Vincke P. A Characterization of PQI Interval Orders, *Discrete Applied Mathematics*, 2003, No. 127 (2), pp. 387–397.
21. Ozturk M., Tsoukias A. Positive and Negative Reasons in the Interval Comparisons: Valued PQI Interval Orders, *Proceedings of the 10th Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU-2004)*, Perugia, Italy, July 4–9, 2004, pp. 983–989.
22. Davydov D. V. Identification of Parameters of Linear Interval Controllable Systems with the Interval Observation, *Journal of Computer and Systems Sciences International*, 2008, Vol. 47, No. 6, pp. 861–865.
23. Piyavskiy S. A. Prostoy i universalniy metod prinyatiya resheniy v prostranstve kriteriev “stoimost-effektivnost”, *Ontologiya proektirovaniya*, 2014, No. 3, pp. 89–102.
24. Borgest N. M. Klyuchevye terminy ontologii proektirovaniya, *Ontologiya proektirovaniya*, 2013, No. 3, pp. 9–31.

ЭКСПЕРИМЕНТАЛЬНОЕ ИССЛЕДОВАНИЕ МЕТОДА СИНТЕЗА НЕЙРО-НЕЧЕТКИХ МОДЕЛЕЙ В ПАРАЛЛЕЛЬНОЙ КОМПЬЮТЕРНОЙ СИСТЕМЕ

Решена задача разработки нелинейной модели, описывающей зависимости между характеристиками системы, в которой осуществляется синтез нейро-нечетких сетей, параметрами исследуемого метода и временем, затраченным системой на выполнение синтеза моделей. Объект исследования – процесс синтеза нейро-нечетких моделей для индивидуального прогнозирования состояния больного гипертонической болезнью. Предметом исследования является параллельная компьютерная система, выполняющая метод синтеза нейро-нечетких сетей. Цель работы заключается в повышении эффективности применения параллельных компьютерных систем для решения задач медицинского направления. Предложена нелинейная модель, позволяющая прогнозировать затраченное параллельной системой время на выполнение метода синтеза нейро-нечетких сетей и, таким образом, осуществлять рациональный выбор ресурсов компьютерной системы. Разработано программное обеспечение, которое реализует предложенную модель. Выполнены эксперименты, подтверждающие адекватность предложенной модели. Результаты экспериментов позволяют рекомендовать применение разработанной модели на практике.

Ключевые слова: синтез модели, параллельная система, планирование ресурсов, нейронная сеть, среднеквадратичная ошибка.

НОМЕНКЛАТУРА

CPU – Central Processing Unit;

GPU – Graphical Processing Unit;

G – критерий качества сети;

N_χ – число возможных решений на этапе инициализации метода;

$R^{(i)}$ – множество решений на i -й итерации метода;

$G^{(i)}$ – множество значений целевой функции для решений $\chi_k^{(i)}$ на i -й итерации метода;

$R^{(i,j)}$ – j -е подмножество множества $R^{(i)}$;

$\chi_k^{(i)}$ – k -е решение во множестве $R^{(i)}$;

$G_k^{(i)}$ – значение целевой функции k -го решения множества $R^{(i)}$;

D – обучающая выборка;

X – множество входных признаков, описывающих время выполнения метода синтеза нейро-нечетких сетей в параллельной системе;

x_i – множество значений i -го признака в обучающей выборке D ;

x_{ij} – значение i -го признака для j -го экземпляра, соответствующее значению i -й характеристики j -го эксперимента;

N – количество экземпляров в выборке D ;

Y – множество значений выходного параметра;

x_1 – тип системы (0 – кластер CPU, 1 – GPU);

x_2 – число процессов, на которых выполняется метод;

x_3 – пропускная способность сети параллельной системы, Гб/с;

x_4 – число возможных решений, с которыми работает метод;

y – время, затраченное системой на синтез нейро-нечеткой сети, с;

x_{jn} – нормированное значение i -го признака j -го экземпляра;

$x_{\min}$ и $x_{\max}$ – соответственно, минимальное и максимальное значения i -го признака в обучающей выборке D ;

Φ – аргумент функции активации, представляющий собой дискриминантную функцию;

w – матрица весовых коэффициентов;

x – множество аргументов дискриминантной функции;

w_θ – значение смещения функции $\Phi(w;x)$

$|x|$ – количество аргументов функции $\Phi(w;x)$

$\Psi_{(\mu,\rho)}$ – функция активации ρ -го нейрона μ -го слоя

$\Phi_{(\mu,\rho)}$ – дискриминантная функция ρ -го нейрона μ -го слоя.

ВВЕДЕНИЕ

Артериальная гипертония является распространенным заболеванием в Украине и в мире. При этом широкой причиной смертности является несвоевременное выявление заболевания, особенно среди жителей сельской местности [1]. Для предотвращения опасных последствий важно своевременно выявлять и прогнозировать развитие гипертонии для каждого пациента. Такое прогнозирование успешно выполняется при помощи соответствующих нейро-нечетких моделей [2–4].

Для учета динамики изменения состояния здоровья пациента необходимо с течением времени синтезировать новые модели на основе увеличения выборки наблюдений за показателями заболевания. Это требует больших затрат временных и вычислительных ресурсов. Поэтому в данной сфере нашли применение методы параллельных вычислений [5], которые реализуются при помощи кластеров вычислительных узлов [6] и графических процессоров CUDA [7]. Поскольку использование кластерных ресурсов является дорогостоящим, то важно рационально планировать выбор вычислительных ресурсов для достижения желаемой производительности.

Целью данной работы является повышение эффективности применения параллельных компьютерных систем

для синтеза нейро-нечетких сетей и решения задачи индивидуального прогнозирования состояния больного гипертонической болезнью.

1 ПОСТАНОВКА ЗАДАЧИ

Рассмотрим метод синтеза нейро-нечетких сетей для прогнозирования состояния больного гипертонической болезнью [2]. Пусть имеем выборку $D = \langle X, Y \rangle$ данных о состоянии здоровья пациентов, полученную в г. Запорожье. Задача синтеза нейро-нечеткой модели заключается в идентификации ее параметров таким образом, чтобы обеспечивалось приемлемое значение критерия качества G , например, значения среднеквадратичной ошибки. На рис. 1 показан граф параметрического синтеза нейро-нечетких моделей в ярусно-параллельной форме [2]. Как видно, процесс параметрического синте-

за нейро-нечетких моделей начинается с этапа инициализации, на котором создается начальное множество решений $R^{(0)} = \{\chi_1^{(0)}, \chi_2^{(0)}, \dots, \chi_{N_\chi}^{(0)}\}$, затем это множество разбивается на подмножества $R^{(0,j)}$, каждое из которых оценивается на отдельном процессе параллельной системы. В ходе работы метода поиск решений происходит с помощью стохастического подхода, используя операторы отбора, скрещивания и мутации. При этом элементы подмножества $R^{(i,j)}$ обрабатываются на j -м процессе параллельной системы. Каждая итерация случайного поиска $RS(R^{(i)})$, продемонстрированная в виде графа на рис. 2 [2], требует существенных вычислительных затрат и поэтому распараллеливается по процессам компьютерной системы. Основным параметром метода, влия-

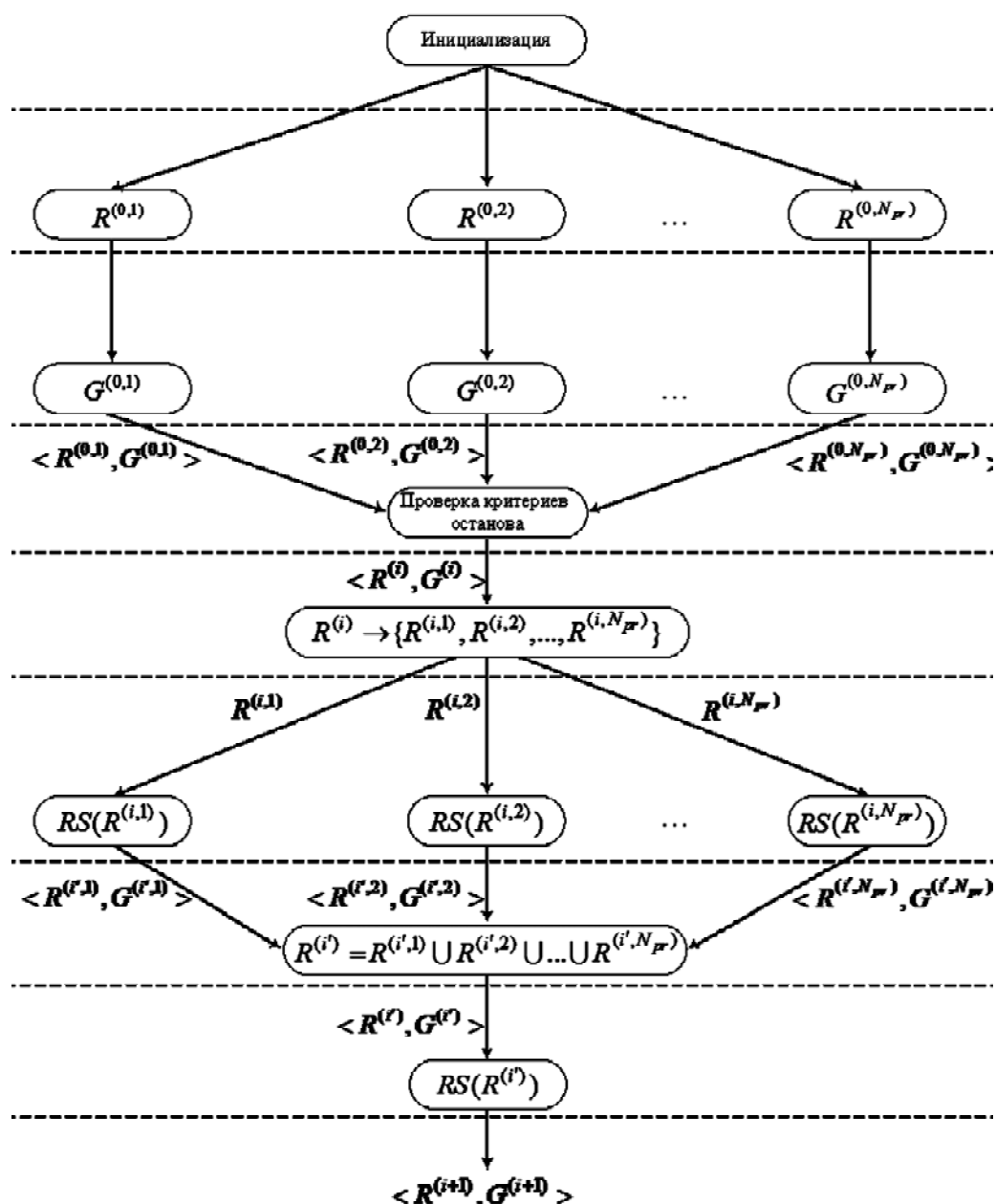


Рисунок 1 – Граф параметрического синтеза нейро-нечетких моделей в ярусно-параллельной форме

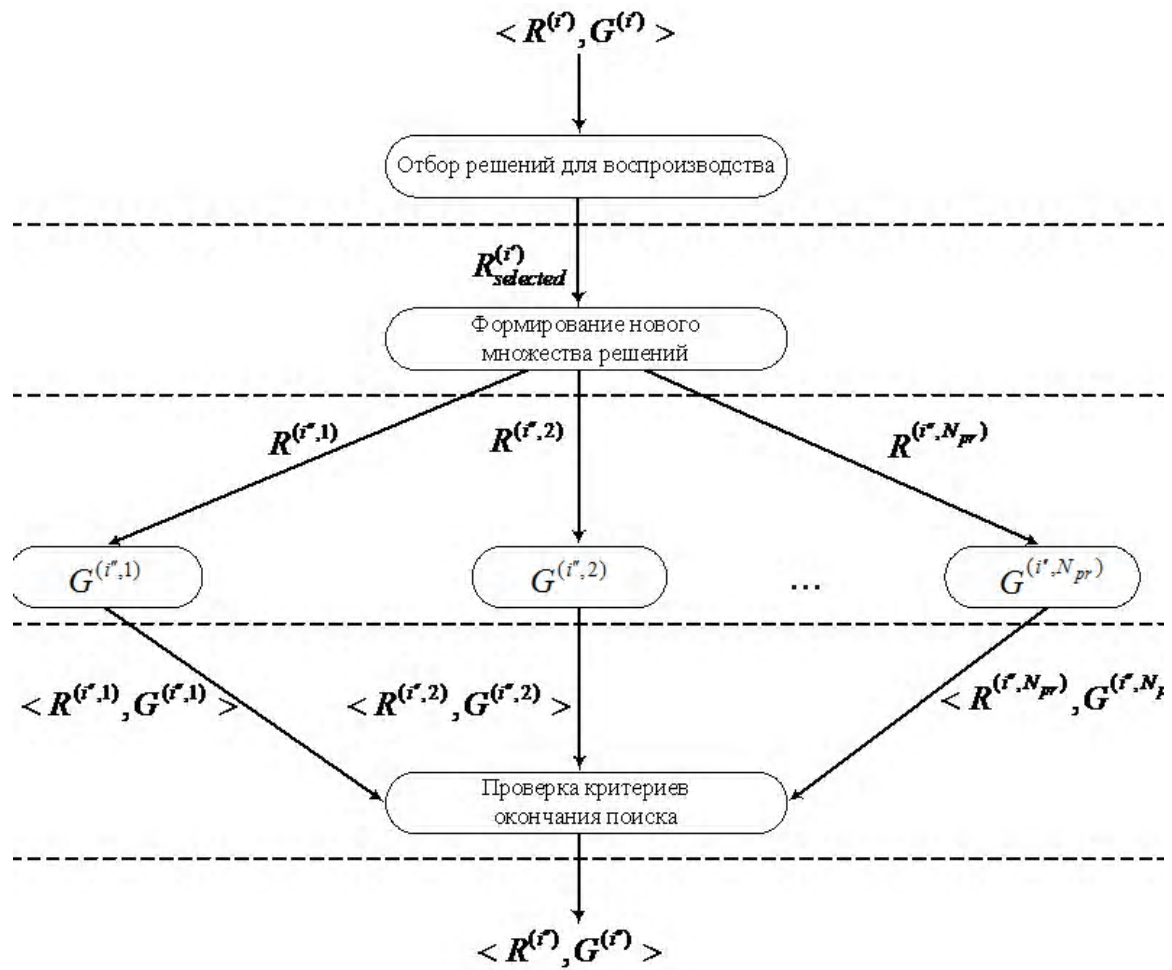


Рисунок 2 – Граф одной итерации случайного поиска

яющим как на точность синтезированных моделей, так и на время, затрачиваемое параллельной системой на выполнение метода, является число возможных решений N_χ на этапе инициализации метода. Чем больше значение N_χ , тем больше вычислительного времени требуется системе, но и тем выше точность синтезированных системой моделей.

В данной работе ставится задача экспериментально исследовать рассмотренный метод в параллельной компьютерной системе и разработать нелинейную модель, описывающую зависимости между характеристиками системы, в которой осуществляется синтез нейро-нечетких сетей, а так же параметрами исследуемого метода и временем, затраченным системой на выполнение синтеза моделей. Это позволит рассчитывать время, необходимое компьютерной системе для индивидуального прогнозирования состояния больного гипертонической болезнью, и таким образом, рационально планировать выбор ресурсов компьютерной системы для достижения желаемой производительности.

2 ОБЗОР ЛИТЕРАТУРЫ

Планирование ресурсов компьютерных систем выполняется на основе специфики параметров решаемых задач и характеристик компьютерных систем. Основным мето-

дом, при помощи которого можно выполнить такое планирование, является моделирование поведения компьютерной системы в решении определенной задачи [8–10].

Методы регрессионного и имитационного моделирования позволяют с приемлемой точностью строить модели планирования ресурсов компьютерных систем, в которых между параметрами такой модели существует хорошая линейная корреляционная зависимость [11]. При отсутствии такой зависимости существенно снижается точность модели, поэтому предпочтительнее применять метод нейросетевого моделирования [12], позволяющий описать нелинейные зависимости между характеристиками компьютерной системы, параметрами задачи, решаемой в такой системе, и временем, затраченным на решение задачи.

Известный инструментальный GridSim [13], применяемый для моделирования поведения компьютерных систем, в процессе построения модели требует ввести оценочные значения временных затрат на выполнения каждого элемента решаемой задачи. В задаче индивидуального прогнозирования состояния больного гипертонической болезнью такие значения отсутствуют для отдельно взятой компьютерной системы, поэтому в данной работе разработано собственное программное обеспечение, на основе которого применен метод нейросетевого моделирования.

3 МАТЕРИАЛЫ И МЕТОДЫ

Основными характеристиками параллельной системы, влияющими на время выполнения метода синтеза нейро-нечетких сетей, являются тип системы (кластер CPU или GPU), число процессов, на которых выполняется задача, и пропускная способность сети. Число возможных решений N_χ как параметр исследуемого метода так же оказывает существенное влияние на затраченное системой время. Рассмотренный метод был применен на кластере CPU и на GPU, в результате чего была сформирована обучающая выборка (1), содержащая, 174 результата выполнения метода, каждый из которых характеризовался четырьмя признаками:

$$D = \langle X, Y \rangle, \quad (1)$$

где $X = \{x_1, x_2, x_3, x_4\}$, $x_i = \{x_{i1}, x_{i2}, \dots, x_{iN}\}$, $N=174$, $Y = \{y_1, y_2, \dots, y_N\}$.

В результате обучающая выборка представляла собой таблицу чисел, состоящую из 174 строк и пяти столбцов, содержащих значения четырех входных признаков и одного выходного для каждого случая применения рассмотренного метода в параллельной системе. Фрагмент обучающей выборки приведен в табл. 1.

Для исключения влияния различного порядка значений признаков на синтезируемую модель выполнялось нормирование признаков, т.е. приведение диапазона их значений к единому интервалу $x_{in} \in [0;1]$ по формуле (2):

$$x_{ijn} = \frac{x_{ij} - x_{i\min}}{x_{i\max} - x_{i\min}}, \quad (2)$$

где $i = \overline{1;4}$, $j = \overline{1;174}$.

Моделирование решения задачи индивидуального прогнозирования состояния больного гипертонической болезнью в параллельной системе осуществлялось на основе построенной обучающей выборки с помощью трехслойного персептрона [14, 15], первый слой которого содержал четыре нейрона, второй слой – три нейрона, третий слой – один нейрон. Все нейроны имели сигмоидную функцию активации (3):

$$\psi(\varphi) = \frac{1}{1 + e^{-\varphi}}, \quad (3)$$

где $\varphi = \varphi(w; x)$.

Таблица 1 – Фрагмент обучающей выборки

Номер эксперимента	Значения признаков				Y
	x_1	x_2	x_3	x_4	
1	0	1	20	50	101,18
2	0	2	20	50	57,67
3	0	3	20	50	40,02
4	0	4	20	50	30,61
5	0	5	20	50	26,23
6	0	6	20	50	23,17
7	0	7	20	50	21,09
8	0	8	20	50	18,09
9	0	9	20	50	16,84
10	0	10	20	50	15,79
...	...	...	...	...	...
174	1	260	64	100	62,81

При синтезе нейромодели в качестве дискриминантной функции [16] использовалась взвешенная сумма (4):

$$\varphi(w; x) = w_0 + \sum_{i=1}^{|x|} w_i x_i, \quad (4)$$

где w_i определяет вес i -го входного параметра x_i в функции $\varphi(w; x)$.

Таким образом, структура синтезируемой трехслойной нейромодели Y_{NN} может быть представлена следующим образом (5):

$$\begin{cases} Y_{NN} = \psi_{(3,1)}(\varphi_{(3,1)}(w_{(3,1)}; \psi_{(2)})); \\ \psi_{(2)} = \{\psi_{(2,1)}; \psi_{(2,2)}; \psi_{(2,3)}\}; \\ \psi_{(2,k)} = \psi_{(2,k)}(\varphi_{(2,k)}(w_{(2,k)}; \psi_{(1)})), k = 1,2,3; \\ \psi_{(1)} = \{\psi_{(1,1)}; \psi_{(1,2)}; \psi_{(1,3)}; \psi_{(1,4)}\}; \\ \psi_{(1,l)} = \psi_{(1,l)}(\varphi_{(1,l)}(w_{(1,l)}; X)), l = 1,2,3,4. \end{cases} \quad (5)$$

Для построения нейромодели и определения значений ее параметров (весовых коэффициентов и смещений каждого нейрона) на ее входы подавались значения пронормированных признаков, на выход – значение времени выполнения метода синтеза нейро-нечетких сетей в параллельной системе. В качестве целевой функции при обучении нейромодели использовался минимум среднеквадратичной ошибки.

Обучение нейронной сети выполнялось на основе метода обратного распространения ошибки [17]. Приемлемым считалось достижение среднеквадратичной ошибки, не превышающей 10^{-4} . Фрагмент матрицы весовых коэффициентов w построенной модели приведен в табл. 2.

После подстановки значений весовых коэффициентов и смещений в (5) с учетом функции активации (3) и дискриминантной функции (4) получаем математическое описание синтезированной нейросетевой модели (6), описывающей зависимости между характеристиками системы, в которой осуществляется синтез нейро-нечетких сетей, параметрами исследуемого метода и временем, затраченным системой на выполнение синтеза сетей. Графическая интерпретация синтезированной нейромодели приведена на рис. 3.

$$\begin{cases} Y_{NN} = \psi_{(3,1)} \left(1 + e^{-(-2,873 + 5,42\psi_{(2,1)} - 10,94\psi_{(2,2)} + 28,35\psi_{(2,3)})} \right)^{-1}; \\ \psi_{(2,1)} = \left(1 + e^{-(-8,65 - 5,2\psi_{(1,1)} - 4,17\psi_{(1,2)} - 4,95\psi_{(1,3)} - 1,74\psi_{(1,4)})} \right)^{-1}; \\ \psi_{(2,2)} = \left(1 + e^{-(-0,43 - 3,63\psi_{(1,1)} + 1,57\psi_{(1,2)} - 4,44\psi_{(1,3)} - 2,46\psi_{(1,4)})} \right)^{-1}; \\ \psi_{(2,3)} = \left(1 + e^{-(-5,37 - 9,56\psi_{(1,1)} - 24,6\psi_{(1,2)} + 2,94\psi_{(1,3)} + 33,79\psi_{(1,4)})} \right)^{-1}; \\ \psi_{(1,1)} = \left(1 + e^{-(-2,38 + 7,7x_1 - 5,1x_2 + 2,2x_3 - 6,85x_4)} \right)^{-1}; \\ \psi_{(1,2)} = \left(1 + e^{-(-6,22 - 5,06x_1 + 22,21x_2 + 18,14x_3 + 6,49x_4)} \right)^{-1}; \\ \psi_{(1,3)} = \left(1 + e^{-(-10,35 - 0,07x_1 + 11,12x_2 - 1,84x_3 + 9,29x_4)} \right)^{-1}; \\ \psi_{(1,4)} = \left(1 + e^{-(-2,07 + 5,13x_1 - 35,66x_2 + 4,49x_3 + 2,06x_4)} \right)^{-1}. \end{cases} \quad (6)$$

Таблица 2 – Фрагмент матрицы весовых коэффициентов w

Номер слоя, μ	Номер нейрона в слое, ρ	Значение смещения w_0	Связи (соединения)	
			узел, от которого идет соединение	значение весового коэффициента
1	1	-2,38	признак x_1	7,7
			признак x_2	-5,1
			признак x_3	2,2
			признак x_4	-6,85
	2	-6,22	признак x_1	-5,06
			признак x_2	22,21
			признак x_3	18,14
			признак x_4	6,49
	3	-10,35	признак x_1	-0,07
			признак x_2	11,12
			признак x_3	-1,84
			признак x_4	9,29
	4	-2,07	признак x_1	5,13
			признак x_2	-35,66
			признак x_3	4,49
			признак x_4	2,06
...	...	...	...	...
3	1	-2,873	нейрон (2, 1)	5,54
			нейрон (2, 2)	-10,94
			нейрон (2, 3)	28,35

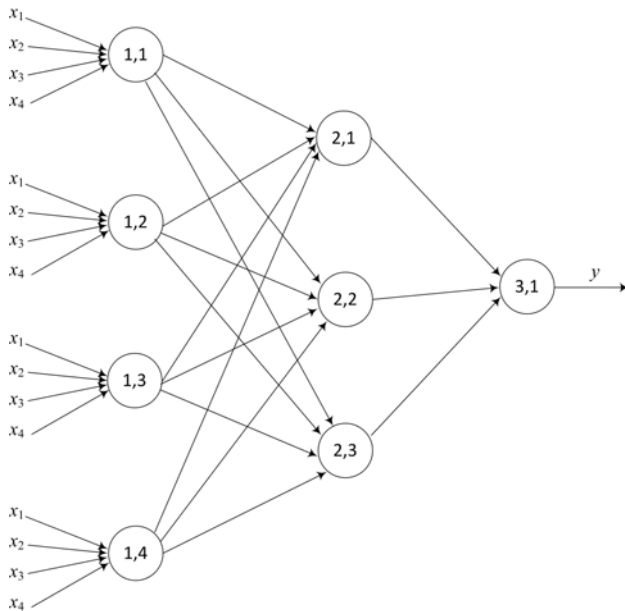


Рисунок 3 – Синтезированная нейросетевая модель

Таким образом, построенная нейросетевая модель представляет собой иерархическую структуру, содержащую нейроны, и позволяет рассчитывать время, необходимое компьютерной системе для индивидуального прогнозирования состояния больного гипертонической болезнью. При этом значение среднеквадратичной ошибки модели составило $2,95 \cdot 10^{-5}$, что является приемлемым для подобного рода задач, решаемых при помощи синтезированной модели.

4 ЭКСПЕРИМЕНТЫ

Для выполнения экспериментального исследования рассмотренного метода и предложенной нейросетевой модели использованы следующие компьютерные системы:

- кластер Института проблем моделирования в энергетике имени Г.Е. Пухова НАН Украины (ИПМЭ) г. Киев: процессоры Intel Xeon 5405, оперативная память – 4×2 Гб DDR-2 на каждый узел, коммуникационная среда InfiniBand 20 Гб/с, middleware Torque и OMPI;

- кластер Запорожского национального технического университета (ЗНТУ) г. Запорожье: процессоры Intel E3200, оперативная память 1 Гб DDR-2 на каждый узел, коммуникационная среда Gigabit Ethernet 1 Гб/с, middleware Torque и MPICH;

- GPU NVIDIA GTX 285+ 240 ядер CUDA;

- GPU NVIDIA GTX 960 1024 ядер CUDA.

В экспериментах число процессов x_2 , на которых выполнялся метод, варьировалось от 1 до 32 для кластеров и от 60 до 260 для GPU. Пропускная способность сети x_3 – от 1 до 20 Гб/с, число возможных решений на этапе инициализации метода N_χ – от 50 до 100. Для проведения экспериментов было разработано программное обеспечение на языке Си с применением библиотеки MPI [18].

5 РЕЗУЛЬТАТЫ

Результаты экспериментов на кластерах ЗНТУ ($N_\chi=50$) и ИПМЭ ($N_\chi=100$) приведены на рис. 4 и рис. 5, соответственно. Рис. 6 демонстрирует экспериментальную проверку предложенной модели на GPU NVIDIA GTX 285+, при этом $N_\chi=50$. Рис. 7 демонстрирует результаты экспериментов на GPU NVIDIA GTX 960 с $N_\chi=100$. Сплошной линией показано время, фактически затраченное системой на выполнение метода синтеза нейро-нечетких сетей, а пунктирной линией – расчетное время при помощи предложенной модели.

6 ОБСУЖДЕНИЕ

Тестовая выборка, состоящая из результатов 53х экспериментов, включала экземпляры решений задачи в параллельной системе, не входящие в обучающую выборку. При проведении экспериментов на более производительном оборудовании число возможных решений задавалось $N_\chi=100$, а на более слабом – $N_\chi=50$. Это было сделано для соблюдения адекватности сложности решаемых задач используемым ресурсам.

Как видно из рис. 4–7, время решения задачи на кластере и на GPU, рассчитанное при помощи предложенной модели, почти всегда несколько меньше, чем фактическое время. Это можно объяснить тем, что время, затраченное на синхронизацию и на пересылки данных между процессами кластера и между потоками GPU, значительно варьируется в зависимости от применяемой среды передачи данных и от числа возможных решений N_χ на этапе инициализации рассмотренного метода. При этом, чем больше задействовано процессов кластера или потоков GPU, тем существеннее влияние синхронизаций и пересылок и тем больше отклонение между фактическим и предсказанным временем решения задачи.

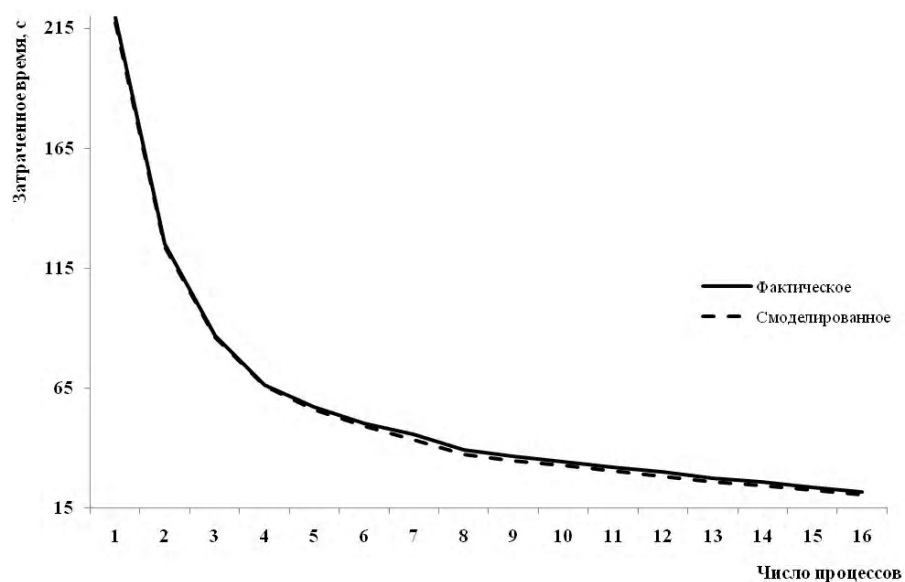


Рисунок 4 – Результаты экспериментов на кластере ИПМЭ ($N_{\chi}=100$)

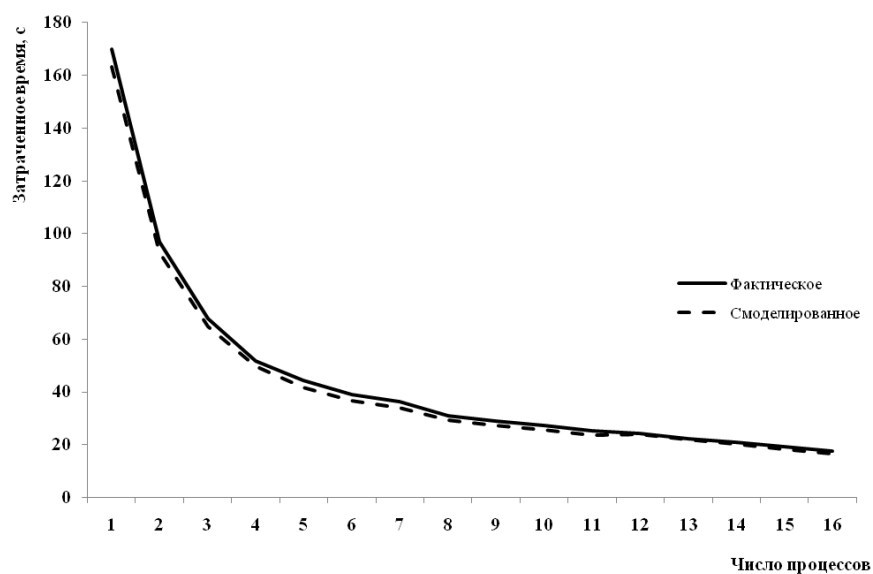


Рисунок 5 – Результаты экспериментов на кластере ЗНТУ ($N_{\chi}=50$)

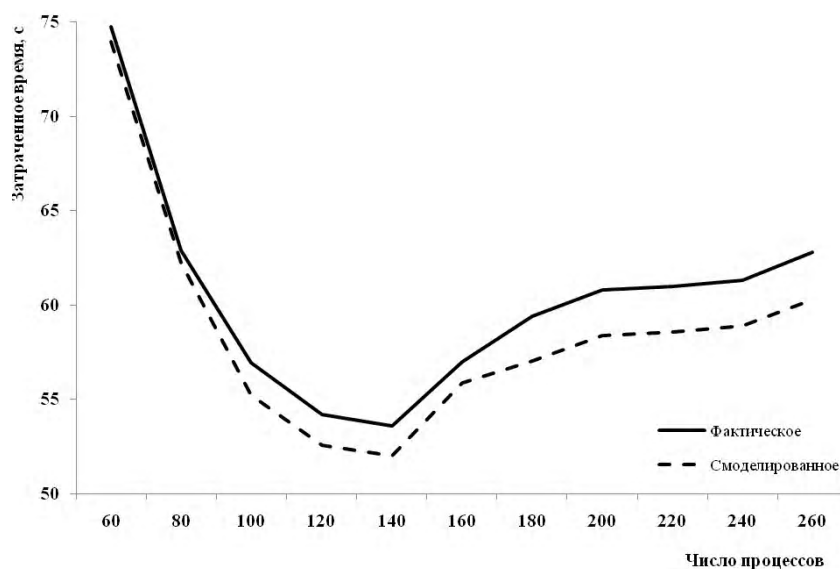
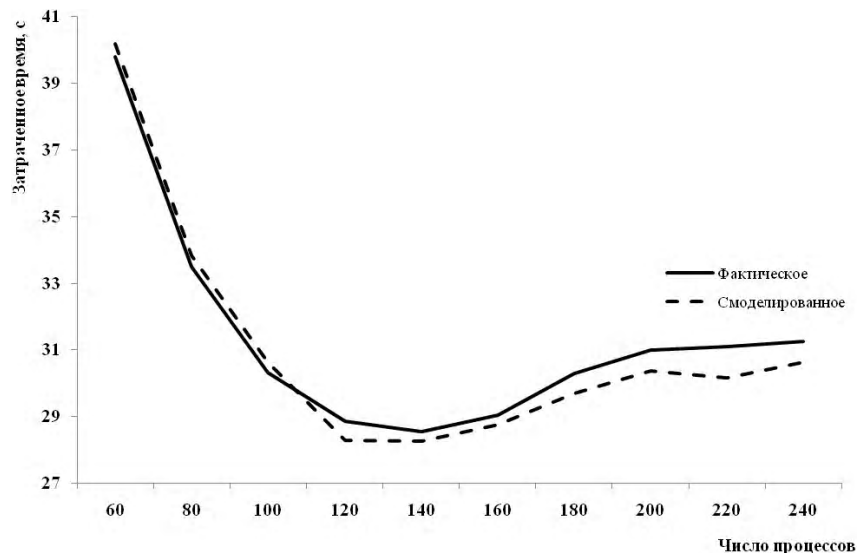


Рисунок 6 – Результаты экспериментов на GPU NVIDIA GTX 960 ($N_{\chi}=100$)

Рисунок 7 – Результаты экспериментов на GPU NVIDIA GTX 285+ ($N_k=50$)

Значение среднеквадратичной ошибки на тестовой выборке составило $7,61 \times 10^{-4}$, что на порядок хуже, чем на обучающей выборке, но остается в пределах приемлемых значений и позволяет рекомендовать предложенную модель на практике для рационального выбора ресурсов компьютерной системы в решении задачи индивидуального прогнозирования состояния больного гипертонической болезнью.

ВЫВОДЫ

В работе повышена эффективность применения параллельных компьютерных систем для синтеза нейро-нечетких сетей и решения задачи индивидуального прогнозирования состояния больного гипертонической болезнью.

Научная новизна заключается в том, что предложена нейросетевая модель, учитывающая тип компьютерной системы, число процессоров, на которых выполняется задача, пропускную способность сети передачи данных и число возможных решений на этапе инициализации метода. Модель дает возможность прогнозировать затраченное параллельной системой время на выполнение метода синтеза нейро-нечетких сетей.

Практическая ценность полученных результатов заключается в разработанном программном обеспечении, которое реализует предложенную модель и позволяет рационально планировать выбор ресурсов компьютерной системы для решения рассмотренной задачи.

БЛАГОДАРНОСТИ

Работа выполнена в рамках научно-исследовательской работы «Дослідження і розробка методів підвищення ефективності комп'ютерних систем та мереж, пошук шляхів удосконалення навчального процесу» кафедри комп'ютерних систем і мереж Запорозького національного технічного університету. Выборка для проведения экспериментального исследования на кластере предоставлена профессором Запорозького национального технического университета С. А. Субботиным.

СПИСОК ЛІТЕРАТУРИ

- Лашкул З. В. Особливості епідеміології артеріальної гіпертензії та її ускладнень на регіональному рівні з 1999 по 2013 роки / З. В. Лашкул // Сучасні медичні технології. – 2014, № 2. – С. 134–141.
- Sergey Subbotin Individual prediction of the hypertensive patient condition based on computational intelligence / Sergey Subbotin, Andrii Oliinyk, Stepan Skrupsky – Information and Digital Technologies 2015 ISBN 978-1-4673-7185-8, 7–9 july 2015, Zilina, Slovakia. – P. 336–344
- Oliinyk A. O. Experimental Investigation with Analyzing the Training Method Complexity of Neuro-Fuzzy Networks Based on Parallel Random Search / Andrii Oliinyk, Stepan Skrupsky, Sergey Subbotin // Automatic Control and Computer Sciences ISSN 0146-4116. – 2015. Vol. 49, No. 1. – P. 11–20. DOI: 10.3103/S0146411615010071
- Субботін С. О. Подання й обробка знань у системах штучного інтелекту та підтримки прийняття рішень : навч. посібник / С. О. Субботін. – Запоріжжя : ЗНТУ, 2008. – 341 с.
- Oliinyk A. O. Using Parallel Random Search to Train Fuzzy Neural Networks / A. O. Oliinyk, S. Yu. Skrupsky, S. A. Subbotin // Automatic Control and Computer Sciences. – 2014. – Vol. 48, Issue 6. – P. 313–323. DOI: 10.3103/S0146411614060078
- Характеристики ГРІД-вузла НАНУ. – Режим доступа: URL: http://www.ipme.kiev.ua/ukr/grid_vuzol/charakter-g.html. – Загл. з екрану.
- Introduction to GPUs. – Режим доступа: URL: <https://www.cs.utexas.edu/~pingali/CS378/2015sp/lectures/IntroGPUs.pdf>. – Загл. з екрану.
- Sulistio A. Simulation of Parallel and Distributed Systems: A Taxonomy - and Survey of Tools / A. Sulistio, C. S. Yeo, R. Buyya // International Journal of Software Practice and Experience. Wiley Press. – 2002. – P. 1–19.
- Методи и модели планирования ресурсов в GRID-системах : монография / [В. С. Пономаренко, С. В. Листровой, С. В. Минухин, С. В. Знахур]. – Х. : ИД «ИНЖЭК», 2008. – 408 с.
- Петренко А. І. Комп'ютерне моделювання ГРІД-систем / А. І. Петренко // Електроніка і зв'язь 5' Тематический выпуск «Електроніка і нанотехнології». – 2010. – С. 40–48.
- Скрупский С. Ю. Имитационные модели распределенных систем компрессии видеoinформации / С. Ю. Скрупский, Р. К. Кудерметов // Наукові праці Донецького національного технічного університету. Серія: «Інформатика, кібернетика та обчислювальна техніка». – 2012. – № 15 (203). – С. 190–202.
- Кучер В. О. Нейросетевая модель выбора эвристики предоставления ресурсов на уровне потока заданий в grid-системе / В. О. Кучер // 14-th International conference on System Analysis and Information Technologies SAIT 2012 Institute for Applied System Analysis, National Technical University of Ukraine «KPI», Kyiv, Ukraine, April 24, 2012 ISBN 978-966-2748-07-9. – P. 287–288.
- Buyya R. Gridsim: a toolkit for the modeling and simulation of distributed resource management and scheduling for grid computing / R. Buyya, M. Murshed // Concurrency and computation: practice and experience. – 2002. – Vol. 14. – P. 1175–1220.
- Руденко О. Г. Штучні нейронні мережі / О. Г. Руденко, Є. В. Бодяньський. – Х. : Компанія СМІТ, 2006. – 404 с.
- Интеллектуальные информационные технологии проектирования автоматизированных систем диагностирования и распознавания об-

- разов : монографія / С. А. Субботин, Ан. А. Олейник, Е. А. Гофман, С. А. Зайцев, Ал. А. Олейник ; под ред. С. А. Субботина. – Харьков : ООО «Компания Смит», 2012. – 317 с.
16. Рассел С. Искусственный интеллект: современный подход / С. Рассел, П. Норвиг. – М. : Вильямс, 2006. – 1408 с.
17. Субботин С. О. Неітеративні, еволюційні та мультиагентні методи синтезу нечіткологічних і нейромережних моделей: монографія С. Ю.

- нографія / С. О. Субботин, А. О. Олійник, О. О. Олійник ; під заг. ред. С. О. Субботина. – Запоріжжя : ЗНТУ, 2009. – 375 с.
18. Quinn M. J. Parallel Programming in C with MPI and OpenMP / M. J. Quinn. – New York, NY : McGraw-Hill, 2004. – 529 p.

Статья поступила в редакцию 01.02.2016.

Канд.техн.наук, доцент кафедри комп'ютерних систем та мереж, Запорізький національний технічний університет, Запоріжжя, Україна

ЕКСПЕРИМЕНТАЛЬНЕ ДОСЛІДЖЕННЯ МЕТОДУ СИНТЕЗУ НЕЙРО-НЕЧІТКИХ МОДЕЛЕЙ В ПАРАЛЕЛЬНІЙ КОМП'ЮТЕРНІЙ СИСТЕМІ

Вирішено задачу розробки нелінійної моделі, що описує залежність між характеристиками системи, в якій здійснюється синтез нейро-нечітких мереж, параметрами досліджуваного методу і часом, що витрачається системою на виконання синтезу моделей. Об'єкт дослідження – процес синтезу нейро-нечітких моделей для індивідуального прогнозування стану хворого гіпертонічною хворобою. Предметом дослідження є паралельна комп'ютерна система, що виконує метод синтезу нейро-нечітких мереж. Мета роботи полягає у підвищенні ефективності використання паралельних комп'ютерних систем для вирішення задач медичного призначення. Запропоновано нелінійну модель, що дозволяє прогнозувати витрачений паралельною системою час на виконання методу синтезу нейро-нечітких мереж і, таким чином, здійснювати раціональний вибір ресурсів комп'ютерної системи. Розроблено програмне забезпечення, що реалізує запропоновану модель. Виконано експерименти, що підтверджують адекватність запропонованої моделі. Результати експериментів дозволяють рекомендувати використання розробленої моделі на практиці.

Ключові слова: синтез моделі, паралельна система, планування ресурсів, нейронна мережа, середньоквадратична помилка.

Skrupsky S. Yu.

PhD, Associate Professor of Computer systems and networks department, Zaporizhzhia National Technical University, Zaporizhzhia, Ukraine

EXPERIMENTAL INVESTIGATION OF METHOD FOR THE SYNTHESIS OF NEURO-FUZZY MODELS IN A PARALLEL COMPUTER SYSTEM

The article deals with the problem of the development of the non-linear model describing dependences between the characteristics of a system, in which synthesis of neuro-fuzzy networks is realized, the parameters of the investigated method and the time spent on execution of the models synthesis. The object of research is a synthesis of neuro-fuzzy models for individual prediction of the hypertensive patient state. The subject of research is a parallel computer system that performs the method of neuro-fuzzy networks synthesis. The purpose of the work is to improve the efficiency of parallel computer systems solving the problems of medical direction. A non-linear model to predict the time used by a parallel system to perform the method of neuro-fuzzy network synthesis and thus to execute a rational choice of the computer system resources has been proposed. The software that implements the proposed model has been developed.

Experiments confirming the adequacy of the proposed model have been executed. The experimental results allow us to recommend the application of the developed model in practice.

Keywords: synthesis of model, parallel system, resource planning, neural network, mean-squared error.

REFERENCES

- Lashkul Z. V. Osoblyvosti epidemiologii arterialnoi hipertenzii ta yii uskladnen na rehionalnomu rivni z 1999 po 2013 roky, *Suchasni medychni tekhnologii*, 2014, No. 2, pp. 134–141.
- Sergey Subbotin, Andrii Oliinyk, Stepan Skrupsky Individual prediction of the hypertensive patient condition based on computational intelligence, *Information and Digital Technologies 2015 ISBN 978-1-4673-7185-8, 7–9 July 2015*. Zilina, Slovakia, pp. 336–344
- Andrii Oliinyk, Stepan Skrupsky, Sergey Subbotin Experimental Investigation with Analyzing the Training Method Complexity of Neuro-Fuzzy Networks Based on Parallel Random Search, *Automatic Control and Computer Sciences ISSN 0146-4116*, 2015, Vol. 49, No. 1, pp. 11–20. DOI: 10.3103/S0146411615010071
- Subbotin S.O. Podannia y obrobka znan u systemakh shtuchnoho intelektu ta pidtrymky pryiniattia rishen : navch. posibnyk. Zaporizhzhia, ZNTU, 2008, 341 p.
- Oliinyk A. O., Skrupsky S. Yu., Subbotin S. A. Using Parallel Random Search to Train Fuzzy Neural Networks, *Automatic Control and Computer Sciences*, 2014, Vol. 48, Issue 6, pp. 313–323. DOI: 10.3103/S0146411614060078
- Kharakterystyky HRID-vuzla NANU. Rezhym dostupa: URL: http://www.ipme.kiev.ua/ukr/grid_vuzol/charakter-g.html. – Zasl. z ekranu.
- Introduction to GPUs. Rezhym dostupa, URL: <https://www.cs.utexas.edu/~pingali/CS378/2015sp/lectures/IntroGPUs.pdf>. Zasl. z ekranu.
- Sulistio A., Yeo C. S., Buyya R. Simulation of Parallel and Distributed Systems: A Taxonomy – and Survey of Tools, *International Journal of Software Practice and Experience*. Wiley Press, 2002, pp. 1–19.
- Ponomarenko S., Listrovoj S. V., Minuxin S. V., Znaxur S. V. Metody i modeli planirovaniya resursov v GRIDsystemax, monografiya. Xark'kov, ID «INZhE'K», 2008, 408 p.
- Petrenko A. I. Kompiuterne modeliuvannia HRID-system, *E'lektronika i svyaz' 5' Tematicheskij vypusk «E'lektronika i nanotekhnologii»*, 2010, pp. 40–48.
- Skrupskij S. Yu., Kudermetov R. K. Imitacionnye modeli raspredelennykh sistem kompressii videoinformacii, *Naukovi pratsi Donetskoho natsionalnoho tekhnichnoho universytetu. Seriya: «Informatyka, kibernetika ta obchysluvalna tekhnika»*, 2012, No. 15 (203), pp. 190–202.
- Kucher V. O. Nejrosetevaya model' vybora e'vristiki predstavleniya resursov na urovne potoka zadaniy v grid-sisteme, *14-th International conference on System Analysis and Information Technologies SAIT 2012 Institute for Applied System Analysis, National Technical University of Ukraine «KPI»*. Kyiv, Ukraine, April 24, 2012 ISBN 978-966-2748-07-9, pp. 287–288.
- Buyya R., Murshed M. Gridsim: a toolkit for the modeling and simulation of distributed resource management and scheduling for grid computing, *Concurrency and computation: practice and experience*, 2002, Vol. 14, pp. 1175–1220.
- Rudenko O. H., Bodianskyi Ye. V. Shtuchni neuronni merezhi. Xark'kov, Kompaniia SMIT, 2006, 404 p.
- Subbotin S. A., Olejnik An. A., Gofman E. A, Zajcev S. A., Olejnik Al. A.; pod red. S. A. Subbotina Intel'ktual'nye informacionnye tekhnologii proektirovaniya avtomatizirovannykh sistem diagnostirovaniya i raspoznavaniya obrazov : monografiya. Xark'kov, ООО «Компания Смит», 2012, 317 p.
- Rassel S., Norvig P. Iskusstvennyj intellekt: sovremennyy podhod. Moscow, Vil'yams, 2006, 1408 p.
- Subbotin S. O., Oliinyk A. O., Oliinyk O. O.; pid zah. red. S. O. Subbotina Neiterativni, evoliutsiini ta multyahentni metody syntezy nechitkolohichnykh i neiromerezhnykh modelei: monohrafiya. Zaporizhzhia, ZNTU, 2009, 375 p.
- Quinn M. J. Parallel Programming in C with MPI and OpenMP. New York, NY, McGraw-Hill, 2004, 529 p.

МЕТОД СИНТЕЗА ДИАГНОСТИЧЕСКИХ МОДЕЛЕЙ НА ОСНОВЕ РАДИАЛЬНО-БАЗИСНЫХ НЕЙРОННЫХ СЕТЕЙ С ПОДДЕРЖКОЙ ОБОБЩАЮЩИХ СВОЙСТВ

В работе решена актуальная проблема автоматизации синтеза радиально-базисных нейронных сетей на основе набора прецедентов для принятия решений в диагностировании. Предложен метод синтеза радиально-базисных нейронных сетей, который формирует в начале по одному эталону класса, которые при необходимости дополняет новыми эталонами, формируемыми на основе ошибочно распознанных экземпляров, а далее оперирует расстояниями от экземпляров до эталонов кластеров. На основе полученных координат эталонов далее в автоматическом режиме синтезируется структура и настраиваются параметры сети, которые дополнительно для повышения обобщающих свойств и интерпретируемости подвергается контрастированию весов. Предложенный метод не требует задания пользователем числа кластеров, не имеет неопределенности выбора числа нейронов в первом слое и выбора начальных значений весов сети, стремится минимизировать размер сети, характеризуется приемлемым временем обучения, благодаря использованию процедуры оптимизации сети позволяет получать безыбыточные контрастные, интерпретируемые нейромодели. Разработано программное обеспечение, реализующее предложенный метод, а также проведены эксперименты, подтвердившие работоспособность разработанного математического обеспечения и позволяющие рекомендовать его для использования на практике при решении задач построения диагностических моделей по прецедентам для автоматизации принятия решений в технической и биомедицинской диагностике.

Ключевые слова: нейронная сеть, радиально-базисная сеть, обучение, синтез, диагностика.

НОМЕНКЛАТУРА

РБНС – радиально-базисная нейронная сеть;

C – множество эталонов;

C^q – q -й эталон;

C_j^q – j -я координата эталона q -го кластера;

E – среднеквадратическая ошибка;

$f()$ – пользовательский критерий, характеризующий качество аргумента относительно решаемой задачи;

$F()$ – структура РБНС;

F_g – показатель обобщения нейромодели;

K – число классов;

M – число слоев РБНС ($M=2$);

N – число диагностических признаков;

N_M – число выходных признаков;

N_w – число параметров (весов) РБНС;

$N_{w=0}$ – число весов РБНС, равных нулю;

N_3 – число узлов в 3-м слое;

opt – оптимальное (желаемое или приемлемое) значение функционала $f()$ для решаемой задачи;

Q – число кластеров;

$R(x^s, C^q)$ – расстояние от экземпляра x^s до эталона C^q ;

S' – число экземпляров в конфликтном наборе;

S – число прецедентов в выборке;

S^q – число экземпляров, попавших в q -й кластер;

S'^q – число экземпляров q -го класса в конфликтном наборе;

$t_{об.}$ – время обучения модели;

w – параметры РБНС;

w – множество управляемых (настраиваемых) параметров (весов) РБНС;

$w_j^{(3,i)}$ – значение j -го настраиваемого параметра или веса j -го входа i -го узла 3-го слоя;

$x_j^{\min}, x_j^{\max}$ – минимальное и максимальное значения j -го признака, соответственно;

x_j^s – значение j -го диагностического признака x_j^s , характеризующее прецедент (экземпляр) x^s ;

x'^s – s -й экземпляр конфликтного набора;

x'^s_j – значение j -го диагностического признака x_j^s , характеризующее прецедент (экземпляр) x'^s ;

X' – конфликтный набор;

y^s – значение выходного признака, сопоставленное прецеденту x^s ;

y^{s*} – расчетное значение на i -м выходе РБНС для экземпляра x^s , поданного на ее входы;

y^s_i – значение i -го выходного признака для экземпляра x^s ;

$y^{s}_{расч.}$ – расчетный номер класса для s -го экземпляра относительно сформированного множества эталонов;

β_q – коэффициент, регулирующий величину q -го кластера;

$\varphi^{(\eta,i)}, \psi^{(\eta,i)}$ – соответственно, дискриминантная (весовая) и активационная функции i -го узла η -го слоя.

ВВЕДЕНИЕ

Искусственные нейронные сети [3–11] благодаря своим способностям способностями обучаться по примерам, посредством обобщения и неявного извлечения знаний из данных, отображая их в структуру и параметры (веса) сети, получили широкое распространение на практике при решении задач диагностирования и распознавания образов.

Объектом исследования являлся процесс синтеза искусственных нейронных сетей.

Среди известных моделей нейронных сетей и методов их обучения [3–11] для задач автоматизации приня-

тия решений целесообразно использовать преимущественно сети прямого распространения сигнала, которые не содержат боковых и обратных связей и легче поддаются последующему анализу [6, 9, 10]. В свою очередь, среди сетей данного класса особо можно выделить радиально-базисные нейронные сети (РБНС), которые легко интерпретируются в терминологии кластерного анализа, что упрощает последующий анализ полученной на их основе модели или ее решений [3, 6, 11, 12].

Предметом исследования являлись методы построения РБНС.

Известные методы синтеза РБНС обладают такими недостатками, как зависимость качества получаемой модели от заданных человеком параметров, а также низкий уровень обобщения получаемых моделей и большие затраты времени на процесс построения моделей.

Целью работы являлось создание метода структурно-параметрического синтеза РБНС, обеспечивающего сокращение вовлечения человека в процесс построения нейромодели, а также гарантирующего поддержку обобщающих свойств модели.

1 ПОСТАНОВКА ЗАДАЧИ

Пусть задана обучающая выборка прецедентов $X = \langle x, y \rangle$, где $x = \{x^s\}$, $y = \{y^s\}$, $x = \{x_j\}$, $x^s = \{x_j^s\}$, $s = 1, 2, \dots, S$, $j = 1, 2, \dots, N$.

Тогда задача структурно-параметрического синтеза РБНС заключается в том, чтобы получить $\langle f(), w \rangle$: $y^{s*} = F(w, x^s), f(F(), w, \langle x, y \rangle) \rightarrow opt$.

Здесь $\langle F(), w \rangle$ – диагностическая модель на основе РБНС, которая задается кортежем $\langle M, \{N_{\eta}\}, \{y^{(\eta,i)}(x^{(\eta,i)})\} \rangle$ и описывается функционально формулами:

$$y_i^s = y^{(M,i)}(y^{(M-1,i)}(\dots y^{(1,i)}(x^s))), \quad y^{(\eta,i)} = \psi^{(\eta,i)}(\varphi^{(\eta,i)}),$$

$$i = 1, 2, \dots, N_{\eta-1}; \quad w^{(\eta,i)} = \{w_j^{(\eta,i)}\}, \quad w = \{w^{(\eta,i)}\} = \{w_j^{(\eta,i)}\},$$

$$y^{(0,i)} = \psi^{(0,i)} = x_j^s, \quad N_0 = N, \quad x_j^{(1,i)} = x_j^s,$$

$$i = 1, 2, \dots, N_{\eta-1}, \quad \eta = 1, 2, \dots, M, \quad j = 1, 2, \dots, N;$$

$$y^s = \{y_{ij}^s\}, y^{s*} = \psi^{(M,i)}(x^s), i = 1, 2, \dots, N_M.$$

Для задач классификации: $y^s \in \{q\}$, $q = 1, 2, \dots, K$, $K > 1$.
В простейшем случае f определяют: $f = E$, где

$$E = \frac{1}{2} \sum_{s=1}^S (y^s - \psi^{(M,i)}(x^s))^2 \rightarrow \min.$$

2 ЛИТЕРАТУРНЫЙ ОБЗОР

Известные методы построения РБНС [8, 10–12] можно разделить на две группы.

Методы первой группы предполагают, что число нейронов РБНС в первом слое задается пользователем, а число нейронов во втором слое определяется размерностью выхода сети. После чего РБНС рассматривается как частный случай многослойной нейронной сети прямого распространения сигнала (многослойного персептрона), которую обучают с помощью градиентных методов многомерной нелинейной безусловной оптими-

зации [8, 12], а частные производные целевой функции определяют на основе техники обратного распространения ошибки [8, 10, 11]. При этом существует неопределенность выбора числа нейронов в первом слое, что может привести к получению избыточной сети или, наоборот, к невозможности построения модели, обладающей требуемой точностью. Другим недостатком данных методов является неопределенность выбора начальных значений весов сети, что может привести к невозможности решить задачу обучения РБНС за ограниченное время.

Методы второй группы предполагают, в структуру РБНС отображается обучающая выборка. Это, как правило, реализуется простым запоминанием всей выборки, либо на основе кластер-анализа, в результате которого определяется число и координаты центры кластеров, которые заносятся в память РБНС. Число нейронов в первом слое РБНС задают равным числу кластеров, а в веса нейронов первого слоя заносят координаты центров кластеров. Далее РБНС обучают, корректируя веса нейронов второго слоя на основе методов многомерной нелинейной безусловной оптимизации [8, 12]. Если сеть получают путем отображения выборки, то она, как правило, не обеспечивает обобщение. Если же используют кластер-анализ, то качество полученной нейромодели существенно зависит от качества результатов кластер-анализа, сеть может оказаться избыточной и проявлять низкие обобщающие свойства из-за чрезмерной детализации разбиения признакового пространства.

Поэтому необходимо разработать метод структурно-параметрического синтеза РБНС, обеспечивающего сокращение вовлечения человека в процесс построения нейромодели, а также гарантирующего поддержку обобщающих свойств модели.

3 МАТЕРИАЛЫ И МЕТОДЫ

Прямое отображение выборки в структуру РБНС не обеспечивает обобщающих свойств. Для того, чтобы нейросетевая модель обладала обобщающими свойствами необходимо, чтобы число ее параметров (весов) N_w было меньше размерности обучающей выборки NS .

Число весов в радиально-базисной сети определяется числом кластеров Q , на которые разбивается обучающая выборка $\langle x, y \rangle$, числом признаков N и числом классов K и рассчитывается по формуле: $N_w = (N+1)Q + (Q+1)K$.

Здесь первое слагаемое определяет число параметров нейронов первого слоя РБНС, каждый из которых в весах хранит координаты центра соответствующего кластера, а второе слагаемое определяет параметры связей между нейронами первого и второго слоя.

Исходя из этого выражения, определим Q при заданных N , K и S :

$$(N+1)Q + (Q+1)K < NS \Rightarrow Q < \frac{NS - K}{N + K + 1}.$$

В то же время, очевидно, что число кластеров Q должно быть не меньше, чем число классов K . Тогда получим диапазон допустимых значений числа кластеров:

$$K \leq Q < \frac{NS - K}{N + K + 1}.$$

Для того, чтобы обеспечить высокие обобщающие свойства нейромодели, необходимо в процессе ее построения стремиться минимизировать число используемых кластеров и контролировать соблюдение приведенного выше условия при добавлении новых эталонов.

С другой стороны, многие известные методы кластер-анализа являются высоко итеративными и требуют расчета расстояний между всем экземплярами выборки, а затем перебора большого числа различных вариантов разбиений выборки, что приводит к большим затратам времени на обучение, а также может требовать начального задания числа кластеров пользователем. Для устранения данного недостатка предлагается вначале формировать по одному эталону класса (предполагая классы компактными и состоящими из одного кластера), которые при необходимости дополнять новыми эталонами, формируемыми на основе ошибочно распознанных экземпляров. Это позволит сразу найти грубое решение задачи, не выполняя расчета расстояний между всеми экземплярами выборки, а далее оперировать расстояниями от экземпляров до эталонов кластеров, что позволит сократить время расчетов.

На основе полученных координат эталонов предлагается в автоматическом режиме синтезировать РБНС, которую затем для повышения обобщающих свойств и интерпретируемости предлагается отконтрастировать, удалив малозначимые связи между нейронами.

На основе данных идей формально изложим метод синтеза РБНС.

Этап 1. Инициализация. Задать допустимый порог ошибки ε , $0 \leq \varepsilon \ll 1$, а также обучающую выборку прецедентов $\langle x, y \rangle$. Определить $x_j^{\min}, x_j^{\max}, j=1, 2, \dots, N$. Положить $Q=K$. Сформировать Q эталонов классов, координаты центров которых определяются по формуле:

$$C_j^q = \frac{1}{S^q} \sum_{j=1}^S \{x_j^s | y^s = q\}, q=1, 2, \dots, K; j=1, 2, \dots, N.$$

Этап 2. Контроль числа кластеров. Если $Q-1 < \left\lfloor \frac{NS-K}{N+K+1} \right\rfloor$, т.е. имеется возможность добавить еще один кластер при сохранении обобщающих свойств, тогда перейти к этапу 3, в противном случае – к этапу 4.

Этап 3. Выделение конфликтного набора экземпляров. Относительно имеющегося множества эталонов $C=\{C^q\}$, $C^q=\{C_j^q\}$, для экземпляров выборки определить расстояния до эталонов:

$$R(x^s, C^q) = R(C^q, x^s) = \sqrt{\sum_{j=1}^N (x_j^s - C_j^q)^2}, s=1, 2, \dots, S; q=1, 2, \dots, Q.$$

Определить расчетный номер класса для каждого экземпляра относительно сформированного множества эталонов:

$$y_{расч.}^s = \arg \min_{q=1, 2, \dots, Q} \{R(x^s, C^q)\}.$$

Все экземпляры, для которых $y^s \neq y_{расч.}^s$, выделить в конфликтный набор:

$$X' = \bigcup_{s=1}^S \{x^s, y^s | y^s \neq y_{расч.}^s\}.$$

Для экземпляров конфликтного набора определить общее число экземпляров S' , а также число экземпляров в q -м классе S'^q .

Этап 4. Расширение набора эталонов. Если достигнута требуемая точность ($S'/S < \varepsilon$), тогда перейти к этапу 5, в противном случае – положить $Q=Q+1$ и сформировать новый эталон для наиболее частотного класса в конфликтном наборе:

$$C_j^Q = \frac{1}{S'^Q} \sum_{j=1}^{S'} \{x_j^s | y^s = q\},$$

после чего перейти к этапу 2.

Этап 5. Определить значение коэффициента, регулирующего величину каждого кластера:

$$\beta_q = \frac{S^q}{\max_{\substack{q=1, 2, \dots, Q; \\ p=q+1, \dots, Q}} \left\{ \sqrt{\sum_{j=1}^N (C_j^q - C_j^p)^2} \right\}}, q=1, 2, \dots, Q.$$

Этап 6. Синтез структуры РБНС. Задать на первом (скрытом) слое РБНС Q нейронов, использующих в качестве весовых (постсинаптических функций) функции расстояния

$$\varphi^{(1,i)}(x^{(1,i)}, w^{(1,i)}) = w_0^{(1,i)} \sqrt{\sum_{j=1}^N (x_j^{(1,i)} - w_j^{(1,i)})^2},$$

где $w_0^{(1,i)}$ – параметр, регулирующий ширину радиально-базисной функции i -го нейрона первого слоя, а в качестве активационных функций – функции Гаусса: $\psi^{(1,i)}(\varphi^{(1,i)}) = \exp(-\varphi^{(1,i)})$.

На втором (выходном) слое РБНС задать число нейронов, равное размерности выходного вектора. Определить в качестве весовых функций нейронов второго слоя сети взвешенные суммы

$$\varphi^{(2,i)}(x^{(2,i)}, w^{(2,i)}) = \sum_{j=1}^{N_M} w_j^{(2,i)} x_j^{(2,i)} + w_0^{(2,i)},$$

где $w_0^{(1,i)}$ – порог, а в качестве активационных функций – линейные функции: $\psi^{(2,i)}(\varphi^{(2,i)}) = \varphi^{(2,i)}$.

Этап 6. Настройка параметров РБНС.

Занести в j -й вес q -го нейрона первого слоя РБНС значение j -й координаты центра q -го кластера: $w_j^{(1,q)} = C_j^q, j=1, 2, \dots, N, q=1, 2, \dots, Q$. Занести в нулевые веса нейронов первого слоя параметры, регулирующие ширину кластеров: $w_0^{(1,q)} = \beta_q, q=1, 2, \dots, Q$.

Определить значения весов нейронов второго слоя РБНС путем решения системы линейных алгебраических уравнений вида $y_i = \psi^{(1,i)}(\varphi^{(1,i)}(x))w^{(2,i)}$, где $y_i = (\{y_i^s\})^T$ – вектор выходных значений, $w^{(2,i)}$ – вектор весов i -го выходного нейрона [9], либо на основе обратного распространения ошибки [8, 10, 11].

Этап 7. Оптимизация (контрастирование) РБНС. Выполнить распознавание обучающей выборки с помощью построенной РБНС и определить ошибку сети E . До тех пор, пока ошибка является приемлемой ($E \leq \varepsilon$) выполнять последовательно в цикле контрастирование весов второго слоя сети: среди весов второго слоя определить вес с наименьшим абсолютным значением, неравным нулю, положить его равным нулю, выполнить распознавание обучающей выборки, оценить ошибку распознавания E . В результате выполнения этих действий часть весов будет обнулена, т.е. можно считать, что будут удалены соответствующие связи, что позволит повысить интерпретируемость полученной модели, а также ее уровень обобщения по отношению к исходным данным и дообучение.

Предложенный метод позволяет автоматизировать процесс синтеза РБНС, сокращая зависимость от пользователя, а также обеспечивает поддержание приемлемых обобщающих свойств сети при приемлемой точности.

4 ЭКСПЕРИМЕНТЫ

Для проверки работоспособности и практической применимости предложенного метода он был программно реализован как дополнение к автоматизированной системе диагностирования [13].

Таблица 1 – Характеристики решавшихся задач

Задача	Характеристики задачи		
	N	S	K
Диагностирование лопаток газотурбинных авиадвигателей по спектрам свободных затухающих колебаний после ударного возбуждения	100	32	2
Прогнозирование прочности лопаток газотурбинных авиадвигателей после упрочняющей обработки	16	57	2
Диагностирование хронического обструктивного бронхита	28	205	2

Разработанная программа использовалась для построения диагностических моделей в задачах технического и медицинского диагностирования [13]: диагностирование лопаток газотурбинных авиадвигателей по спектрам свободных затухающих колебаний после ударного возбуждения, прогнозирование прочности лопаток газотурбинных авиадвигателей после упрочняющей обработки, диагностирование хронического обструктивного бронхита. Характеристики решавшихся задач приведены в табл. 1.

5 РЕЗУЛЬТАТЫ

В табл. 2. приведены результаты, полученные в ходе вычислительных экспериментов.

Для оценивания обобщающих свойств полученных моделей в таблице использован показатель:

$$F_g = \frac{NS}{N_w - N_{w=0}}.$$

Чем больше будет значение данного показателя, тем выше уровень обобщения сетью данных обучающей выборки, тем меньше и компактнее модель, тем она более интерпретируемая.

6 ОБСУЖДЕНИЕ

Как видно из табл. 1, предложенный метод характеризуется меньшими затратами времени на построение РБНС по сравнению с использованием четкого переборного кластерного анализа и по сравнению с отображением выборки в кластеры. При этом предложенный метод обеспечивает в среднем более высокую точность моделей и существенно более высокий уровень обобщения по сравнению с известными методами.

Более высокий уровень обобщения моделей, синтезированных на основе предложенного метода, обеспечивается тем, что кроме обобщения данных эталонов в процессе кластер-анализа, после синтеза сети ее веса контрастируются, а структура сети оптимизируется путем удаления малозначимых связей между нейронами. Это позволяет упростить сеть, снизив ее структурную и параметрическую сложность, а также повысить интерпретируемость нейромодели.

ВЫВОДЫ

В работе решена актуальная проблема автоматизации синтеза РБНС на основе набора прецедентов для принятия решений в диагностировании.

Научная новизна работы состоит в том, что впервые предложен метод синтеза РБНС, который, в отличие от

Таблица 2 – Характеристики методов синтеза РБНС

Задача	Методы синтеза РБНС								
	Прямое отображение выборки в кластеры, обратное распространение ошибки			Кластер-анализ, обратное распространение ошибки			Предложенный метод		
	$t_{об., с}$	E	F_g	$t_{об., с}$	E	F_g	$t_{об., с}$	E	F_g
Диагностирование лопаток	1,3	$0,98 \cdot 10^{-6}$	3,48	4,9	$0,94 \cdot 10^{-6}$	4,05	3,73	$0,92 \cdot 10^{-6}$	4,63
Прогнозирование прочности лопаток	0,8	$0,96 \cdot 10^{-6}$	3,37	3,1	$0,89 \cdot 10^{-6}$	3,62	2,95	$0,83 \cdot 10^{-6}$	4,21
Диагностирование бронхита	1,6	$0,91 \cdot 10^{-6}$	4,07	6,2	$0,93 \cdot 10^{-6}$	5,15	4,96	$0,94 \cdot 10^{-6}$	6,88

известных методов, не требует задания пользователем числа кластеров, не имеет неопределенности выбора числа нейронов в первом слое и выбора начальных значений весов сети, стремится минимизировать размер сети, характеризуется приемлемым временем обучения, благодаря использованию процедуры оптимизации сети позволяет получать безызыточные контрастные, интерпретабельные нейромодели.

Практическая значимость результатов работы заключается в том, что разработано программное обеспечение, реализующее предложенный метод, а также проведены эксперименты, подтвердившие работоспособность разработанного математического обеспечения и позволяющие рекомендовать его для использования на практике при решении задач построения диагностических моделей по прецедентам для автоматизации принятия решений в технической и биомедицинской диагностики.

Перспективы дальнейших исследований состоят в исследовании предложенного математического обеспечения на более широком наборе практических задач диагностики и распознавания образов.

БЛАГОДАРНОСТИ

Работа выполнена в рамках в рамках госбюджетных научно-исследовательских тем Запорожского национального технического университета «Интеллектуальные информационные технологии диагностирования и автоматической классификации» и «Интеллектуальные методы диагностирования систем управления удаленными техническими объектами» при поддержке международного проекта «Centers of Excellence for young REsearchers» программы «Темпус» Европейской Комиссии (№ 544137-TEMPUS-1-2013-1-SK-TEMPUS-JPHES).

Субботін С. О.

Д-р техн. наук, професор, професор кафедри програмних засобів Запорізького національного технічного університету, Запоріжжя, Україна

МЕТОД СИНТЕЗУ ДІАГНОСТИЧНИХ МОДЕЛЕЙ НА ОСНОВІ РАДІАЛЬНО-БАЗИСНИХ НЕЙРОННИХ МЕРЕЖ З ПІДТРИМКОЮ УЗАГАЛЬНЮВАЛЬНИХ ВЛАСТИВОСТЕЙ

У роботі вирішено актуальну проблему автоматизації синтезу радіально-базисних нейронних мереж на основі набору прецедентів для прийняття рішень у діагностуванні. Запропоновано метод синтезу радіально-базисних нейронних мереж, що формує на початку по одному еталону класу, які за необхідності доповнює новими еталонами, формованими на основі помилково розпізнаних екземплярів, а далі оперує відстанями від екземплярів до еталонів кластерів. На основі отриманих координат еталонів далі в автоматичному режимі синтезується структура і налаштовуються параметри мережі, які додатково для підвищення узагальнювальних властивостей та інтерпретабельності піддається контрастуванню. Запропонований метод не вимагає задавання користувачем числа кластерів, не має невизначеності вибору кількості нейронів у першому шарі та вибору початкових значень ваг мережі, прагне мінімізувати розмір мережі, характеризується прийнятним часом навчання, завдяки використанню процедури оптимізації мережі дозволяє одержувати безнадлишкові контрастні, інтерпретабельні нейромоделі. Розроблено програмне забезпечення, яке реалізує запропонований метод, а також проведені експерименти, що підтвердили працездатність розробленого математичного забезпечення і дозволяють рекомендувати його для використання на практиці при вирішенні задач побудови діагностичних моделей за прецедентами для автоматизації прийняття рішень у технічній і біомедицинській діагностиці.

Ключові слова: нейронна мережа, радіально-базисна мережа, навчання, синтез, діагностика.

Subbotin S. A.

Dr.Sc., Professor, Professor of department of software tools, Zaporizhzhya National Technical University, Zaporizhzhya, Ukraine

THE METHOD OF DIAGNOSTIC MODEL SYNTHESIS BASED ON RADIAL BASIS NEURAL NETWORKS WITH THE SUPPORT OF GENERALIZATION PROPERTIES

Urgent problem of automation of radial basis neural network synthesis based on a set of precedents for decision-making in the diagnosis is solved in the paper. The method for the synthesis of radial basis neural network is proposed. It forms at the beginning one class pattern, which, if necessary, supplemented with new patterns formed on the basis of wrongly recognized instances, and then operates with the distance from the instances to the patterns of the clusters. On the basis of the obtained pattern coordinates it further automatically synthesizes structure and

СПИСОК ЛІТЕРАТУРИ

1. Intelligent fault diagnosis and prognosis for engineering systems / [G. Vachtsevanos, F. Lewis, M. Roemer et al.]. – New Jersey: John Wiley & Sons, 2006. – 434 p.
2. Price C. Computer based diagnostic systems / C. Price. – London: Springer, 1999. – 136 p.
3. Дли М. И. Нечеткая логика и искусственные нейронные сети / М. И. Дли. – М.: Физматлит, 2003. – 225 с.
4. Рутковская Д. Нейронные сети, генетические алгоритмы и нечеткие системы / Д. Рутковская, М. Пилинский, Л. Рутковский; пер. с польск. И. Д. Рудинского. – М.: Горячая линия – Телеком, 2004. – 452 с.
5. Rao V. B. C++ neural networks and fuzzy logic / V. B. Rao. – Foster City: M&T Books, 1995. – 549 p.
6. Bishop C. Neural networks for pattern recognition / C. Bishop. – New York: Oxford University Press, 1995. – 482 p.
7. Круглов В. В. Искусственные нейронные сети. Теория и практика / В. В. Круглов, В. В. Борисов. – М.: Горячая линия – Телеком, 2001. – 382 с.
8. Осовский С. Нейронные сети для обработки информации / С. Осовский. – М.: Финансы и статистика, 2004. – 344 с.
9. Руденко О. Г. Штучні нейронні мережі / О. Г. Руденко, Є. В. Бодяньський. – Харків: Компанія СМІТ, 2006. – 404 с.
10. Хайкин С. Нейронные сети: полный курс / С. Хайкин. – СПб: Вильямс, 2005. – 1104 с.
11. Субботін С. О. Нейронні мережі: навч. посіб. / С. О. Субботін, А. О. Олійник; за ред. С. О. Субботіна. – Запоріжжя: ЗНТУ, 2014. – 132 с.
12. Олійник А. О. Інтелектуальний аналіз даних: навчальний посібник / А. О. Олійник, С. О. Субботін, О. О. Олійник. – Запоріжжя: ЗНТУ, 2012. – 278 с.
13. Интеллектуальные информационные технологии проектирования автоматизированных систем диагностирования и распознавания образов: монография / [С. А. Субботин, Ан. А. Олейник, Е. А. Гофман и др.]; под ред. С. А. Субботина. – Харьков: ООО «Компания Смит», 2012. – 317 с.

Статья поступила в редакцию 21.01.2016.

adjust the weights of the network, which is further optimized to improve the generalizing and interpretability properties by weights contrasting. The proposed method does not require the user specify the number of clusters, has no uncertainty in selection the number of neurons in the first layer and in the choice of the initial values of the network weights, seeks to minimize the size of the network, and characterized by an acceptable time of learning through the use of network optimization procedure allows to obtain nonredundant, contrast, and interpretable neural models. The software implementing proposed method has been developed. The experiments confirming efficiency of developed software have been conducted. They allow to recommend the proposed method for use in practice in solving the problems of diagnostic model constructing by precedents to automate the decision-making in technical and biomedical diagnostics.

Keywords: neural network, radial base network, training, synthesis, diagnostics.

REFERENCES

1. Vachtsevanos G., Lewis F., Roemer M. et al Intelligent fault diagnosis and prognosis for engineering systems. New Jersey, John Wiley & Sons, 2006, 434 p.
2. Price C. Computer based diagnostic systems. London, Springer, 1999, 136 p.
3. Dli M. I. Nechotkaya logika i iskusstvennyye neyronnyye seti. Moscow, Fizmatlit, 2003, 225 p.
4. Rutkovskaya D., Pilin'skiy M., Rutkovskiy L.; per. s pol'sk. I. D. Rudinskogo Neyronnyye seti, geneticheskiye algoritmy i nechotkiye sistemy. Moscow, Goryachaya liniya, Telekom, 2004, 452 p.
5. Rao V. B. C++ neural networks and fuzzy logic. Foster City: M&T Books, 1995, 549 p.
6. Bishop C. Neural networks for pattern recognition. New York, Oxford University Press, 1995, 482 p.
7. Kruglov V. V., Borisov V. V. Iskusstvennyye neyronnyye seti. Teoriya i praktika. Moscow, Goryachaya liniya, Telekom, 2001, 382 p.
8. Osovskiy S. Neyronnyye seti dlya obrabotki informatsii. Moscow, Finansy i statistika, 2004, 344 p.
9. Rudenko O. H., Bodyanskyy YE. V. Shtuchni neyronni merezhi. Kharkiv, Kompaniya SMIT, 2006, 404 p.
10. Khaykin S. Neyronnyye seti: polnyy kurs. Sankt-Peterburg, Vil'yams, 2005, 1104 p.
11. Subbotin S. O., Oliynyk A. O.; za red. S. O. Subbotina. Neyronni merezhi : navch. posib. Zaporizhzhya, ZNTU, 2014, 132 p.
12. Oliynyk A. O., Subbotin S. O., Oliynyk O. O. Intelektualnyy analiz danykh : navchal'nyy posibnyk. Zaporizhzhya, ZNTU, 2012, 278 p.
13. Subbotin S. A., Oleynik An. A., Gofman Ye. A. i dr. pod red. S. A. Subbotina Intelektual'nyye informatsionnyye tekhnologii proyektirovaniya avtomatizirovannykh sistem diagnostirovaniya i raspoznavaniya obrazov : monografiya. Khar'kov, OOO «Kompaniya Smit», 2012, 317 p.

КРИТЕРІЇ НАВЧАННЯ НЕЧІТКОГО КЛАСИФІКАТОРА НА ОСНОВІ ВІДСТАНИ МІЖ ГОЛОВНИМИ КОНКУРЕНТАМИ

Класифікація – це віднесення об'єкта за деякими ознаками до одного з класів. До класифікації зводяться різноманітні задачі прийняття рішень в інженерії, економіці, медицині, соціології та в інших областях. В нечітких класифікаторах залежність «входи – вихід» описуються за допомогою лінгвістичних правил <Якщо – тоді>, антецеденти яких містять нечіткі терми «низький», «середній», «високий» тощо. Для підвищення безпомилковості нечіткий класифікатор навчають за експериментальними даними. В даній роботі запропоновано нові критерії навчання нечіткого класифікатора, які враховують різницю належностей нечіткого висновку лише до головних конкурентів. За правильної класифікації головним конкурентом прийнятого рішення є клас, що має другий за величиною ступінь належності. У випадку неправильної класифікації помилково прийняте рішення є головним конкурентом правильного класу. Проведені комп'ютерні експерименти із навчання нечіткого класифікатора для розпізнавання трьох сортів італійських вин засвідчили суттєву перевагу нових критеріїв. Серед нових критеріїв помірну перевагу має критерій на основі квадратичної відстані між головними конкурентами з штрафом за помилкове рішення. Нові критерії можуть застосовуватися не лише для навчання нечітких класифікаторів, але і для навчання деяких інших моделей, наприклад, нейронних мереж.

Ключові слова: класифікація, нечітка база знань, навчання, критерії навчання, головні конкуренти.

НОМЕНКЛАТУРА

$\tilde{a}_{ij}$ – нечіткий терм, яким оцінюється ознака x_i в j -му правилі, $i = \overline{1, n}$, $j = \overline{1, k}$;

$\Delta_r(\mathbf{K})$ – мітка коректності класифікації r -го об'єкту за нечіткою моделлю з параметрами $\mathbf{K}$;

$\mu_j(X^*)$ – ступінь виконання j -го правила з бази знань;

$\mu_{l_s}(y^*)$ – ступінь належності вхідного вектору $\mathbf{X}^*$ до класу l_s ;

$\mu_j(x_i^*)$ – ступінь належності значення x_i^* нечіткому терму $\tilde{a}_{ij}$;

$\mu_{l_s}(y_r)$ – ступінь належності r -го об'єкту навчальної вибірки до класу l_s ;

$\mu_{l_s}(\mathbf{K}, \mathbf{X}_r)$ – ступінь належності висновку до класу l_s , який розраховано за нечіткою моделлю з параметрами $\mathbf{K}$ для вхідного вектора $\mathbf{X}_r$;

$\mu_{win}(\mathbf{X}_r)$ – ступінь належності вхідного вектора $\mathbf{X}_r$ класу переможцю;

$\mu_{vicewin}(\mathbf{X}_r)$ – ступінь належності вхідного вектора $\mathbf{X}_r$ до класу з другим рангом;

$\mu(x)$ – функція належності для змінної x ;

$\wedge$ – t -норма;

b – координата максимуму гаусової функції належності;

c – коефіцієнт концентрації гаусової функції належності.

$Crit_1$ – перший критерій навчання;

$Crit_2$ – другий критерій навчання;

$Crit_3$ – третій критерій навчання;

$Crit_4$ – четвертий критерій навчання;

$Crit_5$ – п'ятий критерій навчання;

$D_r(\mathbf{K})$ – відстань між бажаною та дійсною вихідними нечіткими множинами при класифікації r -го об'єкту на основі нечіткої бази знань з параметрами;

d_j – категоріальне значення консеквента j -го правила;

F – нечітка модель;

k – кількість правил;

$\mathbf{K}$ – вектор параметрів нечіткої бази знань, які налаштовуються;

l_m – мітка класу з номером m ;

M – обсяг навчальної вибірки;

p – штрафний коефіцієнт;

$\mathbf{P}$ – вектор параметрів функцій належності термів бази знань;

smax – операція знаходження другого за величиною елемента множини;

T – обсяг тестової вибірки;

$Vicewin$ – клас з другим за величиною ступенем належності;

$\mathbf{W}$ – вектор вагових коефіцієнтів правил нечіткої бази знань;

win – клас-переможець з максимальним ступенем належності;

w_j – ваговий коефіцієнт j -го правила;

$\mathbf{X}$ – вектор вхідних атрибутів;

$\mathbf{X}_r$ – вхідні атрибути r -го об'єкту;

y – результат класифікації;

y_r – клас r -го об'єкту.

ВСТУП

Класифікація – це віднесення об'єкта за деякими ознаками до одного з класів. До класифікації зводяться різноманітні задачі прийняття рішень в інженерії, економіці, медицині, соціології, політиці, спорті та в інших областях.

Останнім часом все більш популярними стають нечіткі класифікатори, тобто класифікатори, в процесі функціонування або навчання яких використовуються нечіткі множини [1]. Сьогодні переважно застосовуються нечіткі класифікатори на основі логічного виведення по базі продукційних правил, антецеденти яких містять нечіткі терми «низький», «середній», «високий» тощо. Кожне правило задає область входних атрибутів, в межах якої об'єкти належать одному класу. Границі цих областей нечіткі, тому один і той же об'єкт може одночасно належати декільком класам, але з різним ступенем.

Для підвищення безпомилковості нечітких класифікатор навчають по експериментальним даним. Для цього змінюють його параметри, щоб мінімізувати відстань між експериментальними даними та результатами нечіткого виведення. Цю відстань, яку назвемо критерієм навчання, можна визначити у різний спосіб [2]. Критерії навчання впливають на зміну параметрів нечіткого класифікатора на кожній ітерації алгоритму налаштування і, відповідно призводять до різних результатів. Тому, метою статті є знаходження такого критерія, використання якого забезпечує найкращу результативність навчання нечіткого класифікатора.

1 ПОСТАНОВКА ЗАДАЧІ

Нечіткий класифікатор являє собою відображення $\mathbf{X} = (x_1, x_2, \dots, x_n) \xrightarrow{F} y \in \{l_1, l_2, \dots, l_m\}$, яке реалізується логічним виведенням по базі нечітких правил. Навчання нечіткої бази знань здійснимо за експериментальними даними. Відповідно до принципу зовнішнього доповнення [3] експериментальні дані розіб'ємо на навчальну вибірку з M рядків «входи – вихід» та тестову вибірку з T таких рядків:

$$(\mathbf{X}_r, y_r), r = \overline{1, M}, \quad (1)$$

$$(\mathbf{X}_r, y_r), r = \overline{1, T} \quad r = \overline{1, T}, \quad (2)$$

де $\mathbf{X}_r = (x_{r1}, x_{r2}, \dots, x_{rn})$; $y_r \in \{l_1, l_2, \dots, l_m\}$.

Задача полягає у знаходженні такого критерію навчання $Crit$, застосування якого під час налаштування нечіткої бази знань F на вибірці (1) забезпечує мінімальну частоту помилок класифікації на тестовій вибірці (2).

2 ОГЛЯД ЛІТЕРАТУРИ

Класифікація на основі нечітких множин була запропонована 50 років тому в статті [4]. Перші роботи із ідентифікації залежностей за допомогою нечітких класифікаторів з'явилися в середині 90-х років. Вперше задача структурної ідентифікації шляхом відбору правил нечіткого класифікатора за показниками безпомилковості та складності бази знань розглянута в роботі [5]. Пізніше в [6] на основі критерію безпомилковості здійснили і параметричну ідентифікацію шляхом налаштування ваг правил. При зміні лише ваги правил можна отримати швидкі алгоритми навчання [7], але вони не гарантують високої безпомилковості нечіткого класифікатора. Для підвищення безпомилковості нечіткого класифікатора на етапі параметричної ідентифікації налаштовують не лише ваги правил, а й функції належності. Першими роботами з налашту-

вання функцій належності нечіткого класифікатора є статті [8, 9, 10]. В них критерієм навчання виступає квадратична нев'язка між двома нечіткими множинами – бажаними та реальними результатами класифікації. В сучасних роботах навчання нечітких класифікаторів здійснюють саме за цим критерієм (дивись, наприклад, [11]).

В роботі [7] запропоновано новий критерій навчання, який поєднує частоту помилок та квадратичну нев'язку між нечіткими бажаними та дійсними результатами логічного виведення. У випадку помилкової класифікації квадратична нев'язка зважується штрафним коефіцієнтом. Комп'ютерні експерименти в роботах [2, 7] показали перевагу цього критерію навчання над частотою помилок та над квадратичною нев'язкою. Але ця перевага є незначною і навчання нечітких класифікаторів не завжди є результативним. Тому виникає зацікавленість у нових критеріях, навчання за якими забезпечує кращу безпомилковість нечіткого класифікатора.

Зміна параметрів нечіткого класифікатора може відбуватися за логічним висновком для одного об'єкту навчальної вибірки або за результатами логічного виведення за усією навчальною вибіркою [6]. Ми розглядаємо пакетний режим навчання, коли параметри класифікатора модифікуються за логічними висновками за усією вибіркою (1).

Грунтуючись на [2] базу правил нечіткого класифікатора запишемо так:

Якщо $(x_1 = \tilde{a}_{1j} \text{ та } x_2 = \tilde{a}_{2j} \text{ та } \dots \text{ та } x_n = \tilde{a}_{nj} \text{ з вагою } w_j)$,

$$\text{тоді } y = d_j, \quad j = \overline{1, k}, \quad (3)$$

де $w_j \in [0, 1]$, $j = \overline{1, k}$; $d_j \in \{l_1, l_2, \dots, l_m\}$.

Класифікація об'єкта з атрибутами $\mathbf{X}^* = (x_1^*, x_2^*, \dots, x_n^*)$ здійснюється таким чином. Спочатку розраховується ступінь виконання j -го правила з бази (3):

$$\mu_j(\mathbf{X}^*) = w_j \cdot (\mu_j(x_1^*) \wedge \mu_j(x_2^*) \wedge \dots \wedge \mu_j(x_n^*)),$$

$$j = \overline{1, k}. \quad (4)$$

Ступінь належності $\mathbf{X}^*$ до класів $l_1, l_2, \dots, l_m$ розраховується так:

$$\mu_{l_s}(\mathbf{y}^*) = \max_{\forall j: d_j = l_s} (\mu_j(\mathbf{X}^*)), \quad s = \overline{1, m}. \quad (5)$$

Нечітким рішенням задачі класифікації буде нечітка множина

$$\tilde{\mathbf{y}}^* = \left(\frac{\mu_{l_1}(\mathbf{y}^*)}{l_1}, \frac{\mu_{l_2}(\mathbf{y}^*)}{l_2}, \dots, \frac{\mu_{l_m}(\mathbf{y}^*)}{l_m} \right). \quad (6)$$

Кінцевим результатом виведення оберемо ядро нечіткої множини (6), тобто клас з максимальним ступенем належності:

$$y^* = \arg \max_{\{l_1, l_2, \dots, l_m\}} (\mu_{l_s}(\mathbf{y}^*)).$$

Навчання нечіткого класифікатора полягає в знаходженні такого вектора $\mathbf{K} = (\mathbf{P}, \mathbf{W})$, який мінімізує частоту помилок класифікації на тестовій вибірці. При цьому для зміни координат вектора $\mathbf{K}$ використовується лише навчальна вибірка (1). Координати вектора $\mathbf{K}$ змінюємо на кожній ітерації алгоритму оптимізації за відстанню між результатами логічного виведення та експериментальними даними з вибірки (1). Цю відстань, яку назвемо критерієм навчання, можна визначити у різний спосіб.

Сьогодні найвідомішими є 3 критерії навчання нечіткого класифікатора: на основі частоти помилок [6], на основі відстані між нечіткими результатами класифікації і експериментальними даними [8–11], та їх комбінації [2, 7].

Критерій 1 – частота помилок класифікації:

$$Crit_1 = \frac{1}{M} \sum_{r=1, M} \Delta_r(\mathbf{K}), \quad (7)$$

$$\text{де } \Delta_r(\mathbf{K}) = \begin{cases} 1, & \text{якщо } y_r \neq F(\mathbf{K}, \mathbf{X}_r) \\ 0, & \text{якщо } y_r = F(\mathbf{K}, \mathbf{X}_r) \end{cases}.$$

Переваги критерію полягають в його простоті та ясній змістовній інтерпретації. Але цільова функція в задачі оптимізації за цим критерієм приймає дискретні значення, що ускладнює застосування швидких градієнтних методів оптимізації, особливо за малих вибірок даних.

Критерій 2 – квадратична нев'язка між двома нечіткими множинами – бажаними та реальними результатами класифікації. Для її розрахунку значення вихідної змінної y в навчальній вибірці фазифікують таким чином:

$$\left. \begin{aligned} \tilde{y} &= \left(\frac{1}{l_1}, \frac{0}{l_2}, \dots, \frac{0}{l_m} \right), & \text{якщо } y = l_1 \\ \tilde{y} &= \left(\frac{0}{l_1}, \frac{1}{l_2}, \dots, \frac{0}{l_m} \right), & \text{якщо } y = l_2 \\ &\vdots \\ \tilde{y} &= \left(\frac{0}{l_1}, \frac{0}{l_2}, \dots, \frac{1}{l_m} \right), & \text{якщо } y = l_m \end{aligned} \right\}. \quad (8)$$

Критерій навчання враховує відстань між логічним висновком у формі нечіткої множини (6) та бажаним нечітким значенням вихідної змінної (8):

$$Crit_2 = \sum_{r=1, M} D_r(\mathbf{K}). \quad (9)$$

Для розрахунку $D_r(\mathbf{K})$ використовується евклідова метрика:

$$D_r(\mathbf{K}) = \sum_{s=1, m} (\mu_{l_s}(y_r) - \mu_{l_s}(\mathbf{K}, \mathbf{X}_r))^2, \quad (10)$$

Перевага критерію $Crit_2$ полягає в урахуванні міри впевненості в прийнятому рішенні на основі ступенів належності об'єкту різним класам. В критерії $Crit_1$ вважається, що результат класифікації об'єкту є абсолютно достовірним, тобто неважливо наскільки ступінь належ-

ності у рішення більший, ніж у інших альтернатив – на 0,0001 чи на 1. Крім того, цільова функція в задачі навчання за критерієм (9) не має довгих плато, тому вона придатна до оптимізації градієнтними методами. Але близькі до границь класів об'єкти вносять майже однаковий вклад в критерій навчання (9) як за правильною, так і за помилковою класифікацією, тому навчання може бути нерезультативним.

Критерій 3 – квадратична нев'язка між нечіткими бажаними та реальними результатами класифікації з додатковим штрафом за помилкове рішення. Ідея полягає в збільшенні відстані D для помилково класифікованих об'єктів:

$$Crit_3 = \sum_{r=1, M} (\Delta_r(\mathbf{K}) \cdot p + 1) \cdot D_r(\mathbf{K}), \quad (11)$$

де $p > 0$.

3 МАТЕРІАЛИ І МЕТОДИ

Нижче пропонуються нові критерії навчання нечіткого класифікатора, які враховують різницю належностей нечіткого висновку лише до головних конкурентів.

Критерій 4 – відстань між головними конкурентами з штрафом за помилкове рішення. Ідея цього нового критерію полягає у врахуванні різниці належностей нечіткого висновку лише до головних конкурентів. За алгоритмом виведення рішенням обирається клас з максимальним ступенем належності. Позначимо цей клас-переможець через win . У випадку правильної класифікації головним конкурентом прийнятого рішення є $vicewin$ – клас з другим за величиною ступенем належності. Чим більша різниця між належністю до класів win та $vicewin$, тим більша впевненість у логічному висновку, і тим далі об'єкт знаходиться від границі розділу класів.

Для r -го об'єкту з вибірки (1) $\mu_{win}(\mathbf{X}_r) = \max_{s=1, m} (\mu_{l_s}(\mathbf{X}_r))$ та $\mu_{vicewin}(\mathbf{X}_r) = \max_{s=1, m} (\mu_{l_s}(\mathbf{X}_r))$. Відповідно, різниця між головними конкурентами дорівнює $\mu_{win}(\mathbf{X}_r) - \mu_{vicewin}(\mathbf{X}_r)$.

В критерії навчання враховуватимемо відносні показники, розділивши різницю на ступінь належності класу-переможцю. За правильної класифікації відносна різниця дорівнює $D_r^1 = \frac{\mu_{win}(\mathbf{X}_r) - \mu_{vicewin}(\mathbf{X}_r)}{\mu_{win}(\mathbf{X}_r)}$, а за непра-

вильної – $D_r^0 = \frac{\mu_{win}(\mathbf{X}_r) - \mu_{y_r}(\mathbf{X}_r)}{\mu_{win}(\mathbf{X}_r)}$. Крім того, аналогічно до (11), за помилкової класифікації зважимо різницю штрафним коефіцієнтом $p \geq 1$. Математично критерій навчання запишемо таким чином:

$$Crit_4 = p \cdot \sum_{\substack{y_r \neq F(\mathbf{K}, \mathbf{X}_r) \\ r=1, M}} D_r^0(\mathbf{K}) - \sum_{\substack{y_r = F(\mathbf{K}, \mathbf{X}_r) \\ r=1, M}} D_r^1(\mathbf{K}). \quad (12)$$

Критерій 5 – квадратична відстань між головними конкурентами з штрафом за помилкове рішення. Цей

критерій є модифікацією попереднього. Відмінність полягає у використанні не відносних різниць, а їх квадратів:

$$Crit_5 = p \cdot \sum_{r=1, M}^{y_r \neq F(\mathbf{K}, \mathbf{X}_r)} D_r^0(\mathbf{K})^2 - \sum_{r=1, M}^{y_r = F(\mathbf{K}, \mathbf{X}_r)} D_r^1(\mathbf{K})^2.$$

Піднесення до квадрату в $Crit_5$ дозволяє, як і в методі найменших квадратів, збільшити важливість великих різниць – тобто грубих промахів.

4 ЕКСПЕРИМЕНТИ

Метою експериментів є виявлення критерію навчання, який забезпечує найкращу безпомилковість. Розглядається тестова задача Wine Dataset з UCI Machine Learning Repository. Вона полягає у виявленні сорту винограду (y), з якого виготовлено вино. База даних містить результати лабораторних аналізів по 13-ти показникам 178 зразків італійських вин, виготовлених в одному регіоні. Для кожного зразка вказано, з якого із з трьох сортів винограду виготовлено вино.

Навчальну вибірку сформуємо з рядків бази даних з граничними значеннями кожного із 13 атрибутів. Додатково в навчальну вибірку включимо всі непарні рядки бази даних. Всі інші дані занесемо в тестову вибірку. В результаті отримаємо навчальну вибірку з 100 рядків і тестову – з 78. Побудуємо нечіткий класифікатор вин за трьома ознаками: x_7 – flavanoids, x_{10} – color intensity та x_{13} – proline. Експерименти проведемо для нечіткого класифікатора з базою знань з табл. 1. Нечіткі терми задамо гаусовою функцією належності:

$$\mu(x) = \exp\left(-\frac{(x-b)^2}{2c^2}\right),$$

де $c > 0$.

Параметри функцій належності початкового нечіткого класифікатора наведено в табл. 2.

Таблиця 1 – Нечітка база знань класифікатора вин [6]

№	x_7	x_{10}	x_{13}	y
1	–	–	Високий	Сорт 1
2	Високий	Високий	Середній	Сорт 1
3	–	Низький	Низький	Сорт 2
4	Низький	Низький	Середній	Сорт 2
5	Низький	Високий	–	Сорт 3

Таблиця 3 – Вплив штрафного коефіцієнта в критерії навчання на безпомилковість класифікатора на тестовій вибірці (статистика 200 експериментів)

t -норма	Критерій	Середня безпомилковість				
		$p = 1$	$p = 3$	$p = 5$	$p = 7$	$p = 9$
min	$Crit_3$	0,0856	0,0915	0,0912	0,0910	0,0960
	$Crit_4$	0,0732	0,0640	0,0662	0,0736	0,0770
	$Crit_5$	0,0690	0,0604	0,0602	0,0688	0,0657
prod	$Crit_3$	0,0727	0,0656	0,0665	0,0681	0,0649
	$Crit_4$	0,0565	0,0512	0,0493	0,0499	0,0527
	$Crit_5$	0,0448	0,0443	0,0456	0,0486	0,0446

Таблиця 2 – Початкові параметри функцій належності термів нечіткого класифікатора

Вхідна змінна	Терм	Параметри	
		b	c
x_7	Низький	2	0,34
	Високий	2	5,08
x_{10}	Низький	6	1,28
	Високий	6	13
x_{13}	Низький	3	2,78
	Середній	3	10
	Високий	3	16,8

Для кожного критерію проведемо 1000 експериментів із навчання нечіткої бази знань на основі квазіньютонівського алгоритму. Після навчання кожний класифікатор перевіримо на тестовій вибірці за частотою помилок (критерій $Crit_1$). Під час навчання налаштуємо вагові коефіцієнти перших чотирьох правил. Достовірність п'ятого правила не викликає сумнівів, тому його ваговий коефіцієнт залишимо рівним 1. Налаштуємо також коефіцієнти концентрації (c) функцій належності кожного нечіткого терма. Для збереження інтерпретованості бази знань згідно до [11] налаштуємо координати максимумів (b) функцій належності лише не крайніх термів. В базі знань 1 некрайній терм – «Середній», координату максимуму якого і будемо змінювати. Таким чином, загальна кількість параметрів для налаштування становить $4 + 7 + 1 = 12$. Початкові точки для навчання оберемо випадково – для вагових коефіцієнтів правил з діапазону $[0, 1]$, а для параметрів функцій належності в межах $\pm 30\%$ від значень з табл. 2.

Проведемо 2 серії експериментів. Першу серію для нечіткого класифікатора з реалізацією t -норми мінімумом (min), а другу – з реалізацією t -норми добутком (prod). В експериментах з критеріями $Crit_3$, $Crit_4$ та $Crit_5$ спочатку визначимо прийнятний рівень штрафного коефіцієнту. Для цього проведемо по 200 оціночних експериментів для $p = 1, 3, \dots, 9$. Результати цих експериментів наведено в табл. 3. В ній напівжирним шрифтом виділено за яких значень штрафного коефіцієнту навчання відбувається краще. Саме за таких значень штрафного коефіцієнту проведемо решту експериментів.

5 РЕЗУЛЬТАТИ

Результати експериментів наведено на рис. 1–3 та в табл. 4.

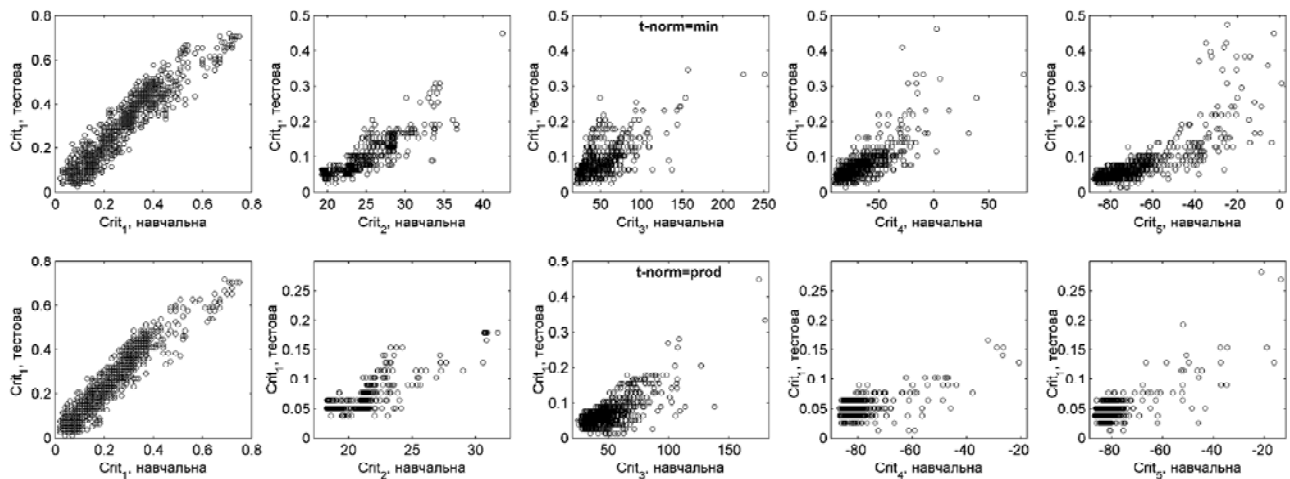


Рисунок 1 – Розподіл результатів навчання нечіткого класифікатора (статистика 1000 експериментів)

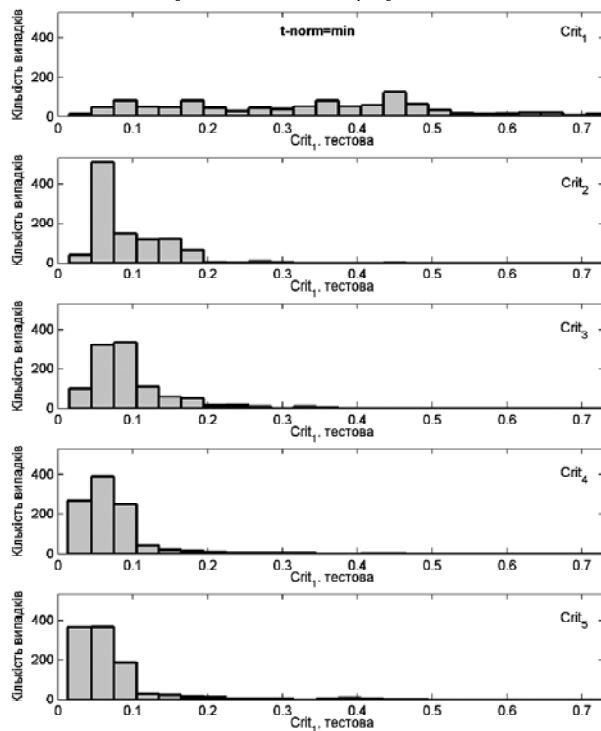


Рисунок 2 – Розподіл якості навчання класифікатора, t -норму якого реалізовано мінімумом

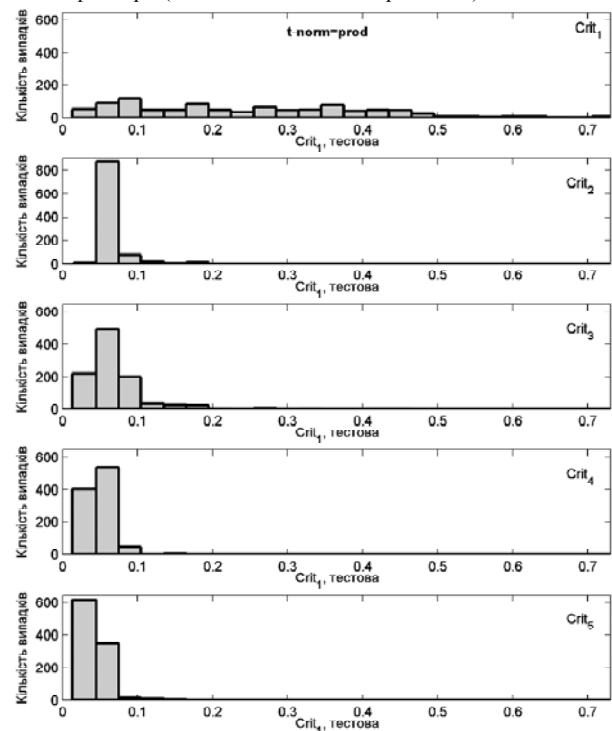


Рисунок 3 – Розподіл якості навчання класифікатора, t -норму якого реалізовано добутком

Таблиця 4 – Статистика навчання нечітких класифікаторів (напівжирним виділено найкращі результати)

t -норма	Критерій навчання	Частота помилок ($Crit_1$) на тестовій вибірці			
		мінімальне	середнє	медіанне	максимальне
min	$Crit_1$	0,0256	0,3194	0,3333	0,7179
	$Crit_2$	0,0256	0,0900	0,0641	0,4487
	$Crit_3$	0,0256	0,0921	0,0769	0,3462
	$Crit_4$	0,0128	0,0680	0,0513	0,4615
	$Crit_5$	0,0128	0,0680	0,0513	0,4744
prod	$Crit_1$	0,0128	0,2459	0,2308	0,7179
	$Crit_2$	0,0385	0,0631	0,0641	0,1795
	$Crit_3$	0,0128	0,0663	0,0641	0,4487
	$Crit_4$	0,0128	0,0503	0,0513	0,1667
	$Crit_5$	0,0128	0,0454	0,0385	0,2821

6 ОБГОВОРЕННЯ

Результати експериментів на рис. 1 вказують на корельованість значень критеріїв $Crit_1 - Crit_5$ на навчальній вибірці з частотою помилок на тестовій вибірці. Відповідно ці критерії можна застосовувати для навчання нечіткого класифікатора. Щодо результативності навчання (див. табл. 4 та рис. 2, 3), то вона суттєво краща при використанні нових критеріїв $Crit_4$ та $Crit_5$. Нові критерії забезпечують кращу безпомилковість як в середньому, так і за кількістю найкращих випадків навчання. Серед нових критеріїв помірну перевагу має $Crit_5$.

ВИСНОВКИ

У роботі запропоновано нові критерії навчання нечіткого класифікатора, які враховують різницю належностей нечіткого висновку лише до головних конкурентів. За правильної класифікації головним конкурентом прийнятого рішення є клас з другим за величиною ступенем належності. У випадку неправильної класифікації помилково прийняте рішення є головним конкурентом правильного класу.

Проведені нами експерименти із навчання нечіткого класифікатора для UCI-задачі із розпізнавання італійських вин засвідчили суттєву перевагу нових критеріїв. Серед нових критеріїв помірну перевагу має критерій на основі квадратичної відстані між головними конкурентами з штрафом за помилкове рішення. Відповідно, мету дослідження досягнуто – встановлено, що для налаштування нечітких класифікаторів доцільно використовувати нові критерії навчання на основі різниці належності рішення до головних конкурентів. Нові критерії можуть застосовуватися не лише для настроювання нечітких класифікаторів, але і для навчання деяких інших моделей, наприклад, нейронних мереж.

ПОДЯКИ

Публікація містить результати досліджень, проведених при грантовій підтримці Державного фонду фундаментальних досліджень за конкурсним проектом №62/201–2015.

СПИСОК ЛІТЕРАТУРИ

1. Kuncheva L. I. Fuzzy classifier design / L. I. Kuncheva // Studies in Fuzziness and Soft Computing. – Berlin – Heidelberg: Springer-Verlag, 2000. – Vol. 49. – 314 p.
2. Shtovba S. Analyzing the criteria for fuzzy classifier learning / S. Shtovba, O. Pankevich, A. Nagorna // Automatic Control and Computer Sciences. – 2015. – Vol. 49, № 3. – P. 123–132.
3. Madala H. R. Inductive learning algorithms for complex systems modeling / H. R. Madala, A. G. Ivakhnenko. – Boca Raton : CRC Press, 1994. – 368 p.
4. Bellman R. Abstraction and pattern classification / R. Bellman, R. Kalaba, L. Zadeh // Journal of Mathematical Analysis and Applications. – 1966. – Vol. 13, № 1. – P. 1–7.
5. Construction of fuzzy classification systems with rectangular fuzzy rules using genetic algorithms / [Ishibuchi H., Nozaki K., Yamamoto N., Tanaka H.] // Fuzzy sets and systems. – 1994. – Vol. 65, № 2. – P. 237–253.
6. Ishibuchi H. Classification and modeling with linguistic information granules: advanced approaches / H. Ishibuchi, T. Nakashima, M. Nii. – Berlin-Heidelberg : Springer-Verlag, 2005. – 307 p.
7. Штовба С. Д. Порівняння критеріїв навчання нечіткого класифікатора / С. Д. Штовба // Вісник Вінницького політехнічного інституту. – 2007. – № 6. – С. 84–91.
8. Abe S. Tuning of a fuzzy classifier derived from data / S. Abe, M. S. Lan, R. Thawonmas // International Journal of Approximate Reasoning. – 1996. – Vol. 14. – P. 1–24.
9. Nauck D. A neuro-fuzzy method to learn fuzzy classification rules from data / D. Nauck, R. Kruse // Fuzzy Sets and Systems. – 1997. – Vol. 89, № 3. – P. 277–288.
10. Rotshtein A. P. Design and Tuning of Fuzzy If – Then Rules for Automatic Classification / A. P. Rotshtein, D. I. Katelnikov // Proc. of NAFIPS'98 – International Conf. «Annual Meeting of North American Fuzzy Information Processing Society», Tampa, USA, 1998. – P. 50–55.
11. Rudzinski F. A multi-objective genetic optimization of interpretability-oriented fuzzy rule-based classifiers / F. Rudzinski // Applied Soft Computing. – 2016. – Vol. 38. – P. 118–133.
12. Shtovba S. Ensuring accuracy and transparency of Mamdani fuzzy model in learning by experimental data / S. Shtovba // Journal of Automation and Information Sciences. – 2007. – Vol. 39, № 8. – P. 39–52.

Стаття надійшла до редакції 21.12.2015.

Після доробки 26.01.2016.

Штовба С. Д.¹, Галушак А. В.²

¹Д-р техн. наук, професор, професор кафедри комп'ютерних систем управління, Вінницький національний технічний університет, Вінниця, Україна

²Асистент кафедри комп'ютерних систем управління, Вінницький національний технічний університет, Вінниця, Україна

ОБУЧЕНИЕ НЕЧЕТКОГО КЛАССИФИКАТОРА НА ОСНОВЕ РАССТОЯНИЯ МЕЖДУ ГЛАВНЫМИ КОНКУРЕНТАМИ

Классификация – это отнесение объекта по некоторым признакам к одному из классов. К классификации сводятся разнообразные задачи принятия решений в инженерии, экономике, медицине, социологии и других областях. В нечетких классификаторах зависимость «входы – выход» описываются с помощью лингвистических правил «Если – то», антецеденты которых содержат нечеткие термы «низкий», «средний», «высокий» и т. п. Для повышения безошибочности нечеткий классификатор обучают по экспериментальным данным. В данной работе предложены новые критерии обучения нечеткого классификатора, которые учитывают разницы принадлежностей нечеткого вывода только к главным конкурентам. При правильной классификации главным конкурентом принятого решения является класс со второй по величине степенью принадлежности. В случае неправильной классификации ошибочно принятое решение является главным конкурентом правильного класса. Компьютерные эксперименты по обучению нечеткого классификатора для распознавания трех сортов итальянских вин показали существенное преимущество новых критериев. Среди новых критериев обучения небольшое преимущество имеет критерий в форме квадратичного расстояния между главными конкурентами со штрафом за ошибочное решение. Новые критерии могут применяться не только для обучения нечетких классификаторов, но и для обучения других моделей, например, нейронных сетей.

Ключевые слова: классификация, нечеткая база знаний, обучение, критерии обучения, главные конкуренты.

Shtovba S. D.¹, Galushchak A. V.²

¹Prof., Dr.Sc., Vinnytsia National Technical University, Vinnytsia, Ukraine

²Assistant, Vinnytsia National Technical University, Vinnytsia, Ukraine

FUZZY CLASSIFIER LEARNING BASED ON DISTANCE BETWEEN THE MAIN COMPETITORS

The classification problem is the assignment an object with certain features to one of classes. Various engineering, management, economic, political, medical, sport, and other problems are reduced to classification. In fuzzy classifiers «inputs – output» relation is described by linguistic <If – then> rules. Antecedents of these rules contain fuzzy terms «low», «average», «high» etc. To increase the correctness it is necessary to tune the fuzzy classifier on experimental data. The new criteria for fuzzy classifier learning that take into account the difference of membership degrees to the main competitors only are proposed. When the classification is correct, the main competitor of the decision is the class with the second largest membership degree. In cases of misclassification the wrong decision is the main competitor to the correct class. Computer experiments with learning the fuzzy classifier of 3 kinds of Italian wines recognition showed a significant advantage of the new criteria. Among new learning criteria the criterion in the form of squared distance between main competitors with the penalty for wrong decision has minor advantage. New criteria can be used not only for tuning fuzzy classifiers but for tuning some other models, such as neural networks.

Keywords: classification, fuzzy knowledge base, tuning, learning criteria, main competitors.

REFERENCES

1. Kuncheva L. I. Fuzzy classifier design, *Studies in Fuzziness and Soft Computing*. Berlin, Heidelberg, Springer-Verlag, 2000, Vol. 49, 314 p.
2. Shtovba S. Pankevich O., Nagorna A. Analyzing the criteria for fuzzy classifier learning, *Automatic Control and Computer Sciences*, 2015, Vol. 49, No. 3, pp. 123–132.
3. Madala H. R., Ivakhnenko A. G. Inductive learning algorithms for complex systems modeling. Boca Raton, CRC Press, 1994, 368 p.
4. Bellman R., Kalaba R., Zadeh L. Abstraction and pattern classification, *Journal of Mathematical Analysis and Applications*, 1966, Vol. 13, No. 1, pp. 1–7.
5. Ishibuchi H., Nozaki K., Yamamoto N., Tanaka H. Construction of fuzzy classification systems with rectangular fuzzy rules using genetic algorithms, *Fuzzy sets and systems*, 1994, Vol. 65, No. 2, pp. 237–253.
6. Ishibuchi H., Nakashima T., Nii M. Classification and modeling with linguistic information granules: advanced approaches advanced approaches to linguistic data mining. Ishibuchi, Berlin-Heidelberg, Springer-Verlag, 2005, 307 p.
7. Shtovba S. D. Porivnjannja kriteriiv navchannja nechitkogo klasifikatora, *Visnik Vinnic'kogo politehnichnogo institutu*, 2007, No. 6, pp. 84–91.
8. Abe S., Lan M. S., Thawonmas R. Tuning of a fuzzy classifier derived from data, *International Journal of Approximate Reasoning*, 1996, Vol. 14, pp. 1–24.
9. Nauck D., Kruse R. A neuro-fuzzy method to learn fuzzy classification rules from data, *Fuzzy Sets and Systems*, 1997, Vol. 89, No. 3, pp. 277–288.
10. Rotshtein A. P., Katelnikov D. I. Design and Tuning of Fuzzy If–Then Rules for Automatic Classification, *Proc. of NAFIPS'98 – International Conf. «Annual Meeting of North American Fuzzy Information Processing Society»*. Tampa, USA, 1998, pp. 50–55.
11. Rudzinski F. A multi-objective genetic optimization of interpretability-oriented fuzzy rule-based classifiers, *Applied Soft Computing*, 2016, Vol. 38, pp. 118–133.
12. Shtovba S. Ensuring accuracy and transparency of Mamdani fuzzy model in learning by experimental data, *Journal of Automation and Information Sciences*, 2007, Vol. 39, No. 8, pp. 39–52.

ПРОГРЕСИВНІ ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ

ПРОГРЕССИВНЫЕ ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

PROGRESSIVE INFORMATION TECHNOLOGIES

УДК 004.056.55

Гальченко А. В.¹, Козіна Г. Л.²

¹Ст. гр. РТ – 710М, магістр кафедри Захисту інформації Запорізького національного технічного університету,
Запоріжжя, Україна

²Канд. фіз.-мат. наук, доцент кафедри Захисту інформації Запорізького національного технічного університету,
Запоріжжя, Україна

МОДИФІКАЦІЯ АЛГОРИТМУ ЗАПЕРЕЧУВАНОГО ШИФРУВАННЯ МЕНГА

В статті обговорюється проблема стійкості сучасних криптографічних систем до атак на основі примушування стосовно абонентів криптографічних систем. У зв'язку зі стрімким розвитком галузі інформаційних технологій ця проблема є актуальною в сфері інформаційної безпеки. Для вирішення проблеми стійкості сучасних криптографічних систем авторами запропоновано використання алгоритмів заперечуваного шифрування, які гарантують, що зломисник не має змоги отримати будь-яку цінну інформацію від абонентів. Вирішення проблеми полягає у використанні алгоритму заперечуваного шифрування Менга, який гарантує захист не лише інформації, а й самих учасників обміну.

Основна мета статті полягає у виконанні модифікації первісного алгоритму заперечуваного шифрування Менга [1] шляхом використання протоколу «непомітної» передачі OT_n^1 , використання якого запропоновано Моні Наором [2]. Застосування протоколу «непомітної» передачі OT_n^1 дозволяє суттєво скоротити час, необхідний для виконання алгоритму заперечуваного шифрування Менга, та спрощує його реалізацію для вирішення прикладних завдань в сфері інформаційної безпеки.

За результатами експериментів авторами статті було підтверджено, що використання протоколу «непомітної» передачі OT_n^1 є ефективним вирішенням проблеми генерації і розподілу ключів в алгоритмі заперечуваного шифрування Менга.

Ключові слова: протокол, заперечуване шифрування, неоднозначність, розподіл ключів, фіктивне повідомлення, зломисник, розширена схема Рабіна, непомітна передача, факторизація, дискретне логарифмування.

НОМЕНКЛАТУРА

AMD – марка виробника центральних процесорів;
BCP – протокол шифрування розроблений E. Bresson, D. Catalano і D. Pointcheval;
RD – PKE – Receiver – Deniable Public Key Encryption Protocol;
RSA – протокол шифрування розроблений Rivest, Shamir і Adleman;
Гб – одиниця вимірювання об'єму пам'яті;
ГГц – одиниця вимірювання частоти (10^{-9} секунди);
КЗІ – криптографічний захист інформації;
мс – одиниця вимірювання часу (0,001 секунди);
ОЗУ – оперативний запам'ятовуючий пристрій (внутрішня пам'ять);
ОС – операційна система;

ПЗ – програмне забезпечення;
ЦП – центральний процесор.
 (A_1, A_2) – пара числових значень A_1 і A_2 ;
 a – секретний ключ відправника;
 α – елемент мультиплікативної групи;
 $C(r_2, m)$ – контрольна функція від числових значень r_2 і m ;
 D – проміжне числове значення;
 $\langle (\partial, \wp), B \rangle$ – криптограма в алгоритмі заперечуваного шифрування Менга;
 $(e; n)$ і $(d; n)$ – публічний та секретний ключ в алгоритмі RSA;

$\langle (e, n), x \rangle$ – трійка публічний параметрів в протоколі OT_n^1 ;

$f_{\text{дл}}(\dots)$ – функція дискретного логарифмування;

g – генератор мультиплікативної групи;

$\text{hash}(r_2)$ – значення хеш – функції від r_2 ;

$k < n$ – випадкове числове значення;

k' – проміжне числове значення;

m – секретне повідомлення, яке належить до групи Z_N ;

m' – фіктивне повідомлення, яке належить до групи Z_N ;

(m_1, m_2) – пара числових значень m_1 і m_2 ;

(m, r) – пара числових значень, еквівалентна m і r_2 ;

N – модуль утворений добутком чисел p і q ;

$\langle (N, g, h), (p, q) \rangle$ – пара публічного та секретного ключів одержувача;

$\text{ord}(\dots)$ – порядок будь-якого числового елемента групи;

OT_n^1 – Oblivious Transfer Protocol;

p – модуль перетворень еквівалентний N^2 ;

p і q – великі прості сильні числа (p' і q');

r – випадковий елемент мультиплікативної групи, еквівалентний r_2 ;

r_1 – випадковий елемент мультиплікативної групи;

$r_2 \in \langle g \rangle$ – випадковий елемент мультиплікативної групи, побудований на g ;

$\{T_i, M_i\}$ – пара секретного та фіктивного повідомлень в алгоритмі Рабіна;

t_p – час виконання будь-якої операції при даному розмірі модуля p ;

$t_{\text{ГК}}$ – числове значення часу генерації ключів в алгоритмі;

v – проміжне числове значення;

$X(t)$ – числове значення псевдовипадкової послідовності;

$x < n$ – випадкове числове значення;

$\langle (y, g, p), a \rangle$ – пара публічного та секретного ключів відправника;

γ_1 – проміжне числове значення;

$Z_{N^2}^*$ – мультиплікативна група для числа N^2 .

ВСТУП

Сучасні алгоритми шифрування не можуть забезпечувати абсолютну стійкість до атак злоумисників як зовні, так і з середини системи. Атаки на основі примушування можна застосовувати для отримання будь-якої секретної інформації. Вирішення цієї проблеми полягає у застосуванні алгоритмів «неоднозначного» або заперечуваного шифрування.

Наукова новизна полягає у функціональній заміні задачі дискретного логарифмування, яка використовується в процедурі розподілу ключів, на протокол «непомітної» передачі OT_n^1 в алгоритмі заперечуваного шифрування Менга.

Актуальність теми статті полягає у зниженні ефективності захисту інформації сучасними криптографічними системами та необхідність пошуку нових рішень для виправлення цієї проблеми.

Об'єкт дослідження – процедура розподілу ключів між абонентами в алгоритмі заперечуваного шифрування Менга.

Актуальність теми статті полягає у зниженні ефективності захисту інформації сучасними криптографічними системами та необхідність пошуку нових рішень для виправлення цієї проблеми.

Об'єкт дослідження – процедура розподілу ключів між абонентами в алгоритмі заперечуваного шифрування Менга.

Предмет дослідження – це властивостей алгоритму заперечуваного шифрування Менга та особливостей протоколу «непомітної» передачі OT_n^1 .

Мета роботи полягає в модифікації алгоритму заперечуваного шифрування Менга шляхом огляду та аналізу існуючих алгоритмів і протоколів безпечної передачі інформації. Для досягнення цієї мети необхідно виконати наступні завдання:

- зробити огляд аналогічних рішень у подібних алгоритмах шифрування персональних даних;

- виконати аналіз структури та особливостей захисту інформації подібними алгоритмами заперечуваного шифрування;

- виконати модифікацію процедури розподілу ключів у алгоритмі заперечуваного шифрування Менга;

- виконати порівняння швидкодії процедур розподілу ключів в первісному та модифікованому алгоритмах заперечуваного шифрування Менга, зробити висновки щодо результатів модифікацій алгоритму та їх вплив на ефективність захисту інформації.

1 ПОСТАНОВКА ЗАДАЧІ

Основна проблема при використанні алгоритму заперечуваного шифрування Менга полягає в умовності його структури, що унеможливає його практичне застосування для захисту інформації в реальних комп'ютерних мережах. Оскільки вона виникає на етапі генерації ключів, то необхідно виконати модифікацію процедури розподілу ключів, але не робити значних змін у структурі самого алгоритму шифрування.

Суть проблеми полягає у неможливості виконання розподілу ключів між абонентами через використання задачі дискретного логарифмування для передачі секретного ключа a відправника одержувачеві. Оскільки задача дискретного логарифмування за складністю обчислень еквівалентна до задачі з факторизації, то при обчисленні одержувачем секретного ключа a відправника з публічного параметру h і модуля $p \sim N^2$ час виконання процедури розподілу ключів можна описати виразом $t_{p \sim N^2} \rightarrow \infty$.

Для забезпечення оптимального часу виконання процедури генерації та розподілу ключів використати протокол «непомітної» передачі OT_n^1 , який описано в [2] і застосовано для вирішення подібної проблеми в алгоритмі заперечуваного шифрування Ібрахіма [3]. Оскільки стійкість протоколу «непомітної» передачі ґрунтується на вирішенні задачі з факторизації, то необхідно виконати аналіз швидкодії даного протоколу із швидкістю вирішення еквівалентної задачі – дискретного логарифмування, в оригінальному алгоритмі заперечуваного шифрування Менга.

2 ОГЛЯД ЛІТЕРАТУРИ

Для надійного захисту інформації, як на локальних комп'ютерах так і тієї, що передається по мережі, використовуються криптографічні засоби захисту інформації. Одним із напрямів криптографічного захисту інформації є шифрування. В зв'язку з цим проводиться безперервні дослідження існуючих криптографічних схем шифрування та створення нових криптографічних схем, на їх базі. Як новий напрям, в криптографії виділяють розробку та дослідження алгоритмів заперечуваного шифрування. В сучасній літературі для цього напрямку використовують такі ж поняття, як «неоднозначне», «суперечливе» або «заперечуване».

Існує не так багато джерел, які описують алгоритми заперечуваного шифрування, на відміну від інших. Така тенденція пов'язана з одним із основним аспектів заперечуваного шифрування, який гарантує надійність захисту – таємність алгоритмів заперечуваного шифрування. В результаті досліджень даного напрямку було знайдено деякі публікації і статті, які описують перспективність алгоритмів заперечуваного шифрування для застосування в сфері криптографічного захисту інформації.

Досліджені алгоритми заперечуваного шифрування, здебільшого, побудовані на основі криптографічних систем з відкритим ключем, оскільки останні набули популярності в останні роки. До таких алгоритмів можна віднести алгоритми заперечуваного шифрування розроблені: Раном Канетті, Хамадою Ібрахімом, Джин – Квін Вангом і Бо Менгом, М. А. Молдов'яном і О.О. Горячевим, Шаффі Голвасером і С. Мікалі [4] та ін.

Алгоритм заперечуваного шифрування Рана Канетті [4] є криптографічною системою з відкритим ключем, яка дозволяє користувачам не виконувати попередній обмін секретними параметрами для шифрування/дешифрування інформації. Алгоритм заперечуваного шифрування Канетті передбачає реалізацію механізму побітового шифрування даних, тому він є менш продуктивним, але має досить надійний рівень захисту.

Алгоритм заперечуваного шифрування Хамади Ібрахіма [3] побудовано на базі протоколу RD – PKE, який дозволяє повністю захистити одержувача секретної інформації від застосування примушування. Така змога надається шляхом поділу секретного ключа між одержувачем та довіреною стороною безпеки за допомогою протоколу [3]. Оскільки одержувач не має достатньої кількості інформації для самостійного розшифрування криптограми, то застосування примушування стосовно нього не є ефективним.

Алгоритм заперечуваного шифрування М. А. Молдов'яна і О. О. Горячева [5] побудовано на базі розширеної криптографічної схеми Рабіна з відкритим ключем. Даний алгоритм заперечуваного шифрування на достатньому рівні вирішує завдання пов'язані зі стійкістю криптографічної системи до атак заснованих на базі примушування. Модифікація алгоритму шифрування Рабіна передбачає відновлення чотирьох різних пар секретних/фіктивних повідомлень $\{T_i, M_i\}$, три з яких мають випадковий характер. Оскільки даний алгоритм відноситься до криптографічних систем з відкритим ключем, то основою захисту є обчислювальна складність, яка базується на вирішенні задачі факторизації. Таким чином, секретність власне алгоритму не є критичною.

Алгоритм заперечуваного шифрування Шафі Голдвасера і Сильвіо Мікалі [4] побудовано на базі комбінованого алгоритму ймовірнісного шифрування [6]. В літературі даний алгоритм отримав назву – асоційований алгоритм ймовірнісного шифрування. Основним критерієм захисту є неоднозначність дешифрованих повідомлень, тобто ймовірність того, що зловмисник має змогу дешифрувати криптограму та відрізнити результати від випадкових значень досить низька. Проте його недоліком є можливість дешифрування лише одного повідомлення.

Алгоритм заперечуваного шифрування Джин-Квін Ванга і Бо Менга [1] побудовано на базі протоколу BCP [7] та ідей запропонованих Клоновські [8]. В даному алгоритмі комбіновано кращі сторони алгоритмів RSA та Ель-Гамала, саме тому він має більш кращі показники захищеності та продуктивності роботи. Представлений алгоритм заперечуваного шифрування передбачає використання двох алгоритмів розшифрування захищеного повідомлення, який катастрофічно знижував ефективність захисту. Але використання двох пар секретних ключів і обчислювальна складність криптографічних систем з відкритим ключем забезпечують досить надійний захист секретної інформації та досить високий рівень продуктивності.

3 МАТЕРІАЛИ І МЕТОДИ

Алгоритм заперечуваного шифрування Менга вирішує проблему щодо атак на основі примушування, яке застосовується як до відправника та одержувача, так і до обох одночасно. Таким чином виключається будь – яка можливість отримання секретної інформації зловмисником. Структурна схема алгоритму заперечуваного шифрування Менга приведена в Додаток А.

Оскільки на етапі генерації ключів виникає проблема пов'язана з розподілом ключів між абонентами, то її вирішення полягає у модифікації первісного алгоритму (Додаток А) шляхом функціональної заміни задачі дискретного логарифмування на протокол «непомітної» передачі OT_n^1 (рис. 1).

Згідно з результатами експериментів проведених у пункті 4 можна зробити висновок, що застосування протоколу «непомітної» передачі OT_n^1 в процедурі розподілу ключів алгоритму заперечуваного шифрування Менга

га має більш оптимальний час і спрощує його технічну реалізацію. Таким чином модифікований варіант алгоритму заперечуваного шифрування Менга приведено в Додаток Б, а приклад його роботи приведено нижче (в якості тестових повідомлень використані «Security» і «United Ukraine» відповідно секретне та фіктивне тестові повідомлення).

На етапі підготовки було виконано генерацію секрет-

них і публічних ключів одержувача $\langle(N, g, h), (p, q)\rangle$ та відправника $\langle(y, g, p), a\rangle$, які приведено в табл. 1.

На етапі шифрування було виконано обчислення криптограми $\langle(\partial, \wp), B\rangle$, яку відправник передає по відкритому каналу одержувачеві. Результати обчислень приведено в табл. 2.

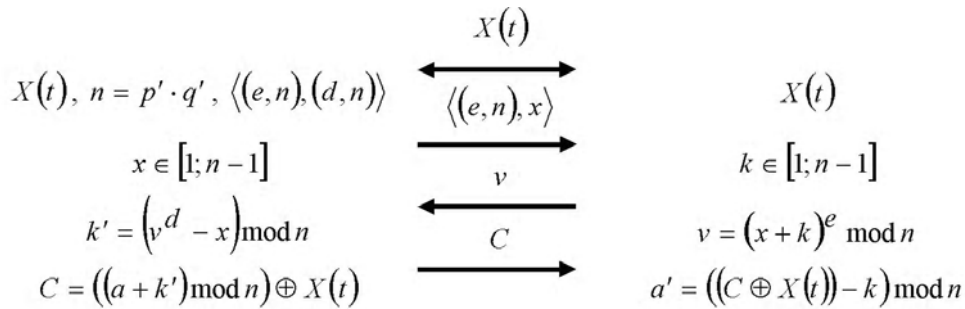


Рисунок 1 – Структурна схема протоколу «непомітної передачі» OT_n^1

Таблиця 1 – Секретний та відкритий ключі відправника та одержувача

Абоне нти	Позначенн я	Значення
Відправник	N	340282393626195344062954022172595310137
	g	457692035475152616848644
	h	109066888395335831186793921949405725400121626796378019323902565298338926914653
	p	18446744847862403543
	q	18446744747251547759
Одержувач	$X(t)$	454517681881
	(e, d, n)	(325514516927, 137402428150799401427903, 663108052854589966200253)
	x	233168199764
	k	365161775673
	v	210858578369542822886041
	a	775224541535
	C	1513289100609
	a'	775224541535
	y	109066888395335831186793921949405725400121626796378019323902565298338926914653
	g	457692035475152616848644
	p	115792107411972949718521283553852949453058139797758426188712183109287214958769

Таблиця 2 – Обчислення криптограми $\langle(\partial, \wp), B\rangle$

Позначення параметрів	Значення
r_2	976111453564
m	8751735916204352851
$C(r_2, m)$	49441283279314548035286181750154525218013730928204068249020928122723077961488
m'	2057271081577553248481314125475413
r_1	27445582010364396438306983190780524779013243571405838002487480018640703378326
$C(r_1, m')$	49441283279314548035286181750154525218013730928204068249020928122723077961488
$hash(r_2)$	976111453573
B	49441283279314548035286181750154525218013730928204068249020928122723077961488
∂	7120757359751600539380457168433360239215956805097502870744764841488557454648
$\wp$	89839296842757781726541293528853401707394111864091289320280465719410134676193

На етапі розшифрування одержувачем було відновлено секретне повідомлення m з криптограми $\langle (\partial, \varphi), B \rangle$, результати обчислень якого приведено в табл. 3.

В результаті виконаних обчислень одержувач отримав секретне повідомлення «Security», яке було адресоване йому.

На етапі дешифрування, в результаті застосування примусу, одержувач виконав відновлення фіктивного повідомлення m' з криптограми $\langle (\partial, \varphi), B \rangle$, яке призначене власне для зловмисника, результати обчислень якого приведено в табл. 4.

В результаті виконаних обчислень одержувач відновив фіктивне повідомлення «United Ukraine» та передав його зловмисникові.

Таким чином було виконано моделювання роботи модифікованого алгоритму заперечуваного шифрування Менга за допомогою 128-бітного ключа та продемонстровано його особливості, які забезпечують надійний захист інформації від атак на основі примусу. Для оцінки ефективності виконаних модифікацій алгоритму авторами проведено експерименти щодо визначення оптимальної швидкодії його роботи шляхом порівняння швидкодії первісного та модифікованого алгоритмів.

Оскільки модифікація полягала в покращенні часових характеристик процедури генерації та розподілу ключів, то для спрощення експериментів, у пункті 4, було досліджено швидкодії процедур генерації ключів $t_{ГК}$: на основі задачі дискретного логарифмування та протоколу «непомітної» передачі OT_n^1 .

4 ЕКСПЕРИМЕНТИ

Для покращення часових характеристик процедури генерації та розподілу ключів авторами було проведено експерименти з дослідження швидкодії процедур генерації ключів $t_{ГК}$: на основі задачі дискретного логарифмування та протоколу «непомітної» передачі OT_n^1 (рис. 2).

Для виконання експериментів було використано структурні схеми первісного та модифікованого алгоритмів заперечуваного шифрування Менга (Додаток А і Додаток Б) і персональний комп'ютер з технічними характеристиками:

- ОС: Windows 7;
- ЦП: AMD Athlon(tm) II P360 Dual – Core Processor 2.3 ГГц;
- ОЗУ: 3 Гб;
- ПЗ: пакет математичних обчислень Maple 14.

Основний показник, який досліджувався в ході експериментів – це швидкодія виконання процедури генерації ключів $t_{ГК}$. За результатами проведених експериментів автори виконали порівняльну характеристику швидкодії процедур генерації ключів і визначили оптимальний час $t_{ГК}$ для виконання процедури генерації ключів, а також метод за допомогою якого вона виконана.

5 РЕЗУЛЬТАТИ

В ході проведених експериментів (рис. 2) автори отримали експериментальні дані щодо генерації ключів із використанням задачі дискретного логарифмування (табл. 5).

При досяжності розрядності модуля p у 256 біт і більше, в першому експерименті спостерігається експоненціальне зростання часу відведеного на генерацію ключів. Оскільки за вимогами сучасних криптографічних систем безпечна розрядність ключа повинна складати 1024–2048 біт, то час необхідний для генерації ключів можна описати виразом $t_p \rightarrow \infty$.

В ході проведення другого експерименту (рис. 2) експериментальні дані, отримані авторами, носять практично константний характер, тобто одне й те саме значення для ключів, яке складає 15–16 мс для ключів розрядністю до 2048 біт і більше (коливання часу на рівні 1 мс).

Таблиця 3 – Відновлення секретного повідомлення m

Позначення параметрів	Значення
D	2960510531747690092524546602264372753048454847654850496789135147941202697542
π	58422416232404436471085961576632679733
m	8751735916204352851

Таблиця 4 – Відновлення секретного повідомлення m'

Позначення параметрів	Значення
r_2	976111453564
r_1	27445582010364396438306983190780524779013243571405838002487480018640703378326
A_1	86120330409262410220532378382346931244890834587769244209422058121590708003399
A_2	42368364285961378337324940846616483686887069999828052556155055445614458883516
m_1	2057271081577553248481314125475413
m_2	8751735916204352851
m	8751735916204352851
r	976111453564
m'	2057271081577553248481314125475413

«Відправник»

«Одержувач»

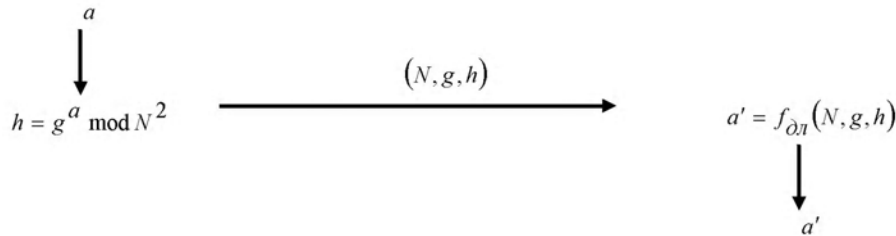
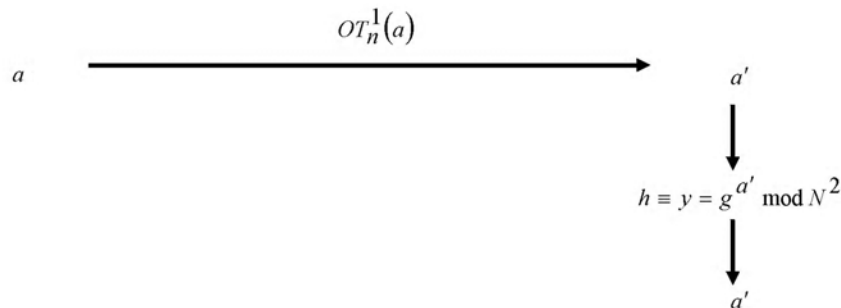
Експеримент №1 – Задача дискретного логарифмування
 (запропонована розробником алгоритму)

Експеримент №2 – Використання протоколу «непомітної»
передачі OT_n^1 (запропонована авторами статті)


Рисунок 2 – Схема експериментів

Таблиця 5 – Дані отримані з першого експерименту

Модуль p , біт	8	16	32	64	128	256
Час $t_{ГК}$, секунд	0,051	0,46	0,61	6,19	598,92	8329,3

6 ОБГОВОРЕННЯ

Проблема практичної реалізації алгоритму заперечуваного шифрування Менга на початку статті полягала обчислювальній стійкості задачі дискретного логарифмування, при якій час генерації виконання усього алгоритму в цілому описується виразом $t_p \rightarrow \infty$. Використовуючи експериментальні дані з табл. 5 автори зробили теоретичний прогноз щодо зростання часу генерації ключів, який відображає графік на рис. 3.

Згідно із прогнозом авторів у першому експерименті, для генерації ключів в алгоритмі заперечуваного шифрування Менга розрядністю до 4096 біт, комп'ютеру із запропонованими тестовими характеристиками необхідно витратити до 596 днів машинного часу. Що згідно зі стандартами криптографічних систем, які декларують безпечну розрядність ключа в 1024–2048 біт, є не оптимальним застосуванням алгоритму для вирішення прикладних завдань.

За результатами другого експерименту авторами було встановлено, що оптимальний час для виконання всього алгоритму заперечуваного шифрування Менга складає 15–16 мс для ключів розрядністю до 4096 біт і

Оцінка часу генерації ключів

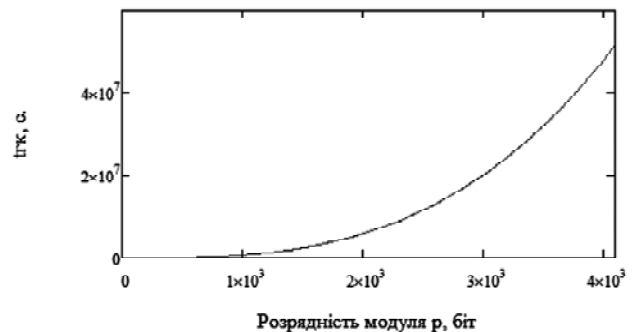


Рисунок 3 – Прогнозована оцінка зростання часу виконання процедури генерації ключів в першому експерименті

більше. Таким чином, дані отримані в результаті проведення другого експерименту (рис. 2), із використанням протоколу «непомітної» передачі OT_n^1 , свідчать про ефективність виконаних модифікацій в алгоритмі заперечуваного шифрування Менга. Отже підхід запропонований авторами для вирішення проблеми у первісного алгоритму є прийнятним з точки зору інформаційної безпеки та практичної реалізації власне алгоритму, а отже застосування протоколу «непомітної» передачі в достатній мірі вирішує проблему пов'язану зі структурою первісного алгоритму заперечуваного шифрування Менга [7].

ВИСНОВКИ

Згідно з метою, поставленою на початку статті, авторами вирішено проблему практичного застосування алгоритму заперечуваного шифрування Менга. В результаті проведених досліджень та експериментів було вирішено наступні завдання:

- виконано огляд рішень проблеми, яку автори визначили структурі алгоритму заперечуваного шифрування Менга, на основі аналогічних алгоритмів заперечуваного шифрування персональних даних;

- виконано аналіз структури та особливостей подібних алгоритмів заперечуваного шифрування та визначено методи для вирішення проблеми;

- виконано модифікацію алгоритму заперечуваного шифрування Менга за допомогою використання протоколу «непомітної» передачі OT_n^1 ;

- на основі проведених експериментів було проведено аналіз ефективності виконаних модифікацій і доведено ефективність використання протоколу «непомітної» передачі OT_n^1 для усунення недоліків первісного алгоритму заперечуваного шифрування Менга.

За результатами досліджень автори виконали модифікацію алгоритму заперечуваного шифрування Менга (Додаток Б). Для вирішення проблеми було використано новий підхід заснований на використанні протоколу «непомітної» передачі OT_n^1 , що дозволило суттєво зменшити час роботи алгоритму Менга та зробити можливим його використання для вирішення практичних завдань в сфері інформаційної безпеки.

Наукова новизна полягає у використанні протоколу «непомітної» передачі для розподілу ключів між абонентами, що є принципово новим підходом до захисту інформації в алгоритмі заперечуваного шифрування Менга.

Робота виконана на кафедрі захисту інформації в рамках НДКР №04515 «Дослідження і розробка криптографічних та технічних засобів захисту інформації».

Гальченко А. В.¹, Козина Л.²

¹Ст. гр. РТ – 710М, магістр кафедри захисту інформації Запорозького національного технічного університету, Запорозьке, Україна

²Канд. физ.-мат. наук, доцент кафедри захисту інформації Запорозького національного технічного університету, Запорозьке, Україна

МОДИФИКАЦИЯ АЛГОРИТМА ОТРИЦАЕМОГО ШИФРОВАНИЯ МЕНГА

В статье обсуждается проблема устойчивости современных криптографических систем к атакам на основе принуждения в отношении абонентов криптографических систем. В связи со стремительным развитием отрасли информационных технологий эта проблема актуальна в сфере информационной безопасности. Для решения проблемы устойчивости современных криптографических систем авторы предлагают использовать алгоритмы отрицаемого шифрования, которые гарантируют, что злоумышленник не имеет возможности получить какую – либо ценную информацию от абонентов. Решение проблемы заключается в использовании алгоритма отрицаемого шифрования Менга, который гарантирует защиту не только информации, но и самих участников обмена.

Основная цель статьи состоит в модификации первоначального алгоритма отрицаемого шифрования Менга [1] с помощью протокола «незаметной» передачи OT_n^1 , использование которого предложено Мони Наором [2]. Применение протокола «незаметной» передачи позволяет существенно сократить время, необходимое для выполнения алгоритма отрицаемого шифрования Менга, и упрощает его реализацию для решения прикладных задач в области информационной безопасности.

По результатам проведенных экспериментов авторами статьи было подтверждено, что использование протокола «незаметной» передачи является эффективным решением проблемы генерации и распределения ключей в алгоритме отрицаемого шифрования Менга.

Ключевые слова: протокол, отрицаемое шифрование, неоднозначность, распределение ключей, фиктивное сообщение, злоумышленник, расширенная схема Рабина, незаметная передача, факторизация, дискретное логарифмирование.

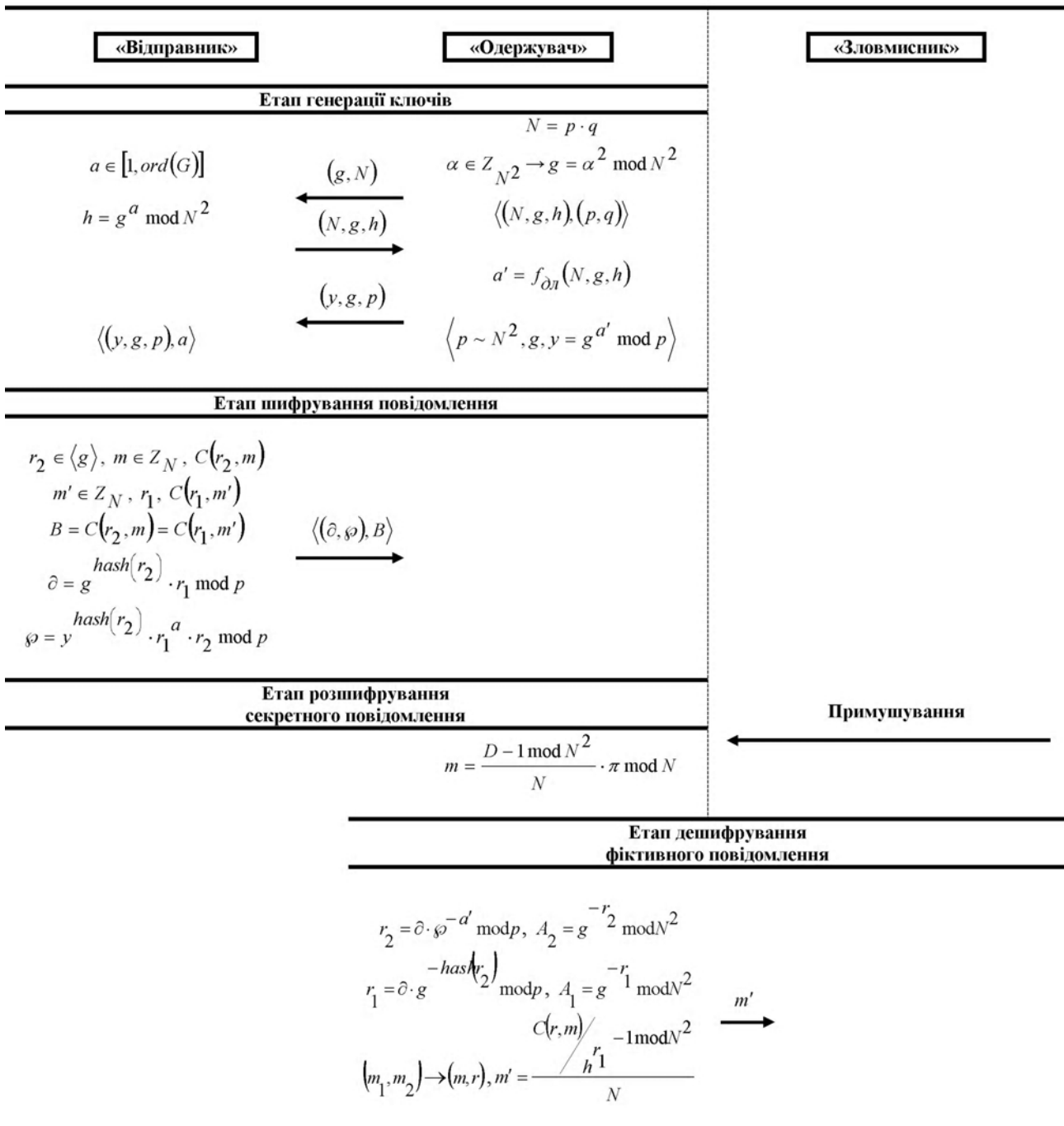
СПИСОК ЛІТЕРАТУРИ

1. Wang J. A Receiver Deniable Encryption Scheme / J. Wang, Bo Meng // Proceedings of the 2009 International Symposium on Information Processing (ISIP'09), 21–23 August 2009: proceedings. – Huangshan : P. R. China, 2009. – P. 254–257.
2. Naor M. Efficient oblivious transfer protocols / M. Naor, B. Pinkas // Proceedings of SIAM Symposium on Discrete Algorithms (SODA '01), 2001: proceedings. – Society for Industrial and Applied Mathematics, 2001. – P. 448–457.
3. Ibrahim H. Receiver–deniable Public–Key Encryption / H. Ibrahim // International Journal of Internet Security. – 2009. – Vol. 8, № 2. – P. 159–165.
4. Козіна Г. Л. Заперечуване шифрування / Г. Л. Козіна, А. В. Гальченко // Тиждень науки – 2015: Тези доповідей щорічної наук. – практ. конф. викладачів, науковців, молодих учених, аспірантів, студентів ЗНТУ, Запоріжжя, 13–17 квітня 2015 р. – Запоріжжя : ЗНТУ, 2015.
5. Canetti R. Deniable Encryption / [R. Canetti, C. Dwork, M. Naor, R. Ostrowsky] // Advances in Cryptology. – CRYPTO, 1997, Proceedings. – P. 90–104.
6. Молдовян Н.А. Расширение криптосхемы Рабина: алгоритм отрицаемого шифрования по открытому ключу / Н. А. Молдовян, А. А. Горячев, М. А. Вайчикаускас // ВЗИ. Журнал по вопросам защиты информации. – ФГУП «ВИМИ», 2014. – № 2. – С. 12–16.
7. Фисун С.Н. Комбинированный алгоритм вероятностного шифрования / С. Н. Фисун, О. И. Куржиевская // Изд-во СевНТУ, 2010. – № 101. – С. 37–40.
8. Bresson E. A Simple Public–Key Cryptosystem with a Double Trapdoor Decryption Mechanism and its Applications / E. Bresson, D. Catalano, D. Pointcheval // Advances in Cryptology – ASIACRYPT, 2003. LNCS, Vol. 2894. Springer, Heidelberg, 2003. – P. 37–54.
9. Klonowski M. Practical Deniable Encryption // M. Klonowski, P. Kubiak, and M. Kutylowski // SOFSEM 2008: Theory and Practice of Computer Science, 34th Conference on Current Trends in Theory and Practice of Computer Science, Slovakia : Novy Smokovec, 19–25 January 2008: proceedings. – 2008. – P. 599–609.
10. Николенько С. Поиск дискретного логарифма [Электронный ресурс] / С. Николенько. – Режим доступа: https://compcenter.ru/media/slides/cryptoprotocols2014_2015_spring/2015_03_11_cryptoprotocols2014_2015_spring_IGKdY5s.pdf.
11. Горбенко И. Д. Методы распараллеливания алгоритма Полларда решения задачи дискретного логарифмирования для систем с общей памятью / И. Д. Горбенко, Е. Г. Качко, К. А. Погребняк // Изд-во ХНУРЕ, 2012. – С. 1–6.

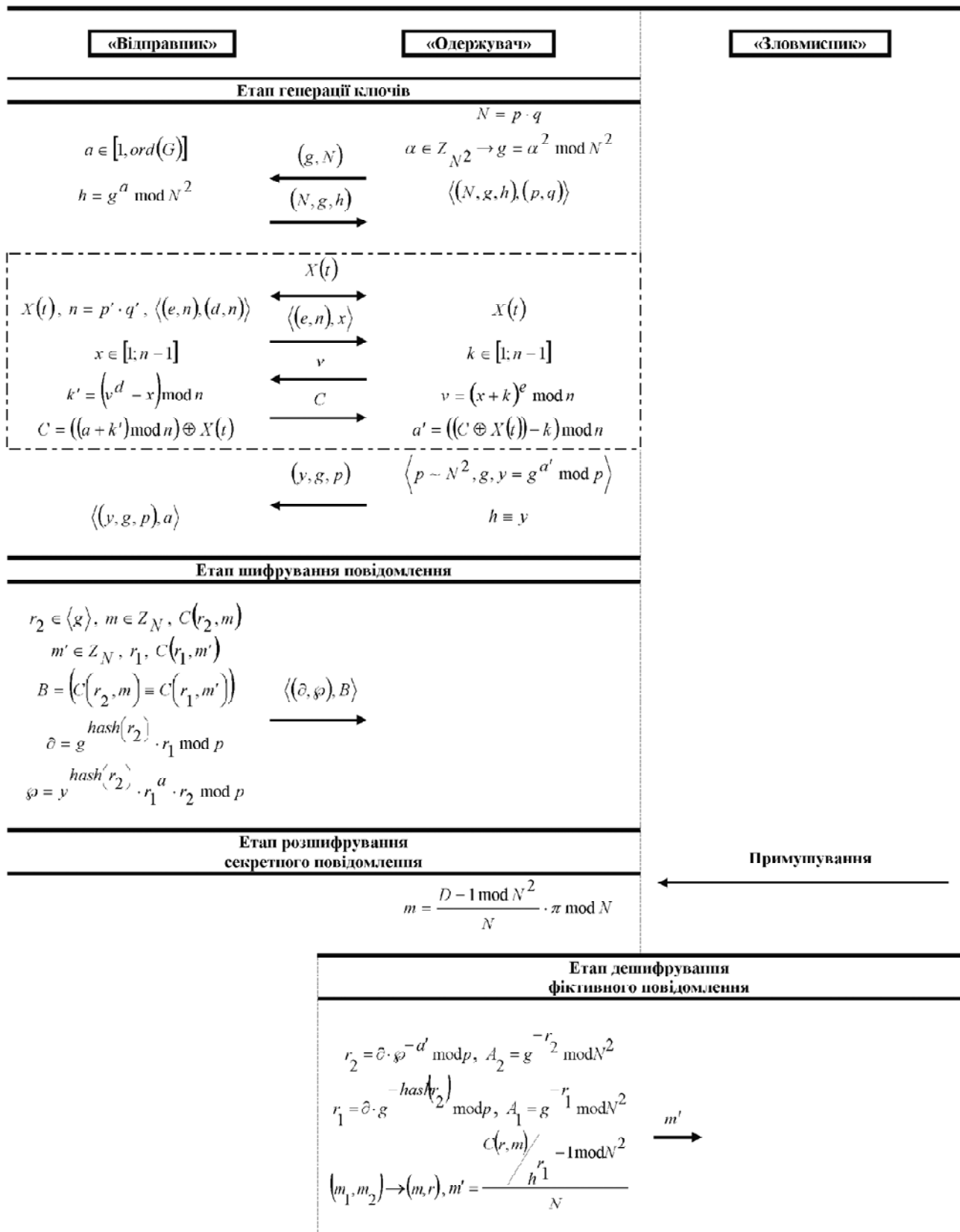
Стаття надійшла до редакції 16.12.2015.

Після доробки 20.12.2015.

Додаток А – Структурна схема первісного алгоритму заперечуваного шифрування Менга



Додаток Б – Структурна схема модифікованому алгоритму заперечуваного шифрування Менга



Galchenko A. V.¹, Kozina G. L.²

¹Gr. RT – 710M, Master of the Department of Information Protection of Zaporozhzhya National Technical University, Zaporizhzhya, Ukraine

²Candidate of Phys.-Math. sciences, Associate Professor of the Department of Information Protection of Zaporozhzhya National Technical University, Zaporizhzhya, Ukraine

MODIFICATION OF MENG'S DENIABLE ENCRYPTION ALGORITHM

The article discusses the stability of modern cryptographic systems to attack from coercion in respect of subscribers cryptographic systems. Due to the rapid development of information technology, this problem is relevant in the field of information security. To address the sustainability of modern cryptographic systems use algorithms offered by the authors deniable encryption ensures that the attacker is unable to get any valuable information from subscribers. Solving the problem is to use Meng's deniable encryption algorithm, which guarantees protection not only information, but also the participants of the exchange.

The main purpose of the article is performed by modifying the initial Meng's deniable encryption algorithm [1] with using Oblivious Transfer Protocol, which prompted by Moni Naor [2]. Oblivious Transfer Protocol to significantly reduce the time required to perform the Meng's deniable encryption algorithm, and facilitates its implementation to solve applied problems in the field of information security.

As a result of experiments, the authors confirmed that Oblivious Transfer Protocol using is an effective solution to the problem of generation and distribution of keys in the Meng's deniable encryption algorithm.

Keywords: protocol, deniable encryption, ambiguity, distribution keys, fake messages, an extended Rabin's scheme, oblivious transfer, factorization, discrete logarithm.

REFERENCES

1. Wang J., Meng Bo A Receiver Deniable Encryption Scheme, *Proceedings of the 2009 International Symposium on Information Processing (ISIP'09), 21–23 August 2009: proceedings*, Huangshan, P. R. China, 2009, pp. 254–257.
2. Naor M., Pinkas B. Efficient oblivious transfer protocols, *Proceedings of SIAM Symposium on Discrete Algorithms (SODA '01), 2001: proceedings. – Society for Industrial and Applied Mathematics*, 2001, pp. 448–457.
3. Ibrahim H. Receiver-deniable Public-Key Encryption, *International Journal of Internet Security*, 2009, Vol. 8, No. 2, P. 159–165.
4. Kozina G. L., Galchenko A. V. Deniable encryption, *Week of Science – 2015: Abstracts annual scientific – practical conference of teachers, scientists, young scientists, graduate students ZNTU, Zaporizhzhya*, 13–17 April 2015. Zaporizhzhya, ZNTU, 2015.
5. Canetti R., Dwork C., Naor M., Ostronsky R. Deniable Encryption. *Advances in Cryptology – CRYPTO, 1997, Proceedings*, pp. 90–104.
6. Moldovyan N. A., Goryachew A. A., Wichikaukas M. A. Extended Rabins's cryptographic scheme: deniable encryption by public key, *VZI. Journal of information security*. FSUE «VIMI», 2014, No. 2, pp. 12–16.
7. Fisun S. N., Kurzhiyevskaya O. I. Combined probabilistic encryption algorithm, *SevNTU*, 2010, No. 101, pp. 37–40.
8. Bresson E., Catalano D., Pointcheval D. A Simple Public – Key Cryptosystem with a Double Trapdoor Decryption Mechanism and its Applications, *Advances in Cryptology – ASIACRYPT, 2003. LNCS, Vol. 2894*. Springer, Heidelberg, 2003, pp. 37–54.
9. Klonowski M., Kubiak P., and Kutylowski M. Practical Deniable Encryption, *SOFSEM 2008: Theory and Practice of Computer Science, 34th Conference on Current Trends in Theory and Practice of Computer Science, Slovakia: Novy Smokovec, 19–25 January 2008: proceedings. – 2008. P. 599–609*.
10. Nikolenko S. Search the discrete logarithm [Electronic resource]. Access: https://compscicenter.ru/media/slides/cryptoprotocols2014_2015_spring/2015_03_11_cryptoprotocols2014_2015_spring_IGKdY5s.pdf.
11. Gorbenko I. D., Musced E. G., Pogrebnyak K. A. Methods of algorithm parallelization Pollard solutions discrete logarithm problem for systems with shared memory, *KNURE Publishing House*, 2012, pp. 1–6.

Гороховатский В. А.¹, Берестовский А. Е.², Передрий Е. О.³

¹Д-р. техн. наук, профессор, профессор кафедры информационных технологий, Харьковский учебно-научный институт
ГВУЗ «Университет банковского дела», Харьков, Украина

²Аспирант кафедры информатики, Харьковский национальный университет радиоэлектроники, Харьков, Украина

³Канд. техн. наук, преподаватель кафедры информатики и компьютерной техники, Харьковский национальный
экономический университет имени Семена Кузнеца, Харьков, Украина

СИСТЕМАТИЗАЦИЯ ПРОСТРАНСТВА СТРУКТУРНЫХ ПРИЗНАКОВ НА ОСНОВЕ МЕТОДОВ САМООБУЧЕНИЯ В ЦЕЛЯХ РЕЗУЛЬТАТИВНОГО РАСПОЗНАВАНИЯ ИЗОБРАЖЕНИЙ

Работа посвящена исследованию вопросов кластеризации для множеств характерных признаков изображений. Для построения массива характерных признаков использован метод Speeded Up Robust Features. Реализованы алгоритмы кластеризации структурных описаний изображений на основе самоорганизующейся нейронной сети Кохонена и метода разностного группирования. Объектом исследования есть методы кластеризации применительно к множествам структурных признаков. Целью работы является построение векторных представлений описаний на основе кластеризации, что повышает быстродействие распознавания. Предметом исследования является систематизация множеств структурных признаков визуальных объектов.

Обсуждаются результаты применения методов кластеризации для структурных описаний изображений в виде множеств характерных признаков с целью повышения быстродействия распознавания визуальных объектов. Для систематизации и сжатия пространства признаков предложено осуществить самообучение с применением методов разностного группирования и сетей Кохонена.

Проведено моделирование и экспериментальные исследования методов кластеризации на примерах конкретных множеств характерных признаков. Результаты исследований доказывают возможность эффективного представления описаний в виде вектора с целочисленными элементами. Данный подход может использоваться для решения задач распознавания и поиска изображений.

В результате построено компактное векторное описание эталонов, получены количественные оценки ошибки кластеризации, подтверждена работоспособность методов для прикладной базы изображений.

Ключевые слова: компьютерное зрение, распознавание изображений, характерные признаки, структурное описание изображения, метод SURF, кластеризация, нейронная сеть, метод разностного группирования, сеть Кохонена, ошибка квантования.

НОМЕНКЛАТУРА

SOM – Self-Organizing Map (самоорганизующиеся карты);

SURF – Speeded up robust features (ускоренное устойчивое выделение особенностей);

WTA – Winner takes all (победитель получает все);

ПФ – пиковая функция;

РГ – разностное группирование;

СК – сеть Кохонена;

ХП – характерный признак;

$\rho(z, c_i)$ – расстояние между z и c_i ;

r – константа, которая определяет «сферу соседства»;

b – показатель степени многомерной функции Гаусса;

s – мощность множества Z ;

ε – доля от значения пика $D(c_i)$;

$z \in C_i$ – точка данных кластера C_i с центром c_i ;

k – число кластеров;

S – мощность обучающего множества Z ;

$0 < \eta < 1$ – коэффициент обучения;

$z(h)$ – входной вектор $z \in Z$ на шаге h .

ВВЕДЕНИЕ

Современные системы компьютерного зрения нацелены на автоматизацию решения актуальных прикладных задач искусственного интеллекта: распознавание человеческих лиц в целях идентификации; поиск визуальных объектов заданного вида в базах (коллекциях) изображений; распознавание объектов на изображениях сцен, включая условия влияния перекрытий объектов,

фона и помех; идентификация и определение координат движущихся объектов в видео-потоке. Перспективным подходом к построению систем распознавания в таких сложных ситуациях есть структурный анализ [1].

Процесс извлечения знаний в системах интеллектуальной обработки условно трактуют как ряд этапов [2–5]:

1. Отбор и предварительная обработка данных.
2. Редукция/проекция данных.
3. Поиск, оценка и интерпретация закономерностей.
4. Применение приобретенных знаний.

Способность к обучению и обобщению накопленных знаний (этап 3) считают одним из критериев, определяющих уровень интеллектуальности системы в рамках современной теории интеллектуальных вычислений. Обучение позволяет не только адаптировать распознавание к имеющимся данным и конкретным условиям, но и выявляет новые закономерности в целях дальнейшего обобщения знаний и улучшения результативности. В ходе обучения часто определяются ключевые признаки, в наибольшей степени отличающие образы объектов, что в целом систематизирует пакет признаков и дает возможность образовать надежный фундамент для принятия качественных решений о классификации образов. Система распознавания совершенствует свою эффективность через процесс обучения на основе данных из окружающей среды [2].

1 ПОСТАНОВКА ЗАДАЧИ

Целью статьи есть изучение возможности и оценивание эффективности применения методов разностного группирования и аппарата сетей Кохонена для осуществления результативного самообучения системы струк-

турного розпознавання зображень в плані побудови кластерного компресійного представлення даних в пространстві ознак прикладної бази зображень. За рахунок кластеризації-класифікації на множині структурних елементів забезпечується перехід до векторному представленню еталонів, що значительно скорочує обсяг обчислювальних витрат і сприяє покращенню швидкодії розпознавання.

Задачі дослідження – вивчення особливостей і удосконалення інформаційних технологій самообучення применительно до описання в вигляді множин дескрипторів структурних ознак зображень, а також оцінювання якості кластеризації на прикладних зразках.

2 ОБЗОР ЛИТЕРАТУРЫ

Обучение и самообучение, которые находят применение в структурном распознавании изображений, успешно осуществляют систематизацию признаков путем кластер-анализа (этап 2), что способствует снижению размерности признакового пространства [5–8]. Данные, сгруппированные в рамках кластера, представляются его центром и размером, что пропорционально сокращает как объем памяти, так и вычислительные затраты на распознавание [1].

Для успешной кластеризации требуется инициализация – тщательная предварительная обработка данных с целью подготовки начальных условий функционирования методов обучения. Например, для группирования данных необходимо установить число кластеров и начальные значения их центров в пространстве данных. Затем метод обучения, например, сеть Кохонена, оттачиваясь от этих значений, уточняет и завершает группировку объектов. Заметим, что, k -степени зависят от начальных установок. Задачи инициализации могут быть решены методами пикового и разностного группирования [3], которые можно считать самостоятельными средствами получения знаний. В сравнении с возможностями аппарата сетей Кохонена эти методы позволяют автономно осуществить самоорганизацию центров образовавшихся кластеров с оцениваемой точностью.

В литературе описан ряд теоретических и практических аспектов построения универсальных методов самоорганизации данных, таких как разностное группирование [3] и сети Кохонена [3, 5, 9]. Основы теории кластеризации изложены в работах [6, 7]. В то же время теория и практика применения этого аппарата к множествам структурных признаков в целях распознавания визуальных объектов на изображениях только начинают свое развитие [8]. Особый интерес представляет изучение специфики технологий обучения и выбор наиболее подходящих среди них. Важно также оценить влияние ошибки обучения на результат распознавания с применением модифицированного пространства признаков.

3 МАТЕРИАЛЫ И МЕТОДЫ

Разностное группирование (РГ) наиболее эффективно в применении к векторам большой размерности, которая для структурных признаков SURF равна 64 и выше [4]. Рассмотрим особенности РГ применительно к пакетам векторов-дескрипторов SURF.

Структурное описание изображения – это конечное множество $Z \subset R_1^n$, $R_1^n = \{z \mid z \in R^n, \|z\| \approx 1\}$, где $R_1^n \subset R^n$ – пространство n -мерных векторов с евклидовой нормой:

$\|z\| = \sqrt{\sum_{k=1}^n z_k^2} \approx 1$ [1]. Выполнение условия нормировки к 1 позволяет напрямую применять вектора описания в процедурах обучения без дополнительной обработки.

В методе РГ кластеры строятся последовательным урезанием исходного множества способом агломерации. Вначале для каждого $z_i \in Z$ осуществляют вычисление значения пиковой функции (ПФ)

$$D(z_i) = \sum_{j=1}^s \exp \left\{ \frac{-\rho^{2b}(z_i, z_j)}{(r/2)^2} \right\}, \quad (1)$$

Идея РГ лежит в русле развития метода потенциальных функций как одного из наиболее общих подходов к классификации образов [6,7]. Значение (1) пропорционально числу векторов из окрестности центра z_i . Малое значение $D(z_i)$ свидетельствует, что центр располагается в зоне сосредоточения незначительной группы векторов z_j . Считается, что коэффициент r практически не оказывает влияния на итоговые пропорции между $D(z_i)$, поэтому его подбор не критичен [3].

После расчета значений ПФ $\forall z_i \in Z$ отбирается вектор z с наибольшей мерой $D(z)$. Эта точка – первый центр: $c_1 = \arg \max_{z \in Z} D(z)$. Перед поиском следующего центра исключаем c_1 из Z и все точки в пределах сферы r . Все они образуют первый кластер.

Отбор элементов сферы соседства для кластера C_1 можно осуществить, например, путем анализа близости значений ПФ:

$$C_1 = \{z \in Z \mid D(c_1) - D(z) \leq \varepsilon D(c_1)\}, \quad (2)$$

Далее переопределяем ПФ для оставшихся точек:

$$D_{new}(z_i) = D(z_i) - D(c_1) \exp \left\{ \frac{-\rho^{2b}(z_i, c_1)}{(r_1/2)^2} \right\}, \quad (3)$$

ПФ $D_{new}(z_i)$ принимает нулевое значение при $z_i = c_1$ и близка к нулю для элементов C_1 .

После модификации (3) определяется следующая точка z с максимальным $D_{new}(z)$, $z \in Z$, $z \notin C_1$. Она образует центр c_2 . Поиск следующего центра возобновляется после исключения компонентов, включенных в уже отобранные кластеры. Кластеризация завершается при фиксации всех центров, предусмотренных начальными условиями.

Метод РГ реализует самоорганизацию множества векторов, суть которой – нахождение центров, представляющих множество данных с минимальной погрешностью. Заметим, что метод РГ инвариантен к нумерации списка входных точек, что важно для задач обработки изображений.

Критерием качества кластеризации (погрешность квантования) выступает функционал усредненной по

числу записей суммы квадратов расстояний между центрами кластеров и включенными в них данными [3]

$$E = \frac{1}{s} \sum_{i=1}^k \sum_{z \in C_i} \rho^2(z, c_i), \quad (4)$$

Значение (4) – это усредненная ошибка в течение s шагов обучения. Выражение (4) можно применить как ко всей базе, так и к отдельному эталону. Если в (4) убрать усреднение – получим величину мгновенной ошибки на очередном шаге обучения.

Важно не только значение ошибки, но и ее динамика изменения в процессе обучения. В частности, представляет интерес событие стабилизации значений центров, что соответствует незначительным колебаниям (4). Методы, минимизирующие (4), называют группировкой с минимальной дисперсией [6]. Оптимизация критерия (4) в процессе кластеризации вскрывает внутреннюю структуру пространства данных в виде концентрированных сгущений точек.

Самоорганизующиеся карты (SOM – Self-Organizing Map) Кохонена (СК) реализуют конкурентное обучение на основе нейробиологического принципа по схеме WTA (Winner Takes All – победитель получает все) [9]. СК – однослойная искусственная нейронная сеть с прямой передачей информации. Преимуществами СК, кроме объединения операций кластеризации и проецирования, считаются простота архитектуры и независимость самоорганизации от размерности задачи [9]. В результате самообучения СК пространство Z разбивается на k подобластей $Z_v \subset Z$ так, что вектор-образу $z \in Z_v \subset Z$ соответствует точка-нейрон пространства $Y: Z = \bigcup_{v=1}^k Z_v$, $Z_v \cap Z_\tau = \emptyset, v \neq \tau$.

Схема соединения и трансформации нейронов в пространстве признаков SURF показана на рис. 1.

Результаты классификации-кластеризации посредством СК в значительной степени зависят от начальных весов нейронов, которые зачастую инициализируют случайным образом. Одним из путей уменьшения степени эвристичности есть применение непосредственно векторов из обучающей выборки [3, 8].

Отметим существенное отличие принципов построения методов РГ и СК: РГ выбирает центры кластеров из имеющегося списка, в то время как СК создает и подстраивает сеть центров под имеющиеся данные. Один из

вариантов коллективного использования СК и РГ предполагает применение на первом этапе РГ для инициализации и определения таких параметров сети, как центры и число кластеров. На втором этапе употребляется СК для завершения самоорганизации нейронов. Обучающее множество используется повторно на каждом шаге обработки.

Формализуем обучение СК для распознавания. Имеем конечное множество $Z = \{Z^i\}_{i=1}^J$ структурных описаний

базы из J эталонов. Эталон Z^i – это конечное множество дескрипторов SURF. Поставим задачу осуществить кластеризацию Z на k кластеров. В качестве обучающей выборки используем множество Z , а нейроны $W = (W_1, W_2, \dots, W_k)$ (центры кластеров) представим пакетом из k векторов $\{W_i = (W_{i,1}, W_{i,2}, \dots, W_{i,64})\}, i = 1, k$.

По принципу обработки WTA на очередном шаге для обучающего вектора $z \in Z$ определим номер q нейрона-победителя:

$$q = \arg \min_i \rho(W_i, z). \quad (5)$$

Подмножество $W^* \subseteq W$ соседей W_q определяется как $W^* = \{W_i \mid \rho(W_i, z) \leq r\}$, где r – порог. В обучении по Кохонену применяется линейная схема подстройки подмножества W^* в направлении вектора z [3]

$$W_i(h+1) = W_i(h) + \eta(h)[z(h) - W_i(h)]. \quad (6)$$

Выражение (6) отвечает градиентному методу оптимизации, а значение $[z(h) - W_i(h)]$ определяет направление в многомерном пространстве, в соответствии с которым осуществляется уточнение вектора весов. Для случая нормализованных к единице входных векторов определение нейрона-победителя по максимуму скалярного произведения равнозначно критерию наименьшего евклидова расстояния. Следствием конкуренции становится самоорганизация в ходе обучения, а нейроны-победители приобретают свойство различения «своей» категории входных данных.

Метод самообучения Кохонена включает следующие шаги [3, 5].

1. Инициализация. Для k нейронов сети устанавливаются начальные нормализованные веса, скорость η и радиус r обучения.

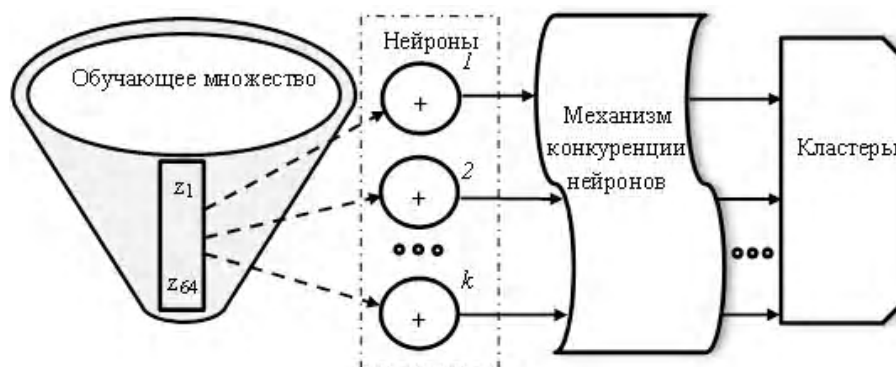


Рисунок 1 – Схема обучения СК на множестве Z

2. Возбуждение. На входной слой подается вектор z обучающей выборки.

3. Конкуренция. Для каждого нейрона W_v вычисляется расстояние $\rho[W_v, z]$, определяется нейрон-победитель.

4. Определяется подмножество нейронов в пределах радиуса обучения, их веса подстраиваются в соответствии с (6).

5. Коррекция параметров.

Натренированная сеть функционирует как векторный классификатор, основанный на сопоставлении входного вектора с набором сформированных векторов-нейронов [3]. Если обучение осуществить на группе похожих векторов с коэффициентом $\eta < 1$, то веса нейронов примут значения, усредненные по этим векторам. В режиме классификации СК лучше всего реагирует на векторы, близкие к средним значениям векторов обучающей группы. Сеть реализует функцию векторного квантования: за счет самоорганизации произвольное количество многомерных данных, образующих кластер, отображается вектором весов центра. Пространственное размещение нейронов определяет зоны концентрации данных. Представление структурного описания базы в виде конечного множества центров кластеров – это компрессия данных с потерями, которая сопровождается определенной погрешностью квантования.

Результат кластеризации для множества ХП прикладной базы изображений всецело определяется используемыми технологиями обучения, которые в настоящее время строго не формализованы и существенно зависят от опыта исследователя. Считается, что независимо от способа обучения самоорганизующихся сетей значима избыточность обучающих данных, без которой качественное обучение невозможно [3].

Процесс обучения сводится к формированию кластеров с фиксированным механизмом конкуренции нейронов при подборе параметров k , η , r . Например, для предотвращения расходимости обучения вблизи локального минимума отслеживают значение целевой функции $E(h)$ на шаге h с тем, чтобы не допустить ее возрастания сверх фиксированного ограничения, например в 5% [3]. Если выполнено условие $E(h+1) < 1,05E(h)$, то изменения игнорируют, приращение считается несущественным, веса не изменяются. Если же $E(h+1) \geq 1,05E(h)$, очередной шаг считается целесообразным, и уточнение проводится. Подбор коэффициентов требует значительного числа экспериментов и часто зависит от специфики проблемы.

Значимым моментом для обеспечения высокой статистической точности есть возможность многократного использования данных с изменяющимися параметрами обучения. Множество обучающих векторов предъявляют несколько раз, причем в случайной последовательности (бутстреп-обучение [9]). Технология многократного предъявления выборки способствует стабилизации весов и ошибки сети. Обучение завершают, если изменение вектора весов становится меньше принятого значения толерантности.

Грамотный подбор коэффициента η , как правило, изменяющегося в ходе обучения, оказывает огромное влияние на сходимость к минимуму целевой функции. Слишком малое значение не позволяет быстро минимизировать целевую функцию и требует многократных итераций. Большой шаг может привести к «перепрыгиванию» через минимум и повторных возвращений к нему. Практически приемлемые результаты в плане сходимости достигаются при η с течением обучения [9]. Фиксация η на весь период обучения упрощает обработку. В то же время более эффективный метод основан на адаптивном подборе коэффициента с учетом фактической динамики величины целевой функции.

Если обозначить погрешности соответственно e_{i-1} , e_i на последующих шагах обучения, а η_{i-1} , η_i – коэффициенты обучения, то в случае $e_i > \gamma e_{i-1}$ (γ – коэффициент допустимого роста погрешности) значение η уменьшают по формуле $\eta_{i+1} = \alpha \eta_i$, где α – коэффициент уменьшения. Если же $e_i \leq \gamma e_{i-1}$, то осуществляют увеличение $\eta_{i+1} = \beta \eta_i$, где β – коэффициент увеличения. Такой адаптивный метод подбора η сильно зависит от вида целевой функции и значений α , β , γ . Оптимальные для функций одного вида значения могут замедлить процесс обучения для других функций. В одной из задач квадратичной аппроксимации удачно использованы значения коэффициентов $\gamma = 1,4$, $\alpha = 0,7$, $\beta = 1,05$ [3]. Другие виды адаптации предполагают задание прямой зависимости снижения η от шага. Один из вариантов изменения η в течение N шагов может иметь вид $\eta(t) = 0,9(1 - t/N)$. Другой распространенный вариант – $\eta(t) = A/(t + B)$, где A , B – константы [3].

Иногда применяют разновидность обработки, где нейрон-победитель уточняется по формуле (6), а ближайшие к нему центры – в противоположном направлении с коэффициентом $-\eta$, что позволяет отдалить близкие центры и тщательно обследовать пространство данных. Исследователи отмечают, что ни один из алгоритмов не гарантирует абсолютную сходимость к глобальному оптимуму, обеспечивая лишь локальную оптимизацию, зависящую от начальных условий и параметров [3]. В итоге нейросетевые методы, включая SOM, не дают однозначно определяемых результатов [9]. Адаптивные процессы могут повести себя совершенно неожиданным образом в зависимости от выбранных значений их параметров и хода обучения. Следует использовать проверенные рекомендации и программные средства, чтобы обеспечить контроль над ходом самоорганизации и качество результатов в приложениях.

Обратим также внимание на тот факт, что ошибка обучения (4), вычисляемая пошагово в процессе обучения, не совпадает со значением, вычисленным после завершения обучения (пост-ошибка), т.к. нейроны-центры изменяются. Если для обучения значима динамика ошибки, то после обучения уже важна ее абсолютная величина как итог обучения. Для получения абсолютно-го значения необходимо пересчитывать (4) либо при каждом изменении нейрона-центра, либо по результату фиксированного числа шагов.

С точки зрения результативности классификации-кластеризации интересен экспериментальный анализ следующих стратегий обучения.

1. Сравнение качества кластеризации раздельно и совместно для двух обсуждаемых методов (РГ и СК) при фиксированном коэффициенте обучения.

2. Анализ качества методов при изменяющемся по некоторому закону коэффициенте обучения.

3. Изучение функционирования методов с применением адаптации и многократного употребления множеств входных данных.

Эксперименты проведены для структурных описаний, полученных методом SURF для базы из 25 изображений гербов городов Украины [8]. Моделирование метода РГ показало, что при постоянном $r=1$ значение максимума ПФ монотонно снижается (рис. 2).

Моделирование для изображения герба Харькова (342 ХП) показало, что погрешность квантования (4) для метода РГ находится в диапазоне 0,4...0,5, в то время как для метода СК она составила значение 0,24. Рис. 3 содержит

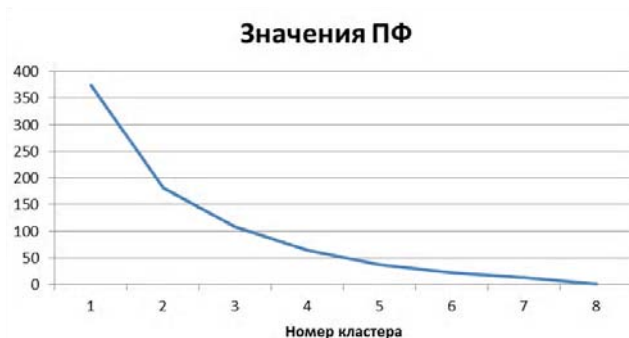


Рисунок 2 – Значения максимума ПФ в процессе кластеризации



Рисунок 3 – Изображение герба Харькова и координаты 30 ХП его описания

жит изображение герба Харькова и координаты 30 ХП из его описания SURF. Число образованных кластеров изменяется в пределах от 1 до 26 с изменением порога ε в соотношении (2) от 0,1 до 0,9. При малых ε наблюдается сосредоточение в 2-3-х кластерах, что впоследствии может сказаться на значении вероятности распознавания. По этой причине для данного набора ХП можно рекомендовать $\varepsilon = 0,7$, при котором образовано $q = 11$ кластеров, где элементы по кластерам распределены достаточно равномерно, а ошибка квантования составила 0,41.

Исследования для пакета из 5 изображений гербов городов Украины [8], содержащей 1543 дескриптора SURF, практически подтверждают эти пост-экспериментальные рекомендации: при пороге $\varepsilon = 0,6$ получено 7 кластеров, при $\varepsilon = 0,7$ – 14 кластеров с более равномерным распределением элементов, ошибка квантования $E = 0,50$. Дальнейшее увеличение ε приводит еще к более равномерному распределению по кластерам, однако, увеличивается их число и ошибка: при $\varepsilon = 0,9$ значения $q = 42$, $E = 0,77$. Выбор оптимальных значений параметров переносится в область применений. Рисунок 4 демонстрирует зависимость числа кластеров от порога ε .

Исследования показали возможность управления параметром ε для достижения нужного числа кластеров методом РГ, которые впоследствии обеспечат результативное распознавание для произвольной базы изображений. Так для рассматриваемого пакета из 5 гербов при пороге $\varepsilon = 0,61$ имеем $q = 8$ кластеров с ошибкой $E = 0,47$.

Путем применения СК для базы из 5 гербов городов при пороге $\varepsilon = 0,61$ получено распределение по 8-ми кластерам с ошибкой $E = 0,236$.

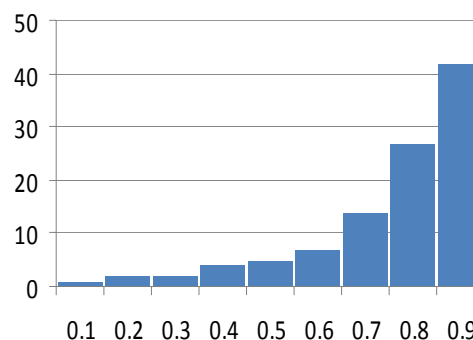


Рисунок 4 – Зависимость числа кластеров от порога кластеризации

Таблица 1 – Количества ХП в кластерном представлении эталонов

Изображения гербов	Номер кластера								Ошибка E
	1	2	3	4	5	6	7	8	
Днепропетровск	40	75	24	9	5	61	5	34	0,271
Львов	14	151	21	14	0	7	1	121	0,253
Киев	10	91	38	27	5	66	4	63	0,229
Харьков	102	113	20	17	4	14	3	69	0,256
Кременчуг	83	51	57	10	5	43	5	61	0,173
База	249	481	160	77	19	191	18	348	0,236

Из табл. 1 (матрица данных) наблюдаем практически равномерное распределение ошибки (4) по множеству эталонов и неравномерное распределение числа элементов по кластерам. Например, кластер 2 содержит 481 ХП (31,2% от общего числа элементов), а кластер 7 – всего 18 ХП, что соответствует 1,2%. В то же время есть мнение, что элементы малочисленных кластеров 5, 7 также влияют на результат распознавания. Альтернативой есть исключение элементов этих кластеров из эталонного описания базы.

Заметим, что в плане результативности распознавания, на наш взгляд, более значимую роль играет уровень существенного преобладания значений отдельных элементов в столбцах матрицы табл. 1. Это наблюдается, например, для кластеров 1–3, 8. Такое преобладание дает возможность с большей степенью уверенности или даже однозначно относить распознаваемый элемент к определенному эталону.

График экспериментально полученного значения ошибки (4) для метода СК при фиксированном $\eta = 0,5$ приведен на рис 5. Как видим, при $s = 300$ величина (4) равна 0,205, а в дальнейшем колеблется в среднем на уровне 0,23. Это характерно для разных вариантов начальных условий. Делаем вывод, что при числе элементов обучения более 300 процесс кластеризации дескрипторов стабилизируется. Ошибка в процессе и после кластеризации находится в пределах интервала 0,2–0,3.

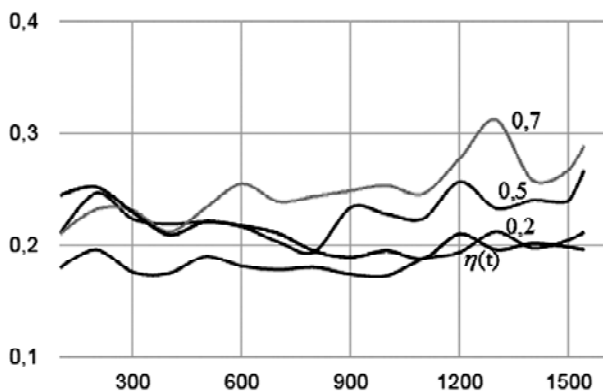


Рисунок 5 – Значения ошибки кластеризации при разных η

5 РЕЗУЛЬТАТЫ

Основным результатом работы есть формализация, моделирование и сравнение полученных экспериментальных оценок для современных методов кластеризации применительно к множествам характерных признаков для прикладной базы изображений.

Самоорганизация на основе сети Кохонена дает примерно в два раза меньшую ошибку квантования на множестве признаков, чем метод разностного группирования, который можно рекомендовать на этапе инициализации. Комбинирование методов разностного группирования и сети Кохонена снижает ошибку квантования в пределах 40%. Восемь кластеров вполне достаточно для осуществления уверенного распознавания в прикладной базе изображений гербов.

6 ОБСУЖДЕНИЕ

Наши эксперименты показали, что с изменением η от 0 до 1 ошибка возрастает. Конечное значение (4) при $\eta = 0,2$ составило 0,211, при $\eta = 0,5$ – 0,267, при $\eta = 0,7$ – 0,289. Кроме того, при постоянном η ошибка начинает медленно возрастать, примерно начиная с шага обучения с номером 1000.

Из графика рис. 5 видим, что изменение $\eta(t) = 0,9(1 - t/N)$ приводит к процессу уменьшения ошибки до значения 0,196. Это несколько меньше, чем при постоянном η , где наименьшее из значений равно 0,211.

Применение повторного обучения на материале базы ХП показало существенное снижение итоговой ошибки обучения от числа повторений. В соответствии с числом повторений экспериментальная ошибка составила: 1) 0,236; 2) 0,045; 3) 0,013; 4) 0,001; 5) 0,0002. Этот факт говорит об имеющихся резервных возможностях при необходимости уменьшать ошибку обучения, особенно для баз мало различающихся изображений, например, человеческих лиц.

Проведены эксперименты по адаптации процесса обучения путем отмены подстройки нейронов при условии $E(h+1) < 1,05E(h)$. Это реализовано для двух вариантов обучения: $\eta = 0,5$ и $\eta(t) = 0,9(1 - t/N)$. Моделирование показало, что в первом случае итоговая ошибка несколько уменьшилась (0,254 вместо 0,267), а во втором – возросла (0,243 вместо 0,196). Как видим, управление процессом обучения путем анализа значений текущей ошибки и игнорирования СК незначимых ее изменений в некоторых ситуациях может привести к возрастанию конечной ошибки.

Применение метода РГ для инициализации СК в сравнении со случайной инициализацией привело к снижению итоговой ошибки квантования (4) в методе СК до уровня 0,142, что соответствует около 40% улучшению. Это подтверждает необходимость и эффективность совместного применения разноплановых подходов кластеризации.

ВЫВОДЫ

Систематизация пространства характерных признаков в задаче распознавания изображений сводится к переходу к кластерному представлению с последующим применением трансформированной меры подобия в новом пространстве. Результирующее распределение признаков эталонов по кластерам определяет качество распознавания. Обучение дает возможность адаптировать структурный анализ к данным эталонного множества признаков прикладной базы изображений, что улучшает показатели распознавания.

Управление обучением путем изменения функцией коэффициента обучения снижает конечную ошибку. Применение адаптации нейронной сети путем анализа значений текущей ошибки не приводит к значимому уменьшению итоговой ошибки квантования. В целом процессу обучения для прикладной базы изображений свойственна стабилизация ошибки квантования при числе шагов, превышающем 65% общего числа признаков.

Ключевым критерием эффективности с использованием самообучения остается вероятность правильного распознавания, а ошибка обучения (векторного квантования) непосредственно отражает лишь качество и свойства самого процесса обучения.

Научная новизна исследования состоит в эффективном применении самообучения системы структурного распознавания изображений путем построения кластерного сжатого представления в пространстве признаков. Это позволяет перейти к векторному описанию пространства эталонов, и как результат, существенно увеличивается быстродействие распознавания.

Практическая ценность работы – получение экспериментальных оценок результативности кластеризации-классификации множества признаков для прикладных примеров баз изображений.

Перспективой обучения на множестве структурных описаний из характерных признаков может быть обучение с учителем, т.к. в рассматриваемой постановке класс характерного признака в составе эталона считается известным. Дальнейшее снижение ошибки квантования может быть получено путем построения нечеткой или гибридной сети [3].

СПИСОК ЛІТЕРАТУРИ

1. Гороховатський В. А. Структурний аналіз і інтелектуальна обробка даних в комп'ютерному зорі /

В. А. Гороховатський. – Харків : Компанія СМІТ, 2014. – 316 с.

2. Контурная обработка динамических изображений / [Л. И. Тимченко, А. А. Поплавский, Н. И. Кокряцкая и др.]. – Киев : Наукова думка, 2013. – 239 с.
3. Осовский С. Нейронные сети для обработки информации / С. Осовский. – М. : Финансы и статистика, 2002. – 344 с.
4. Bay H. Surf: Speeded up robust features / H. Bay, T. Tuytelaars, L. Van Gool // Computer Vision : Ninth European Conference on Computer Vision, Graz, 7–13 May, 2006: proceedings. – Berlin : Springer, 2006. – P.404–417.
5. Паклин Н. Б. Бизнес-аналитика: от данных к знаниям / Н. Б. Паклин, В. И. Орешков. – СПб. : Питер, 2013. – 704 с.
6. Duda R. O. Pattern classification. Second edition / R. O. Duda, P. E. Hart, D. G. Stork. – New York : Wiley, 2000. – 738 p.
7. Прикладная статистика: Классификация и снижение размерности / [С. А. Айвазян, В. М. Бухштабер, И. С. Енюков, Л. Д. Мешалкин; под ред. С. А. Айвазяна.]. – М. : Финансы и статистика, 1989. – 607 с.
8. Берестовский А. Е. Нейросетевые технологии самообучения в системах структурного распознавания визуальных объектов / А. Е. Берестовский, А. Н. Власенко, В. А. Гороховатский // Реєстрація, зберігання і обробка даних. – 2015. – № 1. – С. 108–120.
9. Кохонен Т. Самоорганизующиеся карты / Т. Кохонен. – М. : БИНОМ, Лаборатория знаний, 2013. – 655 с.

Статья поступила в редакцию 02.12.2015.

После доработки 14.12.2015.

Гороховатський В. О.¹, Берестовський А. Е.², Передрій О. О.³

¹Д-р техн. наук, професор, професор кафедри інформаційних технологій, Харківський навчально-науковий інститут ДВНЗ «Університет банківської справи», Харків, Україна

²Аспірант кафедри інформатики, Харківський національний університет радіоелектроніки, Харків, Україна

³Канд. техн. наук, викладач кафедри інформатики і комп'ютерної техніки, Харківський національний економічний університет ім. Семе́на Кузне́ця, Харків, Україна

СИСТЕМАТИЗАЦІЯ ПРОСТОРУ СТРУКТУРНИХ ОЗНАК НА ОСНОВІ МЕТОДІВ САМОНАВЧАННЯ З МЕТОЮ РЕЗУЛЬТАТИВНОГО РОЗПІЗНАВАННЯ ЗОБРАЖЕНЬ

Робота присвячена дослідженню питань кластеризації для множин характерних ознак зображень. Для побудови масиву характерних ознак використаний метод Speeded Up Robust Features. Реалізовані алгоритми кластеризації структурних описів зображень на основі самоорганізуючої нейронної мережі Кохонена та методу різницевого групування. Об'єктом дослідження є методи кластеризації стосовно до множин структурних ознак. Метою роботи є побудова векторних уявлень описів на основі кластеризації, що підвищує швидкість розпізнавання. Предметом дослідження є систематизація множин структурних ознак візуальних об'єктів.

Обговорюються результати застосування методів кластеризації для структурних описів зображень у вигляді множин характерних ознак з метою підвищення швидкості розпізнавання візуальних об'єктів. Для систематизації та стиснення простору ознак запропоновано здійснити самонавчання із застосуванням методів різницевого групування і мережі Кохонена.

Проведено моделювання та експериментальні дослідження методів кластеризації на прикладах конкретних множин характерних ознак. Результати досліджень доводять можливість ефективного представлення описів у вигляді вектора з цілочисельними елементами. Даний підхід може використовуватися для вирішення задач розпізнавання і пошуку зображень.

У результаті побудовано компактний векторний опис еталонів, отримані кількісні оцінки помилки кластеризації, підтверджені працездатність методів для прикладної бази зображень.

Ключові слова: комп'ютерний зір, розпізнавання зображень, характерні ознаки, структурний опис зображення, метод SURF, кластеризація, нейронна мережа, метод різницевого групування, мережа Кохонена, помилка квантування.

Gorokhovatsky V. A.¹, Berestovskiy A. E.², Peredrii E. O.³

¹Dr.Sc., Professor, Professor of the information technologies department, Kharkiv Educational and Scientific Institute SHEI "The University of banking", Kharkiv, Ukraine

²Post-graduate student of the informatics department, Kharkiv National University of Radioelectronics, Kharkiv, Ukraine

³Ph.D., teacher of the department of computer science and computer engineering, Simon Kuznets Kharkiv National University of Economics, Kharkiv, Ukraine

SYSTEMATIZATION OF SPACE OF STRUCTURAL FEATURES BASED ON SELF-LEARNING METHODS FOR EFFECTIVE IMAGE RECOGNITION

The work deals with issues of clustering sets of characteristic features of images. For the construction of array of the characteristic features is used method Speeded Up Robust Features. Implemented algorithms for clustering structural descriptions of images on the

basis of a self-organizing Kohonen neural network and method of grouping the difference. The object of the research are clustering methods which applied to the set of structural features. The aim is to construct a vector representations of descriptions based on clustering, which increases the speed of recognition. The subject of research is systematization a set of structural features of visual objects.

Discussing the results of the application of clustering methods for structural descriptions of images in the form of sets of characteristic features to improve the performance of visual recognition of objects. For systematization and compression the feature space proposed to carry out self-study using the methods of differential grouping and Kohonen networks.

The simulation and experimental study of clustering methods on examples of specific sets of characteristic features were done. The research results proves the possibility of effective representation of the descriptions in the form of a vector with integer elements. This approach can be used to solve problems of recognition and retrieval of images.

As a result compact vector description of etalon images is built, quantitative estimates of clustering error are estimated, efficiency of proposed method during processing of real image database is confirmed.

Keywords: computer vision, image recognition, characteristic signs, structural description of image, method SURF, clusterization, neural network, differential grouping method, Kohonen's neural network, quantization error.

REFERENCES

1. Gorokhovatsky V. Strukturny'j analiz i intellektual'naya obrabotka danny'x v komp'yuternom zrenii: monografiya. Kharkiv, Kompaniya SMIT, 2014, 316 p.
2. Timchenko L. I., Poplavskij A.A., Kokryackaya N. I. i dr. Konturnaya obrabotka dinamicheskix izobrazhenij : monogr. Kyiv, Naukova dumka, 2013, 239 p.
3. Osovskij S. [per. s pol'skogo] Nejronny'e seti dlya obrabotki informacii. Moscow, Finansy' i statistika, 2002, 344 p.
4. Bay H., Tuytelaars T., Van L. Gool Surf: Speeded up robust features, *Computer Vision : Ninth European Conference on Computer Vision, Graz, 7–13 May, 2006: proceedings*. Berlin, Springer, 2006, pp.404–417. DOI: 10.1007/11744023.
5. Paklin N. B., Oreshkov V. I. Biznes-analitika: ot danny'x k znaniyam: uch posob. Sankt-Peterburg, Piter, 2013, 704 p.
6. Duda R. O., Hart P. E., Stork D. G. Pattern classification. Second edition. N'ju-Jork, Wiley-Interscience, 2000, 738 p.
7. Ajvazyan S. A., Buxstaber V. M., Enyukov I. S., Meshalkin L. D.; pod red. S. A. Ajvazyana. Prikladnaya statistika: Klassifikaciya i snizhenie razmernosti: sprav. izd. Moscow, Finansy' i statistika, 1989, 607 p.
8. Berestovskiy A., Vlasenko A. N., Goroxovatskij V. A. Nejrosetevy'e texnologii samoobucheniya v sistemax strukturnogo raspoznavaniya vizual'ny'x ob'ektov, *Reestraciya, zberigannya i obrobka danix*, 2015, Vol. 17, No. 1, pp. 108–120.
9. Koxonen T. Samoorganizuyushhiesya karty' per. 3-go angl. izd. Moscow, BINOM. Laboratoriya znaniy, 2013, 655 p.

МОДИФІКОВАНИЙ ВІКОННИЙ МЕТОД ОДНОКРАТНОГО МНОЖЕННЯ ТОЧКИ ЕЛІПТИЧНОЇ КРИВОЇ НА СКАЛЯР У ПОЛІ $GF(p)$

При реалізації багатьох криптографічних додатків виникає потреба у швидких алгоритмах множення точки еліптичної кривої на число. У даній статті запропоновано модифікований віконний метод однократного множення точки еліптичної кривої на скаляр у полі $GF(p)$. Об'єктом дослідження є процеси виконання операцій у еліптичних криптосистемах. Предметом дослідження є методи та алгоритми виконання операцій однократного множення точки еліптичної кривої на число у полі $GF(p)$. Метою даного дослідження є розроблення та оптимізація методів і алгоритмів виконання операцій множення точки еліптичної кривої на скаляр у полі $GF(p)$ для поліпшення часових характеристик. Існуючі та запропоновані алгоритми реалізовані на мові програмування C# у середовищі розробки Visual Studio 2013. У даній статті проведено дослідження існуючих алгоритмів скалярного множення точки еліптичної кривої та розроблено три модифікації LR-алгоритму віконного методу і узагальнену модифікацію. Експериментальні дослідження реалізованих алгоритмів проводились згідно запропонованої нами методики, яка дозволяє нівелювати вплив на результати дослідження множника та точки еліптичної кривої. Проведене експериментальне дослідження віконних методів та їх модифікацій показало збільшення швидкодії роботи модифікованих алгоритмів у порівнянні з існуючими в середньому на 13%.

Ключові слова: ЕОМ, еліптична криптографія, скалярне множення, таблиця передобчислень, еліптична крива, скінченне поле.

НОМЕНКЛАТУРА

ЕОМ – електронно-обчислювальна машина;

ЕК – еліптична крива;

$GF(p)$ – Galois field, поле Галуа або скінченне поле, де

p – просте число, що є кількістю елементів поля;

LR – left-to-right;

RL – right-to-left;

NAF – a non-adjacent form;

$wNAF$ – a window non-adjacent form;

p – модуль;

a, b – параметри еліптичної кривої;

x, y – змінні еліптичної кривої;

k – множник, скаляр;

P, Q – точка еліптичної кривої;

w – довжина вікна;

n – бітова довжина скаляра k ;

$Table$ – таблиця передобчислень;

for – початок циклу з лічильником;

$end\ for$ – кінець циклу з лічильником;

$while$ – початок циклу з передумовою;

$end\ while$ – кінець циклу з передумовою;

do – виконати;

to – до;

$downto$ – вниз до;

if – умовний оператор;

$then$ – гілка умовного оператора за якою виконується перехід, якщо умова справедлива;

$else$ – гілка умовного оператора за якою відбувається перехід, якщо умова не виконується;

$return$ – видати результат на вихід алгоритму;

$(k_{l-1}, \dots, k_1, k_0)_r$ – l розрядів подання числа k у системі числення з основою r .

ВСТУП

Задача забезпечення конфіденційності інформації та захист її від злоумисників стає дедалі складнішою і в той

же час найбільш актуальною. Більшість систем захисту інформації будується на основі асиметричних криптосистем. Одним з розділів криптографії, який вивчає асиметричні криптосистеми, що засновані на еліптичних кривих над скінченними полями, є еліптична криптографія. Вона бере свій початок ще з 80-х років XX ст. (запропонована Віктором Міллером і Нілом Кобліцем) [1, 2]. Перевага використання еліптичних кривих в криптографічних цілях базується на складності розв'язання задач дискретного логарифмування у групі точок еліптичної кривою. Еліптична криптографія забезпечує набагато вищий рівень криптостійкості використовуючи ключі, що мають меншу довжину у порівнянні з іншими популярними криптосистемами, що засновані на факторизації цілих чисел та проблемі дискретного логарифмування у мультиплікативній групі кільця лишків за певним модулем. У даному дослідженні буде розглянута еліптична крива над полем $GF(p)$. Елементами цього поля є цілі додатні числа від 0 до $p-1$, де p – модуль. У випадках коли p не дорівнює 2 або 3 буде-яке рівняння еліптичної кривої можна звести до форми Вєрштраса [3]. Тому дана робота ґрунтуватиметься на несингулярній еліптичній кривій у формі Вєрштраса: $y^2 = x^3 + ax + b$, де $p \neq 2, 3$.

Найбільш обчислювально витратною операцією в еліптичних криптосистемах є операція множення точки еліптичної кривої на скаляр. У зв'язку з цим актуальною є задача прискорення роботи існуючих алгоритмів скалярного множення точки еліптичної кривої.

1 ПОСТАНОВКА ЗАДАЧІ

Прискорення операції скалярного множення на еліптичній кривій викликає інтерес багатьох дослідників у галузі криптографії [3–9]. З цієї метою була запропонована низка методів виконання даної операції [4–9], що в основному полягають в представленні скаляра у деякій з відповідних форм та виконанні різних операцій на еліп-

тичний кривій (додавання, подвоєння, зменшення у два або три рази).

Множення цілого числа k на точку P еліптичної кривої можна представити наступним чином:

$$[k]P = \underbrace{P + P + \dots + P}_k.$$

Зазвичай число k представляють у двійковому вигляді та за допомогою методів додавання і подвоєння знаходять $[k]P$. Для чисел великих порядків здійснення множення таким чином буде виконуватися занадто довго, тому актуальним є дослідження та модифікація віконних методів, які аналізують одночасно кілька розрядів подання множника у певній системі числення.

Таким чином, метою даного дослідження є модифікація віконного алгоритму виконання операції множення точки еліптичної кривої на скаляр у полі $GF(p)$ для поліпшення часових характеристик алгоритму.

2 ОГЛЯД ЛІТЕРАТУРИ

Вперше ідея подання скаляра у певній системі числення була використана Дональдом Кнудом [3], для мультиплікативної групи, тобто для піднесення до степеня відносно операції множення. Пізніше Хенкерсон та Менезес [4] використали цю ідею для побудови алгоритмів піднесення до степеня в адитивній групі. Скалярне множення точки еліптичної кривої є піднесенням до степеня в адитивній групі точок еліптичної кривої.

Загалом всі алгоритми множення точки еліптичної кривої на число ґрунтуються на поданні множника у певній системі числення та розгляді розрядів цього подання зліва направо, тобто від старшого до молодшого – LR або справа наліво, від молодшого до старшого – RL .

В літературі [4–8] розглядаються різні модифікації віконних алгоритмів скалярного множення, які завдяки побудові таблиць передобчислень (precomputation table) на початкових стадіях алгоритмів дають хороші показники швидкодії.

Віконні методи отримали таку назву через те, що в них розглядається не по одному двійковому розряду, а по w розрядів, де w – довжина вікна. В роботі [6] Метью Рівайн розглядає два види віконних алгоритмів: LR - та RL -алгоритми.

Нехай k – деякий скаляр на який виконується множення точки еліптичної кривої, тоді його подання у системі числення за основою 2^w матиме вигляд

$$k = \sum_{i=0}^{l-1} k_i 2^{iw},$$

де $k_i \in \{0, 1, \dots, 2^w - 1\}$, $k_{l-1} \neq 0$, $l = \left\lceil \frac{n}{w} \right\rceil$ та n – бітова довжина скаляра k .

Віконний LR -алгоритм полягає у виконанні обчислень за наступними формулами:

$$T_{l-1} = [k_{l-1}]P$$

$$T_i = [2^w]T_{i+1} + [k_i]P, i = l-2..0.$$

Після проведення обчислень за цими формулами значення $[k]P$ буде знаходитися у T_0 .

Віконний RL -алгоритм передбачає виконання обчислень за такою формулою:

$$[k]P = \sum_{i=0}^{l-1} [k_i \cdot 2^{iw}]P.$$

Для обох алгоритмів на початковій стадії будується таблиця передобчислень, яка складається з $2^w - 1$ елементів.

Алгоритм 1 – Бінарний віконний LR -алгоритм

Вхід: $P \in E(GF(p))$, $k = (k_{l-1}, \dots, k_1, k_0)_{2^w} \in \mathbb{N}$

Вихід: $Q = [k]P$

1. for $i = 1$ to $2^w - 1$ do
 - 1.1. $Table[i] \leftarrow i \cdot P$
2. end for
3. $Q \leftarrow 0$
4. for $i = l-1$ downto 0 do
 - 4.1. $Q \leftarrow 2^w \cdot Q$
 - 4.2. if $k_i > 0$ then $Q \leftarrow Q + Table[k_i]$
5. end for
6. return Q

Алгоритм 2 – Бінарний віконний RL -алгоритм

Вхід: $P \in E(GF(p))$, $k = (k_{l-1}, \dots, k_1, k_0)_{2^w} \in \mathbb{N}$

Вихід: $Q = [k]P$

1. for $i = 1$ to $2^w - 1$ do
 - 1.1. $Table[i] \leftarrow i \cdot P$
2. end for
3. $Q \leftarrow 0$
4. for $i = 0$ to $l-1$ do
 - 4.1. if $k_i > 0$ then $Q \leftarrow Q + Table[k_i]$
 - 4.2. $Table \leftarrow 2^w \cdot Table$
5. end for
6. return Q

Наприклад, таблиця передобчислень для вікна довжиною 3, тобто $w = 3$ матиме вигляд (P – точка еліптичної кривої):

№ елементу	Значення точки
1	$001 \cdot P$
2	$010 \cdot P$
3	$011 \cdot P$
4	$100 \cdot P$
5	$101 \cdot P$
6	$110 \cdot P$
7	$111 \cdot P$

При аналітичному аналізі алгоритмів 1 та 2 стає зрозумілим, що RL -алгоритм буде повільнішим за LR -алгоритм, оскільки у RL -алгоритмі на кожній ітерації циклу виконується переобчислення значень, що записані у таблицю, тому далі ми будемо розглядати лише LR -алгоритми реалізації віконних методів.

Наступним методом скалярного множення точки ЕК, що розглядається в роботі [8] є метод з пересувним вікном. Даний метод дістав таку назву через те, що згідно цього методу необхідно виділяти вікно довжиною w біт тільки якщо старший біт дорівнює 1, в іншому випадку виконують дії передбачені відповідним бінарним алгоритмом.

При розробці віконного методу з пересувним вікном ставилось за мету досягти компромісу між кількістю додавань і подвоєнь точки. На початковій стадії цього методу будеться таблиця передобчислень, що містить елементи $[t]P$ для $t = \{2^{w-1}, 2^{w-1} + 1, \dots, 2^w - 1\}$. LR -представлення для скаляра k буде мати наступний вигляд: $k = k_0 + 2k_1 + 2^2k_2 + \dots + 2^mk_m$.

Алгоритм 3 – Бінарний LR -алгоритм з пересувним вікном

Вхід: $P \in E(GF(p))$, $k \in \mathbb{N}$

Вихід: $Q = [k]P$

1. $Q \leftarrow 0, i \leftarrow \log_2 k$
2. *while* $i \geq 0$ *do*
 - 2.1. *if* $(k_i = 0)$ *then* $Q = 2 \cdot Q$
 - 2.2. *else*
 - 2.2.1. *if* $i \geq w-1$ *then*
 - a) $t \leftarrow (k_i, \dots, k_{i-w+1})$
 - b) $Q \leftarrow 2^w \cdot Q$
 - c) $Q \leftarrow Q + t \cdot P$
 - 2.2.2. *else*
 - a) Викликати бінарний LR -алгоритм
 - 2.2.3 $i \leftarrow i - w$
3. *end while*
4. *return* Q

Одним з напрямків прискорення методів скалярного множення точки ЕК є переведення множника у NAF представлення [4]. Подання множника k у формі NAF виражається формулою $k = \sum_{i=0}^{l-1} k_i 2^i$, де $k_i \in \{0; \pm 1\}$ та $k_{l-1} \neq 0$.

Метод з поданням множника у вигляді NAF , є ефективнішим (при наявності NAF -розкладення множника) ніж звичайні віконні методи через те, що два сусідні розряди не можуть бути одночасно не нульовим, а це скорочує кількість операцій додавання і віднімання точки. Віконна реалізація даного методу дістала назву метод з поданням множника у вигляді $wNAF$. Подання множника у $wNAF$ формі виражається формулою $k = \sum_{i=0}^{n-1} t_i 2^i$, де кожне ненульове t_i є непарним та таким, що $|t_i| < 2^{w-1}$, $t_{n-1} \neq 0$ і хоча б один з w послідовних біт є не нульовим.

На початковій стадії реалізації LR -алгоритму цього методу множення потрібно крім побудови таблиці передобчислень перевести скаляр у $wNAF$ представлення використовуючи алгоритм 4.

Алгоритм 4 – Переведення додатного цілого числа k у $wNAF$ подання

Вхід: позитивне ціле k , ширина вікна w

Вихід: $wNAF(k)$

1. $i \leftarrow 0$
2. *while* $k \geq 1$ *do*
 - 2.1. *if* k є непарним *then*
 - 2.1.1. $t_i \leftarrow k \bmod 2^w$
 - 2.1.2. $k \leftarrow k - t_i$
 - 2.2. *else* $t_i \leftarrow 0$
 - 2.3. $k \leftarrow \frac{k}{2}, i \leftarrow i + 1$
3. *end while*
4. *return* $\{t_{n-1}, t_{n-2}, \dots, t_1, t_0\}$

Алгоритм 5 – LR -алгоритм з поданням множника у вигляді $wNAF$

Вхід: позитивне ціле k , $P \in E(GF(p))$, ширина вікна w

Вихід: kP

1. Використати алгоритм 4 для обчислення $wNAF(k) = \sum_{i=0}^{n-1} t_i 2^i$
2. *for* $i = 1$ *to* $2^{w-1} - 1$ *з кроком* 2 *do*
 - 2.1. $Table[i] = i \cdot P$
3. *end for*
4. $Q \leftarrow 0$
5. *for* $i = n-1$ *downto* 0 *do*
 - 5.1. $Q \leftarrow 2 \cdot Q$
 - 5.2. *if* $t_i \neq 0$ *then*
 - 5.2.1. *if* $t_i > 0$ *then* $Q \leftarrow Q + Table[t_i]$
 - 5.2.2. *else* $Q \leftarrow Q - Table[t_i]$
6. *end for*
7. *return* Q

Таким чином метод з поданням множника у вигляді $wNAF$ є аналогом віконного методу, але оперуючи знаком NAF подання дозволяє скоротити об'єм необхідної для зберігання таблиці передобчислень пам'яті вдвічі порівняно з класичним віконним методом. Враховуючи особливості $wNAF$ подання цілих додатних чисел таблиця передобчислень буде складатися з елементів $[i]P$, де $i = 1, 3, \dots, 2^{w-1} - 1$, тобто порівняно з бінарним віконним методом отримуємо скорочення об'єму таблиці передобчислень у чотири рази.

Логічним вдосконаленням методу з поданням множника у вигляді $wNAF$ є метод з пересувним вікном та поданням множника у вигляді NAF . На відміну від попереднього алгоритму з пересувним вікном, де ми переводимо число k у $wNAF$ представлення, у даному алгоритмі потрібно перевести його у NAF форму, використовуючи алгоритм 6, а потім виконувати скалярне множення точки еліптичної кривої за алгоритмом 7.

Алгоритм 6 – Переведення додатного цілого числа k у NAF подання

Вхід: позитивне ціле k

Вихід: $NAF(k)$

```

1.  $i \leftarrow 0$ 
2. while  $k \geq 1$  do
2.1. if  $k$  is odd then
2.1.2.  $k_i \leftarrow 2 - (k \bmod 4)$ 
2.1.3.  $k \leftarrow k - k_i$ 
2.2. else  $k_i \leftarrow 0$ 
2.3.  $k \leftarrow \frac{k}{2}$ 
2.4.  $i \leftarrow i + 1$ 
3. end while
4. return  $\{k_{l-1}, k_{l-2}, \dots, k_1, k_0\}$ 

```

Алгоритм 7 – LR-алгоритм з пересувним вікном та поданням множника у вигляді NAF

Вхід: $P \in E(GF(p)), k \in \mathbb{N}$

Вихід: $[k]P$

1. Використати алгоритм 6 для обчислення

$$NAF(k) = \sum_{i=0}^{l-1} k_i 2^i$$

2. for $i = 1$ to $(2^w - (-1)^w) / 3 - 1$ з кроком 2 do

2.1. $Table[i] = i \cdot P$

3. end for

4. $Q \leftarrow 0, i \leftarrow l - 1$

5. while $i \geq 0$ do

5.1. if $k_i = 0$ then $t \leftarrow 1, u \leftarrow 0$

5.1.1. else знайти $\max(t) \leq w$ таке що $u \leftarrow (k_i, \dots, k_{i-t+1})$

є непарним

5.2. $Q \leftarrow 2^t \cdot Q$

5.3. if $u > 0$ then $Q \leftarrow Q + Table[u]$

5.4. else if $u < 0$ then $Q \leftarrow Q - Table[u]$

5.5. $i \leftarrow i - t$

6. end while

7. return Q

Таблиця передобчислень для цього методу буде містити такі елементи $[i]P$, де $i = 1, 3, \dots, \frac{2(2^w - (-1)^w)}{3} - 1$.

Як зазначалося вище, час роботи віконних алгоритмів прискорюється завдяки побудові на першій стадії їх роботи таблиць передобчислень. У таблиці 1 наведено перелік точок, які необхідно обчислити на початковій стадії кожного з розглянутих алгоритмів.

Таблиця 1 – Наперед обчисленні точки для методів скалярного множення

Методи	Значення множника	Кількість наперед обчислених точок
Бінарний віконний метод	$\{1, 2, \dots, 2^w - 1\}$	$2^w - 1$
Бінарний метод з пересувним вікном	$\{2^{w-1}, 2^{w-1} + 1, \dots, 2^w - 1\}$	$2^w - 2^{w-1}$
Метод з поданням множника у вигляді $wNAF$	$\{1, 3, 5, \dots, 2^{w-1} - 1\}$	2^{w-2}
Метод з пересувним вікном та поданням множника у вигляді NAF	$\{1, 3, \dots, \frac{2(2^w - (-1)^w)}{3} - 1\}$	$\frac{1}{3}(2^w - (-1)^w)$

3 МАТЕРІАЛИ ТА МЕТОДИ

Статистично показано, що двійкові подання чисел довжиною понад 100 біт містять довгі послідовності нулів, тому нами запропонована модифікація віконного методу згідно якої буде виділятися вікно зі старшим одинарним бітом та іншими нульовими. Дану модифікацію назовемо модифікація №1. При такому підході таблиця передобчислень буде складатися з таких елементів $[2^i]P$, де $i = 0, 1, \dots, w - 1$.

Алгоритм 8 – Модифікований віконний LR-алгоритм № 1

Вхід: $P \in E(GF(p)), k \in \mathbb{N}$

Вихід: $[k]P$

1. $Table[0] \leftarrow P$

2. for $i = 1$ to $w - 1$

2.1. $Table[i] \leftarrow 2 \cdot Table[i - 1]$

3. end for

4. $Q \leftarrow 0; i \leftarrow \log_2 k$

5. while $i \geq 0$ do

5.1. $Q = 2 \cdot Q$

5.2. if $k_i = 1$ then

5.2.1. if $k_{i-1} = 0$ then

a. знайти $\max(t) \leq w$ таке що

$$k_{i-1} = k_{i-2} = \dots = k_{i-t+1} = 0$$

b. $Q \leftarrow 2^{t-1} \cdot Q$

c. $Q \leftarrow Q + Table[t - 1]$

d. $i \leftarrow i - t$

5.2.2. else

a. $Q \leftarrow Q + P$

b. $i \leftarrow i - 1$

6. end while

7. return Q

Наприклад, таблиця передобчислень для $w = 5$ матиме вигляд (P – точка еліптичної кривої):

№ елементу	Множник	Значення точки
0	2^0	$1 \cdot P$
1	2^1	$10 \cdot P$
2	2^2	$100 \cdot P$
3	2^3	$1000 \cdot P$
4	2^4	$10000 \cdot P$

У випадку якщо множник k складатиметься переважно з нулів, такий спосіб побудови таблиці передобчислень суттєво збільшить швидкість віконного методу. Але в протилежному випадку, коли біти скаляра будуть тільки одиничні такий метод не дасть хороших результатів. Тому актуальним є побудова нової таблиці множника якої будуть складатись тільки з одиничних біт. Такий метод назовемо модифікація №2. Множник k таблиці передобчислень в даному методі буде рівним $2^i - 1$, де $i = 1, 2, \dots, w$.

Алгоритм 9 – Модифікований віконний LR-алгоритм № 2

Вхід: $P \in E(GF(p)), k \in \mathbb{N}$

Вихід: $[k]P$


```

1. for  $i = 1$  to  $w$ 
1.1.  $Table[i] \leftarrow (2^i - 1) \cdot P$ 
2. end for
3.  $Q \leftarrow 0; i \leftarrow \log_2 k$ 
4. while  $i \geq 0$  do
4.1.  $Q = 2 \cdot Q$ 
4.2. if  $k_i = 1$  then
4.2.1. знайти  $\max(t) \leq w$  таке що
 $k_{i-1} = k_{i-2} = \dots = k_{i-t+1} = 1$ 
4.2.2.  $Q \leftarrow 2^{t-1} \cdot Q$ 
4.2.3.  $Q \leftarrow Q + Table[t-1]$ 
4.2.4.  $i \leftarrow i - t + 1$ 
4.3.  $i \leftarrow i - 1$ 
5. end while
6. return  $Q$ 

```

Для вікна $w = 5$ таблиця передобчислень матиме вигляд:

i	Множник	Значення точки
1	$2^1 - 1$	$1 \cdot P$
2	$2^2 - 1$	$11 \cdot P$
3	$2^3 - 1$	$111 \cdot P$
4	$2^4 - 1$	$1111 \cdot P$
5	$2^5 - 1$	$11111 \cdot P$

Також можна сподіватися на приріст швидкодії, коли з розрядів числа k буде виділятися частина біт, що починається і закінчується одиницею, а між ними міститиметься певна кількість нулів (модифікація №3). Множник k у таблиці передобчислень в модифікації №3 буде рівним $2^{i-1} + 1$, де $i = 2, 3, \dots, w$.

Алгоритм 10 – Модифікований віконний LR-алгоритм №3

Вхід: $P \in E(GF(p)), k \in \mathbb{N}$

Вихід: $[k]P$

```

1.  $Table[1] \leftarrow P$ 
2. for  $i = 2$  to  $w$ 
2.1.  $Table[i] \leftarrow (2^{i-1} + 1) \cdot P$ 
3. end for
4.  $Q \leftarrow 0; i \leftarrow \log_2 k$ 
5. while  $i \geq 0$  do
5.1.  $Q = 2 \cdot Q$ 
5.2. if  $k_i = 1$  then
5.2.1. if  $k_{i-1} = 0$  then
a. знайти  $\max(t) \leq w$  таке що
 $k_i = k_{i-t+1} = 1$  та  $k_{i-1} = k_{i-2} = \dots = k_{i-t+2} = 0$ 
b.  $Q \leftarrow 2^{t-1} \cdot Q$ 
c.  $Q \leftarrow Q + Table[t]$ 
d.  $i \leftarrow i - t + 1$ 
5.2.2. else
a.  $Q \leftarrow 2 \cdot Q$ 
b.  $Q \leftarrow Q + Table[2]$ 

```

```

c.  $i \leftarrow i - 1$ 
5.2.3.  $i \leftarrow i - 1$ 
6. end while
7. return  $Q$ 

```

Оскільки за допомогою виразу $2^{i-1} + 1$ неможливо отримати 1, то до початку циклу побудови таблиці передобчислень у елемент таблиці з індексом 1, необхідно записати значення точки P . Приклад такої таблиці передобчислень для вікна $w = 5$:

№ елементу	Множник	Значення точки
1	1	$1 \cdot P$
2	$2^1 + 1$	$11 \cdot P$
3	$2^2 + 1$	$101 \cdot P$
4	$2^3 + 1$	$1001 \cdot P$
5	$2^4 + 1$	$10001 \cdot P$

У таблиці 2 наведено значення множників для таблиць передобчислень запропонованих модифікацій. Зрозуміло, що кожна із запропонованих модифікацій в певному частковому випадку буде збільшувати швидкодію віконного алгоритму скалярного множення.

Модифікації №1–3 дають приріст швидкодії тільки в окремих випадках, тому актуальним є побудувати узагальнений модифікований віконний метод скалярного множення точки ЕК та об'єднати переваги кожної із запропонованих модифікацій таким чином, щоб прискорити віконний алгоритм для будь-якого вигляду двійкового представлення множника k . Доцільним буде об'єднувати таблиці передобчислень з модифікації №2 та №3, оскільки модифікація №1 покривається модифікацією №3. Особливістю побудови таблиць передобчислень для другої та третьої модифікації є те, що номер елемента у таблиці відповідає бітовій довжині множника для якого обчислене відповідне значення.

Алгоритм 11 – Узагальнений модифікований віконний LR-алгоритм скалярного множення

Вхід: $P \in E(GF(p)), k \in \mathbb{N}$

Вихід: $[k]P$

```

1.  $Table1[1] \leftarrow P$ 
2.  $Table2[1] \leftarrow P$ 
3. for  $i = 2$  to  $w$  do
3.1.  $Table1[i] \leftarrow (2^i - 1) \cdot P$ 
3.2.  $Table2[i] \leftarrow (2^{i-1} + 1) \cdot P$ 
4. end for
5.  $Q \leftarrow 0, i \leftarrow \log_2 k$ 

```

Таблиця 2 – Наперед обчислені точки для модифікованих віконних методів

Метод	Значення множника	Кількість наперед обчислених точок
Модифікація №1	2^i , де $i \in [0; w-1]$	w
Модифікація №2	$2^i - 1$, де $i \in [1; w]$	w
Модифікація №3	1 та $2^{i-1} + 1$, де $i \in [2; w]$	w

```

6. while  $i \geq 0$  do
6.1. if  $k_i = 1$  then
6.1.1. if  $k_{i-1} = 0$  then
a. знайти  $\max(t) \leq w$  такий, що
 $k_i = k_{i-t+1} = 1$  and  $k_{i-1} = k_{i-2} = \dots = k_{i-t+2} = 0$ 
b.  $Q \leftarrow 2^t \cdot Q$ 
c.  $Q \leftarrow Q + \text{Table1}[t]$ 
6.1.2. else
a. знайти  $\max(t) \leq w$  такий,
що  $k_{i-1} = k_{i-2} = \dots = k_{i-t+1} = 1$ 
b.  $Q \leftarrow 2^t \cdot Q$ 
c.  $Q \leftarrow Q + \text{Table2}[t]$ 
6.1.3.  $i \leftarrow i - t$ 
6.2. else
6.2.1.  $Q = 2 \cdot Q$ 
6.2.2.  $i = i - 1$ 
7. end while
8. return  $Q$ 

```

На кроці 11 при реалізації алгоритму 11 потрібно передбачити перевірку чи є праворуч від i -го біта одиничні біти. Якщо всі біти молодші i -го є нульовими, то необхідно виконати такі дії: $Q = 2 \cdot Q$, $Q = Q + P$, $i = i - 1$ та перейти на наступну ітерацію циклу *while*.

4 ЕКСПЕРИМЕНТИ

З метою проведення експериментальних досліджень було розроблено програмний продукт на мові програмування C# у середовищі розробки «Visual Studio 2013». Експериментальне дослідження проводилося на EOM з операційною системою Windows 8.1, об'ємом оперативної пам'яті 2Gb, процесором Pentium Dual-Core 2,30 Hz.

Даний програмний продукт дозволяє проводити тестування коректності роботи алгоритмів та проводити дослідження розглянутих віконних методів скалярного множення точки ЕК. Завдяки тестуванню віконних алгоритмів для різних значень довжини вікна w було отримано оптимальне значення w .

У розробленому програмному продукті, окрім відомих алгоритмів реалізовано запропоновані нами модифікації віконних LR -алгоритмів, оскільки, як показало дослідження LR -алгоритми показують кращі результати ніж RL -алгоритми. Для збільшення швидкодії алгоритмів на початковій стадії роботи кожного з них будується таблиця передобчислень, значення таблиць передобчислень наведені у таблиці 1 та таблиці 2. При аналізі літературних джерел [5, 7] було встановлено, що оптимальними параметрами еліптичної кривої, які забезпечують високу крипостійкість є: $a = 79$, $b = -3$. Тому ці значення параметрів було використано у дослідженні.

Замір швидкодії роботи алгоритмів проводився за наступною методикою:

- 1) обирається довжина модуля p (64, 128, 256 та 512 біт);
- 2) для обраного модуля p формується множина з 25 випадкових точок еліптичної кривої;
- 3) випадковим чином генерується 25 множників довжиною від 32 до 64 біт;

4) виконується множення кожної точки з п. 2 на кожен множник з п. 3, отримані часові показники усереднюються.

Пошук оптимальних значень довжини вікна (рис. 1) виконувався для модуля p що має довжину 128 біт та довжин вікна $w \in [2; 9]$, оскільки після збільшення довжини вікна до 10 біт час виконання алгоритмів зростає. Значення множників змінювалось від 10 до 1000 з кроком 100, оскільки для проведення дослідження при більших значеннях множника було не достатньо обчислювальної потужності комп'ютера.

5 РЕЗУЛЬТАТИ

За наведеною методикою для існуючих віконних LR -та RL -алгоритмів було побудовано таблицю 3 та для запропонованих модифікацій і узагальненого модифікованого віконного методу таблицю 4.

Як можна поміти з таблиці 3, LR -алгоритми дають значно кращі результати ніж RL -алгоритми, що підтверджує результати проведеного аналітичного дослідження, тому доцільним є пошук оптимального значення довжини вікна w для існуючих LR -алгоритмів.

На рис. 1 наведено залежність часу роботи LR -алгоритмів від довжини вікна, де 1, 3, 5, 7 – номери відповідних алгоритмів у таблиці 3.

Як видно з рисунку 1 найкращі часові характеристики показують алгоритми 1 і 3, а саме бінарний віконний LR -алгоритм та бінарний LR -алгоритм з пересувним вікном при довжині вікна 9 біт, що підтверджує доцільність їх модифікацій. Швидкодія роботи модифікованих віконних LR -алгоритмів та узагальненого алгоритму наведена в таблиці 4.

Таблиця 3 – Часові характеристики методів скалярного множення, мс

№	Методи	Довжина модуля, біт			
		64	128	256	512
1	Бінарний віконний LR -алгоритм	5,5	9,0	32,1	84,4
2	Бінарний віконний RL -алгоритм	54,7	120,8	403,6	1125,7
3	Бінарний LR -алгоритм з пересувним вікном	6,8	10,3	32,5	84,6
4	Бінарний RL -алгоритм з пересувним вікном	33,8	73,7	257,0	696,2
5	LR -алгоритм з пересувним вікном та поданням множника у вигляді NAF	4,3	8,6	30,5	84,2
6	RL -алгоритм з пересувним вікном та поданням множника у вигляді NAF	19,1	42,1	144,7	390,8
7	LR -алгоритм з поданням множника у вигляді $wNAF$	4,1	8,9	32,1	87,0
8	RL -алгоритм з поданням множника у вигляді $wNAF$	14,7	33,1	110,8	311,1

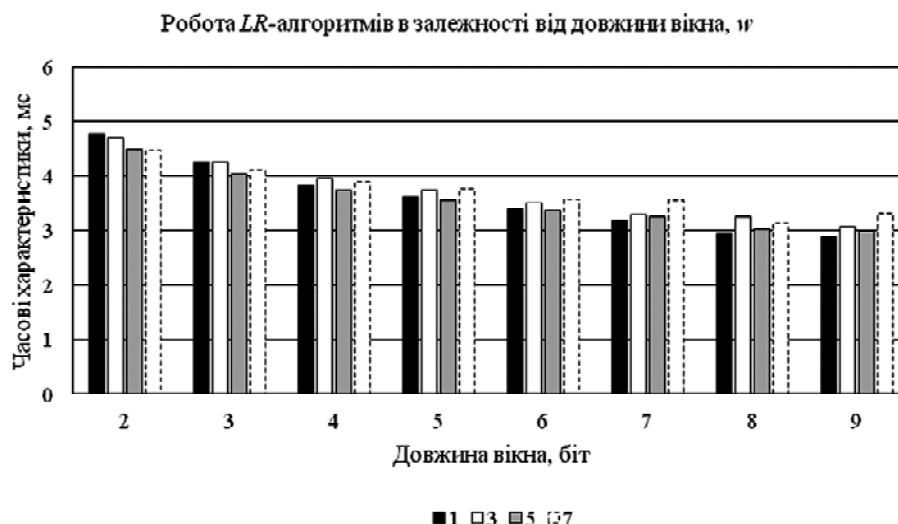


Рисунок 1 – Часові характеристики віконних LR-алгоритмів в залежності від довжини вікна

Таблиця 4 – Часові характеристики модифікованих методів в залежності від довжини модуля

№	Метод	Довжина модуля, біт			
		64	128	256	512
1	Модифікація №1	4,7	11,1	39,1	106,6
2	Модифікація №2	4,3	9,2	34,4	91,1
3	Модифікація №3	4,3	9,0	31,2	86,4
4	Узагальнений модифікований віконний метод	4,1	8,2	31,1	82,0

6 ОБГОВОРЕННЯ

За результатами наведеними в таблиці 3, найкращі часові характеристики для модулів будь-якої довжини показують LR-алгоритм з поданням множника у вигляді $wNAF$ та бінарний LR-алгоритм з пересувним вікном і поданням множника у вигляді NAF . При правильно підбраній довжині вікна (рис. 1) та довжині модуля 128 біт найкращі часові характеристики показують віконний LR-алгоритм та LR-алгоритм з пересувним вікном. З таблиці 4 видно, що побудовані модифікації методу з пересувним вікном дають приріст швидкодії для модулів довжиною 64 біт на 20%. Модифікація №3 порівняно з попередніми модифікаціями покращує часові характеристики майже на 3% для модулів довжиною 128 і 256 біт.

Найнефективнішою є побудова узагальненої модифікації віконного методу та використання таблиць передобчислень зі значеннями $[t]P$ для $t = 2^i + 1$, де $i = 1..w$ та $t = 2^i - 1$, де $i = 2..w$. Побудований LR-алгоритм на її основі дає приріст швидкодії в середньому на 13% для модулів 64, 128, 256 та 512 біт. Запропоновані таблиці передобчислень використовують

порівняно з бінарними віконними методами у $\frac{2^w - 1}{w}$ разів менше оперативної пам'яті.

ВИСНОВКИ

У роботі вирішено актуальну задачу вдосконалення існуючих методів скалярного множення точки еліптичної кривої у полі $GF(p)$.

Наукова новизна роботи полягає у тому, що дістав подальшого розвитку науковий підхід модифікації бінарних методів множення точки еліптичної кривої на число, що заснований на розгляді кількох розрядів двійкового подання множника одночасно на кожній ітерації циклу.

Проведений аналіз віконних методів множення точки еліптичної кривої на скаляр у полі $GF(p)$ показав, що при правильному виборі довжини вікна та побудові на початковій стадії таблиць передобчислень можна суттєво збільшувати швидкість алгоритмів. За допомогою розробленого програмного забезпечення для аналізу і тестування методів множення точки ЕК на скаляр, проведено експериментальне дослідження, яке довело практичну доцільність використання запропонованих методів, замість існуючих.

Розроблені три модифікації віконного LR-алгоритму з пересувним вікном, які відрізняються від існуючого методу способом побудови таблиць передобчислень, забезпечують приріст швидкодії для множників спеціальної структури. Побудована узагальнена модифікація віконного LR-алгоритму з пересувним вікном, що полягає у комбінації, на початковій стадії роботи алгоритму, двох таблиць передобчислень з модифікації №2 та №3, яка забезпечує приріст швидкодії порівняно з існуючими методами в середньому на 13%.

Практична цінність отриманих результатів полягає у тому, що розроблено програмне забезпечення, яке реалізує запропоновані модифікації та існуючі методи множення точки еліптичної кривої на число і дозволяє проводити аналіз алгоритмів, що реалізують зазначені методи. За допомогою даного програмного забезпечення може бути вирішена практична задача вибору найкращого алгоритму скалярного множення точки еліптичної кривої для використання у алгоритмі цифрового підпису на еліптичних кривих, який широко використовується.

Перспективою для подальшого дослідження є побудова модифікованого методу з пересувним вікном та поданням множника у вигляді NAF та модифікованого методу з поданням множника у вигляді $wNAF$.

ПОДЯКИ

Дослідження виконано у межах держбюджетної науково-дослідної теми «Розроблення та дослідження високоєфективних архітектур спеціалізованих комп'ютерних систем для реалізації обчислень у скінченних полях «Національного технічного університету України «Київський політехнічний інститут» (номер державної реєстрації 0115U000319).

СПИСОК ЛІТЕРАТУРИ

1. Miller V. Use of elliptic curves in cryptography / V. Miller // *Lecture Notes in Computer Science. Advances in cryptology – CRYPTO 85*. – Springer, 1986. – P. 417–426. 10.1007/3-540-39799-X_31
2. Koblitz N. Introduction to Elliptic Curves and Modular Forms / Neal Koblitz. – New York : Springer, 1984. – 248 p. 10.1007/978-1-4684-0255-1
3. Knuth, D. The Art of Computer Programming. Volume 2 Seminumerical Algorithms, Third Edition / D. E. Knuth. – Massachusetts: Addison-Wesley, 1997. – 762 p.
4. Hankerson D. Guide to Elliptic Curve Cryptography / D. Hankerson, A. Menezes, S. Vanstone. – New York : Springer, 2004. – 311 p.
5. Crandall R. Prime Numbers. A Computational Perspective. Second Edition / Richard Crandall, Carl Pomerance. – New York : Springer, 2005. – 604 p. 10.1007/978-1-4684-9316-0
6. Rivain M. Fast and Regular Algorithms for Scalar Multiplication over Elliptic Curves / Matthieu Rivain. – IACR Cryptology ePrint Archive, 2011. – 338 p.
7. Болотов, А. А. Алгоритмические основы эллиптической криптографии / А. А. Болотов. – М. : Изд-во, 2004. – 499 с.
8. Elliptic Curve Point Multiplication [Electronic resource] // December 31, 2015: Proceedings. – Mode of access: https://en.wikipedia.org/wiki/Elliptic_curve_point_multiplication. – Last access: 2016. – Title from the screen.
9. Pathak H. Speeding Up Computation of Scalar Multiplication in Elliptic Curve Cryptosystem / H. Pathak, M. Sanghi // *International Journal on Computer Science and Engineering*. – 2010. – No. 04. – P. 1024–1028.

Стаття надійшла до редакції 15.02.2016.

Після доробки 01.03.2016.

Дичка І. А.¹, Онай Н. В.², Дрозда Т. П.³

¹Професор, доктор технічних наук, декан факультета прикладної математики НТУУ «КПІ», Київ, Україна

²Старший преподаватель кафедры программного обеспечения компьютерных систем факультета прикладной математики НТУУ «КПІ», Київ, Україна

³Магістрант кафедри программного обеспечения компьютерных систем факультета прикладной математики НТУУ «КПІ», Київ, Україна

МОДИФИЦІРОВАННИЙ ОКОННИЙ МЕТОД ОДНОКРАТНОГО УМНОЖЕННЯ ТОЧКИ ЕЛЛІПТИЧЕСКОЙ КРИВОЙ НА СКАЛЯР В ПОЛЕ GF(P)

При реализации многих криптографических приложений возникает потребность в быстрых алгоритмах умножения точки эллиптической кривой на число. В данной статье предложен модифицированный оконный метод однократного умножения точки эллиптической кривой на скаляр над полем GF(p). Объектом исследования являются процессы выполнения операций в эллиптических криптосистемах. Предметом исследования являются методы и алгоритмы выполнения операций однократного умножения точки эллиптической кривой на число над полем GF(p). Целью данного исследования является разработка и оптимизация методов и алгоритмов выполнения операции умножения точки эллиптической кривой на скаляр над полем GF(p) для улучшения временных характеристик. Существующие и предложенные алгоритмы реализованы на языке программирования C# в среде разработки Visual Studio 2013. В данной статье проведено исследование существующих алгоритмов скалярного умножения точки эллиптической кривой и разработаны три модификации LR-алгоритма оконного метода и обобщенная модификация. Экспериментальные исследования реализованных алгоритмов проводились согласно предложенной нами методики, которая позволяет нивелировать влияние на результаты исследования множителя и точки эллиптической кривой. Проведенное экспериментальное исследование оконных методов и их модификаций показало увеличение быстродействия работы модифицированных алгоритмов по сравнению с существующими в среднем на 13%.

Ключевые слова: ЭВМ, эллиптическая криптография, скалярное умножение, таблица предвычислений, эллиптическая кривая, конечное поле.

Dychka I. A., Onai M. V., Drozda T. P.

¹Professor, Dc.Sc., Dean of the Faculty of Applied Mathematics NTUU «KPI», Kyiv, Ukraine

²Senior Lecturer, Department of Computer Systems Software of Faculty of Applied Mathematics, NTUU «KPI», Kyiv, Ukraine

³Master student, Department of Computer Systems Software of Faculty of Applied Mathematics, NTUU «KPI», Kyiv, Ukraine

MODIFIED METHOD FOR ELLIPTIC CURVE SCALAR POINT MULTIPLICATION OVER GF(P)

During development of many cryptographic applications, we need to perform fast algorithms of scalar multiplication. In this paper we propose a modified window method of elliptic curve point multiplication over the GF(p). The object of the research are the processes of performing operations in elliptic cryptosystems. The subject of the research are the methods and the algorithms of elliptic curve point multiplication over the GF(p). The goal of the research is to develop and optimize the methods and the algorithms of performing elliptic curve point multiplication operation over the GF(p) for improving the time characteristics. Existing and proposed algorithms were implemented with C# programming language and integrated development environment – Visual Studio 2013. In this article we did an investigation of the existing algorithms of elliptic curve point multiplication and developed three versions of the window method LR-algorithm and generalized modification. Experimental studies of the implemented algorithms were performed according to the proposed methodology, which allows us to explore the impact of the multiplier and elliptical curve point on the results of the research. The experimental research of window methods and their modifications showed an increase speed of the modified algorithms compared to the existing algorithms in average of 13%.

Keywords: computers, elliptic curve cryptography, scalar multiplication, precomputation table, elliptic curve, finite field.

REFERENCES

1. Miller V. Use of elliptic curves in cryptography, *Lecture Notes in Computer Science. Advances in cryptology – CRYPTO 85*. Springer, 1986, pp. 417–426. 10.1007/3-540-39799-X_31
2. Koblitz N. Introduction to Elliptic Curves and Modular Forms. New York, Springer, 1984, 248 p. 10.1007/978-1-4684-0255-1
3. Knuth D. The Art of Computer Programming. Volume 2 Seminumerical Algorithms, Third Edition. Massachusetts, Addison-Wesley, 1997, 762 p.
4. Hankerson D., Menezes A., Vanstone S. Guide to Elliptic Curve Cryptography. New York, Springer, 2004, 311 p.
5. Crandall R., Pomerance C. Prime Numbers. A Computational Perspective. Second Edition. New York : Springer, 2005, 604 p. 10.1007/978-1-4684-9316-0
6. Rivain M. Fast and Regular Algorithms for Scalar Multiplication over Elliptic Curves. IACR Cryptology ePrint Archive, 2011, 338 p.
7. Bolotov A. A. Algoritmicheskie osnovy e'llipticheskoy kriptografii. Moscow, Izd-vo, 2004, 499 p.
8. Elliptic Curve Point Multiplication [Electronic resource]. December 31, 2015: Proceedings. Mode of access: https://en.wikipedia.org/wiki/Elliptic_curve_point_multiplication, Last access: 2016, Title from the screen.
9. Pathak H., Sanghi M. Speeding Up Computation of Scalar Multiplication in Elliptic Curve Cryptosystem, *International Journal on Computer Science and Engineering*, 2010, No. 04, pp. 1024–1028.

ОЦІНЮВАННЯ ЕФЕКТИВНОСТІ ІНФОРМАЦІЙНОГО ПОШУКУ В СИСТЕМАХ КОНСОЛІДОВАНОЇ ІНФОРМАЦІЇ

Для оцінювання ефективності інформаційного пошуку запропоновано поділяти знайдені документи на пертинентні, релевантні та нерелевантні. Ефективність пошуку визначати відношенням кількості пертинентних та релевантних документів до кількості нерелевантних документів, а властивості пошукової системи повинні бути подані відповідним коефіцієнтом. Основною метою цього дослідження є розробка інтегрального критерію оцінювання ефективності інформаційного пошуку за результатами видачі в системах консолідованої інформації. Введено поняття консолідованої інформації.

Розроблено метод оцінювання ефективності інформаційного пошуку. Він демонструє використання поділу знайдених і виданих документів на пертинентні, релевантні та нерелевантні. Введено інтегральний показник релевантності документу пошуковому запиту, який враховує негативну та позитивну оцінку. Оцінку ефективності подано як сумарне значення різних компонентів. Експеримент виконано на підставі проведення інформаційного пошуку в одному або в кількох інформаційних фондах і на різних пошукових системах за одного набору ключових слів.

Розроблений підхід до побудови оцінки інформаційного пошуку має практичне значення, оскільки отримані кількісні значення локальних оцінок дають підстави для оптимізації набору ключових слів, та визначення найбільш відповідних інформаційних фондів і пошукових систем.

Ключові слова: інформаційна система, інформаційний пошук, ефективність, пертинентність, релевантність.

НОМЕНКЛАТУРА

P – клас релевантних документів;

Π – клас пертинентних документів;

H – клас нерелевантних документів;

D – множина інформаційних джерел;

K – користувач, який формує запит для інформаційного пошуку в джерелах;

N – кількість документів у видачі;

f_{Π} – відносна частоту появи пертинентних документів;

f_P – відносна частоту появи релевантних документів;

f_H – відносна частоту появи нерелевантних документів;

$f_{P+\Pi}$ – відносна частоту появи релевантних і пертинентних документів;

γ – відношення загальної кількості релевантних і пертинентних документів до кількості нерелевантних документів.

ВСТУП

Розвиток комп'ютерної техніки та інформаційних технологій значною мірою стимулював створювання і наповнення різноманітною інформацією як загальні, так і спеціалізовані бази даних, забезпечуючи управління ними. Проте, з іншої сторони, величезні обсяги даних практично унеможливають безпосередню роботу користувача з ними, що у свою чергу стимулювало розвиток відповідних пошукових систем, основною метою яких є своєчасне і повне забезпечення користувача необхідними йому даними. Тому найкритичнішою проблемою, з якою зустрічаються користувачі, – це забезпечення надійного, постійного та повнофункціонального доступу до актуальних даних.

1 ПОСТАНОВКА ЗАДАЧІ

Проблема побудови критеріїв оцінювання функціональної ефективності інформаційного пошуку в системах консолідованої інформації полягає в тому, що шукана інформація зберігається в різних джерелах, створених в різний час і з різною метою; вона є складно структурованою, для різних задач має різну інформаційну цінність, і різними користувачами сприймається по-різному. Натомість, за високої надійності і стабільності апаратного та програмного забезпечення вся відповідальність за результати пошуку покладена на людський фактор в сенсі укладання пошукового запиту. В цьому плані, об'єктивна оцінка ефективності пошуку, а в даному випадку ще й консолідації знайдених і виданих документів може бути зроблена саме на підставі виданих документів.

Історично, а в певному сенсі і політично (з метою захисту інформації) різні джерела інформації (електронні бібліотеки, загальні та локальні бази, сховища, простори даних) мають свої особливості стосовно організації форм збереження, пошуку, виявлення, видачі потрібної інформації, які в основному полягають у видах і тонкощах мов запитів та кодування збереженої інформації. Очевидно, що вихід з такої ситуації для користувача є і йому немає потреби вивчати премудрості мов різних запитів, оскільки потрібний пошук здійснюють спеціальні пошукові системи. Робота з однією чи навіть декількома базами даних практично полягає у правильному формуванні запиту і тут існуюча пошукова система допомагає знайти необхідну інформацію. Наприклад, локальні бази даних навіть великих підприємств досить швидко дають інформацію про виготовлені вироби, товари, зарплату працівників тощо. Проте, пошук даних в «чужих» базах

даних може стати складною проблемою [11]. Тут найкращим прикладом є пошукова система Google та аналогічні з нею, які видають десятки тисяч документів, з яких вибирають лише декілька, витрачаючи величезну кількість часу на пошук потрібних серед наданих пошуковою системою.

Не меншою є проблема інформаційного пошуку в системах консолідованої інформації. Термін консолідована інформація означає одержані з декількох інформаційних джерел системно інтегровані різноманітні інформаційні ресурси в сукупності наділені ознаками повноти, цілісності та несуперечності. Вони фактично подаються у формі адекватної інформаційної моделі проблемної області для її аналізу, опрацювання та ефективного використання в процесах підтримки прийняття рішень. Як правило, такі системи є результатом інтеграції різноманітних джерел інформації, які були створені в різний час і за різними принципами та мовами запитів, а головне за різними фаховими ознаками та онтологіями. Досить часто основні техніко-економічні дані зосереджені в системах, які реалізують численні офісні, адміністративні і технологічні процеси, а в результаті такі дані не можуть спільно використовуватись в масштабах всього підприємства.

2 ЛІТЕРАТУРНИЙ ОГЛЯД

Поняття ефективності має широке тлумачення і переважно в економічному аспекті. В роботі [1] для оцінки якості роботи пошукової системи використовуються такі оцінки: точність, повнота, акуратність, помилка, F -міра, які визначаються як метрики на множинах документів і фактично дають кількісну характеристику самого пошуку. З результатів аналізу існуючих пошукових систем в [2] робиться висновок, що для пошуку документів гіпертекстових баз даних існуючі загальновизнані оцінки мають певні обмеження. Запропоновано використовувати додаткові характеристики, до яких відносять M -різновид вибірки та U -впорядкованість вибірки. На цій підставі наводиться коефіцієнт впорядкованості та коефіцієнт пошукового шуму. Виділена низка факторів, що впливають на успішність пошуку. Оцінці ефективності інформаційних систем, як одній з проблем інформаційного суспільства присвячена стаття [3], в якій на основі аналізу практичного застосування інформаційних систем показано, що в оцінці ефективності інформаційних систем можна виділити три типи ефектів: врахування додаткової інформації, нормування та врахування організаційних процесів та планування, оптимізації, управління процесами та ресурсами. Підкреслено роль врахування витрат, які ділять на дві складові: капітальні (бюджетні) або прямі витрати і позабюджетні, пов'язані з користувачами. Для оцінки трудовитрат приведена модифікована формула, яка враховує модель оцінки вартості розробки програмного забезпечення. Кількісні показники – оцінки функціональної ефективності інформаційно-пошукових систем приведені в [4]. До них віднесено такі: повноту, точність, акуратність, помилки. Для оцінювання функціональної ефективності інформаційно-пошукових систем запропоновано використовувати методи теорії статистичних рішень. Значна увага приділена модифікації відомого

критерію зваженої комбінації, та показано його ефективність на прикладі експериментального пошуку в масиві патентів США. У роботі [5] розглянута проблема пошуку інформації в Інтернет, її зв'язок з традиційною проблемою пошуку інформації. Описано нові завдання, відрізняють проблему пошуку в Інтернет від традиційної проблеми пошуку інформації, даний огляд існуючих методів пошуку інформації в Інтернет. Модель розв'язку задачі інформаційного пошуку, яка включає математичний опис послідовного та бінарного пошуків приведена в [6]. Зміст послідовного пошуку полягає в проведенні порівнянь записів. Для бінарного пошуку використовується бінарне дерево. Показано, що ефективність пошуку визначається принаймні двома основними – точністю і повнотою, та чотирма додатковими – специфічністю, вибірковістю, коефіцієнтом втрати інформації та коефіцієнтом пошукового шуму – показниками. Зазначено, що для оцінки роботи пошукової системи потрібна репрезентативна кількість запитів. У [7] формулюються принципи оцінки ефективності функціонування сучасних інформаційно-пошукових систем Інтернету. Наводяться результати тестування шести інформаційно-пошукових систем на основі методу визначення глибини користувацького пошуку.

На підставі аналізу існуючих підходів до оцінювання ефективності інформаційного пошуку можна зробити такі висновки.

1. В теоретичному плані оцінювання ефективності проводиться на підставі математичних моделей інформаційного пошуку. Для цього використовують переважно теоретико-множинний апарат, рідше ймовірнісний, і розглядають відношення множин релевантних та нерелевантних документів у видачі та інколи у інформаційному фонді.

2. В практичному використанні використовують критерії точності і повноти, рідше включають і частку нерелевантних документів у видачі.

3. Відсутність інтегрального критерію ефективності інформаційного пошуку.

Основною метою дослідження є розробка інтегрального критерію оцінювання ефективності інформаційного пошуку за результатами видачі в системах консолідованої інформації.

Такий інтегральний показник повинен враховувати не лише позитивний результат пошуку, але і негативний – частку нерелевантних документів та частку релевантних, але не виданих документів. Релевантні невидані документи за одним запитом можуть бути знайдені і включені у видачу або за рахунок іншого (нового) запиту або за рахунок модифікації даного запиту. Проте в першу чергу базове оцінювання ефективності пошуку має здійснюватись виключно на підставі видачі першого запиту, а вже далі такий інтегральний показник можна уточнювати додатковими оцінками.

3 МАТЕРІАЛИ І МЕТОДИ

Пошук в системі консолідованої інформації. Розглянемо роботу системи консолідованої інформації як діяльність користувача, пов'язану з відбором відповідної інформації стосовно поставленої задачі. В результаті зроб-

леного запиту інформаційно-пошукова система здійснює видачу знайдених документів. Як правило, не всі видані документи відповідають зробленому запиту і потребам користувача. З точки зору його задачі видані документи можуть бути поділені принаймні на три класи: релевантні P , пертинентні Π та не релевантні H .

Позначимо, через D множину різноманітних інформаційних джерел $D = \{d_1, d_2, \dots, d_n, d_i \in D, d_i \cap d_j \neq \emptyset, i, j = 1, 2, \dots, n\}$, які можуть мати спільні фонди інформаційного ресурсу; K – користувач, який формує запит для інформаційного пошуку в різнотипних джерелах.

Тоді, процес інформаційного пошуку в системах консолідованої інформації можна подати у вигляді схеми цілеорієнтованої роботи трьох блоків – «Опрацювання запитів», «Консолідація даних» та «Опрацювання даних» зображеної на рис. 1. Перший з цих блоків функціонально забезпечує переклад мови запиту користувача на мову запитів кожного з інформаційних джерел $d_i \in D$. В результаті, кожне таке джерело розуміє отриманий запит і процес пошуку може здійснюватися переважно його власною пошуковою системою.

Функціонально другий блок здійснює консолідацію знайдених даних, тобто приводить дані різних форматів у формат користувача, тобто вирішує обернену задачу – приведення різнотипних даних до типу запиту, сформованого користувачем. Консолідовані дані передаються в блок «Опрацювання даних», робота якого полягає у ранжуванні даних за частотою використання, часовими характеристиками, важливістю, доступністю, терміном використання тощо. Іншими словами, знайдена в результаті пошуку інформація має бути подана користувачеві у тій самій формі, у якій він сформував свій запит або в іншій, зрозумілій для нього формі. У такій ситуації кори-

стувач може отримати надзвичайно велику кількість, випадково перемішаних, як релевантних так і не релевантних документів. Для зменшення кількості нерелевантних документів у цьому блоці здійснюється відповідний логічний аналіз наявності збігів виданих документів з визначеними в запиті. Тут, фактично здійснюється фільтрація виданих документів, шляхом використання відповідних критеріїв, попередньо заданих користувачем. У цьому плані, інформаційний пошук в системах консолідованої інформації суттєво відрізняється від пошуку в звичайних базах чи сховищах даних. Тому, оцінювання ефективності інформаційного пошуку в системах консолідованої інформації має враховувати і особливості нормалізації різнотипних даних, тобто приведення їх до форми запиту користувача.

Відповідність видачі запиту. Найскладнішим моментом в оцінюванні ефективності будь-якого інформаційного пошуку є встановлення відповідності між знайденими і виданими документами і документами, а точніше пошуковими ознаками документів, поданих у запиті. Справа в тому, що ступінь відповідності, тобто чи є релевантними видані документи чи ні, є вельми суб'єктивним. Крім того, якщо можна точно відповісти даний документ є релевантний або нерелевантний, то чітко вказати, чи даний документ є пертинентним чи ні, оскільки він може бути пертинентним різною мірою.

Зі змісту понять релевантності та пертинентності випливає, що оцінювання ефективності пошуку має принципові дві складові. Нагадаємо, що поняття релевантності означає відповідність інформаційного пошуку, зробленому користувачем запиту, а пертинентність – відповідність інформаційній потребі користувача.

Перша з них це оцінювання, а точніше розуміння пошуковою системою складеного користувачем запиту. В цьому плані інформаційно-пошукова система відбирає ті документи, ознаки яких вказані у запиті. Очевидно, що в такому разі семантичний аналіз виявлених документів в базі або сховищі даних, у файлах чи бібліотеках не проводиться, а лише здійснюється зіставлення ознак виявлених документів і, за умови повного чи часткового збігу, документи подаються у видачу.

Друга складова це оцінювання документів у видачі, отриманих користувачем, у результаті інформаційного пошуку. Тут користувач розділяє документи на три групи: релевантні (P), пертинентні (Π) та нерелевантні (H).

Документи у видачі як правило сортуються інформаційно-пошуковою системою за певними критеріями: за датою (власна дата документа або остання дата звертання до нього), за рейтингом користування (скільки разів даний документ фігурував у запитах різних користувачів загалом чи за певний період). Можливі і інші критерії, наприклад за обсягом. Отримавши видачу, тобто перелік знайдених документів користувач послідовно або вибірково ознайомлюється з документами відбираючи релевантні та пертинентні і відкидаючи нерелевантні. Послідовність релевантних, пертинентних та нерелевантних документів у кожній конкретній видачі практично завжди є випадковою. Перевірка цього факту здійснена експериментально в такий спосіб.

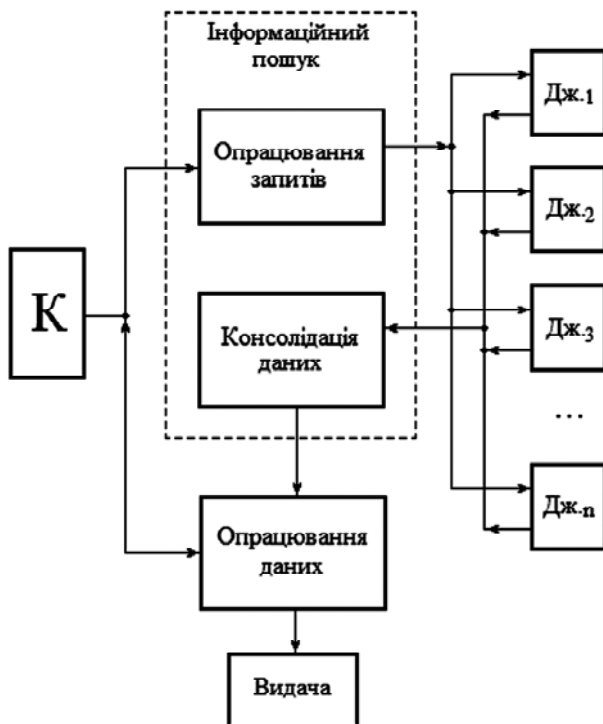


Рисунок 1 – Інформаційний пошук в різнотипних системах

Організація експериментів. В процесі пошуку необхідної інформації для проведення наукових досліджень поряд з відбором pertinentних документів також фіксувалися релевантні та нерелевантні. Зміст експерименту поданий планом дослідження.

4 ЕКСПЕРИМЕНТИ

Відбір ключових слів. Для цього були сформульовані такі ключові слова, а точніше словосполучення: інформаційний пошук; моделі інформаційного пошуку; інформаційно-пошукова система; ефективність інформаційно-пошукових систем; оцінювання ефективності інформаційного пошуку.

Уточнення понять. Pertinentність – документи, які за змістом максимально відповідають потребі користувача і мають усі реквізити для посилання на них у магістерській кваліфікаційній роботі (тобто документи, що є електронними копіями паперових монографій, статей у наукових журналах та збірниках праць, тезисах та працях наукових форумів та статті подані в енциклопедіях та довідниках).

Релевантність – документи, які за змістом цілком відповідають потребі користувача, але не мають реквізитів своїх паперових оригіналів і для посилання на них треба використовувати їхню електронну адресу, яка в деяких випадках є або громіздкою або неточною і для виявлення цього документу необхідно провести додатково ще й окремий спеціальний пошук, причому результат не гарантується.

Усі інші документи визнаються як нерелевантні.

Хід експерименту. Для експериментальних досліджень використано інформаційно-пошукові системи Google, Yandex, Meta, Rambler та Yahoo, які за ключовими словами видають веб-сторінки знайдених документів. Налаштуван-

ня пошуку забезпечило оптимальний варіант видачі результату – 10 електронних документів на кожній сторінці. На основі попередніх результатів пошуку і з власного досвіду, відомо, що потрібна інформація стосовно даного питання буде знаходитись на перших п'яти сторінках. Тому для експериментів вибрано обмеження 5 повних сторінок, тобто обсяг видачі становив 50 документів.

Для кожної сторінки, за результатами перегляду, кожному з десяти виданих документів присвоювалися індекси P , Π та H .

У табл. 1 приведені результати одноразової видачі знайдених документів для вказаних пошукових систем за ключовим словом «Оцінювання ефективності інформаційного пошуку».

Попередні результати. Задача користувача полягає в тому, щоб серед цієї множини вибрати саме ті, які йому потрібні. Очевидно за будь-якого пошуку перегляд отриманих документів буде аналогічним. Оскільки надана вибірка є скінченна, можемо оцінити ефективність пошукової системи відношенням сприятливих подій до всіх можливих, тобто відношенням, наприклад, кількості релевантних документів до кількості всіх наданих документів, отриманих за даним запитом. Якщо документи класифікувати як в даному прикладі, то можна отримати три частоти появи документів кожного класу:

$$f_P = \frac{\sum_{i=1}^n P_i}{N}, \quad f_H = \frac{\sum_{j=1}^m H_j}{N}, \quad f_{\Pi} = \frac{\sum_{k=1}^l \Pi_k}{N},$$

де N – кількість документів у видачі.

Таблиця 1 – Оцінювання ефективності інформаційного пошуку

Ключове слово «Оцінювання ефективності інформаційного пошуку»					
	Google	Yandex	Meta	Rambler	Yahoo
1	Р Р Р Р Н	Р Р Н Р Н	Н Н Н Н Н	Н Р Н Р Н	Н Р Р Р Н
2	Р Р П Н Н	Р Р Н Н Н	Н Р Н Р Н	Н Р Р Р Р	Н Н Р Р Н
3	Р Н П Р Н	Р П Н Р Н	Н Р Р Р Н	Р Р Р Р Н	Р Р Р Р Н
4	Н Н Н Р Н	Н Р Н Р Н	Р Н Н Н Н	Р Р П Р Н	Р Р П Р Н
5	Н Н Р Р Н	Р Н Р Р Н	Р Н Н Н Н	Р Н Н Р Р	Р Н Н Н Н
6	Н Н Н Р Н	Р Р Р Р Н	Р Н Р Р Н	Р Н Н Н Н	Р Н Р Р Н
7	Н Н Н Н Н	Р Р Н Н Н	Н Н Р Р Н	Н Н Н Н Н	Н Н Н Р Н
8	Н Р Н Н Н	Р Н Н Н Н	Н Н Н Н Н	Н Н Н Н Н	Н Н Н Н Н
9	Н П Н Н Н	Р Н Н Н Н	Н Н Н Н Н	Н Н Р Р Н	Н Н Н Н Н
10	Н Р Н Р Н	Р Р Р Р Н	Р Н Н Н Н	Р Р Н Н Н	Р Н Н Н Н

Таблиця 2 – Зведена таблиця експериментів

Ключові слова	Google	Yandex	Meta	Rambler	Yahoo
Оцінка ефективності інформаційного пошуку	П – 4 Р – 13 Н – 33	П – 1 Р – 21 Н – 28	П – 0 Р – 11 Н – 39	П – 1 Р – 19 Н – 30	П – 1 Р – 15 Н – 34
Інформаційний пошук	П – 7 Р – 11 Н – 32	П – 2 Р – 12 Н – 36	П – 2 Р – 14 Н – 34	П – 3 Р – 15 Н – 32	П – 4 Р – 10 Н – 36
Моделі інформаційного пошуку	П – 3 Р – 11 Н – 36	П – 2 Р – 18 Н – 30	П – 1 Р – 9 Н – 40	П – 9 Р – 20 Н – 21	П – 1 Р – 21 Н – 28
Інформаційно-пошукова система	П – 1 Р – 21 Н – 28	П – 0 Р – 22 Н – 28	П – 0 Р – 5 Н – 45	П – 1 Р – 18 Н – 31	П – 1 Р – 11 Н – 38
Ефективність інформаційно-пошукових систем	П – 1 Р – 16 Н – 33	П – 2 Р – 22 Н – 26	П – 0 Р – 17 Н – 33	П – 1 Р – 16 Н – 33	П – 3 Р – 21 Н – 26

На практиці як правило інформаційний пошук здійснюється за різними запитами, в залежності від поставлених задач.

У свою чергу, задачі можуть стосуватися різних предметних областей, обсягу їх онтологій, специфіки конкретних об'єктів, що потребують їхнього розв'язку. З другої сторони, не можна бути впевненим в тому, що інформаційні джерела мають усю необхідну інформацію з будь-якої області знань та діяльності людини. А тому кількість наданих користувачам документів є різними. Зазвичай пошук в джерелах інформації здійснюється пошуковою системою, яка працює за певним алгоритмом і визначеними формальними критеріями відповідності, а тому, можна припустити, що результати різних пошуків в одному і тому ж джерелі інформації будуть статистично однорідні, тобто матимуть певні статистичні закономірності, які можуть відбитися, принаймні, на співвідношенні частоти появи розглянутих вище класів.

5 РЕЗУЛЬТАТИ

Послідовність документів у видачі можна зобразити графічно, у вигляді діаграми приведеної на рис. 2.

В якості кількісної оцінки використано відносну частоту появи того чи іншого виду документів. Для поданого результату пошуку маємо такі співвідношення:

пертинентних $f_{\Pi} = \frac{\Pi}{50}$, релевантних $f_P = \frac{P}{50}$, нерелевантних $f_H = \frac{H}{50}$,

релевантних і пертинентних (корисних) $f_{P+\Pi} = \frac{P+\Pi}{50}$, а також відношення загальної кількості релевантних і пертинентних документів до кількості нерелевантних документів

$$\gamma = \frac{P+\Pi}{H}.$$

Очевидно, що усі ці значення значною мірою залежать від обсягу документів в інформаційній системі (базі, сховищі даних, папках з файлами, бібліотеці), можливостей інформаційно-пошукової системи, форми запиту, а також від інформаційної потреби користувача – наскільки глибоко він розуміє завдання, для вирішення якого він здійснює даний пошук.

Оцінювання ефективності інформаційного пошуку. Сформовані, практично на відповідних пошукових мовах, властивих тому чи іншому інформаційному фонду, запити мають досить обмежену кількість пошукових ознак – ключових слів, певного типу розширень та пояснень чи обмежень. Алгоритми інформаційно-пошуко-

вих систем використовуючи ці дані в процесі сканування-пошуку існуючого каталогу переважно використовують в якості даних автора, назву та анотацію документів, хоча можливим є і сканування самого документа. Оскільки ключові слова в залежності від контексту можуть мати декілька значень у видачу потрапляють абсолютно нерелевантні документи.

У загальному оцінювання ефективності базується на визначенні, як було сказано вище, на оцінках точності і повноти. Спроба використати додаткові показники пошуку вимагає врахування не лише обсягу самого інформаційного фонду, але і обсягу релевантних та нерелевантних стосовно даного запиту документів. Отримати такі дані практично неможливо, оскільки для одної задачі документи можуть бути релевантними, а для другої вже ні. З другої сторони, якщо знати всі релевантні документи у фонді то можна здійснити пошук лише для них і тоді у видачі будуть лише релевантні документи, а це здійснити практично не можливо, принаймні з двох причин: ніхто не буде з багатотисячного інформаційного фонду відбирати релевантні для даної задачі окремого користувача документи, присутність конфіденційної інформації та відсутність інформації про сам фонд, за винятком лише загальних його характеристик. Тому найбільш правомірним є оцінювання ефективності пошуку за його результатами, тобто на основі документів, які є у видачі.

За наявності трьох типів документів оцінити ефективність інформаційного пошуку можна в такий спосіб логічного виведення. Очевидно, що пертинентні документи мають найбільшу цінність для користувача.

Очевидним є те, що для своєї задачі користувач використовує лише релевантні та пертинентні документи, тому ефективність пошуку у загальному випадку є пропорційна кількості релевантних і пертинентних документів, що є сприятливою подією для користувача, тобто

$$E_{\text{пош}} = \frac{P+\Pi}{\Pi+P+H}. \quad (1)$$

Наявність у видачі не релевантних документів є обернено-пропорційною подією до кількості пертинентних і релевантних документів у видачі, а тому, ефективність відносно не релевантних документів можна подати як

$$E_{\text{пош}} = \frac{P+\Pi}{H}. \quad (2)$$

Враховуючи особливості форми запиту, яка тісно пов'язана з даною конкретною інформаційною системою, тобто з її інформаційним фондом та його системою індексування необхідно ввести деякий коректую-

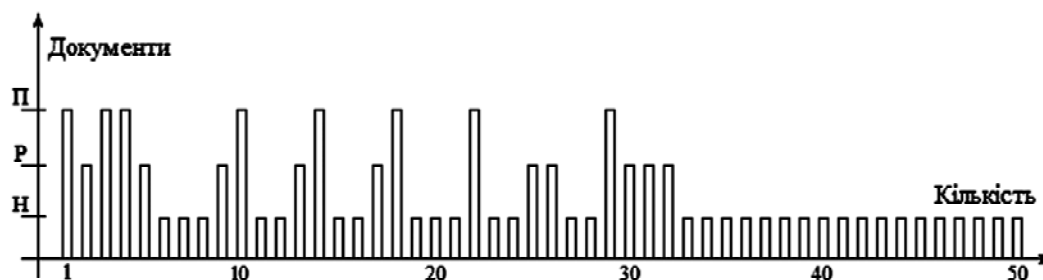


Рисунок 2 – Розподіл документів у видачі: П – пертинентні, Р – релевантні, Н – нерелевантні

чий множник – коефіцієнт пропорційності β , в результаті чого отримаємо

$$E_{\text{пош}} = \beta \cdot \frac{P + P}{P + P + H} \quad (3)$$

Оцінка ефективності пошуку у вигляді (3) дає характеристику здійсненого інформаційного пошуку для даного конкретного ключового слова та конкретної пошукової системи за результатами отриманої видачі, обмеженої, наприклад, 50 документами.

За кількістю релевантних документів стосовно вибраних ключових слів та інформаційно-пошукових систем результати пошуку наведені в табл. 3.

У поданій оцінці ефективності, залишається невідомим показник β . Оцінити цей показник можна лише на підставі даних про пошуковий алгоритм і саму пошукову систему, яка використовується даним інформаційним фондом.

Якщо припустити, що пошуковий алгоритм інформаційної системи переглядає всі документи інформаційного фонду або, принаймні, усі документи конкретної рубрики згідно з інформацією поданою у запиті, тоді можна прийняти значення показника $\beta = 1$. Проте докладну інформацію про характеристики цього алгоритму отримати практично неможливо.

Очевидним є той факт, що чим більший обсяг інформаційного фонду, тим більше релевантних документів буде знайдено. Однак, тут треба мати на увазі і популярність чи розвиненість даної тематики, оскільки саме вона, тобто її популярність і затребуваність визначають обсяг документів у фонді [9–10].

6 ОБГОВОРЕННЯ

Вирази (2) і (3) дають об'єктивну оцінку ефективності інформаційного пошуку, але в перших (лівих) варіантах, лише за умови, що у видачі будуть присутніми і не релевантні документи – принаймні, хоча б один. Невиконання цієї умови означає ділення на нуль. Тобто, буде невірний результат оцінювання. Така ситуація може виникнути тоді, коли кількість релевантних документів в інформаційному фонді перевищує обсяг видачі.

У цьому випадку оцінюється ефективність за видачею для одного чи декількох запитів. Якщо такий показник використати для кожного з декількох запитів, але таких, що стосуються конкретної теми можна оцінити якість і самого запиту, точніше встановити, який з запитів чи які ключові слова є найбільш ефективними і вже за ними модифікувати наступні запити.

ВИСНОВКИ

Ефективність інформаційного пошуку в системах консолідованої інформації в сенсі побудови інтегрального показника практично не може бути визначена, оскільки крім двох показників – повноти і точності, усі інші вимагають знання кількості

релевантних та нерелевантних документів у даному інформаційному фонді стосовно даної задачі. Отримати такі дані для великих за обсягами фондів є неможливо, оскільки: по-перше здійснити такий підрахунок означає перегляд кожного документа, по-друге, у великих базах даних перехід від нерелевантних документів до релевантних практично за будь-яким запитом є нечітким і розмитим, по-третє – для різних задач поняття релевантності документів різняться.

Найпростішим способом побудови оцінки ефективності пошуку є використання логічного підходу, який подається відношенням – кількості отриманих потрібних і замовлених документів до кількості документів у даній видачі. В цьому плані, на ефективність пошуку впливає не лише наявність в інформаційному фонді потрібних документів, але й правильність побудови самого запиту згідно з вимогами даної пошукової системи.

Наведений приклад оцінювання ефективності інформаційного пошуку демонструє використання поділу знайдених і виданих документів на пертинентні, релевантні та нерелевантні. В результаті якого, оцінку ефективності можна подати як усереднену, або сумарну, на підставі проведення інформаційного пошуку в одному або в кількох інформаційних фондах і на різних пошукових системах за одного набору ключових слів.

Розроблений підхід до побудови оцінки інформаційного пошуку має практичне значення, оскільки отримані кількісні значення локальних оцінок дають підстави для оптимізації набору ключових слів, та визначення найбільш відповідних інформаційних фондів і пошукових систем.

ПОДЯКИ

Роботу виконано в рамках держбюджетної науково-дослідної теми «Методи та засоби консолідації баз даних в інформаційних системах електронного урядування», тематика кафедри інформаційних систем та мереж Національного університету «Львівська політехніка», 2010/2012, № держреєстрації 0110U005022.

СПИСОК ЛІТЕРАТУРИ

1. Агеев М. Официальные метрики РОМИП 2010 / М. Агеев, И. Кураленок, И. Некрестьянов // Российский семинар по Оценке Методов Информационного Поиска. Труды РОМИП 2010, Казань, 15 октября 2010 г. – Казань, 2010. – С. 172–187.
2. Целых А.Н. Оценка эффективности информационного поиска / А. Н. Целых, Э. М. Котов // Известия ТРТУ. Тематический выпуск «Управление в математических системах». – Таганрог : Изд-во ТРТУ. – 2006. – № 10 (65). – С. 43–45.
3. Яхина Е.П. Методы оценки информационных систем / Е. П. Яхина // В мире научных открытий. – 2010. – № 3 (09). – Часть 1. – С. 63–66.
4. Попов С. В. Оценка функциональной эффективности систем текстового поиска на примере поиска патентных документов / С. В. Попов // Патентная информация сегодня. – 2010. – № 1. – С. 22–25.
5. Козлов Д. Д. Информационно-поисковые системы в Internet: текущее состояние и пути развития / Д. Д. Козлов // Техноло-

Таблиця 3 – Оцінка ефективності інформаційно-пошукових систем за кількістю релевантних документів

Ключові слова	Google	Yandex	Meta	Rambler	Yahoo
Оцінка ефективності інформаційного пошуку	0,34	0,44	0,22	0,40	0,32
Інформаційний пошук	0,36	0,28	0,32	0,36	0,28
Модель інформаційного пошуку	0,28	0,40	0,20	0,58	0,44
Інформаційно-пошукова система	0,44	0,44	0,10	0,38	0,24
Ефективність інформаційно-пошукових систем	0,34	0,48	0,34	0,34	0,48
Усереднений показник ефективності	0,35	0,40	0,23	0,41	0,35

гический обзор [Электронный ресурс]. – Режим доступа: lvk.cs.msu.su/~ddk/ir_and_ia_review.pdf

6. Тьякин И. В. Математическая модель информационного поиска и оценка эффективности поисковой системы / И. В. Тьякин, В. М. Тютюнник // Вестник ТГТУ. – 2008. – Том 14. – № 3. – С. 478–481.
7. Козлов М. В. Метод оценки эффективности функционирования современных информационно-поисковых систем Интернета / М. В. Козлов, В. А. Яцко [Электронный ресурс]. – Режим доступа: <http://www.dialog-21.ru/dialog2006/materials/html/Kozlov.htm>.
8. Лекции по введению в информатику и информационные системы. – Лекция 13. Эффективность информационных систем.

Ныч Л. Я.¹, Шаховска Н. Б.², Каминский Р. М.³

¹Ассистент, кафедра информационных систем и сетей, Национальный университет «Львовская политехника», Львов, Украина

²Д-р техн. наук, профессор, кафедра информационных систем и сетей, Национальный университет «Львовская политехника», Львов, Украина

³Д-р техн. наук, доцент, кафедра информационных систем и сетей, Национальный университет «Львовская политехника», Львов, Украина

ОЦЕНКА ЭФФЕКТИВНОСТИ ИНФОРМАЦИОННОГО ПОИСКА В СИСТЕМАХ КОНСОЛИДИРОВАННОЙ ИНФОРМАЦИИ

Для оценки эффективности информационного поиска предложено разделять найденные документы на pertinent, релевантные и нерелевантные. Эффективность поиска определяется отношением количества pertinent и релевантных документов к количеству нежелательных документов, а свойством поисковой системы должно быть соответствующие коэффициенты. Основной целью данного исследования является разработка интегрального критерия оценки эффективности информационного поиска по результатам выдачи в системах консолидированной информации. Введено понятие консолидированной информации.

Разработан метод оценки эффективности информационного поиска. Он демонстрирует использование разделения найденных и выданных документов на pertinent, релевантные и нерелевантные. Введено интегральный показатель релевантности документа поисковому запросу, который учитывает негативную и положительную оценку. Оценку эффективности определено как суммарное значение различных компонентов. Эксперимент выполнен на основании проведения информационного поиска в одном или в нескольких информационных фондах и на разных поисковых системах с одним набором ключевых слов.

Разработанный подход к построению оценки информационного поиска имеет практическое значение, поскольку полученные количественные значения локальных оценок дают основания для оптимизации набора ключевых слов, и определение наиболее подходящих информационных фондов и поисковых систем.

Ключевые слова: информационная система, информационный поиск, эффективность, pertinentность, релевантность.

Nych L. Ya.¹, Kaminsky R. M.², Shakhovska N. B.³

¹Assistant professor, department of information systems and networks, Lviv Polytechnic National University, Lviv, Ukraine

²Dr.Sc., Professor, department of information systems and networks, Lviv Polytechnic National University, Lviv, Ukraine

³Dr.Sc., Professor, department of information systems and networks, Lviv Polytechnic National University, Lviv, Ukraine

EFFECTIVENESS EVALUATION OF SEARCH IN INFORMATION SYSTEMS WITH CONSOLIDATED INFORMATION

To evaluate the effectiveness of information retrieval there is proposed to share the found documents on pertinent, relevant and irrelevant. Search Performance is ratio to determine the number of pertinent and relevant documents to the number of irrelevant documents and search engine properties have been submitted by the coefficient. The goal of this paper is to develop integrated criterion of evaluating the effectiveness of information retrieval on the results of the issuance of consolidated information systems. The concept of consolidated information is given.

The method of evaluating the effectiveness of information retrieval is built. It demonstrates the usage of the division found and published documents on pertinent, relevant and irrelevant. There is given integral indicator of the relevance of the document search query that takes into account the negative and positive features. Evaluation of effectiveness presented as the total value of the different components. The experiment was performed on the basis of information search in one or several search machines and information on the various search engines for one set of keywords.

The approach to building assessment information retrieval is of practical importance because quantitative values obtained local assessments give grounds to optimize the set of keywords and determine the most appropriate information collection and search engines.

Keywords: information system, information search, efficiency, pertinence, relevance, irrelevance.

REFERENCES

1. Aheev M., Kuralenok Y., Nekrestianov Y. Ofytsyalnye metryky ROMYP 2010, *Rosyiskiy semynar po Otsenke Metodov Informatsyonnoho Poiska. Trudy ROMYP 2010. (Kazan, 15 october 2010.)* Kazan, 2010, pp. 172–187.
2. Tselykh A. N., Kotov E. M. Otsenka efektyvnosti ynformatsyonnoho poyska, *Yzvestyia TRTU. Tematycheskyi vypusk «Upravlenye v matematycheskykh systemakh»*. Tahanroh, Yzd-vo TRTU, 2006, No. 10 (65), pp. 43–45.
3. Yakhyna E. P. Metody otsenky ynformatsyonnykh system, *V myre nauchnykh otkrytiy*, 2010, No. 3 (09), Chast 1, pp. 63–66.
4. Popov S. V. Otsenka funktsionalnoi efektyvnosti system tekstovoho poyska na prymere poyska patentnykh dokumentov, *Patentnaia informatsiya sehodnia*, 2010, No. 1, pp. 22–25.
5. Kozlov D. D. Ynformatsyonno-poyskovyye systemy v Internet: tekushchee sostoyaniye i puty razvityia, *Tekhnolohycheskyi obzor*. Access mode: lvk.cs.msu.su/~ddk/ir_and_ia_review.pdf
6. Tiavkyn Y. V., Tiutiunyk V. M. Matematycheskaia model informatsyonnoho poiska i otsenka yeffektyvnosti poyskovoi sistemy, *Vestnyk THTU*, 2008, Vol 14, № 3, pp. 478–481.
7. Kozlov M. V., Yatsko V. A. Metod otsenky efektyvnosti funktsyonyrovaniya sovremennykh ynformatsyonno-poyskovykh system Interneta, Access mode: <http://www.dialog-21.ru/dialog2006/materials/html/Kozlov.htm>.
8. Lektsyy po vvedeniyu v informatyku i ynformatsyonnyye systemy. Lektsiya 13. Yeffektyvnost ynformatsyonnykh system. Access mode: <http://informling.narod.ru/lectures.html>
9. Kirchgassner G., Wolters J. Introduction to Modern Time Series Analysis. Springer Berlin Heidelberg, New York, 2007, 274 p.
10. Hegger R., Kantz H., Schreiber T. Practical implementation of nonlinear time series methods: The TISEAN package. *CHAOS* 9, 1999, pp. 413–435.
11. Kuhlthau, C. C. Seeking Meaning: A Process Approach to Library and Information Services, 2nd. ed. Westport, CT, Libraries Unlimited, 2004, 342 p.

Стаття надійшла до редакції 10.02.2016.

Після доробки 15.02.2016.

УПРАВЛІННЯ У ТЕХНІЧНИХ СИСТЕМАХ

УПРАВЛЕНИЕ В ТЕХНИЧЕСКИХ СИСТЕМАХ

CONTROL IN TECHNICAL SYSTEMS

УДК 004.052: 656.61

Kołowrocki K.¹, Kwiatkowska-Sarnecka B.², Soszyńska-Budny J.³¹Dr. hab. Sc., Professor, Professor of Department of Mathematics, Gdynia Maritime University, Gdynia, Poland²PhD, Assistant Professor, Assistant Professor of Department of Mathematics, Gdynia Maritime University, Gdynia, Poland³PhD., Associate Professor of Department of Mathematics, Gdynia Maritime University, Gdynia, Poland

RELIABILITY AND RISK OPTIMIZATION OF MULTISTATE SYSTEMS WITH APPLICATION TO PORT TRANSPORTATION SYSTEM

The complexity of technical systems' operation processes and its influence on the changing in time systems' structures and their components' reliability parameters poses a difficulty to first meet in real and then to fix and analyse those structures and reliability parameters. By constructing a joint model of reliability of complex technical systems at variable operation conditions, which links a semi-markov modelling of system operation processes with multi-state approach to system reliability analysis, we find the system's main reliability characteristics. Consequently, we use linear programming to build a model of complex technical systems reliability optimization. We investigate the model's application in marine transport, specifically in reliability and risk optimization of a bulk cargo transportation system. The tools we develop can be used in reliability evaluation and optimization of a very wide class of real technical systems operating at varying conditions that influence their reliability structures and the reliability parameters of their components. Consequently, the tools we developed can be implemented by reliability practitioners from both maritime transport industry and other industrial sectors.

Keywords: reliability function, risk function, operation process, optimization.

NOMENCLATURE

$Z(t), z_i$ is a system operation process; a system operation state; $i = 1, 2, \dots, v$,

$\dot{p}_b$ is an optimal transient probability;

U is a particular reliability state of the system;

$u = 1, 2, \dots, z$;

$T(u)$ is a lifetime of a system in the reliability state subset $\{u, u+1, \dots, z\}$;

$[R(t, u)]^{(b)}$ is a conditional reliability function of a system at the operational state z_b ;

$\dot{R}(t, u)$ is an optimal unconditional reliability function of a system;

$\mu(u)$ is a mean lifetime of the system in the reliability state subset $\{u, u+1, \dots, z\}$;

$\bar{\mu}(u)$ is a mean lifetime of the system in the reliability particular state u ;

R is a critical state of the multi-state system;

$\dot{r}(t)$ is an optimal risk function of the multi-state system.

INTRODUCTION

Most real technical systems are very complex because they are composed of large numbers of components and subsystems and have high operating complexity. The complexity of the systems' operation processes and its influence on the changing in time systems' structures and their components' reliability parameters poses a difficulty to first meet in real and then to fix and analyse those structures and reliability parameters. A convenient tool to investigate this problem is a semi-markov [2] modelling of the system operation process linked with a multi-state approach for the system reliability analysis [1, 4, 9–10] and a linear programming for the system reliability optimization [3]. Using this approach, it is possible to find this complex system's main reliability characteristics including the system reliability function, the system mean lifetimes in the reliability states subsets and the system risk function [4, 6, 8]. Having those characteristics it is possible to optimize the system operation process to get optimal values [8]. To this end the linear programming [3] can be applied to maximize the mean value of the system lifetime in the subset of the system reliability states, which are not worse than the system critical reliability state.

1 SYSTEM RELIABILITY AT VARIABLE OPERATIONS PROCESS

We suppose that the system has v different operation states during its operation process. Thus, we can define the system operation process $Z(t)$, $t \in \langle 0, +\infty \rangle$, as the process with discrete operation states from the set $Z = \{z_1, z_2, \dots, z_v\}$. Further, we assume that $Z(t)$ is a semi-markov process [2] with its conditional sojourn times θ_{bl} at the operation state z_b when its next operation state is z_l , $b, l = 1, 2, \dots, v$, $b \neq l$. In this case the process $Z(t)$ may be described by:

- the vector of probabilities of the process initial operation states $[p_b(0)]_{1 \times v}$,
- the matrix of probabilities of the process transitions between the operation states $[p_{bl}]_{v \times v}$,
- the matrix of conditional distribution functions $[H_{bl}(t)]_{v \times v}$ of the process sojourn times θ_{bl} , $b \neq l$, in the operation state z_b when the next operation state is z_l .

Under these assumptions, the sojourn times θ_{bl} mean values are given by

$$M_{bl} = E[\theta_{bl}] = \int_0^{\infty} t dH_{bl}(t), \quad b, l = 1, 2, \dots, v, \quad b \neq l. \quad (1)$$

The mean values $E[\theta_b]$ of the unconditional sojourn times θ_b are given by

$$M_b = E[\theta_b] = \sum_{l=1}^v p_{bl} M_{bl}, \quad b = 1, 2, \dots, v, \quad (2)$$

where M_{bl} are defined by (1).

Limit values of the transient probabilities at the operation states are given by

$$p_b = \lim_{t \rightarrow \infty} p_b(t) = \frac{\pi_b M_b}{\sum_{l=1}^v \pi_l M_l}, \quad b = 1, 2, \dots, v,$$

where the probabilities π_b of the vector $[\pi_b]_{1 \times v}$ satisfy the system of equations $[\pi_b] = [\pi_b][p_{bl}]$ and $\sum_{l=1}^v \pi_l = 1$ and

$$p_b(t) = P(Z(t) = z_b), \quad t \in \langle 0, +\infty \rangle, \quad b = 1, 2, \dots, v.$$

We assume that the system is composed of n independent multistate components E_i , $i = 1, 2, \dots, n$, and that the changes of the operation process $Z(t)$ states have an influence on both the system components E_i reliability and on the system reliability structure. Consequently, we denote the component E_i lifetime in the reliability state subset $\{u, u+1, \dots, z\}$ by $T_i^{(b)}(u)$ and by

$$[R_i(t, \cdot)]^{(b)} = [1, [R_i(t, 1)]^{(b)}, [R_i(t, 2)]^{(b)}, \dots, [R_i(t, z)]^{(b)}],$$

where for $t \in \langle 0, \infty \rangle$, $b = 1, 2, \dots, v$, $u = 1, 2, \dots, z$,

$$[R_i(t, u)]^{(b)} = P(T_i^{(b)}(u) > t | Z(t) = z_b),$$

is the conditional reliability function while the system is at the operational state z_b , $b = 1, 2, \dots, v$.

Next, we denote the system lifetime in the reliability state subset $\{u, u+1, \dots, z\}$ by $T^{(b)}(u)$ and by

$$[R(t, \cdot)]^{(b)} = [1, [R(t, 1)]^{(b)}, [R(t, 2)]^{(b)}, \dots, [R(t, z)]^{(b)}],$$

where for $t \in \langle 0, \infty \rangle$, $b = 1, 2, \dots, v$, $u = 1, 2, \dots, z$,

$$[R(t, u)]^{(b)} = P(T^{(b)}(u) > t | Z(t) = z_b),$$

is the conditional reliability function of the system while the system is at the operational state z_b .

In the case when the system operation time is large enough, the unconditional reliability function of the system is given by

$$R(t, \cdot) = [1, R(t, 1), R(t, 2), \dots, R(t, z)], \quad t \geq 0,$$

where

$$R(t, u) \cong \sum_{b=1}^v p_b [R(t, u)]^{(b)}. \quad (3)$$

The mean values of the system lifetimes in the reliability state subset $\{u, u+1, \dots, z\}$ are

$$\mu(u) = E[T(u)] \cong \sum_{b=1}^v p_b \mu_b(u), \quad u = 1, 2, \dots, z, \quad (4)$$

and the mean values of the system lifetimes in the particular reliability state u , are [4]

$$\bar{\mu}(u) = \mu(u) - \mu(u+1), \quad u = 1, 2, \dots, z-1, \quad \bar{\mu}(z) = \mu(z). \quad (5)$$

A probability $r(t) = P(s(t) < r | R(0) = z) = P(T^{(b)}(r) \leq t)$, $t \in (-\infty, \infty)$,

that the system is in the subset of reliability states worse than the critical state r , $r \in \{1, \dots, z\}$ while it was in the state z at the moment $t = 0$ is called a risk function of the multi-state system [4].

Under this definition, from (3), we have

$$r(t) = 1 - R(t, r), \quad t \in (-\infty, \infty), \quad (6)$$

and if τ is the moment when the risk exceeds a permitted level δ , then

$$\tau = r^{-1}(\delta), \quad (7)$$

where $r^{-1}(t)$, if it exists, is the inverse function of the risk function $r(t)$.

2 OPTIMAL TRANSIENT PROBABILITIES MAXIMIZING SYSTEM LIFETIME

Considering the equation (3), it is natural to assume that the system operation process has a significant influence on the system reliability. This influence is also clearly expressed in the equation (4) for the mean values of the system

unconditional lifetimes in the reliability state subsets. From linear equation (4), we can see that the mean value of the system unconditional lifetime $\mu(u)$, $u = 1, 2, \dots, z$, is determined by the limit values of transient probabilities p_b , $b = 1, 2, \dots, v$, of the system operation states and the mean values $\mu_b(u)$, $b = 1, 2, \dots, v$, $u = 1, 2, \dots, z$, of the system conditional lifetimes in the reliability state subsets $\{u, u+1, \dots, z\}$, $u = 1, 2, \dots, z$. Therefore, the system lifetime optimization approach based on the linear programming can be proposed. Namely, we may look for the corresponding optimal values $\dot{p}_b$ of the transient probabilities p_b in the system operation states to maximize the mean value $\mu(u)$ of the unconditional system lifetimes in the reliability state subsets $\{u, u+1, \dots, z\}$ under the assumption that the mean values $\mu_b(u)$ of the system conditional lifetimes in the reliability state subsets are fixed. As a special case of the above formulated system lifetime optimization problem: if r , $r = 1, 2, \dots, z$, is a system critical reliability state, then we want to find the optimal values $\dot{p}_b$ of the transient probabilities p_b in the system operation states to maximize the mean value $\mu(r)$ of the unconditional system lifetime in the reliability state subset $\{r, r+1, \dots, z\}$ under the assumption that the mean values $\mu_b(r)$, $b = 1, 2, \dots, v$, of the system conditional lifetimes in this reliability state subset are fixed. More exactly, we formulate the optimization problem as a linear programming model with the objective function of the following linear form

$$\mu(r) = \sum_{b=1}^v p_b \mu_b(r) \quad (8)$$

for a fixed $r \in \{1, 2, \dots, z\}$ and with the following bound constraints

$$\sum_{b=1}^v p_b = 1, \quad \bar{p}_b \leq p_b \leq \hat{p}_b, \quad b = 1, 2, \dots, v, \quad (9)$$

where $\mu_b(r)$, $\mu_b(r) \geq 0$, $b = 1, 2, \dots, v$, are fixed mean values of the system conditional lifetimes in the reliability state subset $\{r, r+1, \dots, z\}$ and

$\bar{p}_b$, $0 \leq \bar{p}_b \leq 1$ and $\hat{p}_b$, $0 \leq \hat{p}_b \leq 1$, $\bar{p}_b \leq \hat{p}_b$, $b = 1, 2, \dots, v$, (10) are respectively the lower and upper bounds of the unknown transient probabilities p_b .

Now, we can obtain the optimal solution to the formulated by (8)–(10) the linear programming problem, i.e. we can find the optimal values $\dot{p}_b$ of the transient probabilities p_b , $b = 1, 2, \dots, v$, that maximize the objective function given by (8). First, we arrange the system conditional lifetime mean values $\mu_b(r)$, $b = 1, 2, \dots, v$, in non-increasing order

$$\mu_{b_1}(r) \geq \mu_{b_2}(r) \geq \dots \geq \mu_{b_v}(r), \quad b_i \in \{1, 2, \dots, v\} \text{ for } i = 1, 2, \dots, v.$$

Next, we substitute

$$x_i = p_{b_i}, \quad \bar{x}_i = \bar{p}_{b_i}, \quad \hat{x}_i = \hat{p}_{b_i} \text{ for } i = 1, 2, \dots, v \quad (11)$$

and we maximize with respect to x_i , $i = 1, 2, \dots, v$, the linear form (8) that takes the form

$$\mu(r) = \sum_{i=1}^v x_i \mu_{b_i}(r) \quad (12)$$

for a fixed $r \in \{1, 2, \dots, z\}$ with the following bound constraints

$$\sum_{i=1}^v x_i = 1, \quad \bar{x}_i \leq x_i \leq \hat{x}_i, \quad i = 1, 2, \dots, v, \quad (13)$$

where $\mu_{b_i}(r)$, $\mu_{b_i}(r) \geq 0$, $i = 1, 2, \dots, v$, are fixed mean values of the system conditional lifetimes in the reliability state subset $\{r, r+1, \dots, z\}$ arranged in non-increasing order and

$$\bar{x}_i, \quad 0 \leq \bar{x}_i \leq 1 \text{ and } \hat{x}_i, \quad 0 \leq \hat{x}_i \leq 1, \quad \bar{x}_i \leq \hat{x}_i, \quad i = 1, 2, \dots, v, \quad (14)$$

are the lower and upper bounds of the unknown probabilities x_i , $i = 1, 2, \dots, v$, respectively.

We define

$$\bar{x} = \sum_{i=1}^v \bar{x}_i, \quad \hat{y} = 1 - \bar{x} \quad (15)$$

and

$$\bar{x}^0 = 0, \quad \hat{x}^0 = 0 \text{ and } \bar{x}^I = \sum_{i=1}^I \bar{x}_i, \quad \hat{x}^I = \sum_{i=1}^I \hat{x}_i \text{ for } I = 1, 2, \dots, v. \quad (16)$$

Next, we find the largest value $I \in \{0, 1, \dots, v\}$ such that

$$\hat{x}^I - \bar{x}^I < \hat{y} \quad (17)$$

and we fix the optimal solution that maximize (12) in the following way:

i) if $I = 0$, the optimal solution is $\dot{x}_1 = \hat{y} + \bar{x}_1$ and $\dot{x}_i = \bar{x}_i$

$$\text{for } i = 2, 3, \dots, v; \quad (18)$$

ii) if $0 < I < v$, the optimal solution is

$$\dot{x}_i = \bar{x}_i \text{ for } i = 1, 2, \dots, I, \quad \dot{x}_{I+1} = \hat{y} - \bar{x}^I + \bar{x}^I + \bar{x}_{I+1} \text{ and}$$

$$\dot{x}_i = \bar{x}_i \text{ for } i = I+2, I+3, \dots, v; \quad (19)$$

iii) if $I = v$, the optimal solution is $\dot{x}_i = \bar{x}_i$ for $i = 2, 3, \dots, v$. (20)

Finally, after making the inverse to (11) substitution, we get the optimal limit transient probabilities

$$\dot{p}_{b_i} = \dot{x}_i \text{ for } i = 1, 2, \dots, v, \quad (21)$$

that maximize the system mean lifetime $\mu(r)$ in the reliability state subset $\{r, r+1, \dots, z\}$, defined by the linear form (8) giving its maximum value in the following form

$$\dot{\mu}(r) = \sum_{b=1}^v \dot{p}_b \mu_b(r) \text{ for a fixed } r \in \{1, 2, \dots, z\}. \quad (22)$$

From the above, replacing r by u , $u = 1, 2, \dots, z$, we obtain the corresponding optimal solutions for the mean values of the system unconditional lifetimes in the reliability state subsets $\{u, u+1, \dots, z\}$ of the form

$$\dot{\mu}(u) = \sum_{b=1}^v \dot{p}_b \mu_b(u) \text{ for } u = 1, 2, \dots, z. \quad (23)$$

Further, according to (3), the corresponding optimal unconditional multistate reliability function of the system is

$$\dot{R}(t, \cdot) = [1, \dot{R}(t, 1), \dots, \dot{R}(t, z)], \quad (24)$$

where

$$\dot{R}_n(t, \cdot) \cong \sum_{b=1}^v \dot{p}_b [\dot{R}(t, u)]^{(b)} \text{ for } t \geq 0, u = 1, 2, \dots, z, \quad (25)$$

and by (5) the optimal solutions for the mean values of the system unconditional lifetimes in the particular reliability states are of the form

$$\dot{\bar{\mu}}(u) = \dot{\mu}(u) - \dot{\mu}(u+1), u = 0, 1, \dots, z-1, \dot{\bar{\mu}}(z) = \dot{\mu}(z). \quad (26)$$

Moreover, considering (6) and (7), the corresponding optimal system risk function and the moment when the risk exceeds a permitted level δ , respectively are given by

$$\dot{r}(t) = 1 - \dot{R}(t, r) \text{ for } t \in (-\infty, \infty) \text{ and } \dot{\tau} = \dot{r}^{-1}(\delta). \quad (27)-(28)$$

3 OPTIMAL SOJOURN TIMES OF COMPLEX TECHNICAL SYSTEM OPERATION PROCESS

Replacing the limit transient probabilities p_b of the system operation process at the operation states by their optimal values $\dot{p}_b$ and the mean values M_b of the unconditional sojourn times at the operation states by their corresponding unknown optimal values $\dot{M}_b$ maximizing the mean value of the system lifetime in the reliability states subset $\{r, r+1, \dots, z\}$, we get the system of equations

$$\dot{p}_b = \frac{\pi_b \dot{M}_b}{\sum_{l=1}^v \pi_l \dot{M}_l}, b = 1, 2, \dots, v. \quad (29)$$

After simple transformations the above system takes the form

$$\begin{cases} (\dot{p}_1 - 1)\pi_1 \dot{M}_1 + \dot{p}_1 \pi_2 \dot{M}_2 + \dots + \dot{p}_1 \pi_v \dot{M}_v = 0 \\ \dot{p}_2 \pi_1 \dot{M}_1 + (\dot{p}_2 - 1)\pi_2 \dot{M}_2 + \dots + \dot{p}_2 \pi_v \dot{M}_v = 0 \\ \dots \\ \dot{p}_v \pi_1 \dot{M}_1 + \dot{p}_v \pi_2 \dot{M}_2 + \dots + (\dot{p}_v - 1)\pi_v \dot{M}_v = 0, \end{cases} \quad (30)$$

where $\dot{M}_b$ are unknown variables we want to find, $\dot{p}_b$ are optimal transient probabilities and π_b are steady probabilities.

Since the system of equations is homogeneous and it can be proved that the determinant of its main matrix is equal to zero, then it has nonzero solutions and moreover,

these solutions are ambiguous. Thus, if we fix some of the optimal values $\dot{M}_b$ of the mean values M_b of the unconditional sojourn times at the operation states, for instance by arbitrary fixing either one or multiple of them, we may find the values of the remaining ones and using this method arrive at the solution of this equation.

Another very useful and much easier applicable in practice tool that can help in planning the operation processes of complex technical systems are the system operation process optimal mean values of the total system operation process sojourn times $\hat{\theta}_b$ at the particular operation states z_b , $b = 1, 2, \dots, v$, during the fixed system operation time θ . They can be obtained by replacing the transient probabilities p_b at the operation states z_b with their optimal values $\dot{p}_b$. This results in the following expression

$$\dot{M}_b = \dot{E}[\hat{\theta}_b] = \dot{p}_b \theta, b = 1, 2, \dots, v. \quad (31)$$

The knowledge of the optimal values $\dot{M}_b$ of the mean values of the unconditional sojourn times and the optimal mean values $\dot{M}_b$ of the total sojourn times at the particular operation states during the fixed system operation time may be the basis for changing the complex technical systems operation processes in order to ensure that these systems operate both more reliably and more safely. This knowledge may also be useful in these systems operation cost analysis.

4 THE BULK CARGO TRANSPORTATION SYSTEM RELIABILITY AND RISK

The considered bulk cargo terminal placed at the Baltic seaside is designated for storage and reloading of bulk cargo, but its primary activity is loading bulk cargo on board the ships for export. There are two independent transportation systems: the system of reloading rail wagons and the system of loading vessels.

Cargo is brought to the terminal by trains consisting of self-discharging wagons, which are discharged to a hopper and then by means of conveyors transported into one of four storage tanks (silos). Loading of fertilizers from storage tanks on board the ship is done by means of special reloading system which consists of several belt conveyors and one bucket conveyor which allows the transfer of bulk cargo in a vertical direction. Researched system is a system of belt conveyors, referred to as the transport system.

In the conveyor reloading system we distinguish three bulk cargo transportation subsystems, the belt conveyors S_1 , S_2 and S_3 . The conveyor loading system is composed of six bulk cargo transportation subsystems, the dosage conveyor S_4 , the horizontal conveyor S_5 , the horizontal conveyor S_6 , the sloping conveyors S_7 , the dosage conveyor with buffer S_8 , the loading system S_9 .

The bulk cargo transportation subsystems are built, respectively:

- the subsystem S_1 : 1 rubber belt, 2 drums, set of 121 bow rollers, set of 23 belt supporting rollers,
- the subsystem S_2 : 1 rubber belt, 2 drums, set of 44 bow rollers, set of 14 belt supporting rollers,

- the subsystem S_3 : 1 rubber belt, 2 drums, set of 185 bow rollers, set of 60 belt supporting rollers,
- the subsystem S_4 : 3 identical belt conveyors, each composed of 1 rubber belt, 2 drums, set of 12 bow rollers, set of 3 belt supporting rollers,
- the subsystem S_5 : 1 rubber belt, 2 drums, set of 125 bow rollers, set of 45 belt supporting rollers,
- the subsystem S_6 : 1 rubber belt, 2 drums, set of 65 bow rollers, set of 20 belt supporting rollers,
- the subsystem S_7 : 1 rubber belt, 2 drums, set of 12 bow rollers, set of 3 belt supporting rollers,
- the subsystem S_8 : 1 rubber belt, 2 drums, set of 162 bow rollers, set of 53 belt supporting rollers,
- the subsystem S_9 : 3 rubber belts, 6 drums, set of 64 bow rollers, set of 20 belt supporting rollers.

Taking into account the operation process of the considered system we distinguish the following as its three operation states:

- an operation state z_1 – loading fertilizers from rail wagons on board the ship is done using $S_1, S_2, S_3, S_6, S_7, S_8$ and S_9 subsystems.
- an operation state z_2 – discharging rail wagons to storage tanks or hall when subsystems S_1, S_2 and S_3 , are used,
- an operation state z_3 – loading fertilizers from storage tanks or hall on board the ship is done by using S_4, S_5, S_6, S_7, S_8 and S_9 subsystems.

The limit values of the bulk cargo transportation systems operation process transient probabilities $p_b(t)$ at the operation states z_b , $b = 1, 2, 3$, determined in [5], on the basis of data coming from experts are

$$p_1 = 0.2376, p_2 = 0.6679, p_3 = 0.0945. \quad (32)$$

Further, assuming that the system is in the reliability state subset $\{u, u+1, \dots, z\}$ if all its subsystems are in this subset of reliability states, we conclude that the bulk cargo transportation system is a series system [4] of subsystems $S_1, S_2, S_3, S_6, S_7, S_8$ and S_9 .

Under the assumption that changes of the bulk cargo transportation system operation states have an influence on both the subsystem S_i reliability and the entire reliability structure [8], on the basis of expert opinions and statistical data given in [9], [10], the bulk cargo transportation system reliability structures and their components reliability functions at different operation states can be determined.

Additionally, we assume that subsystems S_i , $i = 1, 2, 3, \dots, 9$, are composed of four-state exponential components, with the reliability functions

$$[R_i(t, \cdot)]^{(b)} = [1, [R_i(t, 1)]^{(b)}, [R_i(t, 2)]^{(b)}, [R_i(t, 3)]^{(b)}], \\ t \in <0, \infty), b = 1, 2, 3, u = 1, 2, 3.$$

At the operation state z_1 , at loading of fertilizers from rail wagons on board the ship, the system is composed of seven non-homogenous series subsystems $S_1, S_2, S_3, S_6, S_7, S_8$, and S_9 forming a series structure.

The conditional reliability function of the system while it is at the operation state z_1 is given by

$$[R(t, \cdot)]^{(1)} = [1, [R(t, 1)]^{(1)}, [R(t, 2)]^{(1)}, [R(t, 3)]^{(1)}],$$

where

$$[R(t, u)]^{(1)} = [R_{147}(t, u)]^{(1)} [R_{61}(t, u)]^{(1)} [R_{248}(t, u)]^{(1)} [R_{88}(t, u)]^{(1)}$$

$$[R_{18}(t, u)]^{(1)} [R_{218}(t, u)]^{(1)} [R_{93}(t, u)]^{(1)} \text{ for } t \in <0, \infty), u = 1, 2, 3, \\ \text{i.e.}$$

$$[R(t, 1)]^{(1)} = \exp[-74.426t], [R(t, 2)]^{(1)} = \exp[-93.472t],$$

$$[R(t, 3)]^{(1)} = \exp[-150.206t]. \quad (33)-(35)$$

The expected values of the conditional lifetimes in the reliability state subsets at the operation state z_1 , calculated from the above result given by (33)–(35), are:

$$\mu_1(1) \cong 0.013, \mu_1(2) \cong 0.011, \mu_1(3) \cong 0.007 \text{ years}, \quad (36)$$

and further, using (5), it follows that the conditional lifetimes in the particular reliability states at the operation state z_1 are:

$$\bar{\mu}_1(1) \cong 0.002, \bar{\mu}_1(2) \cong 0.004, \bar{\mu}_1(3) \cong 0.007 \text{ years}.$$

At the operation state z_2 , i.e. at the state of discharging rail wagons to storage tanks or hall, the system is built of three subsystems S_1, S_2 and S_3 forming a series structure [4].

The conditional reliability function of the bulk cargo transportation system at the operation state z_2 is given by

$$[R(t, \cdot)]^{(2)} = [1, [R(t, 1)]^{(2)}, [R(t, 2)]^{(2)}, [R(t, 3)]^{(2)}],$$

where

$$[R(t, u)]^{(2)} = [R_{147}(t, u)]^{(2)} [R_{61}(t, u)]^{(2)} [R_{248}(t, u)]^{(2)}$$

$$\text{for } t \in <0, \infty), u = 1, 2, 3,$$

i.e.

$$[R(t, 1)]^{(2)} = \exp[-39.563t], [R(t, 2)]^{(2)} = \exp[-49.663t],$$

$$[R(t, 3)]^{(2)} = \exp[-64.280t]. \quad (37)-(39)$$

The expected values of the conditional lifetimes in the reliability state subsets at the operation state z_2 , calculated from the above result given by (37)–(39), are:

$$\mu_2(1) \cong 0.025, \mu_2(2) \cong 0.020, \mu_2(3) \cong 0.016 \text{ years}, \quad (40)$$

and further, using (5), it follows that the conditional lifetimes in the particular reliability states at the operation state z_2 are:

$$\bar{\mu}_2(1) \cong 0.005, \bar{\mu}_2(2) \cong 0.004, \bar{\mu}_2(3) \cong 0.016 \text{ years}.$$

At the operation state z_3 , i.e. at the loading of fertilizers from storage tanks or hall on board, the bulk cargo transportation system is built of six subsystems one series-parallel subsystem S_4 and five series subsystems S_5, S_6, S_7, S_8, S_9 forming a series structure [4].

The conditional reliability function of the system while it is at the operation state z_3 is given by

$$[R(t, \cdot)]^{(3)} = [1, [R(t, 1)]^{(3)}, [R(t, 2)]^{(3)}, [R(t, 3)]^{(3)}],$$

where

$$[R(t, u)]^{(3)} = [R_{3,18}(t, u)]^{(3)} \cdot [R_{1,73}(t, u)]^{(3)} \cdot [R_{8,8}(t, u)]^{(3)} \cdot [R_{1,8}(t, u)]^{(3)} \cdot [R_{2,18}(t, u)]^{(3)} \cdot [R_{9,3}(t, u)]^{(3)} \text{ for } t \in (-\infty, \infty), u = 1, 2, 3,$$

i.e.

$$[R(t, 1)]^{(3)} = \exp[-57.758t] - 3\exp[-55.007t] + 3\exp[-52.256t], (41)$$

$$[R(t, 2)]^{(3)} = \exp[-70.974t] - 3\exp[-68.018t] + 3\exp[-65.062t] (42)$$

$$[R(t, 3)]^{(3)} = \exp[-89.416t] - 3\exp[-86.140t] + 3\exp[-82.864t]. (43)$$

The expected values of the conditional lifetimes in the reliability state subsets at the operation state z_3 , calculated from the above result given by (41)–(43), are:

$$\mu_3(1) \cong 0.020, \mu_3(2) \cong 0.016, \mu_3(3) \cong 0.013 \text{ years}, (44)$$

and further, using (5), it follows that the conditional lifetimes in the particular reliability states at the operational state z_3 are:

$$\bar{\mu}_3(1) \cong 0.004, \bar{\mu}_3(2) \cong 0.003, \bar{\mu}_3(3) \cong 0.013 \text{ years}.$$

In the case when the system operation time is large enough, the unconditional reliability function of the bulk cargo transportation system is given by the vector

$$R(t, \cdot) = [1, R(t, 1), R(t, 2), R(t, 3)], t \geq 0,$$

where, according to (3) and after considering the values of $p_b, b = 1, 2, 3$, given by (32), its co-ordinates are as follows:

$$R(t, u) = p_1 \cdot [R(t, u)]^{(1)} + p_2 \cdot [R(t, u)]^{(2)} + p_3 \cdot [R(t, u)]^{(3)} (45)$$

for $t \geq 0, u = 1, 2, 3$, where $[R(t, u)]^{(1)}$ and $[R(t, u)]^{(2)}$ and $[R(t, u)]^{(3)}$ are respectively given by (33)–(35) and (37)–(39) and (41)–(43), i.e.

$$R(t, 1) = 0.6679 \exp[-39.563t] + 0.0945 \exp[-74.426t] + 0.2376[\exp[-57.758t] - 3\exp[-55.007t] + 3\exp[-52.256t]], (46)$$

$$R(t, 2) = 0.6679 \exp[-93.472t] + 0.0945 \exp[-49.663t] + 0.2376[\exp[-70.974t] - 3\exp[-68.018t] + 3\exp[-65.062t]], (47)$$

$$R(t, 3) = 0.6679 \exp[-150.206t] + 0.0945 \exp[-64.280t] + 0.0945[\exp[-89.416t] - 3\exp[-86.140t] + 3\exp[-82.864t]]. (48)$$

The mean values of the system unconditional lifetimes in the reliability state subsets, according to (4) are respectively:

$$\mu(1) \cong 0.016, \mu(2) \cong 0.013, \mu(3) \cong 0.009. (49)$$

The mean values of the system lifetimes in the particular reliability states, (5), are

$$\bar{\mu}(1) = \mu(1) - \mu(2) = 0.003, \bar{\mu}(2) = \mu(2) - \mu(3) = 0.004, \bar{\mu}(3) = \mu(3) = 0.009.$$

If the critical reliability state is $r = 2$, then the system risk function, according to (6), is given by

$$r(t) = 1 - [0.6679 \exp[-93.472t] + 0.0945 \exp[-49.663t] + 0.2376(\exp[-70.974t] - 3\exp[-68.018t] + 3\exp[-65.062t])] \text{ for } t \geq 0. (50)$$

Hence, the moment when the system risk function (Fig. 1) exceeds a permitted level, for instance $\delta = 0.05$, from (7), is

$$\tau = r^{-1}(\delta) \cong 0.000627 \text{ years}. (51)$$

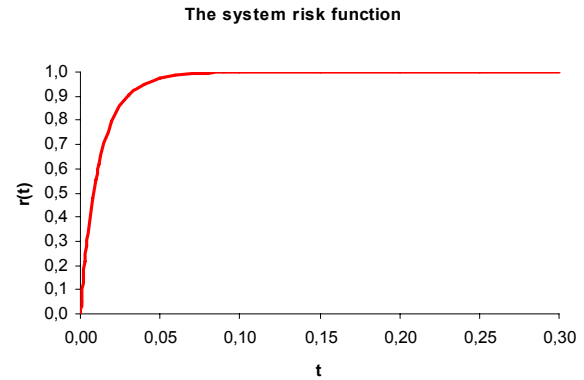


Figure 1 – The graph of the port bulk cargo transportation system risk function

5 OPTIMIZATION OF THE BULK CARGO TRANSPORTATION SYSTEM OPERATION PROCESS

In our case, as the critical state is $r = 2$, then considering the expression for $\mu(2)$, the objective function (8), takes the form

$$\mu(2) = p_1 \cdot 0.011 + p_2 \cdot 0.020 + p_3 \cdot 0.016 = 0.013 \text{ years}. (52)$$

The lower $\check{p}_b$ and upper $\hat{p}_b$ bounds of the unknown transient probabilities $p_b, b = 1, 2, 3$, coming from experts, respectively are [6]:

$$\check{p}_1 = 0.150, \check{p}_2 = 0.005, \check{p}_3 = 0.015,$$

$$\hat{p}_1 = 0.850, \hat{p}_2 = 0.120, \hat{p}_3 = 0.390.$$

Therefore, according to (9)–(10), we assume the following bound constraints

$$\sum_{b=1}^3 p_b = 1, 0.250 \leq p_1 \leq 0.850,$$

$$0.005 \leq p_2 \leq 0.150, 0.050 \leq p_3 \leq 0.550.$$

Now, before we find optimal values $\check{p}_b$ of the transient probabilities $p_b, b = 1, 2, 3$, that maximize the objective function (53), we arrange the system conditional lifetimes mean values $\mu_b(2), b = 1, 2, 3$, in non-increasing order $\mu_2(2) \geq \mu_3(2) \geq \mu_1(2)$.

Next, according to (11), we substitute

$$x_1 = p_2 = 0.0945, x_2 = p_3 = 0.2376, x_3 = p_1 = 0.6679, (53)$$

$$\begin{aligned}\tilde{x}_1 &= 0.005, \quad \tilde{x}_2 = 0.050, \quad \tilde{x}_3 = 0.250, \quad \hat{x}_1 = 0.150, \\ \hat{x}_2 &= 0.550, \quad \hat{x}_3 = 0.850,\end{aligned}\quad (54)$$

where $\tilde{x}_i$ and $\hat{x}_i$ are lower and upper bounds of the unknown limit transient probabilities x_i , $i=1,2,3$, respectively and we maximize with respect to x_i , $i=1,2,3$, the linear form (52) that according to (13) takes the form

$$\mu(2) = x_1 \cdot 0.011 + x_2 \cdot 0.020 + x_3 \cdot 0.016, \quad (55)$$

with the following bound constraints

$$\sum_{i=1}^3 x_i = 1,$$

$$0.005 \leq x_1 \leq 0.12, \quad 0.015 \leq x_2 \leq 0.390, \quad 0.150 \leq x_3 \leq 0.850. \quad (56)$$

According to (15)–(17), we calculate and fix the optimal solution that maximizes linear function (55) according to the rule (19). Namely, we get

$$\dot{x}_1 = \hat{x}_1 = 0.120, \quad \dot{x}^2 = \hat{x}^2 = 0.390,$$

$$\dot{x}_3 = 0.830 - 0.490 + 0.150 = 0.490.$$

Finally, according to (21) after making the inverse to (53) substitution, we get the optimal transient probabilities

$$\dot{p}_1 = \dot{x}_3 = 0.490, \quad \dot{p}_2 = \dot{x}_1 = 0.120, \quad \dot{p}_3 = \dot{x}_2 = 0.390, \quad (57)$$

that maximize the system mean lifetime in the reliability state subset $\{2,3\}$ expressed by the linear form (53) giving, according to (12) and (57), its optimal value

$$\begin{aligned}\dot{\mu}(2) &= \dot{p}_1 \cdot 0.011 + \dot{p}_2 \cdot 0.020 + \dot{p}_3 \cdot 0.016 = 0.49 \cdot 0.011 + \\ &+ 0.12 \cdot 0.020 + 0.39 \cdot 0.016 = 0.014.\end{aligned}\quad (58)$$

6 OPTIMAL RELIABILITY CHARACTERISTICS OF THE BULK CARGO TRANSPORTATION SYSTEM

Further, substituting the optimal solution (57) according to (24), we obtain the optimal solution for the mean value of the system unconditional lifetime in the reliability state subset $\{1,2\}$, $\{3\}$ that respectively amounts:

$$\dot{\mu}(1) \cong 0.0172, \quad \dot{\mu}(3) \cong 0.0104, \quad (59)$$

and according to (26), the optimal solutions for the mean values of the system unconditional lifetimes in the particular reliability states are

$$\dot{\mu}(1) \cong 0.0032, \quad \dot{\mu}(2) \cong 0.0036, \quad \dot{\mu}(3) \cong 0.0104.$$

Moreover, according to (24)–(25), the corresponding optimal unconditional multistate reliability function of the system is of the form

$$\dot{R}(t, \cdot) = [1, \dot{R}(t,1), \dot{R}(t,2), \dot{R}(t,3)] \text{ for } t \geq 0,$$

where according to (3) and after considering the values of $\dot{p}_b$, its co-ordinates are as follows:

$$\begin{aligned}\dot{R}(t,u) &\cong 0.49 \cdot [\dot{R}(t,u)]^{(1)} + 0.12 \cdot [\dot{R}(t,u)]^{(2)} + 0.39 \cdot [\dot{R}(t,u)]^{(3)} \\ &\text{for } t \geq 0, \quad u = 1,2,3,\end{aligned}\quad (60)$$

where $[\dot{R}(t,u)]^{(1)}$, $[\dot{R}(t,u)]^{(2)}$, $[\dot{R}(t,u)]^{(3)}$ are respectively given by (33)–(35) and (37)–(39), (41)–(43).

If the critical reliability state is $r = 2$, then the system risk function, according to (27), is given by

$$\begin{aligned}\dot{r}(t) &= 1 - \dot{R}(t,2) = 1 - [0.49 \exp[-93.472t] + 0.12 \exp[-49.663t] + \\ &+ 0.39(\exp[-70.974t] - 3 \exp[-68.018t] + 3 \exp[-65.062t])] \quad (61)\end{aligned}$$

where $\dot{R}(t,2)$ is given by (60) for $u = 2$.

Hence, considering (28), the moment when the optimal system risk function (Fig. 2) exceeds a permitted level, for instance $\delta = 0.05$, is

$$\tau = \dot{r}^{-1}(\delta) \cong 0.000676 \text{ years.} \quad (62)$$

Comparing the bulk cargo transportation system reliability characteristics after its operation process optimization given by (58)–(62) with the corresponding characteristics before this optimization determined by (45)–(51) justifies this action.

The optimal system risk function

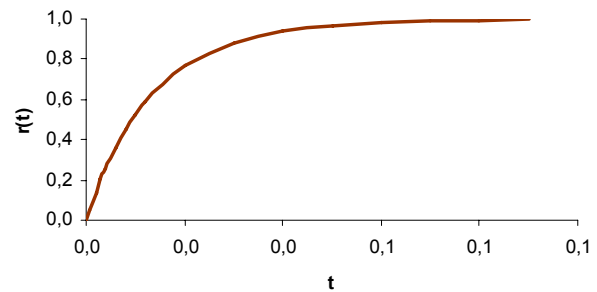


Figure 2 – The graph of the port bulk cargo transportation system optimal risk function

7 OPTIMAL SOJOURN TIMES OF BULK CARGO TRANSPORTATION SYSTEM OPERATION PROCESS AT OPERATION STATES

Having the values of the optimal transient probabilities determined by (57), it is possible to find the optimal conditional and unconditional mean values of the sojourn times of the bulk cargo transportation operation process at the operation states and the optimal mean values of the total unconditional sojourn times of the bulk cargo transportation system operation process at the operation states during the fixed operation time as well.

Substituting the optimal transient probabilities at operation states determined in (57) and the steady probabilities

$$\pi_1 = 0.315, \quad \pi_2 = 0.5, \quad \pi_3 = 0.185,$$

we get the following system of equations

$$\begin{cases} -0.16065\dot{M}_1 + 0.245\dot{M}_2 + 0.09065\dot{M}_3 = 0 \\ 0.0378\dot{M}_1 + (-0.44)\dot{M}_2 + 0.0222\dot{M}_3 = 0 \\ 0.12285\dot{M}_1 + 0.195\dot{M}_2 + (-0.11285)\dot{M}_3 = 0 \end{cases} \quad (63)$$

with the unknown optimal mean values $\dot{M}_b$ of the system unconditional sojourn times in the operation states. Consequently, we get

$$\dot{M}_1, \quad \dot{M}_2 = 0.154286\dot{M}_1, \quad \dot{M}_3 = 1.216216\dot{M}_1.$$

Thus, we may fix $\dot{M}_1$ and determine the remaining ones. In our case, after considering expert opinion, we conclude that it is sensible to assume

$$\dot{M}_1 \cong 2.$$

This way the obtained solutions of the system of equations, are

$$\dot{M}_1 \cong 2, \dot{M}_2 = 0.308571, \dot{M}_3 = 2.432432. \quad (64)$$

It can be seen that these solutions differ substantially from the values M_1, M_2, M_3 .

Other very useful and much easier to apply in practice tools that can help in planning the operation process of the technical system are the system operation process optimal mean values of the total sojourn times at the particular operation states during the fixed system operation time θ . Assuming the system operation time $\theta = 1$ year = 365 days, after applying (32), we get their values

$$\begin{aligned} \dot{M}_1 &= \dot{E}[\hat{\theta}_1] = \dot{p}_1\theta = 0.49 \cdot 365 \cong 179 \text{ days}, \\ \dot{M}_2 &= \dot{E}[\hat{\theta}_2] = \dot{p}_2\theta = 0.12 \cdot 365 = 44 \text{ days}, \\ \dot{M}_3 &= \dot{E}[\hat{\theta}_3] = \dot{p}_3\theta = 0.39 \cdot 365 = 142 \text{ days}. \end{aligned} \quad (65)$$

In practice, the knowledge of the optimal values of $\dot{M}_b$ and $\dot{M}_b$ given respectively by (64)–(65), is important and very helpful in planning and improving the operation process, as it allows for more reliable and safer system operation. From the performed analysis of the results of the bulk cargo transportation system operation process optimization it can be suggested to change the operation process characteristics that result in replacing (or the approaching/convergence to) the unconditional mean sojourn times M_b in the particular operation states before the optimization by their optimal values $\dot{M}_b$ after the optimization. The easiest way of reorganizing the system operation process leads to replacing (or the approaching/convergence to) the total sojourn times, $\hat{M}_b = E[\hat{\theta}_b]$, of the bulk cargo transportation system operation process, and in particular operation states during the operation time $\theta = 1$ year, with their optimal values $\dot{M}_b = \dot{E}[\hat{\theta}_b]$.

CONCLUSIONS

The joint model of reliability of complex technical systems at variable operation conditions linking a semi-markov modelling of the system operation processes with a multi-state approach to system reliability analysis was constructed. Next, the final results obtained from this joint model and linear programming were used to build the model of complex technical systems reliability optimization. These tools can be used in reliability evaluation and optimization of a very wide class of real technical systems operating at

varying conditions that influence their reliability structures and the reliability parameters of their components. The practical application of these tools to reliability and risk evaluation and optimization of a technical system of a bulk cargo transportation system, operating in variable conditions, and the results achieved can be implemented by reliability practitioners from both maritime transport industry and other industrial sectors.

REFERENCES

1. Aven T. Reliability evaluation of multi-state systems with multi-state components / T. Aven // IEEE Transactions on reliability. – 1985. – Vol. 34. – P. 473–479.
2. Grabski F. Semi-Markov Models of Systems Reliability and Operations / F. Grabski. – Warsaw : Systems Research Institute, Polish Academy of Science. – 2015.
3. Klabjan D. Existence of optimal policies for semi-Markov decision processes using duality for infinite linear programming / D. Klabjan, D. Adelman // Siam. J. Control Optim. – 2006. – Vol. 44(6). – P. 2104–2122. DOI: 10.1137/s0363012903437290
4. Kołowrocki K. Reliability of large and complex systems / K. Kołowrocki. – Elsevier, 2014. – 460 p. DOI: 10.1016/b978-0-08-099949-4.00010-6
5. Kołowrocki K. Modelling of operational processes of port bulk cargo system / K. Kołowrocki, B. Kwiatkowska-Sarnecka, J. Soszyńska // Proc. 2nd Summer Safety and Reliability Seminars – SSARS 2008. – Gdansk-Sopot. – 2008. – Vol. 2. – P. 217–222.
6. Kołowrocki K. Reliability and risk analysis of large systems with ageing components / K. Kołowrocki, B. Kwiatkowska-Sarnecka // Reliability Engineering & System Safety – 2008. – Vol. 93. – P. 1821–182. DOI: 10.1016/j.res.2008.03.008
7. Kołowrocki K. Reliability and availability of complex systems / K. Kołowrocki, J. Soszyńska // Quality and Reliability Engineering International. – 2006. – Vol. 22, Issue 1. – P. 79–99. DOI: 10.1002/qre.749
8. Kołowrocki K. Reliability and Safety of Complex Technical Systems and Processes: Modeling-Identification-Prediction-Optimization / K. Kołowrocki, J. Soszyńska-Budny. – London : Springer, 2011. – 405 p. DOI: 10.1007/978-0-85729-694-8
9. Kołowrocki K. Reliability and risk improvement with components quantitative and qualitative redundancy – bulk cargo terminal / K. Kołowrocki, B. Kwiatkowska-Sarnecka, J. Soszyńska // Proc. 9th Summer Safety and Reliability Seminars (SSARS 2015). – Gdansk-Sopot, 2015.
10. Kwiatkowska-Sarnecka B. Analysis of Reservation Efficiency in Series Systems : thesis ... doctor of philosophy / Kwiatkowska-Sarnecka Bożena. – Gdynia Maritime University, 2003.
11. Kwiatkowska-Sarnecka B. Reliability Improvement of Large Multi-state Series-parallel Systems / B. Kwiatkowska-Sarnecka // International Journal of Automation and Computing. – 2006. – Vol. 2. – P. 157–164. DOI: 10.1007/s11633-006-0157-y
12. Lisnianski A. Multi-state System Reliability. Assessment, Optimisation and Applications / A. Lisnianski, G. Levitin. – Singapore : World Scientific Publishing Co., 2003. – 376 p.
13. Meng F. Component-relevancy and characterization in multi-state systems / F. Meng // IEEE Transactions on reliability. – 1993. – Vol. 42. – P. 478–483. DOI: 10.1109/24.257834

Article was submitted 01.12.2015.

After revision 15.12.2015.

Коловровський К.¹, Квапішевська-Сарнецька Б.², Сошинська-Будный Й.³

¹Д-р наук, професор, професор кафедри математики, Морська Академія в Гдині, Гдиня, Польща

²Д-р філософії, доцент, доцент кафедри математики, Морська Академія в Гдині, Гдиня, Польща

³Д-р філософії, ад'юнкт кафедри математики, Морська Академія в Гдині, Гдиня, Польща

ОПТИМИЗАЦИЯ НАДЕЖНОСТИ И РИСКОВ СИСТЕМ С НЕСКОЛЬКИМИ УСТОЙЧИВЫМИ СОСТОЯНИЯМИ В ПРИЛОЖЕНИИ К ТРАНСПОРТНОЙ СИСТЕМЕ ПОРТА

Сложность процессов работы технических систем и их влияние на изменение во времени структур систем и параметров надежности их компонентов обуславливают сложности при первой встрече в реальности, а затем в фиксации и анализе этих структур и

параметров надійності. Путем построения объединенной модели надежности сложных технических систем в различных условиях эксплуатации, связывающей полумарковское моделирование процессов работы системы с подходом нескольких состояний в анализе надежности систем, мы находим основные характеристики надежности системы. Затем мы используем линейное программирование для того, чтобы построить модель оптимизации надежности сложных технических систем. Мы исследуем приложение модели в морском транспорте, в частности, в оптимизации надежности и рисков объемной системы грузоперевозок. Инструменты, разработанные нами, могут быть использованы для оценки надежности и оптимизации очень широкого класса реальных технических систем, работающих в различных условиях, которые влияют на их структуру надежности и параметры надежности их компонентов. Следовательно, разработанные нами инструменты могут быть использованы специалистами-практиками в области надежности как в отрасли морского транспорта, так и в других отраслях промышленности.

Ключевые слова: функция надежности, функция риска, рабочий процесс, оптимизация.

Коловровський К.¹, Квалішевська-Сарнецька Б.², Сошинська-Будний Й.³

¹Д-р наук, професор, професор кафедри математики, Морська Академія в Гдині, Гдиня, Польща

²Д-р філософії, доцент, доцент кафедри математики, Морська Академія в Гдині, Гдиня, Польща

³Д-р філософії, ад'юнкт кафедри математики, Морська Академія в Гдині, Гдиня, Польща

ОПТИМІЗАЦІЯ НАДІЙНОСТІ І РИЗИКІВ СИСТЕМ З КІЛЬКОМА СТІЙКИМИ СТАНАМИ У ЗАСТОСУВАННІ ДО ТРАНСПОРТНОЇ СИСТЕМИ ПОРТУ

Складність процесів роботи технічних систем та їхній вплив на зміну в часі структур систем і параметрів надійності їхніх компонентів обумовлюють складнощі при першій зустрічі у реальності, а потім у фіксації і аналізі цих структур і параметрів надійності. Шляхом побудови об'єднаної моделі надійності складних технічних систем в різних умовах експлуатації, що зв'язує напівмарковське моделювання процесів роботи системи з підходом декількох станів в аналізі надійності систем, ми знаходимо основні характеристики надійності системи. Потім ми використовуємо лінійне програмування для того, щоб побудувати модель оптимізації надійності складних технічних систем. Ми досліджуємо застосування моделі в морському транспорті, зокрема в оптимізації надійності та ризиків об'ємної системи вантажоперевезень. Інструменти, розроблені нами, можуть бути використані для оцінки надійності та оптимізації дуже широкого класу реальних технічних систем, що працюють в різних умовах, які впливають на їх структуру надійності і параметри надійності їхніх компонентів. Отже, розроблені нами інструменти можуть бути використані фахівцями-практиками в галузі надійності як у галузі морського транспорту, так і в інших галузях промисловості.

Ключові слова: функція надійності, функція ризику, робочий процес, оптимізація.

REFERENCES

1. Aven T. Reliability evaluation of multi-state systems with multi-state components, *IEEE Transactions on reliability*, 1985, Vol. 34, pp. 473–479.
2. Grabski F. Semi-Markov Models of Systems Reliability and Operations. Warsaw, Systems Research Institute, Polish Academy of Science, 2015.
3. Klabjan D., Adelman D. Existence of optimal policies for semi-Markov decision processes using duality for infinite linear programming, *Siam. J. Control Optim.*, 2006, Vol. 44(6), pp. 2104–2122. DOI: 10.1137/s0363012903437290
4. Kołowrocki K. Reliability of large and complex systems. Elsevier, 2014, 460 p. DOI: 10.1016/b978-0-08-099949-4.00010-6
5. Kołowrocki K., Kwiatkowska-Sarnecka B., Soszyńska J. Modelling of operational processes of port bulk cargo system, *Proc. 2nd Summer Safety and Reliability Seminars. – SSARS 2008*. Gdansk-Sopot, 2008, Vol. 2, pp. 217–222.
6. Kołowrocki K., Kwiatkowska-Sarnecka B. Reliability and risk analysis of large systems with ageing components, *Reliability Engineering & System Safety*, 2008, Vol. 93, pp. 1821–182. DOI: 10.1016/j.ress.2008.03.008
7. Kołowrocki K., Soszyńska J. Reliability and availability of complex systems, *Quality and Reliability Engineering International*, 2006, Vol. 22, Issue 1, pp. 79–99. DOI: 10.1002/qre.749
8. Kołowrocki K., Soszyńska-Budny J. Reliability and Safety of Complex Technical Systems and Processes: Modeling-Identification-Prediction-Optimization. London, Springer, 2011, 405 p. DOI: 10.1007/978-0-85729-694-8
9. Kołowrocki K., Kwiatkowska-Sarnecka B., Soszyńska J. Reliability and risk improvement with components quantitative and qualitative redundancy – bulk cargo terminal, *Proc. 9th Summer Safety and Reliability Seminars (SSARS 2015)*. Gdansk-Sopot, 2015.
10. Kwiatkowska-Sarnecka B. Analysis of Reservation Efficiency in Series Systems : thesis ... doctor of philosophy. Gdynia Maritime University, 2003.
11. Kwiatkowska-Sarnecka B. Reliability Improvement of Large Multi-state Series-parallel Systems, *International Journal of Automation and Computing*, 2006, Vol. 2, pp. 157–164. DOI: 10.1007/s11633-006-0157-y
12. Lisnianski A., Levitin G. Multi-state System Reliability. Assessment, Optimisation and Applications. Singapore, World Scientific Publishing Co., 2003, 376 p.
13. Meng F. Component-relevancy and characterization in multi-state systems, *IEEE Transactions on reliability*, 1993, Vol. 42, pp. 478–483. DOI: 10.1109/24.257834

МОДЕЛЬ ИНТЕЛЛЕКТУАЛЬНОЙ СИСТЕМЫ УПРАВЛЕНИЯ ГОРОДСКИМИ АВТОБУСНЫМИ ПЕРЕВОЗКАМИ

Решена задача разработки математического обеспечения для интеллектуальной системы управления городскими автобусными перевозками. Объектом исследования является автоматизация процесса принятия решений интеллектуальными системами управления городскими транспортными потоками. Предмет исследования составляют модели диспетчерского управления транспортными потоками автобусных маршрутов. Цель работы: совершенствование моделей для интеллектуальных транспортных систем управляющих городскими потоками автобусов и маршрутных такси. Разработана модель для интеллектуальной транспортной системы управления с учетом влияния наиболее значимых факторов на график движения автобусов по маршруту. Модель позволяет оперативно оценивать влияние возмущающих действий на движение подвижной единицы, в частности, переполнение пассажирами автобусов на маршруте, их сход с линии, отклонение от расписания и др., на показатели качества обслуживания, а также, оптимизировать расписание движения. В качестве критерия оптимизации предложен показатель минимума времени ожидания пассажирами автобусов и маршрутных такси на остановках.

В ходе экспериментов проверена адекватность разработанной модели, которая оценивалась методом однофакторного дисперсионного анализа и полнофакторного эксперимента в реальных городских условиях. Результаты экспериментов позволяют рекомендовать предложенную модель для практического использования в интеллектуальных транспортных системах управления городскими автобусными маршрутами.

Ключевые слова: интеллектуальная транспортная система, управление автобусными перевозками, модель, алгоритм управления.

НОМЕНКЛАТУРА

ВВ – возмущающее воздействие;
ИТС – интеллектуальная транспортная система;
КП – контрольный пункт;
КТС – контур транспортной связи;
ОП – остановочный пункт;
ПЕ – подвижная единица;
ТС – транспортное средство;
ЭВМ – электронная вычислительная машина;
 A_i – вместимость i -й подвижной единицы;
 d_i^j – потребность в перевозке с j -го ОП в момент прибытия i -й ПЕ;
 g_i^j – количество пассажиров, не обслуженных i -й ПЕ j -го ОП;
 K_i – коэффициент наполнения ПЕ;
 l^j – расстояние между j -м и $j+1$ -м остановочными пунктами;
 L – длина маршрута;
 m – количество подвижных единиц на маршруте;
 n – количество остановочных пунктов на маршруте;
 N – количество возможных вариантов распределения перевозок;
 $N_{КП}$ – номер конечного КП;
 P_i – суммарный пассажиропоток, приходящийся на i -ю ПЕ за рейс;
 P_i^j – количество пассажиров, пришедших на j -й ОП за время между прибытием i -й и $i-1$ -й ПЕ;
 $\tilde{P}_i^j$ – количество вошедших на ОП пассажиров;
 q_i^j – количество вышедших на ОП пассажиров;
 Q_i^j – наполнение i -й ПЕ после обслуживания j -го ОП;

r_i^j – резерв мест в i -й ПЕ на j -м ОП;
 S_i^j – средние затраты времени ожидания;
 t_{cj} – среднее время ожидания пассажиров на j -м временном отрезке при сходе на нем ПЕ;
 t_{nj} – среднее время ожидания пассажиров на j -м временном отрезке при переключении ПЕ на другой маршрут;
 t_i^j – время прибытия подвижной единицы на j -й остановочный пункт;
 $t_i^{начр(k)}$ – время начала k -го рейса;
 $t_{inn}^{КП}$ – время прибытия ПЕ на контрольный пункт;
 t_{nj} – среднее время ожидания пассажиров на j -м временном отрезке при работе «переключенной» на нем ПЕ;
 t_{oj} – среднее время ожидания пассажиров на j -м временном отрезке при движении ПЕ на нем с оперативным интервалом τ_o ;
 t_{pj} – среднее время ожидания пассажиров при вводе резервной ПЕ на j -м временном отрезке;
 t_{uj} – среднее время ожидания пассажиров на j -м временном отрезке при увеличении интервала движения ПЕ на нем на величину τ_u ;
 t_{yj} – среднее время ожидания пассажира на j -м временном отрезке при увеличении на нем пассажиропотока;
 t_{2j} – среднее время ожидания пассажиров на j -м временном отрезке без переключения ПЕ на другой маршрут;
 T – среднее время рейса;
 T_i – время рейса i -й ПЕ;
 $T_i^{\partial\theta}$ – время, затрачиваемое на проезд по перегонам между ОП;

T_i^{ocm} – суммарное время стоянок на ОП;

W_{in} – закономерности распределения перевозок по часам суток, заданные в виде весовых оценок;

Z – суммарные затраты времени для всех пассажиров на ожидание транспорта;

γ – допустимая величина уменьшения времени t_0 отстоя на конечном КП;

γ_{kj} – допустимая величина уменьшения планового времени τ_{kj} проезда на участке $k-j$ маршрута;

γ_{kN} – допустимая величина уменьшения планового времени проезда на участке $k-N$ маршрута (времени τ_{kN});

δ_k – опоздание ПЕ на ОП;

ΔJ – общее снижение времени ожидания пассажиров;

$\Delta \tau$ – интервал между ПЕ;

ε – допустимая величина увеличения времени t_0 отстоя на конечном КП;

ε_{kj} – допустимая величина увеличения времени τ_{kj} проезда на участке $k-j$;

ε_{kN} – допустимое увеличение планового времени проезда на участке $k-N$;

η_i – случайная величина, характеризующая интенсивность посадки и высадки для i -й ПЕ;

λ_j – интенсивность пассажиропотока на j -м временном отрезке;

v_i^j – случайная величина, характеризующая количество пассажиров, выходящих из i -й ПЕ на j -м ОП;

$v(U)$ – случайная величина, характеризующаяся законом распределения скорости движения ПЕ при данном признаке управления в ИТС;

v_{cp} – средняя скорость движения ПЕ;

ρ – суммарный результат пассажиропотока маршрута;

$\rho^j(t)$ – случайная величина, характеризующая плотность пассажиропотока на j -м ОП;

τ – интервал времени между приходами n ПЕ;

τ_{ni}^j – время, затрачиваемое на посадку и высадку пассажиров на j -м ОП i -й ПЕ;

τ_i^j – время движения между $i-1$ -й и i -й ПЕ после прохождения j -го ОП;

$\tau_{i-1,i}^j$ – интервал между прибытиями $i-1$ -й и i -й ПЕ;

$\tilde{\tau}_i^{j+1}$ – время, затраченное на движение по перегону маршрута между j -м и $j+1$ -м ОП;

τ_j – плановый момент прибытия на j -й КП;

$\tau_j(t_c)$ – часть j -го временного отрезка, перекрываемая временем схода ПЕ;

$\tau_j(t_n)$ – часть j -го временного отрезка, перекрываемая временем переключения ПЕ (временем τ_n);

$\tau_j(t_n)$ – часть j -го временного отрезка, на котором работает «переключенная» ПЕ;

$\tau_j(t_p)$ – часть j -го временного отрезка, заключенная между моментами схода ПЕ с линии и ввода резервной ПЕ;

$\tau_j(\tau_o)$ – часть j -го временного отрезка, на которой интервал движения ПЕ осуществляется с оперативным интервалом τ_o ;

$\tau_j(\tau_u)$ – часть j -го временного отрезка, на которой интервал движения ПЕ увеличен на τ_u ;

$\tau_z(U)$ – случайная величина, определяющая время задержки как функцию от признака управления в ИТС.

ВВЕДЕНИЕ

Транспортный комплекс больших городов образуют множество пассажирских и грузовых средств, управляющий ими персонал и дорожно-информационно-коммуникационная инфраструктура. Современные интеллектуальные транспортные системы (ИТС) базируются на использовании наукоемких инфокоммуникационных технологий, востребованных необходимостью повышения эффективности дорожного движения [1, 2].

Объектом исследования является автоматизация процесса принятия решений интеллектуальными системами управления городскими транспортными потоками в современных мегаполисах. Следует отметить, что автоматизированные системы управления движением, базирующиеся на детерминированных математических моделях, часто оказываются неадекватными реальным процессам перевозки пассажиров [1–6].

Предмет исследования составляют модели диспетчерского управления транспортными потоками автобусных маршрутов.

Цель работы – совершенствование математических моделей для интеллектуальных транспортных систем управляющих городскими потоками автобусов и маршрутных такси.

1 ПОСТАНОВКА ЗАДАЧИ

В моделировании транспортных потоков существуют две противоположные тенденции. С одной стороны, стремятся, как можно точнее описать исследуемый объект, чтобы добиться полной адекватности объекта и модели. А с другой, – упростить модель, привести ее к виду, удобному для нахождения решения одним из известных методов [3, 4, 6]. Однако, в большинстве задач управления городскими транспортными потоками приходится иметь дело со случайными величинами, что значительно усложняет математическую обработку моделей [7–9].

В этих случаях используется стохастическое моделирование. Его сущность состоит в том, что процессу функционирования объекта ставится в соответствие процесс вычисления функционала $\bar{y} = F(x)$, задаваемого математической моделью объекта.

Элементарным звеном функционирования системы является технологический процесс работы отдельной подвижной единицы (ПЕ), например, городского автобуса или маршрутного такси. Он состоит из выхода ПЕ на маршрут, работы на маршруте и заезда на стоянку. Работа на маршруте разбивается на выполнение рейсов и отстоя [10].

Основная задача ИТС управления движением ПЕ – это удовлетворение потребности пассажиров в поездках. Она решается за счет выполнения расписания, а также подключения в случае необходимости резервных транспортных средств. Критерием качества удовлетворения потребности в поездках является минимизация времени ожидания пассажирами автобуса или маршрутного такси.

2 ЛИТЕРАТУРНЫЙ ОБЗОР

В [6, 7] говорится, что включение ИТС в систему управления городскими перевозками требует от разработчика: 1) наполнить ее программами, связанными с технологией управления; 2) согласовать ее динамические характеристики с динамикой работы реальных объектов; 3) согласовать сигналы (например, от сенсоров движения, системы GPS и т.п.), поступающие от объекта к ИТС и выдаваемые ИТС на объект.

Если же алгоритм управления транспортным потоком не носит жесткого характера, если в процессе принятия решения о выдаче сигналов на объект управления решаются различные оптимизационные задачи, если, наконец, на первом этапе экспериментальной проверки ИТС происходит накопление новой информации, ранее неизвестной заказчику или разработчику, то применение ЭВМ становится оправданным [4, 8, 9].

С учетом вышесказанного, целью исследования является разработка и совершенствование моделей для ИТС управляющих городскими транспортными потоками.

3 МАТЕРИАЛЫ И МЕТОДЫ

Все нарушения движения транспорта – отклонения от расписания, переполнения, являющиеся следствием неравномерной скорости движения отдельных ПЕ и флуктуаций пассажиропотока, – вызывают увеличение времени ожидания. Поэтому и эффективность управляющих воздействий имеет смысл оценивать по тому же критерию.

Управляющие воздействия данного уровня можно разделить на две категории. Первые из них вызваны нарушениями расписания отдельными ПЕ и имеют индивидуальный характер: увеличение скорости (нагон в пути), сокращение отстоя и т.п. Во вторую категорию входят воздействия, вызванные как индивидуальными, так и групповыми нарушениями, сходами, ухудшениями условий движения, выполняемые группой ПЕ или резервными ПЕ: раздвижка интервала, вход резерва, переход на оперативный интервал и т.д. Функцию управления на этом уровне может выполнять ИТС. Для этого необходим объем информации о состоянии транспортной городской системы. При имеющемся контуре транспортной связи (КТС) информация в ИТС поступает с каждой ПЕ в контрольных точках маршрута [10].

Модель функционирования маршрута для разрабатываемой ИТС описывается следующей системой уравнений. Система моделирует временные характеристики движения подвижной единицы (ПЕ) по маршруту и процесс пассажирообразования:

$$\left. \begin{aligned} t_i^1 &= t_i^{nqp(k)}; \\ Q_i^0 &= 0; \\ q_i^1 &= 0; \\ v_i^j &= \frac{i}{n} \eta_i \rho^j(t); \\ q_i^j &= v_i^j Q_i^{j-1}; \\ r_i^j &= A_i - (Q_i^{j-1} - q_i^j); \\ P_i^j &= \int_{t_{i-1}^j}^{t_i^j} \rho^j(t) dt; \\ d_i^j &= P_i^j + g_{i-1}^j; \\ Q_i^j &= \min \{A_i, A_i + d_i^j - r_i^j\}; \\ g_i^j &= \max \{0, d_i^j - r_i^j\}; \\ \tilde{P}_i^j &= d_i^j - g_i^j; \\ q_i^n &= Q_i^{n-1}; \\ P_i^n &= 0; \\ \tau_{ni}^j &= \eta_i (\tilde{P}_i^j + q_i^j); \\ \tau_{zi}^j &= \tau_z(U); \\ \tau_i^j &= \begin{cases} \tau_{ni}^j, & \text{если } \tau_{zi}^j \leq \tau_{ni}^j, \\ \tau_{zi}^j, & \text{если } \tau_{zi}^j > \tau_{ni}^j; \end{cases} \\ v_i^j &= v(U); \\ \tilde{\tau}_i^{j+1} &= l^j / v_i^j; \\ t_i^{j+1} &= t_i^j + \tilde{\tau}_i^{j+1} + \tau_i^j; \\ \tau_i'^j &= t_i^j - t_{i-1}^j. \end{aligned} \right\} i = \overline{1, m}; j = \overline{1, n}, (1)$$

Исходными данными для моделирования являются величины $m; n; l^j; t_i^{nqp(k)}; A_i; v(U); \tau_z(U); \eta_i; \rho^j(t)$. Остальные величины вычисляются рекурсивно в процессе функционирования модели. Случайные величины η_i и $\rho^j(t)$ определяются в результате проведения обследования пассажиропотоков.

Характеристики плотности пассажиропотока формируются на основании суточных объемов перевозок на маршруте A^j и известных общих закономерностей распределения перевозок по часам суток, заданных в виде весовых оценок W_{in} , таких, что $\sum_i W_{in} = 1, i = \overline{1, T}, n = \overline{1, N}$.

Для вычисления величины W_{in} можно использовать алгоритм, описанный в [4, 7].

Моделирование системы сбора информации заключается в присвоении некоторым ОП признаков контрольного пункта (КП). При переходе модели в состояние, в котором i -я ПЕ находится на КП, вычисляются

величины: t_{inn}^{KP} ; $K_i = 25\%$; (здесь $[.]$ – «целая часть»), являющийся аналогом информации, получаемой ИТС при данном КТС, который вместе с номерами ПЕ и КП поступают на вход модели ДУ движением ПЕ.

Цель ИТС – обеспечить выполнение расписания движения с минимальными отклонениями от запланированного с помощью использования соответствующих диспетчерских воздействий. Поэтому алгоритмы ИТС ориентированы на выбор диспетчерского воздействия (ДВ), компенсирующих возмущающие воздействия (ВВ). Выбор оптимальных алгоритмов управления движением является важнейшей задачей построения ИТС.

Входной информацией для модели ИТС служат плановое расписание движения, а также информация, поступающая из модели маршрута: момент прохода подвижной единицы КП, коэффициенты наполнения ПЕ, сообщения о сходах ПЕ с линии. В [9, 10] установлено взаимное соответствие между видами нарушений движения и управляющими воздействиями, которые могут быть направлены на восстановление движения.

Как видно из данных [2, 4, 6, 8, 9], одни и те же нарушения могут устраняться различными управляющими воздействиями. Важнейшей задачей моделирования является количественный анализ управления движением. Модель позволяет оценивать качество управления по выбранному критерию.

Основным критерием качества удовлетворения потребности в поездках принимаем время ожидания пассажирами транспорта. Структура модели позволяет непосредственно оценивать эту величину. Заметим, что в реальных условиях возможна только приблизительная оценка времени ожидания пассажирами транспорта.

Будем считать, что поток пассажиров на остановке является пуассоновским с параметром ρ . Тогда за время τ на остановку в среднем приходят $\rho\tau$ пассажиров. Время ожидания пассажиром автобуса (ПЕ) в среднем равно половине интервала ожидания. Тогда суммарные затраты времени для всех пассажиров на ожидание транспорта, собравшихся за время τ , $Z = \rho\tau^2 / 2$.

Время τ сбора пассажиров принимается равным интервалу времени между прибытиями на остановку двух соседних ПЕ. Если пассажир не был обслужен подошедшей ПЕ, то время ожидания им прихода следующей равно уже не половине, а интервалу между ПЕ. Поэтому суммарные затраты времени на j -м ОП за интервал $\tau_{i-1,i}^j$ между прибытиями $i-1$ -й и i -й ПЕ

$$Z_i^j = P_i^j \frac{\tau_{i-1,i}^j}{2} + g_{i-1}^j \tau_{i-1,i}^j. \quad (2)$$

Средние затраты времени ожидания S_i^j вычисляются по формуле

$$S_i^j = Z_i^j / P_i^j. \quad (3)$$

На основании выражения (2) можно определить затраты для каждой ПЕ за рейс

$$Z_i = \sum_{j=1}^n Z_i^j, \quad (4)$$

$$S_i = \sum_{j=1}^n S_i^j \quad (5)$$

и для каждого ОП за интервал времени между приходами n ПЕ:

$$\tau = t_{K+n}^j - t_K^j; \quad (6)$$

$$Z^j = \sum_{i=K}^{K+n} Z_i^j, \quad (7)$$

$$S^j = \sum_{i=K}^{K+n} S_i^j. \quad (8)$$

Величины Z_i^j могут вычисляться для всех i, j в процессе функционирования модели по формуле (2).

Рассмотрим влияние возмущающих воздействий на величину параметров Z и S .

Отклонение ПЕ от расписания. Допустим, что на достаточно коротком временном интервале плотность пассажиропотока постоянна, $\rho^j(t) = \rho^j$. Тогда

$$P_i^j = \int_{t_{i-1}^j}^{t_i^j} \rho^j dt = \rho^j (t_i^j - t_{i-1}^j) = \rho^j \tau_{i-1,i}^j; \quad (9)$$

$$P_i = \sum_{j=1}^n P_i^j = \sum_{j=1}^n \rho^j \tau_{i-1,i}^j.$$

Пусть на маршруте m ПЕ, а среднее время рейса T . Для ПЕ задано расписание с равными интервалами движения. При этом $\Delta\tau = T / m$ – интервал между ПЕ. Будем считать, что расписание выполняется идеально, $\tau_{i-1,i}^j = \Delta\tau$ для всех $i = \overline{1, m}$. Тогда

$$P_i = \sum_{j=1}^n \rho^j \Delta\tau = \Delta\tau \sum_{j=1}^n \rho^j = \rho \Delta\tau, \quad (10)$$

где $\rho = \sum_{j=1}^n \rho^j$.

Из выражения (10) видно, что все P_i равны между собой и при идеальном выполнении расписания пассажиропоток разделен между ПЕ. Будем считать, что ПЕ достаточно для обслуживания маршрута и при равномерном распределении пассажиропотока переполнений не возникало. Тогда

$$Z_i^j = P_i^j \frac{\tau_{i-1,i}^j}{2} = \rho^j \frac{(\Delta\tau)^2}{2}; \quad (11)$$

$$Z_i = \sum_{j=1}^n \rho^j \frac{(\Delta\tau)^2}{2} = \rho \frac{(\Delta\tau)^2}{2}; \quad (12)$$

$$Z^j = \sum_{i=1}^m \rho^j \frac{(\Delta\tau)^2}{2} = m\rho^j \frac{(\Delta\tau)^2}{2}; \quad (13)$$

$$Z = m\rho \frac{(\Delta\tau)^2}{2}, \quad (14)$$

где $Z = \sum_{j=1}^n Z^j$.

Пусть i -я ПЕ отклоняется от расписания на величину δ_i . Тогда

$$\tau_{i-1,i} = \Delta\tau - \delta_{i-1} + \delta_i; P_i = \rho\Delta\tau + \rho(\delta_i - \delta_{i-1}); \quad (15)$$

$$Z_i = \rho \frac{(\Delta\tau + \delta_i - \delta_{i-1})^2}{2} = \rho \frac{(\Delta\tau)^2}{2} + \rho\Delta\tau(\delta_i - \delta_{i-1}) + \rho \frac{(\delta_i - \delta_{i-1})^2}{2}; \quad (16)$$

$$Z^j = \sum_{i=1}^m \rho^j \frac{(\Delta\tau + \delta_i - \delta_{i-1})^2}{2};$$

$$Z = \sum_{j=1}^n Z^j = m\rho \frac{(\Delta\tau)^2}{2} + \rho \sum_{i=1}^m \frac{(\delta_i - \delta_{i-1})^2}{2}. \quad (17)$$

Если мы обозначим P_i и Z в формулах (10) и (14) через P'_i и Z' , а в формулах (15) и (17) через P''_i и Z'' , то получим

$$\Delta P_i = \rho\Delta\tau + \rho(\delta_i - \delta_{i-1}) - \rho\Delta\tau = P''_i - P'_i = \rho(\delta_i - \delta_{i-1}), \quad (18)$$

при нарушении регулярности движения появляется неравномерность в распределении пассажиропотока между ПЕ даже при постоянной плотности, что может привести к переполнениям

$$\Delta Z = Z'' - Z' = \rho \sum_{i=1}^m \frac{(\delta_i - \delta_{i-1})^2}{2} \geq 0. \quad (19)$$

Таким образом, время ожидания пассажиров при нарушении расписания возрастет. Z будет минимальным, если $\delta_i = \delta_{i-1}$ для всех i из выражения (19) или $\delta_i = 0$.

Средние затраты времени ожидания ПЕ пассажирами

$$\Delta S = \frac{\Delta Z}{P} = \frac{\rho \sum_{i=1}^m \frac{(\delta_i - \delta_{i-1})^2}{2}}{m\rho\Delta\tau} = \sum_{i=1}^m \frac{(\delta_i - \delta_{i-1})^2}{2m\Delta\tau} =$$

$$= \sum_{i=1}^m \frac{(\delta_i - \delta_{i-1})^2}{2T} \geq 0. \quad (20)$$

На основании анализа выражений (19) и (20) можно сделать вывод, что для повышения качества обслуживания пассажиров на маршрутах движения необходим равномерный интервал между ПЕ.

Предположим, что ПЕ прибыла на k -й КП в момент времени t_k с опозданием на время δ_k . Если величина времени опоздания удовлетворяет условию $\delta_k = \gamma_{kj}\tau_{kj}$, то можно формировать управляющее воздействие «на-

гон на участке $k-j$ маршрута» ($j = k+1, k+2, \dots, N_{КП} - 1$). При этом необходимо осуществлять корректировку плановых моментов проследования КП:

$$\tau_j^* = \begin{cases} \tau_j + \delta_k - \gamma_{kj}\tau_{kj}, & \text{если } (\delta_k - \gamma_{kj}\tau_{kj}) > 0; \\ \tau_j & \text{в противном случае.} \end{cases}$$

Если величина времени опоздания такова, что за счет увеличения скорости движения ПЕ обеспечить прибытие ее на конечный КП в плановый момент времени не представляется возможным, можно осуществить сокращение времени отстоя ПЕ на конечном КП. При этом необходимым условием формирования данного управляющего воздействия является $\delta_k \leq \gamma_{kN}\Phi_{kN} + \gamma_{t0}$.

Если величина опоздания такова, что за счет увеличения скорости движения и уменьшения времени отстоя ТЕ на конечном КП обеспечить отправление в очередной рейсов плановый момент времени не представляется возможным, допустимо формировать управляющие воздействия типа: «увеличение времени отстоя на КП» (так, чтобы ПЕ была отправлена в очередной рейс в плановый момент); «укороченный рейс» (такой, чтобы ПЕ закончила его к плановому моменту начала очередного рейса); «удлиненный рейс» (такой, чтобы ПЕ, пропустив один обычный рейс, вернулась из удлиненного к плановому моменту начала следующего рейса).

Предположим, что ПЕ прибыла на k -й КП в момент времени t_k с опережением планового момента прибытия τ_k на величину δ_k . Если величина $\delta_k \leq \varepsilon_{kj}\tau_{kj}$, то следует формировать управляющее воздействие «увеличение времени проезда». При этом необходимо осуществлять корректировку плановых моментов проследования контрольных пунктов:

$$\tau_j^* = \begin{cases} \tau_j - (\delta_k - \gamma_{kj}\tau_{kj}), & \text{если } (\delta_k - \gamma_{kj}\tau_{kj}) > 0; \\ \tau_j & \text{в противном случае.} \end{cases}$$

Если величина времени опережения такова, что за счет уменьшения скорости движения ПЕ обеспечить прибытие ее на конечный КП в плановый момент времени невозможно, то требуется увеличить время отстоя ПЕ на конечном КП. При этом необходимым условием формирования данного управляющего воздействия является

$$\delta_k \leq \varepsilon_{kN}\Phi_{kN} + \varepsilon_{t0}.$$

Сход ПЕ с линии. Пусть все предложения предыдущего пункта остаются в силе. Из формулы (10) следует, что для всех i

$$P_i = \rho\Delta\tau = \frac{\rho\tau}{m}. \quad (21)$$

Предположим, что i -я ПЕ сошла с линии. Тогда

$$P'_{i+1} = P_i + P_{i+1} = 2P_i, \quad (22)$$

пассажиропоток, приходящейся на следующую за сошедшей ПЕ, увеличивается в два раза.

Потери времени на ожидание двух ПЕ

$$Z' = Z_i + Z_{i+1} = \rho \frac{(\Delta\tau)^2}{2} + \rho \frac{(\Delta\tau)^2}{2} = \rho\Delta\tau^2. \quad (23)$$

Потери времени после схода ПЕ

$$Z'' = \rho \frac{2(\Delta\tau)^2}{2} = 2\rho\Delta\tau^2; \quad (24)$$

$$\Delta Z = Z'' - Z' = \rho\Delta\tau^2 > 0. \quad (25)$$

Как видим, и в этом случае ВВ влияют на увеличение критерия Z . Учитывая условия, приведенные в выражении (22), можно сказать, что возникает возможность переполнения на маршруте. В данном случае ИТС формирует следующие управляющие воздействия: «уменьшение времени рейса», «раздвижка интервала», «переход на равномерный интервал», «переключение с маршрута на маршрут» и «ввод резервной ПЕ». Вначале оценим эффективность применения ИТС управляющих воздействий «ввод резервной ПЕ», «раздвижка интервала» и «переход на равномерный интервал». Предположим, что на маршруте выделено пять временных отрезков относительно постоянной интенсивности пассажиропотока (два «пиковых» и три «межпиковых»): $t_0 - t_1, t_1 - t_2, \dots, t_4 - t_5$. В момент времени t_c зафиксирован сход ПЕ с линии. Работоспособное состояние ПЕ будет восстановлено через время τ_e , в момент времени t_e . Через время τ_p в момент времени t_p может быть осуществлен ввод резервной ПЕ.

Суммарное время ожидания пассажиров на маршруте за время τ_e при сходе ПЕ с линии составит

$$J_c = \sum_{j=1}^5 \lambda_j \tau_j ((t_c) t_{cj}).$$

Суммарное время ожидания пассажиров на маршруте за время τ_e при вводе резерва составит

$$J_p = \sum_{j=1}^5 \lambda_j \tau_j (t_p) t_{pj} + \sum_{j=1}^5 \lambda_j \tau_j (t_p) t_{cj}. \text{ Тогда снижение}$$

ΔJ_p суммарного времени ожидания пассажиров за счет ввода резерва при сходе ПЕ с линии определится в виде $\Delta J_p = J_c - J_p$.

Учитывая, что $\tau_j(t_c) = \tau_j(\tau_p + \tau_j(t_p))$, получим следующее выражение для ΔJ_p :

$$\Delta J_p = \sum_{j=1}^5 \lambda_j \tau_j (t_p) t_{cj} - \sum_{j=1}^5 \lambda_j \tau_j (t_p) t_{pj} = \sum_{j=1}^5 \lambda_j \tau_j (t_p) (t_{cj} - t_{pj}).$$

Проведя аналогичный анализ, можно показать, что снижение ΔJ_u суммарного времени ожидания пассажиров за счет раздвижки интервала при сходе ПЕ с линии определяется так $\Delta J_u = \sum_{j=1}^5 \lambda_j \tau_j (\tau_u) (t_{cj} - t_{uj})$. Аналогично

можно показать, что снижение ΔJ_o суммарного времени ожидания пассажиров за счет перехода на равномерный или оперативный интервал определяется в виде

$$\Delta J_o = \sum_{j=1}^5 \lambda_j \tau_j (\tau_o) (t_{cj} - t_{oj}).$$

Перейдем теперь к оценке эффективности применения управляющего воздействия «переключение ПЕ с маршрута на маршрут». Предположим, что анализируется возможность переключения ПЕ с одного маршрута, интенсивность пассажиропотока на котором λ_1 , на другой, где интенсивность пассажиропотока λ_2 в случае схода ПЕ с первого маршрута в момент времени t_c на время τ_e . Начало переключения ПЕ со второго маршрута на первый осуществляется в момент времени t_n . При этом переключаемая ПЕ прибывает на первый маршрут в момент t_n .

Переключение ПЕ с одного маршрута на другой можно отождествлять с вводом резервной ПЕ на первый маршрут. Поэтому снижение ΔJ_n суммарного времени ожидания пассажиров на первом маршруте за счет переключения ПЕ с маршрута на маршрут определится

$$\text{по аналогии с } \Delta J_p: \Delta J_n = \sum_{j=1}^5 \lambda_j \tau_j (t_n) (t_{cj} - t_{nj}). \text{ Пере-}$$

ключение ПЕ со второго маршрута равносильно сходу с него ПЕ, что вызывает увеличение времени ΔJ_1 ожида-

$$\text{ния пассажиров: } \Delta J_1 = \sum_{j=1}^5 \lambda_{2j} \tau_j (t_n) t_{nj} - \sum_{j=1}^5 \lambda_{2j} \tau_j t_{2j}.$$

Перевод ПЕ с маршрута на маршрут будет оправдан только в том случае, если снижение времени ожидания пассажиров на новом маршруте будет больше увеличения времени ожидания пассажиров на прежнем маршруте (в рассматриваемом случае при $\Delta J_n > \Delta J_1$). Общее снижение ΔJ времени ожидания пассажиров $\Delta J = \Delta J_n - \Delta J_1$.

При формировании управляющего воздействия «переключение ПЕ с маршрута на маршрут» в случае схода ПЕ с линии необходимо определить такую пару маршрутов, для которой обеспечивается максимум общего снижения времени ожидания пассажиров – $\max \Delta J$. Выбор наиболее эффективного управляющего воздействия при сходе ПЕ с линии может быть осуществлен при определении максимального снижения ΔJ_c суммарного времени ожидания пассажиров: $\Delta J_c = \max(\Delta J_p, \Delta J_u, \Delta J_o, \max \Delta J)$.

Таким образом, при каждом сходе ПЕ с линии целесообразно формировать такое управляющее воздействие, которое обеспечивает максимальное снижение времени ожидания пассажиров (по сравнению со случаем отсутствия управляющих воздействий).

Ухудшение условий движения. Рассмотрим соотношение между интервалом движения и другими параметрами движения: $T_i = T_i^{\text{дв}} + T_i^{\text{огм}}$, где

$$T_i^{\text{дв}} = \sum_{j=1}^n \tau_i^{j,j+1} = \sum_{j=1}^n l^{j,j+1} / v_i^{j,j+1} = \frac{1}{v_{cp}} \sum_{j=1}^n l^{j,j+1} = L / v_{cp}, \quad (26)$$

$$T_i^{\text{огм}} = \sum_{j=1}^n \tau_{ni}^j = \sum_{j=1}^n \tau_{hi}^j + \sum_{j=1}^n \tau_{zi}^j; \quad (27)$$

$$\sum_{j=1}^n \tau_{hi}^j = \sum_{j=1}^n \eta_i (\tilde{P}_i^j + q_i^j) = \eta_i (\sum_{j=1}^n \tilde{P}_i^j + \sum_{j=1}^n q_i^j) = 2\eta_i \sum_{j=1}^n \tilde{P}_i^j = 2\eta_i \rho \Delta\tau.$$

Следовательно, $T_i = \frac{L}{v_{cp}} + 2\eta_i \rho \Delta\tau + T_{zi} \approx T$; $\Delta\tau = \frac{T}{m}$;

$$m\Delta\tau = L/v_{cp} + 2\eta_i \rho \Delta\tau + T_{zi}.$$

Отсюда

$$\Delta\tau = \frac{L/v_{cp} + T_{zi}}{m - 2\eta_i \rho}. \quad (28)$$

Из отношения

$$Z = m\rho \frac{(\Delta\tau)^2}{2} \quad (29)$$

видно, что с увеличением $\Delta\tau$ растет и Z . При ухудшении условий движения v_{cp} понижается, $v'_{cp} < v_{cp}$.

Тогда

$$\delta = \Delta\tau' - \Delta\tau = \frac{L}{m - 2\rho\eta_i} \frac{\Delta v}{(v_{cp} - \Delta v)}; \quad \Delta v = v_{cp} - v'_{cp};$$

$$\Delta Z = m\rho \frac{(\Delta\tau + \delta)^2}{2} - m\rho \frac{(\Delta\tau)^2}{2} = m\rho \Delta\tau \delta + m\rho \frac{\delta^2}{2} > 0. \quad (30)$$

При ухудшении условий движения на отдельных маршрутах или маршрутной сети в целом, в качестве основного управляющего воздействия в ИТС рекомендуется управление «переход на равномерный интервал». С учетом возможной скорости движения ПЕ рассчитывается интервал, соблюдая который ПЕ отправляются с конечных КП.

Переполнения на маршруте фиксируется, если не все пассажиры могут быть обслужены прибывшей на ОП подвижной единицей, когда $g_i^j > 0$. Возникнув на одной остановке, переполнение может появиться и на других за счет уменьшения резерва мест в ПЕ. При переполнениях

$$Z_i^j = \rho^j \frac{(\Delta\tau)^2}{2} + g_i^j \Delta\tau; \quad (31)$$

$$\Delta Z_i^j = g_i^j \Delta\tau > 0; \quad (32)$$

$$\Delta Z = \sum_{g_i > 1} \Delta Z_i^j > 0. \quad (33)$$

Сумма берется по всем i, j для которых $g_i^j > 0$.

Внеплановое увеличение пассажиропотоков ведет к переполнению ПЕ, к снижению комфортабельности поездок и увеличению временных затрат пассажиров. Основными управляющими воздействиями при этом являются «переключение ПЕ с маршрута на маршрут» и «ввод резервной ПЕ». Целесообразно формировать такое управляющее воздействие, которое обеспечит максимальное снижение времени ожидания пассажиров (по сравнению со случаем отсутствия управляющих воздействий).

Предположим, на части j -го временного отрезка $\tau_j(t_y)$ в момент t_y произошло внеплановое увеличение λ_j на величину $\Delta\lambda_j$. Тогда увеличение суммарного времени ожидания пассажиров на данном маршруте

$$\Delta J_y = \sum_{j=1}^5 \Delta\lambda_j \tau_j(t_n) t_{yj}.$$

Если на данный маршрут переключить ПЕ, то ΔJ_y уменьшится на ΔJ_n , на величину снижения суммарного времени ожидания пассажиров за счет переключения ПЕ с маршрута на маршрут. Если ввести резервную ПЕ, то величина ΔJ_y уменьшится на ΔJ_p , на величину снижения суммарного времени ожидания пассажиров за счет ввода резервной подвижной единицы.

Выбор типа управляющего воздействия при внеплановом увеличении пассажиропотока на маршруте может быть осуществлен при определении максимального уменьшения величины ΔJ_y : $J_y = \max\{(\Delta J_y - \Delta J_n), (\Delta J_y - \Delta J_p)\}$.

Таким образом, установлено, что все возмущающие воздействия ухудшают показатели качества обслуживания пассажиров. На описанной модели – это влияние можно подвергнуть количественной оценке. Моделируя ВВ и вычисляя величины Z и S до нарушения движения и после принятия диспетчерского воздействия, можно определить скорость реакции системы на ВВ и ДВ; эффективность ДВ, а в конечном результате – непосредственный экономический эффект от внедрения ИТС. Все указанные вычисления, а также статистический анализ моделирования производит блок анализа.

4 ЭКСПЕРИМЕНТЫ

Предложенная модель реализована в MatLab. В ходе имитационного моделирования исследовались следующие параметры: η_i ; $\tilde{P}_i^j + q_i^j$; $\tilde{\tau}_i^{j+1}$; $\Delta\tau'$. Для проведения имитационного эксперимента была составлена рабочая таблица планирования в соответствии рекомендациями [8, 10].

5 РЕЗУЛЬТАТЫ

Результаты имитационного моделирования пассажиропотока показаны в таблице 1.

Таблица 1 – Результаты имитационных экспериментов

Номер опыта	η_i	$\tilde{P}_i^j + q_i^j$	$\tilde{\tau}_i^{j+1}$	$\Delta\tau'$
1	0,2	1	1	0,34
2	0,8	14	1	0,17
3	0,2	1	11	2
4	0,8	14	11	1
5	0,5	7,5	6	-0,68

Экспериментальное значение критерия Фишера $K_{fe} = 8,0 < K_{fm} = 16$ при $P = 0,95$, позволило сделать вывод об адекватности разработанной модели. Этот факт дает возможность обоснованно утверждать, что стохастическая имитационная модель является базой для внедрения управленческих действий для создаваемых ИТС управления пассажирским автотранспортном.

6 ОБСУЖДЕНИЕ

Имитационное моделирование с использованием пакета MatLab позволило сделать следующие выводы:

- предложенная математическая модель является работоспособной и позволяет адекватно описывать процесс движения подвижной единицы по маршруту;
- использование пакета MatLab в реальной практике эксплуатации ИТС не представляется возможным, поскольку его применение предполагает определенную квалификацию пользователя.

ВЫВОДЫ

Научная новизна результатов, полученных в статье, состоит в том, что впервые разработана модель для ИТС управления маршрутами городских пассажирских автобусов с учетом влияния наиболее значимых стохастических факторов. Получена система уравнений, которая моделирует функционирование маршрута.

Практическая значимость полученных результатов заключается в том, что разработано программное обеспечение для имитационного моделирования в среде MatLab, реализующее предложенную модель, на основе которой решена практическая задача выбора управляющих воздействий для ИТС городских автобусных маршрутов.

Перспективы дальнейших исследований состоят в том, чтобы програмно реализовать предложенную модель на языках высокого уровня для проектируемых ИТС крупных городов Украины, что позволит оперативно оценивать влияние возмущающих воздействий (отклонение подвижной единицы от расписания, сход с линии, ухудшение условий движения, переполнение на маршруте) на показатели качества обслуживания пассажиров и составить оптимальное расписание движения транспорта.

БЛАГОДАРНОСТИ

Работа выполнена в рамках госбюджетных научно-исследовательских тем «Информационно-аналитические технологии управления в интеллектуальных транспортных системах многокритериальными и многопродуктовыми потоками в условиях неоднородной неопределенности параметров процессов» (номер г/р 0113U000695) Днепропетровского национального университета железнодорожного транспорта им. академика В. Лазаряна и «Разработка программного обеспечения для модулей АСУ предприятий с использованием объектно-ориентированных языков программирования и технологии клиент-сервер» (номер г/р 0107U006840) Луганского национального аграрного университета.

СПИСОК ЛІТЕРАТУРИ

1. Transportation & Logistics 2030. [Electronic resource]. – London: PWC, 2014. – Access mode: <http://www.pwc.com/tl2030>
2. Mikulski J. Modern Transport Telematics / J. Mikulski // 11th International Conference on Transport Systems Telematics, TST 2011. Katowice-Ustron, 19–22 October 2011.: Abstracts. Katowice-Ustron, Poland. – 2011. – P. 8–11.
3. Continuing Evolution of Travel Time Data Information Collection and Processing / [Tarnoff, Philip John, Bullock, Darcy M, Young, Stanley E, et al.] // Transportation Research Board Annual Meeting. TRB. – 2009. – P. 23–45.
4. Tyagi V. Vehicular Traffic Density State Estimation Based on Cumulative Road Acoustics / V. Tyagi, S. Kalyanaraman, R. Krishnapuram // IEEE Transactions on Intelligent Transportation Systems. – 2012. – Vol. 6. – P. 456–468.
5. Horowitz R. Control design of an automated highway system / R. Horowitz, P. Varaiya // In Proceedings of the IEEE. – 2000. – Vol. 88. – P. 913–925.
6. Dynamic Traffic Light Sequence, Science Publications / [Khalid A. S. Al-Khateeb, Jaiz A. Y. Johari, Wajdi F. Al-Khateeb] // Journal of Computer Science. – 2008. – No. 4 (7). – P. 517–524.
7. Автоматизированные системы обработки информации и управления на автомобильном транспорте : учебник / [А. Б. Николаев, С. В. Алексахин, И. А. Кузнецов, В. Ю. Строганов]; под ред. А. Б. Николаева. – М. : ACADEMIA, 2003. – 223 с.
8. Методика имитационного моделирования работы городского транспорта / [В. Н. Галушко, В. Д. Левчук, И. В. Максимей и др.] // Электрон. моделирование. – 2006. – № 2. – С. 79–95.
9. Дубова С. В. Особенности развития пассажирского транспорта в Киеве / С. В. Дубова // Містобудування та терит. планування. – 2003. – Вип. 15. – С. 68–72.
10. Ляхно В. А. Повышение эффективности систем управления автомобильным пассажирским транспортом методами стохастического моделирования : монография / В. А. Ляхно, А. И. Пилипенко. – Луганск : Элтон-2, 2007. – 177 с.
11. Transport Logistics. Shared solution to common challenges [Electronic resource]. – Paris: Organisation for economic co-operation and development, 2002. – Access mode: <http://www.internationaltransportforum.org/pub/pdf/02LogisticsE.pdf>

Статья поступила в редакцию 28.01.2016.

После доработки 10.02.2016.

Ляхно В. А.

Д-р техн. наук, доцент, завідувач кафедри організації комплексного захисту інформації, Європейський університет, Київ, Україна
МОДЕЛЬ ІНТЕЛЕКТУАЛЬНОЇ СИСТЕМИ УПРАВЛІННЯ МІСЬКИМИ АВТОБУСНИМИ ПЕРЕВЕЗЕННЯМИ

Вирішене завдання розробки математичного забезпечення для інтелектуальної системи управління міськими автобусними перевезеннями. Об'єктом дослідження є автоматизація процесу прийняття рішень інтелектуальними системами управління міськими транспортними потоками. Предмет дослідження становлять моделі диспетчерського управління транспортними потоками автобусних маршрутів. Метою роботи є вдосконалення моделей для інтелектуальних транспортних систем керуючих міськими потоками автобусів і маршрутних таксі. Розроблено модель для інтелектуальної транспортної системи управління з урахуванням впливу найбільш значущих чинників на графік руху автобусів за маршрутом. Модель дозволяє оперативно оцінювати вплив збурюючих дій на рух рухомої одиниці, зокрема таких, як переповнення автобусів на маршруті пасажирами, їх схід з лінії, відхилення від розкладу та ін., на показники якості обслуговування, а також, оптимізувати розклад руху. В якості критерію оптимізації запропоновано використовувати показник мінімуму часу очікування пасажирами автобусів і маршрутних таксі на зупинках. Під час експериментів перевірена адекватність розробленої моделі, яка оцінювалася методом однофакторного дисперсійного аналізу і повнофакторного експерименту в реальних міських умовах. Результати експериментів дозволяють рекомендувати запропоновану модель для практичного використання в інтелектуальних транспортних системах управління міськими автобусними маршрутами.

Ключові слова: інтелектуальна транспортна система, управління автобусними перевезеннями, модель, алгоритм управління.

Lakhno V. A.

Dr. Sc., associate professor, Head of Complex Information Security Organization Department, European University, Kyiv, Ukraine
MODEL OF INTELLIGENT MANAGEMENT SYSTEM OF CITY BUS TRANSPORTATIONS

This paper is devoted to improving the mathematical support for the intelligent traffic management system of city buses. It provides an overview of what is the real state of the-art with respect to traffic flow theory. A new mathematical model of the buses motion has been generating in consideration of stochastic factors. The model allows the calculating immediate changes in the city buses schedules connected with speed parameters. The model and program realization make allowance for increasing the efficiency of passenger service when projecting

city passenger transports. With regard to the traffic organization, the automated control system as the element of the intelligent transport systems plays the increasingly important role as a key component of the transport system, which is able to form the right choice for customers across a network, to support safe travel. The software implementing proposed method is developed. The experiments to study the properties of the proposed model are conducted. The experimental results allow to recommend the proposed model for use in practice.

Keywords: modeling, intelligent transport systems, information systems, dispatching control, transport flow, passenger flow.

REFERENCES

1. Transportation & Logistics 2030. [Electronic resource]. London, PWC, 2014. Access mode: <http://www.pwc.com/tl2030>
2. Mikulski J. Modern Transport Telematics, *11 th International Conference on Transport Systems Telematics, TST 2011. Katowice-Ustron, 19–22 October 2011*. Abstracts. Katowice-Ustron, Poland. 2011. – P. 8–11.
3. Tarnoff, Philip John, Bullock, Darcy M, Young, Stanley E, et al. Continuing Evolution of Travel Time Data Information Collection and Processing, *Transportation Research Board Annual Meeting*. TRB, 2009, pp. 23–45.
4. Tyagi V., Kalyanaraman S., Krishnapuram R. Vehicular Traffic Density State Estimation Based on Cumulative Road Acoustics, *IEEE Transactions on Intelligent Transportation Systems*, 2012, Vol. 6, pp. 456–468.
5. Horowitz R., Varaiya P. Control design of an automated highway system, *In Proceedings of the IEEE*, 2000, Vol. 88, pp. 913–925.
6. Khalid A. S., Al-Khateeb, Jaiz A.Y. Johari, Wajdi F. Al-Khateeb Dynamic Traffic Light Sequence, Science Publications, *Journal of Computer Science*, 2008, No. 4 (7), pp. 517–524.
7. Nikolaev A. B., Aleksahin S. V., Kuznetsov I. A., Stroganov V. Yu. Ed. A. B. Nikolaev Automated systems for information processing and management of road transport : textbook. Moscow, Academy, 2003, 223 p.
8. Galushko V. N., Levchuk V. D., Maksimey I. V., Mogila V. S., Chechet P. L. Methods of simulation of urban transport, *Elektron. Modelirovanie*, 2006, No. 2, pp. 79–95.
9. Dubova S. V. Features of development of passenger transport in Kiev, *Mistobuduvannya ta terit. planuvannya*, 2003, Vyp. 15, pp. 68–72.
10. Lahno V. A., Pilipenko A. I. Improving the efficiency of control systems of road passenger transport methods of stochastic modeling, monografija. Lugansk, Elton-2, 2007, 177 p.
11. Transport Logistics. Shared solution to common challenges [Electronic resource]. Paris, Organisation for economic co-operation and development, 2002. Access mode: <http://www.internationaltransportforum.org/pub/pdf/02LogisticsE.pdf>

Наукове видання

**Радіoeлектроніка,
інформатика,
управління**

№ 2/2016

Науковий журнал

Головний редактор – д-р фіз.-мат. наук В. В. Погосов

Заст. головного редактора – д-р техн. наук С. О. Субботін

Комп'ютерне моделювання та верстання
Редактор англійських текстів

С. В. Зуб
С. О. Субботін

Оригінал-макет підготовлено у редакційно-видавничому відділі ЗНТУ

Свідоцтво про державну реєстрацію
КВ № 6904 від 29.01.2003.

*Підписано до друку 21.06.2016. Формат 60×84/8.
Папір офс. Різогр. друк. Ум. друк. арк. 14,88.
Тираж 300 прим. Зам. № 558.*

69063, м. Запоріжжя, ЗНТУ, друкарня, вул. Жуковського, 64

Свідоцтво суб'єкта видавничої справи
ДК № 2394 від 27.12.2005.