



1898

Випуск присвячено
125-річчю
Національного технічного університету
України
«Київський політехнічний інститут
імені Ігоря Сікорського»

НАЦІОНАЛЬНА АКАДЕМІЯ НАУК УКРАЇНИ
НАВЧАЛЬНО-НАУКОВИЙ КОМПЛЕКС
«ІНСТИТУТ ПРИКЛАДНОГО СИСТЕМНОГО АНАЛІЗУ»
НАЦІОНАЛЬНОГО ТЕХНІЧНОГО УНІВЕРСИТЕТУ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ ІМЕНІ ІГОРЯ СІКОРСЬКОГО»

СИСТЕМНІ ДОСЛІДЖЕННЯ ТА ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ

МІЖНАРОДНИЙ НАУКОВО-ТЕХНІЧНИЙ ЖУРНАЛ

№ 2

2023

ЗАСНОВАНО У ЛИПНІ 2001 р.

РЕДАКЦІЙНА КОЛЕГІЯ:

Головний редактор

М.З. ЗГУРОВСЬКИЙ, акад. НАН України

Заступник головного редактора

Н.Д. ПАНКРАТОВА, чл.-кор. НАН України

Члени редколегії:

П.І. АНДОН, акад. НАН України

А.В. АНІСІМОВ, чл.-кор. НАН України

Х. ВАЛЕРО, проф., Іспанія

Г.-В. ВЕБЕР, проф., Турція

П.О. КАСЬЯНОВ, проф., д.ф.-м.н.,
Україна

Й. КОРБИЧ, проф., Польща

О.А. ПАВЛОВ, проф., д.т.н., Україна

Л. САКАЛАУСКАС, проф., Литва

А.М. САЛЕМ, проф., Єгипет

І.В. СЕРГІЄНКО, акад. НАН України

Х.-М. ТЕОДОРЕСКУ, акад. Румунської
Академії

Е.О. ФАЙНБЕРГ, проф., США

Я.С. ЯЦКІВ, акад. НАН України

АДРЕСА РЕДАКЦІЇ:

03056, м. Київ,

просп. Перемоги, 37, корп. 35,

ННК «ІПСА» КПІ ім. Ігоря Сікорського

Тел.: 204-81-44; факс: 204-81-44

E-mail: journal.iasa@gmail.com

http://journal.iasa.kpi.ua

У номері:

• **Теоретичні та прикладні проблеми і методи системного аналізу**

• **Прогресивні інформаційні технології, високопродуктивні комп'ютерні системи**

• **Проблеми прийняття рішень та управління в економічних, технічних, екологічних і соціальних системах**

• **Теоретичні та прикладні проблеми інтелектуальних систем підтримання прийняття рішень**

• **Методи оптимізації, оптимальне управління і теорія ігор**

• **Математичні методи, моделі, проблеми і технології дослідження складних систем**

NATIONAL ACADEMY OF SCIENCES OF UKRAINE
EDUCATIONAL AND SCIENTIFIC COMPLEX
«INSTITUTE FOR APPLIED SYSTEM ANALYSIS»
OF THE NATIONAL TECHNICAL UNIVERSITY OF UKRAINE
«IGOR SIKORSKY KYIV POLYTECHNIC INSTITUTE»

SYSTEM RESEARCH AND INFORMATION TECHNOLOGIES

INTERNATIONAL SCIENTIFIC AND TECHNICAL JOURNAL

№ 2

2023

IT IS FOUNDED IN JULY 2001

EDITORIAL BOARD:

The editor – in – chief

M.Z. ZGUROVSKY, Academician of
NASU

Deputy editor – in – chief

N.D. PANKRATOVA, Correspondent
member of NASU

Associate editors:

F.I. ANDON, Academician of
NASU

A.V. ANISIMOV, Correspondent
member of NASU

E.A. FEINBERG, Prof., USA

P.O. KASYANOV, Prof., Ukraine

J. KORBICH, Prof., Poland

A.A. PAVLOV, Prof., Ukraine

L. SAKALAIUSKAS, Prof., Lithuania

A.M. SALEM, Prof., Egypt

I.V. SERGIENKO, Academician of NASU

H.-N. TEODORESCU, Academician of
Romanian Academy

J. VALERO Prof., Spain

G.-W. WEBER, Prof., Turkey

Ya.S. YATSKIV, Academician of NASU

THE EDITION ADDRESS:

03056, Kyiv,
av. Peremogy, 37, building 35,
Institute for Applied System Analysis
at the Igor Sikorsky Kyiv Polytechnic Institute
Phone: **204-81-44**; Fax: **204-81-44**
E-mail: journal.iasa@gmail.com
<http://journal.iasa.kpi.ua>

In the issue:

- **Theoretical and applied problems and methods of system analysis**
- **Progressive information technologies, high-efficiency computer systems**
- **Decision making and control in economic, technical, ecological and social systems**
- **Theoretical and applied problems of intelligent systems for decision making support**
- **Methods of optimization, optimum control and theory of games**
- **Mathematical methods, models, problems and technologies for complex systems research**

Шановні читачі!

Навчально-науковий комплекс «Інститут прикладного системного аналізу» Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського» видає міжнародний науково-технічний журнал

«СИСТЕМНІ ДОСЛІДЖЕННЯ ТА ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ».

Журнал публікує праці теоретичного та прикладного характеру в широкому спектрі проблем, що стосуються системних досліджень та інформаційних технологій.

Провідні тематичні розділи журналу:

Теоретичні та прикладні проблеми і методи системного аналізу; теоретичні та прикладні проблеми інформатики; автоматизовані системи управління; прогресивні інформаційні технології, високопродуктивні комп'ютерні системи; проблеми прийняття рішень і управління в економічних, технічних, екологічних і соціальних системах; теоретичні та прикладні проблеми інтелектуальних систем підтримання прийняття рішень; проблемно і функціонально орієнтовані комп'ютерні системи та мережі; методи оптимізації, оптимальне управління і теорія ігор; математичні методи, моделі, проблеми і технології дослідження складних систем; методи аналізу та управління системами в умовах ризику і невизначеності; евристичні методи та алгоритми в системному аналізі та управлінні; нові методи в системному аналізі, інформатиці та теорії прийняття рішень; науково-методичні проблеми в освіті.

Головний редактор журналу — ректор Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського», академік НАН України Михайло Захарович Згуровський.

Журнал «Системні дослідження та інформаційні технології» включено до переліку фахових видань ВАК України.

Журнал «Системні дослідження та інформаційні технології» входить до таких наукометричних баз даних: Scopus, EBSCO, Google Scholar, DOAJ, Index Copernicus, реферативна база даних «Україніка наукова», український реферативний журнал «Джерело», наукова періодика України.

Статті публікуються українською та англійською мовами.

Журнал рекомендовано передплатити. **Наш індекс 23918.** Якщо ви не встигли передплатити журнал, його можна придбати безпосередньо в редакції за адресою: 03056, м. Київ, просп. Перемоги, 37, корп. 35.

Завідувачка редакції **С.М. Шевченко**

Редакторка **Р.М. Шульженко**

Молодша редакторка **Л.О. Тарин**

Комп'ютерна верстка, дизайн **А.А. Патіюхи**

Свідоцтво про реєстрацію КВ № 23234–13074 ПР від 22.03.2018 р.

Підписано до друку 30.06.2023. Формат 70x108 1/16. Папір офс. Гарнітура Times.

Спосіб друку – цифровий. Ум. друк. арк. 14,411. Обл.-вид. арк. 28,56. Наклад 100 пр. Зам. № 11/04

Національний технічний університет України

«Київський політехнічний інститут імені Ігоря Сікорського»

Свідоцтво про державну реєстрацію: ДК № 5354 від 25.05.2017 р.

просп. Перемоги, 37, м. Київ, 03056.

ФОП Пилипенко Н.М., вул. Мічуріна, б. 2/7, м. Київ, 01014.

Виписка з Єдиного державного реєстру № 2 070 000 0000 0214697 від 17.05.2019 р.,

тел.: (044) 361 78 68.

Dear Readers!

Educational and Scientific Complex «Institute for Applied System Analysis» of the National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute» is published of the international scientific and technical journal

«SYSTEM RESEARCH AND INFORMATION TECHNOLOGIES».

The Journal is printing works of a theoretical and applied character on a wide spectrum of problems, connected with system researches and information technologies.

The main thematic sections of the Journal are the following:

Theoretical and applied problems and methods of system analysis; theoretical and applied problems of computer science; automated control systems; progressive information technologies, high-efficiency computer systems; decision making and control in economic, technical, ecological and social systems; theoretical and applied problems of intellectual systems for decision making support; problem- and function-oriented computer systems and networks; methods of optimization, optimum control and theory of games; mathematical methods, models, problems and technologies for complex systems research; methods of system analysis and control in conditions of risk and uncertainty; heuristic methods and algorithms in system analysis and control; new methods in system analysis, computer science and theory of decision making; scientific and methodical problems in education.

The editor-in-chief of the Journal is rector of the National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute», academician of the NASU Michael Zaharovich Zgurovsky.

The articles to be published in the Journal in Ukrainian and English languages are accepted. Information printed in the Journal is included in the Catalogue of periodicals of Ukraine.

СИСТЕМНІ ДОСЛІДЖЕННЯ ТА ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ

2 • 2023

ЗМІСТ

ТЕОРЕТИЧНІ ТА ПРИКЛАДНІ ПРОБЛЕМИ І МЕТОДИ СИСТЕМНОГО АНАЛІЗУ	
<i>Pankratova N.D., Grishyn K.D., Barilko V.E.</i> Digital twins: stages of concept development, areas of use, prospects	7
ПРОГРЕСИВНІ ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ, ВИСОКОПРОДУКТИВНІ КОМП'ЮТЕРНІ СИСТЕМИ	
<i>Bidyuk P., Gavrilenko O., Myagkyi M.</i> The algorithm for predicting the cryptocurrency rate taking into account the influence of posts of a group of famous people in social networks	22
<i>Kulik A.S., Chukhray A.G., Havrylenko O.V.</i> Information technology for creating intelligent computer programs for training in algorithmic tasks. Part 2: research and implementation	35
ПРОБЛЕМИ ПРИЙНЯТТЯ РІШЕНЬ ТА УПРАВЛІННЯ В ЕКОНОМІЧНИХ, ТЕХНІЧНИХ, ЕКОЛОГІЧНИХ І СОЦІАЛЬНИХ СИСТЕМАХ	
<i>Zgurovsky M., Yefremov K., Gapon S., Pyshnograiev I.</i> Assessment of the economical dimension of sustainable development of the Ukraine's regions based on the brightness of night lights	49
<i>Кондратенко Н.Р., Снігур О.О., Кондратенко Р.М.</i> Узагальнювальна інтервально-нечітка модель типу-2 для моніторингу станів складних систем з використанням експертних знань	63
ТЕОРЕТИЧНІ ТА ПРИКЛАДНІ ПРОБЛЕМИ ІНТЕЛЕКТУАЛЬНИХ СИСТЕМ ПІДТРИМАННЯ ПРИЙНЯТТЯ РІШЕНЬ	
<i>Петренко А.І., Вохранов І.А.</i> Нейронні мережі: дослідження правил прийняття ними рішень	74
МЕТОДИ ОПТИМІЗАЦІЇ, ОПТИМАЛЬНЕ УПРАВЛІННЯ І ТЕОРІЯ ІГОР	
<i>Nakonechnyi O.G., Kudin G.I., Zinko P.M., Zinko T.P.</i> Guaranteed root-mean-square estimates of the forecast of matrix observations under conditions of statistical uncertainty	86
<i>Romanuke V.V., Romanov A.Y., Malaksiano M.O.</i> A genetic algorithm improvement by tour constraint violation penalty discount for maritime cargo delivery	104
<i>Куреннов С.С., Барахов К.П., Барахова Г.С.</i> Застосування кривих Безье для опису форми конструкції при оптимізації клейового з'єднання	127
МАТЕМАТИЧНІ МЕТОДИ, МОДЕЛІ, ПРОБЛЕМИ І ТЕХНОЛОГІЇ ДОСЛІДЖЕННЯ СКЛАДНИХ СИСТЕМ	
<i>Panibratov R.S., Bidyuk P.I.</i> Estimation of the parameters of generalized linear models in the analysis of actuarial risks	139
Відомості про авторів	149

SYSTEM RESEARCH AND INFORMATION TECHNOLOGIES

2 • 2023

CONTENT

THEORETICAL AND APPLIED PROBLEMS AND METHODS OF SYSTEM ANALYSIS	
<i>Pankratova N.D., Grishyn K.D., Barilko V.E.</i> Digital twins: stages of concept development, areas of use, prospects	7
PROGRESSIVE INFORMATION TECHNOLOGIES, HIGH-EFFICIENCY COMPUTER SYSTEMS	
<i>Bidyuk P., Gavrilenko O., Myagkyi M.</i> The algorithm for predicting the cryptocurrency rate taking into account the influence of posts of a group of famous people in social networks	22
<i>Kulik A.S., Chukhray A.G., Havrylenko O.V.</i> Information technology for creating intelligent computer programs for training in algorithmic tasks. Part 2: Research and implementation	35
DECISION MAKING AND CONTROL IN ECONOMIC, TECHNICAL, ECOLOGICAL AND SOCIAL SYSTEMS	
<i>Zgurovsky M., Yefremov K., Gapon S., Pyshnograiev I.</i> Assessment of the economical dimension of sustainable development of the Ukraine's regions based on the brightness of night lights	49
<i>Kondratenko N.R., Snihur O.O., Kondratenko R.M.</i> Interval type-2 generalizing fuzzy model for monitoring the states of complex systems using expert knowledge	63
THEORETICAL AND APPLIED PROBLEMS OF INTELLIGENT SYSTEMS FOR DECISION MAKING SUPPORT	
<i>Petrenko A.I., Vokhranov I.A.</i> Neural networks: studying their decision-making rules	74
METHODS OF OPTIMIZATION, OPTIMUM CONTROL AND THEORY OF GAMES	
<i>Nakonechnyi O.G., Kudin G.I., Zinko P.M., Zinko T.P.</i> Guaranteed root-mean-square estimates of the forecast of matrix observations under conditions of statistical uncertainty	86
<i>Romanuke V.V., Romanov A.Y., Malaksiano M.O.</i> A genetic algorithm improvement by tour constraint violation penalty discount for maritime cargo delivery	104
<i>Kurennov S.S., Barakhov K.P., Barakhova H.S.</i> Application of bezier curves for description of structure shape in optimization of adhesive joints	127
MATHEMATICAL METHODS, MODELS, PROBLEMS AND TECHNOLOGIES FOR COMPLEX SYSTEMS RESEARCH	
<i>Panibratov R.S., Bidyuk P.I.</i> Estimation of the parameters of generalized linear models in the analysis of actuarial risks	139
Information about the authors	149

DIGITAL TWINS: STAGES OF CONCEPT DEVELOPMENT, AREAS OF USE, PROSPECTS

N.D. PANKRATOVA, K.D. GRISHYN, V.E. BARILKO

Abstract. The results of a review of the digital twin concept development, the areas of their use, and the prospects are highlighted. The history of the emergence and development of the digital twin concept, its definition, and its classification are given. The relevance of the technology under consideration is reflected. The purpose of this review is to provide the most complete, up-to-date information on the current state of the digital twin technology, its application in various fields of human activity, and further prospects for the development of the industry. An extensive bibliography on the topic is provided, which may be helpful for researchers and representatives of various industries.

Keywords: Industry 4.0, digital twin, classification, areas of application, Internet of Things, physical and mathematical models.

INTRODUCTION

The increasing complexity of production management task necessitates the adoption of new management methods and systems. Today, production management requires Industry 4.0 competencies [1], which have emerged relatively recently and are largely interconnected with IT technologies. The traditional pyramidal model for designing production management systems is being replaced by modern approaches to direct interaction between the components of the production system based on the concept of the Internet of Things (IoT). The digital twin refers to a new innovative toolkit that helps to exploit the advanced scenarios of the Internet of Things and other technologies. This toolkit is used to create digital models of the physical environment, and digital analogues will be able to perceive information from the world around them, interact and exchange data. The result is a completely new environment where the intelligence embedded in applications will allow to assess what is happening in the physical world, take into account the accumulated information and experience to support decision-making. This environment creates qualitatively new conditions for business and ensuring environmental safety at work. Thus, the digital and physical worlds are being united, with applications, services, networking components and edge devices being the Internet of Things. Each element of the Internet of Things in production will have a digital twin, i.e. its virtual model for rational production management. Based on a systematic literature review and a thematic analysis of 92 publications on digital

twins over the past ten years, [2] this work describes this technology, identifies knowledge gaps and necessary areas for future research. In the characterisation of a digital twin, the state of the concept, key terminology and related processes are identified, discussed and combined to produce 13 characteristics (physical entity/twin; virtual entity/twin; physical environment; virtual environment; state; implementation; metrology; duplication; duplication rate; physical-virtual connection/twin; virtual-physical connection/twin; physical processes; and virtual processes) and the complete structure of the digital twin and its operation process. Following this characterisation, seven knowledge gaps and topics for future research are identified: perceived benefits; digital twin in the product lifecycle; use cases; technical implementations; confidence levels; data ownership; and integration of virtual entities, each of which is necessary for digital twin implementation. There are also other reviews [3]–[6] on the digital twin application in various industries.

The history of the digital twin concept. The first digital twin in history (at that time, such name and concept did not exist) can be considered a program used by NASA in the 1960S to design the Apollo 13 mission. It was created to test how the future object would behave in the physical world. Later, engineers discovered that the virtual model could be used to control equipment and predict what would happen to it. The sensor readings during the flight were compared with those predicted in the digital model in real time [7]. This idea later became the basis for the creation of modern digital twins.

The emergence of the digital twins concept is associated with the work of Michael Greaves, a professor at the University of Michigan. In 2002, as part of a presentation at the University of Michigan for industry representatives, he proposed a model consisting of three components: a real space; a virtual space; and a mechanism that ensured the exchange of information between them. At the time, this concept was called “ideal PLM” (Product Lifecycle Management) [8]. A similar concept named “Mirror Worlds” [9] was proposed earlier by computer scientist David Gelernter in 1991.

In response to the M. Greaves publication, K. Främling et al. 2003 [10] proposed to modify the PLM concept to solve the problem of organizing the exchange of information between a virtual object and a physical one using the Internet of Things. In particular, the authors argued that the standards already existing at that time were sufficient to organize the computing architecture. In 2003, the idea was not widely supported due to the imperfection of technologies for processing a large data flow in real time, with most data stored in paper form.

Later, Michael Greaves developed this idea and presented it in his course on PLM systems, which he taught at the same university. Despite the fact that the term “digital twin” appeared a little later, the basic concept of having a physical object, a virtual object and ensuring the information exchange between them was developed already then. In the following years, one of the names of this concept was the “model of mirror spaces” [11], in the period 2006–2010 the term “model of information mirror” was used [12].

In 2010, the term “Digital Twin” first appeared in NASA’s technology roadmap, its authorship could be attributed to J. Vickers. This publication also used the term “Virtual Digital Fleet Leader” [13]. Since 2011, the concept of digital twins has been used by the US Air Force Research Laboratory (AFRL) to effectively resolve the design issues, maintenance and forecasting of the aircraft service life [13], [14]. A virtual copy of the aircraft accompanied the real object

throughout its life cycle. In 2013, Germany proposed the concept of Industry 4.0. Its key idea is the cyber-physical system (CPS). CPS combines people, machines and things that provide a continuous exchange of information to monitor, collect data, analyse and optimise production processes. To implement such processes, it is necessary to use a digital twin [15]. The concept of a “digital twin” is a part of the fourth industrial revolution and is designed to help businesses detect physical problems faster, make more accurate predictions, develop better products [16], [17].

M. Greaves gave a detailed description of the digital twin in 2014 in the so-called White Paper, intended for corporate society. Some industrial companies, such as Siemens, almost immediately adopted the terminology and paradigm outlined in the book [18]. In his book “The Origin of Digital Twins”, M. Greaves divided any digital twin into three main parts [19]: a physical product; a virtual product; and data and information that unite the virtual and physical products. In his opinion, “under ideal conditions, all the information that needs to be obtained from a product can be provided by a digital twin”.

The design of digital twins is based on simulation modelling methods that provide the most realistic representation of a physical environment or object in the virtual world. A mathematical description of digital twins can be obtained via statistical modelling, machine learning or analytical modelling methods. The mathematical models of a digital twin for supporting and predicting the functioning of a physical process of a real objects are given in [20], [21]. The parametric uncertainty of the mathematical description of the physical process is taken into account. As an example, the design of a digital twin is given on the example of an analytical model of an electric heater.

In 2023, the concept of a digital twin is evolving into something more subtle and incredibly practical: an executable digital twin (xDT). Simply put, xDT is a digital twin on a chip. xDT uses data from a (relatively) small number of sensors embedded in a physical product to perform real-time simulations using reduced-order models. Based on the data from a small number of sensors, it can predict the physical state anywhere within the object (even in places where sensors cannot be placed) [22].

The main stages of development of the digital twin concept are shown in Fig. 1, from the inception of the idea in 1960 to the current state. Source: compiled by the authors based on [7]–[22].

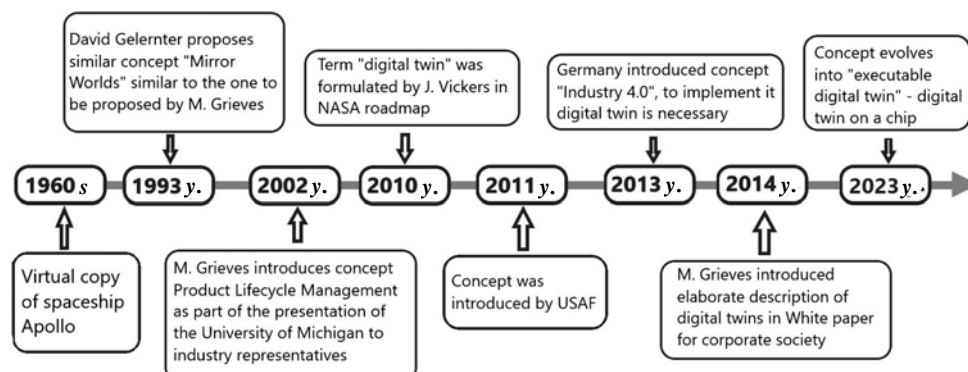


Fig. 1. Key stages in the development of the digital twin concept. Source: compiled by the authors on the basis of [7]–[22]

Definition of a digital twin. The review has shown that there is no single accepted definition of digital twin. The paper “Digital twin paradigm: A systematic literature review” counted 30 definitions given by researchers and organizations [7]. Therefore, we will present the definition proposed by the author of the concept of digital twin and one of the modern ones, which, in our opinion, is the most complete among the existing ones.

Definition by M. Greaves. A set of virtual information constructs that fully describe a potentially or actually produced physical product from microatomic to macrogeometric level [8].

Definition by Asimov et al. 2018. A virtual copy of a physical object that can monitor data consistency, perform data mining to identify existing and predict future problems, and use artificial intelligence to support business decisions [23]. However, the definition of a DT is not finalized. For example, a 2021 meta-analysis that included 24 articles showed that 83% authors agreed that the concept was in the early stages of development [24].

The relevance of the topic of digital twins. According to a study conducted in September 2022, the number of publications in Scopus on the topic of "digital twin" is growing every year [25]. The popularity of this topic can be seen from the graph of the number of publications depending on the year. Interest in the creation of digital twins has been growing significantly since 2016.

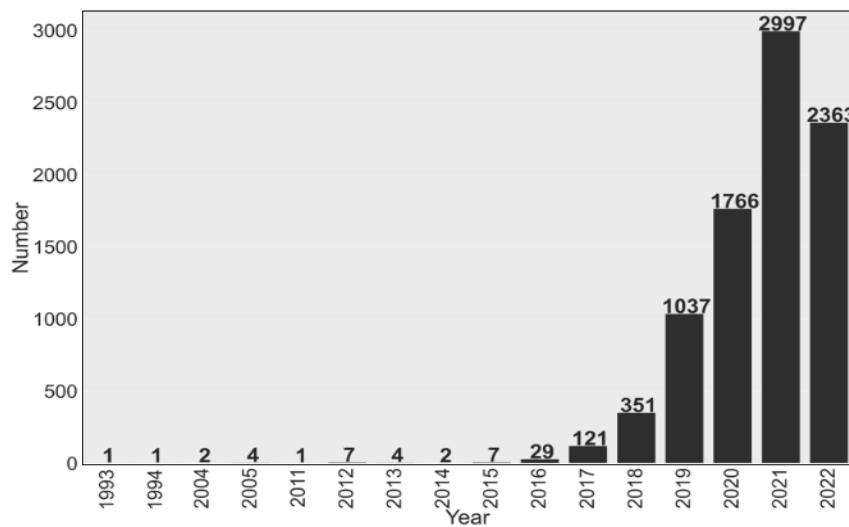


Fig. 2. Number of publications on DT depending on the year. Source [20].

More and more countries are adopting development plans based on technologies that combine physical and information space: “Industry 4.0” in Germany [15], “Advanced Manufacturing” in the United States [26], “Society 5.0” in Japan [27], “Made in China 2025” in China [28], and a similar project for the EU, “The Factory of Future” [29]. Topic relevance is confirmed by numerous highly cited publications. Authors of these publications have high H-index (Table 1).

Table 1. The most cited works on digital twins

Title	Year	Number of citations	H-index of authors
		Scopus / Google Scholar	
Jay Lee, Behrad Bagheri, Hung-An Kao “A cyber-physical systems architecture for industry 4.0-based manufacturing systems” [30]	2014	3003/5088	Jay Lee: 58/73 Behrad Bagheri: 12/19
Fei Tao, Jiangfeng Cheng, Qinglin Qi, Meng Zhang, He Zhang, Fangyuan Sui “Digital twin-driven product design, manufacturing and service with big data” [31]	2018	1433/2084	Fei Tao: 69/81 Jiangfeng Cheng: 14/16 Qinglin Qi: 23/24
Dmitry Ivanov “Predicting the impacts of epidemic outbreaks on global supply chains: A simulation-based analysis on the coronavirus outbreak (COVID-19/SARS-CoV-2) case” [32]	2020	1079/1821	Dmitry Ivanov: 62/74
Werner Kritzing, Matthias Karner, Georg Traar, Jan Henjes, Wilfried Sihm “Digital twin in manufacturing: a categorical literature review and Classification” [33]	2018	1114/1718	Wilfried Sihm: 22/31
Fei Tao, He Zhang, Ang Liu, Andrew Yeh-Ching Nee “Digital Twin in Industry: State-of-the-Art” [34]	2019	1197/1750	Fei Tao: 69/81
Michael Grieves, John Vickers “Digital Twin: Mitigating Unpredictable, Undesirable Emergent Behavior in Complex Systems” [35]	2017	1063/2076	Michael Grieves: 7/16
Abram L. J. Walton, Cynthia L. Tomovic, Michael W. Grieves, “Product Lifecycle Management: Measuring What Is Important – Product Lifecycle Implementation Maturity Model” [36]	2013	6/13	Michael Grieves: 7/16 Tomovic Cynthia: 5/9
F Tao, Q Qi, L Wang, AYC Nee “Digital twins and cyber-physical systems toward smart manufacturing and industry 4.0: correlation and comparison” [37]	2019	487/712	Qinglin Qi: 23/24 Fei Tao: 69/81

Source: compiled by the authors on the basis of [30]–[37]

The study by Semeraro et al. 2021 [7] provides the following facts about the popularity of the term “digital twin” in the Scopus, Elsevier, and ScienceDirect databases: 60 journals have published papers on this topic; it was discussed in 47 conferences; 8 chapters in various books have been found.

Many DT-related technologies have been patented. As part of the review, we searched for patents with the term “digital twin” using the query TAC=“digital twin”. At the time of writing, Google Patents returned 6181 results for this query dated 2003-01-01. The largest patent holders are large corporations: Siemens AG (13.4% of all patents); General Electric (9.8%); Beijing University of Aeronautics and Astronautics (3.4%). A study that conducted a cluster analysis of DT patents from the Webpat and Derwent databases found 140 records by 2018 [38].

The classification of DT. The most complete list of classifications is given in the paper [39]: by application, by level of integration, by hierarchy, by level of maturity/complexity, including the classification introduced by M. Greaves [40].

Reading literature, one may also encounter terms “digital model” and “digital shadow”. They should be distinguished from a digital twin. In particular, the difference between these terms was explained by Kritzinger et al [33]:

- A digital model is a virtual representation of an entity in which there is no automatic data exchange between a physical and a virtual object. All information is transferred manually.
- A digital shadow is a virtual representation of an entity where data is transferred from a physical object to a virtual one, but not vice versa. Thus, a digital shadow is a kind of logbook where all important changes are recorded.
- A digital twin is a virtual representation in which data flows in both directions, from a digital entity to a physical one and vice versa.

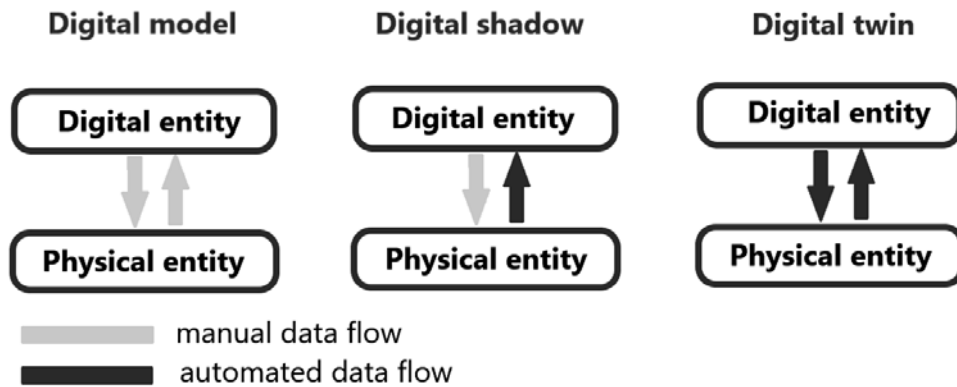


Fig. 3. Data flow in a digital model, shadow and twin. Compiled on the basis of the publication [33]

AREAS OF APPLICATION OF DIGITAL TWINS

1. **Aerospace industry.** As noted in the IEEE publication [40], leading aerospace industry agencies such as NASA, EASA, the US Air Force, and the Royal Canadian Air Force (RCAF) consider DT as a key element in production. This technology enables the implementation of the Operation & Maintenance paradigm, which prioritises the prevention of accidents and breakdowns, and fixing them before they occur.

Digital twins are indispensable in the design of critical components such as engines, which are expensive and technologically complex to test. Computer simulations allow to test different flight modes, altitudes, climatic conditions, emergency situations, and system failures before physical testing. Simulations make possible to study engine behaviour under extreme conditions without physical experiments [40].

In-service DT reduces the need to rely on probabilistic models to determine when a part may need maintenance. Various machine learning algorithms are used in the aerospace industry: Support Vector Machines, autoencoders, convolutional neural networks. A full list of algorithm applications is available in the IEEE publication [40]. More details can be found in a textbook dedicated to the problem of using DT in aircraft [41].

The DTs are used by Rolls-Royce in the design and operation of its aircrafts. Once released, the virtual model accompanies the aircraft during each flight, comparing predicted performance with real-world performance, and building What If scenarios [42].

2. Automotive industry. Digital twins are used to speed up the design of new car models. First, a virtual model is developed and tested, then a physical prototype is built and improved in parallel with the digital prototype. A virtual twin makes it possible to review different design options and run computer simulations of various situations (for example, how a car would behave on a slippery road or in icy conditions). Siemens, for instance, offers software solutions for creating digital prototypes of cars with the function of developing a 3D model and conducting simulations [43].

Maserati has replaced the costly and time-consuming wind tunnel streamlining procedure with a digital twin. Also, the engine sound (which is a symbol of the brand) is tuned in a computer simulation, whereas previously a physical prototype was built and the acoustics were tested on mannequins with microphones. In general, the use of this technology reduces the new car development time by 30% [44]. Renault uses the DT to conduct virtual crash tests, study body aerodynamics, engine efficiency, even check how well air conditioner cools the interior [45].

Digital twins are used not only for design. Cars with a high degree of computerisation form the so-called Internet of Vehicles [46], which provides the city with many advantages: increased safety, traffic optimisation, information about available parking spaces, and the possibility of remote maintenance [47]. A striking example of the use of DT during operation is Tesla, which creates a DT for each car sold. During the trip, sensor readings are continuously taken and sent to servers, artificial intelligence detects problems and forms an individual approach to their solution (such as changing the settings to make the doors close more quietly) [48].

3. Oil and gas industry. The digital twin plays an important role in modern drilling operations throughout the entire life cycle of a borehole, starting with the exploration stage. The DT of borehole is a complex system: the drilling model is made up of four unique elements, each of which is a digital twin itself: wellbore, drill string, drilling mud, and formation. Computer simulation allows to find the best types of tools and rig settings. Real-time monitoring of all processes and forecasting via machine learning methods helps to identify equipment problems that can lead to unplanned downtime or correct process deviations before they lead to a deterioration in the quality of the workflow [49]. Ericsson uses systems to monitor the condition of workers through smartwatches and bracelets to prevent mistakes due to fatigue [50]. The most comprehensive bibliography of the use of digital twins in the oil industry is provided by Cameron et al. 2018 [51].

4. Chemical and pharmaceutical industry. In the chemical and pharmaceutical industries, each batch of products is created with a DT that combines all the necessary information about the batch. Based on these data, it is possible to run simulations and predict the optimal technological parameters for the next production stages to ensure the planned product quality.

Information about changes in product properties depending on the parameters can be stored in databases, which allows to “design” new substances on the computer. This functionality makes DT a key element of the quality management

process [49]. For example, such software was developed by Atos and Siemens [52].

5. Digital Twins of Cities. There are terms like “Smart City” and “Digital City” that do not mean that a city has a digital twin. A smart city is one that follows a strategy of saving resources through innovative solutions. A digital city is a city that has implemented IoT technologies [53].

The smart city concept emerged in response to the Kyoto Protocol, which obliged countries to reduce greenhouse gas emissions [53]. Intensive urbanisation (in particular, according to UN forecasts, 68% of the world’s population will live in cities in 2050 [54]) has led to cities being the main source of pollution. Digital transformation has been proposed as a way to reduce emissions. With the development of 3D modelling technologies, it became possible to visualise infrastructure, but the models were static and used only for infrastructure planning, they could not display real-time processes, nor could they be remotely controlled – they were not digital twins [55]. It was only with the appearance of IoT technologies, cloud computing, the implementation of artificial intelligence and big data that it became possible to set up data exchange between the physical and digital essence of the city. This moment can be considered the emergence of digital twins of cities [53].

The world’s first digital twin city was created for Singapore in 2014 at a cost of US \$73 million. In 2022, the system was replaced with an advanced version that includes data from sensors, drones, government agencies, etc. [56]. The city of Zurich has its own DT with a detailed 3D map of roads, underground and above-ground facilities [57]. According to the World Economic Forum 2022, by 2030, the technology of DT will have saved \$289 billion on city planning and construction; in 2020, investments in DT of Chinese cities exceeded \$380 billion [58].

The most comprehensive bibliography on digital twin cities is provided in the review, which lists the main applications: planning and forecasting, visualisation, engagement, monitoring, data management [59].

6. Digital twins of urban water supply systems. The problem of drinking water is becoming more urgent in the world, as such, there is a need for its rational use. The International Water Association (IWA) has published a paper on digital twins of urban water supply systems as a way to reduce water consumption [60].

An example of a successfully implemented project is the digital twin of water supply system of Valencia (Spain) and its suburbs, developed by Global Omnium. It feeds approximately 1.6 million inhabitants, and the water supply system provides water to 325,000 nodes. The programme contains a detailed model to monitor the system and predict emergencies [61]. The digital twin can predict possible problems 24 hours in advance. Simulations and records of previous accidents are used to train personnel [62].

No less successful is the H2PORTO project, a digital twin of the water supply system in Porto (Portugal). It provides real-time monitoring of the system, uses weather and tidal data to model water levels and warn of flood risk. H2PORTO can run virtual simulations of emergency situations: pipe bursts, valve closures, pumping station shutdowns [63].

7. Energy. Power grids are among the most complex engineering systems in the world. With the appearance of solar panels and wind turbines, power supply

networks are becoming more complex, requiring effective management. For example, Siemens has set up a research centre to develop a digital twin of Australia's power grid. In 2023, €3.4 million was invested in the project. It is planned to be developed via the Siemens programme PSS software [64].

The lack of a single standard for digital twins remains a problem in the energy sector. Management is complicated if several power grids are connected but have different digital twins. In a publication issued by the IEEE, a single standard for digital twins was proposed. The authors note that the architecture is flexible, which will spare rebuilding the system in the future and enable connecting existing digital twins [65]. The most comprehensive bibliography on digital twin power grid structures is provided in the article [66].

8. Construction. Several reviews have already been conducted on this topic [67, 68]. Found publications are classified by the stages of the life cycle of the object where the digital twin is used: design, construction, and operation. In the construction process, complete digital twins are used as part of the implementation of Building information modelling (BIM), a strategy to improve the efficiency of the construction process [69]. Papers have been published that propose the design of the digital twin for the operation of the building. The topics raised include building maintenance, fault detection, scheduling and personnel management, etc. [70, 71].

9. Retail trade. In the book "Advances in computers Vol. 130", there is a separate chapter devoted to DT in retail trade as a tool for effective planning. In its simplest form, a DT can be a virtual copy of a physical supply chain with information on weather, fuel prices, etc. This allows to simulate how, for example, a transport delay due to bad weather can affect the flow of products in the supply chain [72].

According to Forbes, retailers and companies such as Lowes, Kroger, and Tyson Foods are already exploring or implementing digital twins of their supply chains. French supermarket chain Intermarché has already created digital twins of most of its stores. [73].

10. Preservation of cultural heritage. Museums use digital twins to monitor the preservation of exhibits. Some institutions create digital twins of entire buildings of cultural value. Such technologies are developed by Weiss AG. For example, the Forbidden City Museum (Beijing, China) conducts periodic 3D scans of ancient buildings to monitor whether restoration is needed [74]. Digital twins are also used to organise virtual tours. More details about the application can be found in the review dedicated to the use of digital twins in museums [75].

THE PROSPECTS OF DIGITAL TWINS. THE LARGEST PROJECTS OF DIGITAL TWINS.

A digital twin of the port of Rotterdam. Currently, IBM is creating a digital twin of the port of Rotterdam [76], the largest port in the world. The project complexity is comparable to projects that develop digital twin of large city. For example, in the annual report of 2022, it was stated that revenue was 825,700,000 euros [77]. It is planned to organise the port territory into a single digital space through IBM Cloud and IBM Internet of Things systems. The digital twin of the port will provide the following opportunities [76]:

- collection of meteorological and hydrological data to determine the most favourable time for getting to the port;

- autonomous ship control and data exchange between vessels;
- fuel economy, optimisation of approach and mooring speed;

A digital twin of the planet Earth [78]. The European Union plans to become a climate-neutral region by 2050. The key point of this plan is digital transformation, and for this purpose, in particular, a digital twin of the planet Earth is being developed, in which 1 trillion euros have been invested. It will perform the following tasks:

- combining data from satellites, scientific stations, sensors, etc. into a single high-resolution information platform;
- analysis of human impact on nature, such as carbon emissions, water pollution, littering;
- more accurate forecasting of weather and climate in the long-term perspective;
- building what-if scenarios to rationally implement environmental projects.

Today, some companies have developed software products for creating digital twins. Digital twin software is an advanced type of modelling software that generates a digital simulation of a physical object. Here are examples of software for creating digital twins:

1. Siemens Digital Twin allows to create digital twins for equipment and industrial processes. It provides real-time monitoring and control of equipment, as well as the ability to predict possible failures and diagnose problems.

2. ANSYS Twin Builder – allows to create digital twins for various industrial systems, including automotive, electronics, and power generation. It allows to conduct virtual testing and analyse various system scenarios, which helps to reduce costs and improve productivity.

3. Microsoft Azure Digital Twins is a cloud-based solution that allows to create digital twins for different types of buildings and urban environments. It allows to visualise the structure of a building and its various systems, as well as predict and manage energy consumption.

4. Predix Digital Twin is a solution from General Electric that allows to create digital twins for equipment and infrastructure in the energy and transport industries. The software product provides equipment condition monitoring and forecasting of possible failures, which helps to reduce maintenance and repair costs.

5. The computer game Factorio is a game about building and managing factories to produce various goods. One of the key aspects is that players have to collect resources, process them and build new factories and machines to improve production and get more products.

As for the digital twin aspect, Factorio players create a virtual model of the production process. The game allows to create complex mechanisms from a variety of elements, such as conveyors, workers, sorting machines, resource mines, etc. Each element performs a specific function in the production process, and players must properly configure their work to maximise production.

In Factorio, one can create and save templates of his own production processes. This allows players to create digital twins of their factories and mechanisms that can be saved and transferred between different game worlds. In this way,

players can use digital twins to quickly restore the production process in the case of a failure or destruction.

6. “AJAX Systems” is a Ukrainian company specialising in the development and production of wireless security alarm systems and intercoms. AJAX Systems main products are wireless security alarm systems that use advanced technologies and are highly reliable and efficient. AJAX systems are easy to install and configure, as well as the ability to control their operation using a mobile application. In addition to wireless security alarm systems, the company also produces wireless intercom systems that allow to control access to the building and communicate with visitors using a mobile application.

CONCLUSIONS

The article provides a detailed overview of the history of the emergence and development of the digital twin concept, provides definitions, explains the difference between this term and related concepts, and substantiates its relevance and application in various fields of activity.

The topic relevance is substantiated by calculating number of publications in Scopus, Elsevier, analysing the number of patents in Google Patents, and analysing the industrial plans of developed countries such as Germany, the USA, Japan, and China. In particular, it can be concluded that industrial giants such as Siemens and General Electric have a large number of patents for this technology. A detailed review of digital twin applications in various industries was carried out: aerospace, automotive, oil, chemical, energy, urban planning, water supply, construction, retail, and cultural heritage preservation. Links to specifically focused reviews are provided in case of need to learn more about the use of digital twins in a particular industry.

The global digital twin industry was estimated to be worth USD 6.5 billion in 2021 and is predicted to reach USD 125.7 billion by 2030. The rise of IoT and cloud technologies, as well as the motivation to reduce costs and shorten product development time, are key factors contributing to this growth.

REFERENCES

1. H. Lasi, P. Fettke, HG. Kemper et al., “Industry 4.0,” *Business & Information Systems Engineering*, vol. 6, pp. 239–242, Aug 2014. doi: 10.1007/s12599-014 0334-4.
2. D. Jones, C. Snider, A. Nassehi, J. Yon, and B. Hicks, “Characterising the Digital Twin: A systematic literature review,” *CIRP Journal of Manufacturing Science and Technology*, vol. 29, pp. 36–52, 2020. doi: 10.1016/j.cirpj.2020.02.002.
3. M. Liu, S. Fang, H. Dong, and C. Xu, “Review of digital twin about concepts, technologies, and industrial applications,” *Journal of Manufacturing Systems*, 2020. doi: 10.1016/j.jmsy. 2020.06.017.
4. M. Enders and N. Hoßbach, *Dimensions of Digital Twin Applications - A Literature Review*. 2019.
5. Q. Liu, B. Liu, G. Wang, and C. Zhang, “A comparative study on digital twin models,” *AIP Conference Proceedings*, vol. 2073, no. 1, 2019. doi: 10.1063/1.5090745.
6. F. Tao, H. Zhang, A. Liu, and A.Y.C. Nee, “Digital Twin in Industry: State-of-the-Art,” *IEEE Transactions on Industrial Informatics*, vol. 15, pp. 2405–2415, 2019. doi: 10.1109/TII.2018.2873186.
7. C. Semeraro, M. Lezoche, H. Panetto, and M. Dassisi, “Digital twin paradigm: A systematic literature review,” *Computers in Industry*, vol. 130, p. 103–469, 2021. doi: 10.1016/j.compind.2021.103469.

8. M. Grieves and J. Vickers, "Digital Twin: Mitigating Unpredictable, Undesirable Emergent Behavior in Complex Systems," in *Transdisciplinary Perspectives on Complex Systems: New Findings and Approaches*, F.-J. Kahlen, S. Flumerfelt, and A. Alves, Eds. Cham: Springer International Publishing, 2017, pp. 85–113. doi: 10.1007/978-3-319-38756-7_4.
9. David Gelernter, *Mirror Worlds: Or: The Day Software Puts the Universe in a Shoebox...How It Will Happen and What It Will Mean*. Oxford University Press, '233Index' 11, 1991.
10. K. Främling, T. Ala-Risku, M. Kärkkäinen, and J. Holmström, "Agent-based model for managing composite product information," *Computers in Industry*, vol. 57, no. 1, pp. 72–81, 2006. doi: 10.1016/j.compind.2005.04.004.
11. M. Grieves, "Product lifecycle management: the new paradigm for enterprises," *International Journal of Product Development*, vol. 2, 2005. doi: 10.1504/IJPD.2005.006669.
12. "DRAFT Modeling, Simulation, Information Technology & Processing Roadmap Technology Area 11," *NASA Report*, November, 2010.
13. E.J. Tuegel, A.R. Ingrassia, T.G. Eason, and S.M. Spottswood, "Reengineering Aircraft Structural Life Prediction Using a Digital Twin," *International Journal of Aerospace Engineering*, vol. 2011, pp. 154–198, Oct. 2011. doi: 10.1155/2011/154798.
14. E. Glaessgen and D. Stargel, *The digital twin paradigm for future NASA and U.S. air force vehicles*. 2012. doi: 10.2514/6.2012-1818.
15. H. Kagermann, J. Helbig, A. Hellinger, and W. Wahlster, *Recommendations for Implementing the Strategic Initiative INDUSTRIE 4.0: Securing the Future of German Manufacturing Industry; Final Report of the Industrie 4.0 Working Group*. 2013.
16. A. Parrott and L. Warshaw, "Industry 4.0 and the digital twin Manufacturing meets its match," *Deloitte*, pp. 1–17, May 2017. Available: <https://www2.deloitte.com/us/en/insights/focus/industry-4-0/digital-twin-technology-smart-factory.html>
17. C. Monsone, E. Mercier-Laurent, and J. Dr. Jósvei, "The Overview of Digital Twins in Industry 4.0: Managing the Whole Ecosystem," in *Proceedings of the 11th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management*, vol. 01C3K, pp. 271–276, 2019. doi: 10.5220/0008348202710276.
18. M. Grieves, *Digital Twin: Manufacturing Excellence through Virtual Factory Replication*. 2015.
19. M. Grieves, *Origins of the Digital Twin Concept*. 2016. doi: 10.13140/RG.2.2.26367.61609
20. N. Pankratova and I. Golinko, "Digital Twin Simulation for Cyber-Physical Systems on Electric Heater Example," *2022 IEEE 3rd International Conference on System Analysis & Intelligent Computing (SAIC), Kyiv, Ukraine, 2022*, pp. 1–6. doi: 10.1109/SAIC57818.2022.9922971.
21. N.D. Pankratova and I.M. Golinko, "Development of digital twin based on a model with fractional-rational uncertainty," *Computer Modeling and Intelligent Systems 2023. Proceedings of The Sixth International Workshop on Computer Modeling and Intelligent Systems (CMIS 2023) Zaporizhzhia, Ukraine, May 3, 2023*, pp. 11–22. Available: <https://doi.org/10.32782/cmish/3392-2>
22. S. Ferguson, "Five reasons why Executable Digital Twins are set to dominate engineering in 2023", *Siemens*, January 2023. Available: <https://blogs.sw.siemens.com/simcenter/the-executable-digital-twin/>
23. R. Asimov, S. Chernoshey, I. Kruse, and V. Asipovich, *Digital Twin in the Analysis of a Big Data*. 2018.
24. K. Kuehner, R. Scheer, and S. Straßburger, "Digital Twin: Finding Common Ground – A Meta-Review," *Procedia CIRP*, vol. 104, pp. 1227–1232, 2021. doi: 10.1016/j.procir.2021.11.206.
25. K. Kukushkin, Y. Ryabov, and A. Borovkov, "Digital Twins: A Systematic Literature Review Based on Data Analysis and Topic Modeling," *Data*, vol. 7, no. 12, 2022. doi: 10.3390/data7120173.
26. *White House*. 2022. Available: <https://www.whitehouse.gov/wp-content/uploads/2022/10/National-Strategy-for-Advanced-Manufacturing-10072022.pdf>

27. A. Deguchi et al., *What Is Society 5.0?*. 2020. doi: 10.1007/978-981-15-2989-4_1.
28. J. Wang, H. Wu, and Y. Chen, “Made in China 2025 and manufacturing strategy decisions with reverse QFD,” *International Journal of Production Economics*, vol. 224, pp. 107539, 2020, doi: 10.1016/j.ijpe.2019.107539
29. *Factories of the future. Multi-annual roadmap for the contractual PPP under Horizon 2020*. European Factories of the Future Research Association, Report, 2020.
30. J. Lee, B. Bagheri, and H.-A. Kao, “A Cyber-Physical Systems architecture for Industry 4.0-based manufacturing systems,” *SME Manufacturing Letters*, vol. 3, 2014. doi: 10.1016/j.mfglet.2014.12.001.
31. F. Tao, J. Cheng, Q. Qi, M. Zhang, H. Zhang, and F. Sui, “Digital twin-driven product design, manufacturing and service with big data,” *The International Journal of Advanced Manufacturing Technology*, vol. 94, no. 9, pp. 3563–3576, Feb. 2018. doi: 10.1007/s00170-017-0233-1.
32. D. Ivanov, “Predicting the impacts of epidemic outbreaks on global supply chains: A simulation-based analysis on the coronavirus outbreak (COVID-19/SARS-CoV-2) case,” *Transp. Res. E. Logist. Transp. Rev.*, vol. 136, p. 101922, Mar. 2020. doi: 10.1016/j.tre.2020.101922.
33. W. Kritzinger, M. Karner, G. Traar, J. Henjes, and W. Sihn, “Digital Twin in manufacturing: A categorical literature review and classification,” *IFAC-PapersOnLine*, vol. 51, pp. 1016–1022, 2018. doi: 10.1016/j.ifacol.2018.08.474.
34. F. Tao, H. Zhang, A. Liu and A.Y.C. Nee, “Digital Twin in Industry: State-of-the-Art,” in *IEEE Transactions on Industrial Informatics*, vol. 15, no. 4, pp. 2405–2415, April 2019. doi: 10.1109/TII.2018.2873186.
35. M. Grieves and J. Vickers, *Digital Twin: Mitigating Unpredictable, Undesirable Emergent Behavior in Complex Systems*. 2017, pp. 85–113. doi: 10.1007/978-3-319-38756-7_4.
36. A.L.J. Walton, C.L. Tomovic, and M.W. Grieves, “Product Lifecycle Management: Measuring What Is Important – Product Lifecycle Implementation Maturity Model,” in *Product Lifecycle Management for Society*, pp. 406–421, 2013.
37. F. Tao, Q. Qi, L. Wang, and A. Nee, “Digital Twins and Cyber–Physical Systems toward Smart Manufacturing and Industry 4.0: Correlation and Comparison,” *Engineering*, vol. 5, pp. 653–661, 2019. doi: 10.1016/j.eng.2019.01.014.
38. K.-J. Wang, T.-L. Lee and Y. Hsu, “Revolution on digital twin technology—a patent research approach,” *The International Journal of Advanced Manufacturing Technology*, vol. 107, pp. 4687–4704, 2020. doi: 10.1007/s00170-020-05314-w.
39. M. Singh, E. Fuenmayor, E. Hinchy, Y. Qiao, N. Murray, and D. Devine, “Digital Twin: Origin to Future,” *Applied System Innovation*, vol. 4(2), p. 36, 2021. doi: 10.3390/asi4020036.
40. L. Li, S. Aslam, A. Wileman and S. Perinpanayagam, “Digital Twin in Aerospace Industry: A Gentle Introduction,” in *IEEE Access*, vol. 10, pp. 9543–9562, 2022. doi: 10.1109/ACCESS.2021.3136458.
41. T. Wang and Z. Liu, “Digital Twin and Its Application for the Maintenance of Aircraft,” in *Handbook of Nondestructive Evaluation 4.0*, N. Meyendorf, N. Ida, R. Singh, and J. Vrana, Eds. Cham: Springer International Publishing, 2022, pp. 1035–1052. doi: 10.1007/978-3-030-73206-6_7.
42. *Rolls-Royce Official Website*. 2019. Available: <https://www.rolls-royce.com/media/our-stories/discover/2019/how-digital-twin-technology-can-enhance-aviation.aspx>.
43. *Siemens Official Website*. Available: <https://www.siemens.com/global/en/markets/automotive-manufacturing/digital-twin-product.html>
44. *Siemens Official Website*. Available: <https://new.siemens.com/global/en/company/stories/industry/getting-to-market-quickly.html>
45. *Renault Official Website*. Available: <https://www.renaultgroup.com/en/news-on-air/news/vehicle-digital-twin-when-physical-and-digital-models-unite/>
46. M. Sadiku, M. Tembely, and S. Musa, “Internet of Vehicles: an Introduction,” *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 8, p. 11, 2018. doi: 10.23956/ijarsse.v8i1.512.
47. B. Poudel, “Smart Electric Vehicle Charging in the Era of Internet of Vehicles, Emerging Trends, and Open Issues,” *Energies*, vol. 15, 2022. doi: 10.3390/en15051908.

48. B. Marr, "The best examples of digital twins everyone should know about," *Forbes Online*. Available: <https://www.forbes.com/sites/bernardmarr/2022/06/20/the-best-examples-of-digital-twins-everyone-should-know-about/?sh=520cdb3d225f>.
49. S. Malakuti et al., *Digital Twins for Industrial Applications. Definition, Business Values, Design Aspects, Standards and Use Cases*. 2020
50. *Ericsson Official Website*. Available: <https://www.ericsson.com/en/blog/2021/11/how-digital-twins-in-the-oil-and-gas-industry-can-modernize-your-business>.
51. D. Cameron, A. Waaler, and T. Komulainen, "Oil and Gas digital twins after twenty years. How can they be made sustainable, maintainable and useful?" in *Conf. Exergy Analysis for Combined Heat and Power (CHP) Plants, 2018*. doi: 10.3384/ecp181539.
52. *Siemens Official Press*. Available: <https://press.siemens.com/global/en/pressrelease/atos-and-siemens-introduce-digital-twin-solution-within-global-pharmaceutical-industry>
53. A. Cocchia, "Smart and Digital City: A Systematic Literature Review," in *Smart City: How to Create Public and Economic Value with High Technology in Urban Space*, R. P. Dameri and C. Rosenthal-Sabroux, Eds. Cham: Springer International Publishing, 2014, pp. 13–43. doi: 10.1007/978-3-319-06160-3_2.
54. *United Nations (DESA): The 2018 Revision of World Urbanization Prospects*. Available: <https://esa.un.org/unpd/wup>
55. J. Yan, S. Zlatanova, M. Aleksandrov, A. Diakité, and C. Pettit, *Integration of 3D objects and terrain for 3D modelling supporting the digital twin*. 2019.
56. *Singapore government programme "Virtual Singapore"*. Available: <https://www.nrf.gov.sg/programmes/virtual-singapore>
57. G. Schrotter and C. Hürzeler, "The Digital Twin of the City of Zurich for Urban Planning," *PGF – Journal of Photogrammetry Remote Sensing and Geoinformation Science*, vol. 88, 2020. doi: 10.1007/s41064-020-00092-2.
58. "Digital Twin Cities: Framework and Global Practices", *World Economic Forum. Report April 2022*. Available: https://www3.weforum.org/docs/WEF_Global_Digital_Twin_Cities_Framework_and_Practice_2022.pdf
59. E. Shahat, C.T. Hyun, and C. Yeom, "City Digital Twin Potentials: A Review and Research Agenda," *Sustainability*, vol. 13, no. 6, 2021. doi: 10.3390/su13063386.
60. "Digital Water: Operational digital twins in the urban water sector", *International Water Association*, 2021. Available: <https://iwa-network.org/wp-content/uploads/2021/03/Digital-Twins.pdf>
61. P. Conejos, F. Martínez Alzamora, M. Hervás Carot, and J. Alonso Campos, *Development and Use of a Digital Twin for the Water Supply and Distribution Network of Valencia (Spain)*. 2019.
62. P. Conejos, F. Martínez Alzamora, M. Hervás Carot, and J. Alonso Campos, "Building and exploiting a Digital Twin for the management of drinking water distribution networks", *Urban Water Journal*, vol. 17, pp. 1–10, 2020. doi: 10.1080/1573062X.2020.1771382.
63. "Águas do Porto: Douro River Basin," *International Water Association*. Available: <https://iwa-network.org/aguas-do-porto/>
64. *Siemens official website*. Available: <https://assets.new.siemens.com/siemens/assets/api/uuid:4398972e-3e6f-46f5-8e38-077a07ea4af8/HQSIPR202302146651EN.pdf>
65. F. Arraño-Vargas and G. Konstantinou, "Modular Design and Real-Time Simulators Toward Power System Digital Twins Implementation," *IEEE Transactions on Industrial Informatics*, vol. 19, no. 1, pp. 52–61, 2022. doi: 10.1109/TII.2022.3178713.
66. M. Zhou, J. Yan and D. Feng, "Digital twin framework and its application to power grid online analysis," in *CSEE Journal of Power and Energy Systems*, vol. 5, no. 3, pp. 391–398, 2019. doi: 10.17775/CSEEJPES.2018.01460.
67. D.-G.J. Opoku, S. Perera, R. Osei-Kyei, and M. Rashidi, "Digital twin application in the construction industry: A literature review," *Journal of Building Engineering*, vol. 40, p. 102726, 2021. doi: 10.1016/j.jobee.2021.102726.
68. M. El Jazzar, M. Piskernik, and H. Nassereddine, *Digital Twin in Construction: an Empirical Analysis*. 2020.

69. Y. Tchana, G. Ducellier, and S. Remy, "Designing a unique Digital Twin for linear infrastructures lifecycle management," *Procedia CIRP*, vol. 84, pp. 545–549, 2019. doi: 10.1016/j.procir.2019.04.176.
70. F. Tao, J. Cheng, Q. Qi, M. Zhang, H. Zhang, and F. Sui, "Digital twin-driven product design, manufacturing and service with big data," *The International Journal of Advanced Manufacturing Technology*, vol. 94, no. 9, pp. 3563–3576, 2018. doi: 10.1007/s00170-017-0233-1.
71. Z.-Z. Hu, P.-L. Tian, S.-W. Li, and J.-P. Zhang, "BIM-based integrated delivery technologies for intelligent MEP management in the operation and maintenance phase," *Advances in Engineering Software*, vol. 115, pp. 1–16, 2018. Available: <https://doi.org/10.1016/j.advengsoft.2017.08.007>
72. D.S. Vijayakumar, "Chapter Eleven - Digital twin in consumer choice modeling," in *The Digital Twin Paradigm for Smarter Systems and Environments: The Industry Use Cases*, vol. 117. Elsevier, 2020, pp. 265–284. doi: 10.1016/bs.adcom.2019.09.010.
73. R. Shivatsava, "How Digital Twin Technology Can Help Retail CIOs," *Forbes Online*. 2022. Available: <https://www.forbes.com/sites/forbestechcouncil/2022/11/11/how-digital-twin-technology-can-help-retail-cios/?sh=2449caf32836>
74. *Weiss AG company official website*. Available: <https://weiss-ag.com/museumgallery/>
75. W. Luther, N. Baloiian, D. Biella, and D. Sacher, "Digital Twins and Enabling Technologies in Museums and Cultural Heritage: An Overview," *Sensors*, vol. 23, p. 1583, 2023. doi: 10.3390/s23031583.
76. "How the Port of Rotterdam is using IBM digital twin technology to transform itself from the biggest to the smartest," *Official IBM website*. Available: <https://www.ibm.com/blog/iot-digital-twin-rotterdam/>
77. "Highlights annual report 2022 – Port of Rotterdam Authority," *Port of Rotterdam Official Report*, 2022.
78. P. Bauer, B. Stevens, and W. Hazeleger, "A digital twin of Earth for the green transition" *Nature Climate Change*, vol. 11, pp. 1–4, 2021. doi: 10.1038/s41558-021-00986-y.

Received 01.05.2023

INFORMATION ON THE ARTICLE

Nataliya D. Pankratova, ORCID: 0000-0002-6372-5813, Educational and Research Institute for Applied System Analysis of the National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine, e-mail: natalidmp@gmail.com

Kostiantyn D. Grishyn, ORCID: 0009-0006-5950-3739, Educational and Research Institute for Applied System Analysis of the National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine, e-mail: constantine1223h@gmail.com

Viesta E. Barilko, ORCID: 0009-0002-3903-899X, Educational and Research Institute for Applied System Analysis of the National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine, e-mail: vesta8904@gmail.com

ЦИФРОВІ ДВІЙНИКИ: ЕТАПИ РОЗВИТКУ КОНЦЕПЦІЇ, ГАЛУЗІ ВИКОРИСТАННЯ, ПЕРСПЕКТИВНІСТЬ / Н.Д. Панкратова, К.Д. Грішин, В.Є. Барілко

Анотація. Висвітлено результати огляду розвитку концепції цифрових двійників, галузі їх використання, перспективність. Наведено історію виникнення та розвитку концепції цифрового двійника, означення, класифікацію. Відображено актуальність технології, що розглядається. Метою роботи є надання найбільш повної сучасної інформації про поточний стан технології «цифровий двійник», її застосування в різних сферах діяльності людини та подальші перспективи розвитку індустрії. Наведено широку бібліографію з теми, яка може бути корисна дослідникам і представникам різних галузей.

Ключові слова: Індустрія 4.0, цифровий двійник, класифікація, галузі використання, Інтернет речей, фізичні та математичні моделі.

**THE ALGORITHM FOR PREDICTING
THE CRYPTOCURRENCY RATE TAKING INTO ACCOUNT
THE INFLUENCE OF POSTS OF A GROUP OF FAMOUS
PEOPLE IN SOCIAL NETWORKS**

P. BIDYUK, O. GAVRILENKO, M. MYAGKYI

Abstract. This article presents an algorithm for predicting the rate of a selected cryptocurrency, taking into account the posts of a group of famous people in a particular social network. The celebrities chosen as experts, i.e., famous personalities whose posts on social networks were studied, are either familiar with the financial industry, particularly the cryptocurrency market, or some cryptocurrency. The dataset used was the actual rates of the cryptocurrency in question for the selected period and the statistics of expert posts in the selected social network. The study used methods such as the full probability formula and the Bayesian formula. It was found that posts by famous people on social media differently affected cryptocurrency rates. The “main” expert was identified, and his posts were used to forecast the selected cryptocurrency’s rate.

Keywords: cryptocurrency exchange rate, forecasting algorithms, social media posts, group of experts, “main” expert, information technology of intelligent analysis.

INTRODUCTION

The study of cryptocurrency changes is gaining more and more popularity every day due to the relative ease of entry and the abundance of recommendation information regarding the process. Buying and selling cryptocurrencies is a rather interesting process, since, if certain conditions are met, you can increase your wealth several times, or even replace your main job with this business. However, in order to really make money on this process, it is necessary to conduct research on the chosen cryptocurrency, exchange and news related to them.

The relevance is due to the growing popularity of investing in cryptocurrencies. Publications of famous people who have a vested interest in this process have a significant impact on the price formation of certain cryptocurrencies. When traders create forecasts of changes in the exchange rate of certain cryptocurrencies, they will need a recommendation information system that can analyze the impact of such publications on cryptocurrency changes, which will increase the accuracy of the forecast.

The obtained forecasts can be used by financial market participants to obtain high-quality forecasts of cryptocurrency rates, on the basis of which they will make decisions on its purchase (sale).

LITERATURE ANALYSIS AND PROBLEM STATEMENT

The task of analyzing publications from the Internet is very important, since a well analyzed publication can provide a lot of different information.

Article [1] discusses the process of computer based detection and categorization of opinions expressed in a piece of text to determine whether the writer's attitude towards a particular topic, product, etc. is positive, negative, or neutral. A detailed study of sentiment analysis and its cause-and-effect relationship. Using sentiment analysis, you can get a generalized event based on mood and time. On the other hand, the use of causality will be useful not only for determining the causes and effects, respectively, but also for their further forecasting. The main part of the article is an overview of the combination of these two approaches, which degenerates into a model that allows to determine the mood for future events, as well as to create a temporal forecast of the time that will pass between certain events. To assess the accuracy, we used the following statistic: average relative error.

To view publications, you need to choose a place where there are the most of them and they are in a single text format, for this purpose Twitter is a good messenger, work [2] discusses in detail the special linguistic analysis and statistics of Twitter. This study aimed to identify criminal elements in the United States by modeling topics of discussion and then incorporating them into a crime prediction model. Thus, the study was conducted on the impact of social media posts on future crimes.

In [3], methods for predicting user ratings of individual items using probabilistic algorithms were considered. In fact, the article perfectly illustrates the existence of computational patterns in terms of what exactly network users like under certain circumstances. In other words, this study emphasizes the impact of probabilistic algorithms in the field of recommender systems, and provides an overview of key methods that have been successfully applied. The considered algorithms for object classification allow solving the problem of predicting user evaluation of content and its categorization, as well as improving existing methodologies for building information systems.

Article [4] is quite relevant today due to the difficult epidemiological situation in the world. It analyzed microblogs on Twitter and proposed several methods for identifying messages. It was determined that over ten weeks of more than five hundred thousand reports, their best model achieved a correlation of 0.78 with CDC statistics.

It is also necessary to highlight Internet blogs, where many people express their own opinions and visions of certain problems, etc. Therefore, in [5], a study was conducted to identify hate groups. The proposed approach is semi-automatic and consists of four modules, namely: blog spider, information retrieval, network analysis, and visualization. The study was conducted on the Xanga blog site. The results of the analysis were to identify some interesting demographic and topological characteristics in hate groups and to identify at least two large communities in addition to smaller ones. The proposed approach is also appropriate for studying hate groups and other related communities on blogs.

For business, the process of analyzing large amounts of data and understanding the needs of most people is very important, as it directly affects the company's revenue. Article [6] provides a constructive consideration of the problem of business accumulation of large amounts of data and the problems of their intellectual processing. The author provides a clear definition and explanation of the terms "data mining" and "data intelligence". As a result, an objective conclusion was made about the expediency of using data mining to increase the competitiveness of enterprises.

It should be noted that all of the above works describe the methods used in our study, but do not provide the results of forecasting currency rates, including cryptocurrencies. Accordingly, the factors that influence them were not studied.

In [7], the authors study the main macroeconomic indicators of influence on the US dollar exchange rate in Ukraine: purchase/sale of cash currency, purchase/sale of non-cash currency, balance of purchase/sale of cash and non-cash currency, current year inflation, nominal and real GDP, purchase/sale by bank customers, transactions between banks, gross and net international reserves, unemployment rate, discount (interest) rate, balance of foreign exchange interventions, and volume of transactions of nominal value. The main economic components of exchange rate formation were identified using the principal components method. Using the statistical models ARIMA, Exponential Smoothing and SSA, the values of the selected factors of influence are predicted. The values of exchange rates are forecasted using regression models built by Fast Tree, Fast Forest, Fast Tree Tweedie and Gam algorithms, and the obtained values are tested for accuracy. This article did not forecast cryptocurrency rates specifically and did not study the impact of such a factor as publications in social media.

Article [8] analyzes the methods, areas of application, and approaches to analyzing publications and forecasting events based on the collected data, and also gives the concept of the impact of publications on changes in the cryptocurrency rate. The relevance of the topic is substantiated and the possibilities of appropriate application of the results of the work are described. The main stages of working with event forecasting data are identified, namely: data pre-processing, further analysis and forecasting. This article did not investigate the level of influence of celebrity publications on social media on the cryptocurrency rate. In addition, we considered forecasts based on the posts of only one expert.

Within the framework of the studies cited in [7, 8], information systems were created to implement the above tasks of data mining.

Studies have shown that celebrity posts do have a significant impact on cryptocurrency rates. This can be easily verified using classical statistical analysis tools, in particular, by analyzing the correlations between real and predicted cryptocurrency rates. However, it should be noted that each famous personality – hereinafter referred to as an expert – has a different level of awareness in the financial sector, and is also involved in the process of forming cryptocurrency rates in different ways (i.e. some experts are directly related to a particular cryptocurrency, and some are not), so the level of their influence on the forecasted rate will be different. Therefore, it is advisable to study the level of influence of different experts on the forecasted cryptocurrency rates in order to further rank them. This study will improve the accuracy of cryptocurrency rate forecasts.

The following are recommended as ranking parameters:

- 1) the number of posts of a particular famous person in social networks for the period under consideration;
- 2) the accuracy of the forecasts obtained for each expert in relation to the actual cryptocurrency rate;
- 3) deviations from the respective forecasts obtained without taking into account social media posts.

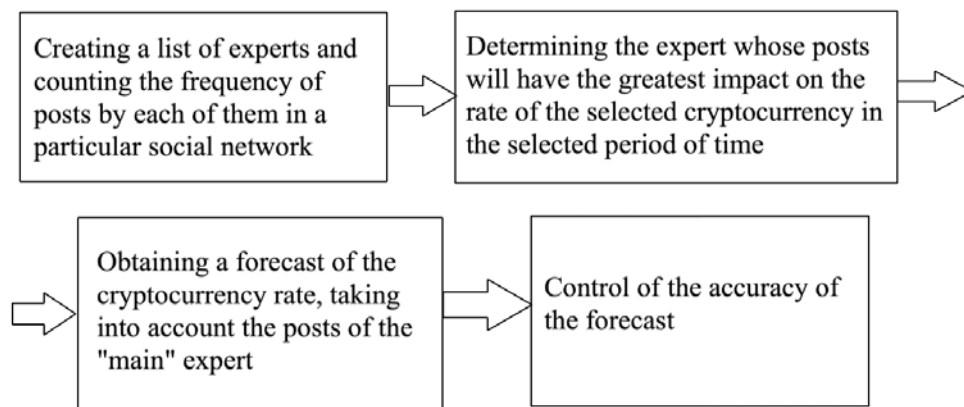
In this article, the number of posts for each of the pre-selected well-known persons in social networks for the period under consideration is taken as such a parameter.

PURPOSE AND OBJECTIVES OF THE STUDY

The purpose of the study is to develop an algorithm for predicting the cryptocurrency rate based on the posts of a group of famous people in social networks.

- This will make it possible to increase the reliability of cryptocurrency rate forecasting.
- To achieve this goal, the following tasks were set:
 - create a list of experts and calculate the frequency of posts by each of them in a particular social network;
 - to identify the expert whose posts will have the greatest impact on the rate of the selected cryptocurrency in a selected period of time;
 - to obtain a forecast of the cryptocurrency rate taking into account the posts of the “main” expert;
 - control the accuracy of the forecast.

Figure shows the process of calculating the rate forecast for the selected cryptocurrency:



The process of calculating the cryptocurrency rate forecast

MATERIALS AND METHODS OF THE STUDY

The object of the study is the forecast of cryptocurrency rates.

The information required to analyze the level of influence of social media posts on cryptocurrency rates is a list of experts whose level of influence will be studied, the time interval of the study, the number of posts made by each of the experts in question during the specified period of time, as well as the actual cryptocurrency rates for the relevant period.

The use of mathematical tools based on the full probability and Bayesian formulas allows us to use this information to determine the expert who is more likely to make posts during the period under consideration. We will call this expert the “main” expert.

So, the following information is required as input: a list of experts, the number of posts by each expert, and the real rates of the selected cryptocurrency for the period under consideration.

Table 1. Fragment of the input data

Hours	Real courses
1	467
2	475
3	516
4	533
5	508
6	510
7	525
8	512
9	514
10	514

Table 2. Fragment of input data (continued)

Expert	Number of posts	Number of posts related to cryptocurrency
Expert 1	9	6
Expert 2	7	5
Expert 3	4	2

Experts were selected as well-known personalities who are either knowledgeable in the field of finance in general and cryptocurrencies in particular, or whose activities are somehow related to a particular cryptocurrency, or not.

A fragment of the dataset is shown in Table 1. Table 1 shows the real rates of the selected cryptocurrencies, which were taken from the website of the Binance crypto exchange [9].

Table 2 shows the number of posts by the selected experts in 10 hours on the social network.

The data generated in this way is the input for this study. As part of the study, it is necessary to:

- create a list of experts and count the number of their posts on social media;
- determine the “main” expert using the Bayesian formula;
- to obtain a forecast of the cryptocurrency rate, taking into account the posts of the “main” expert, based on the approach described in [8].
- to calculate the accuracy of MAPE forecasts.

The use of these methods guarantees reliable results when predicting cryptocurrency rates and studying the level of influence on them by posts of famous people in social networks.

To conduct statistical analysis and obtain results based on these methods, the corresponding software was developed.

ALGORITHM FOR TAKING INTO ACCOUNT THE LEVEL OF INFLUENCE OF POSTS BY SEVERAL FAMOUS PEOPLE IN SOCIAL NETWORKS ON THE CRYPTOCURRENCY RATE

Forming a list of experts and counting the frequency of posts by each of them in a particular social network

Task statement. From the set of users of a social network, we select a subset $A = (a_1, a_2, \dots, a_n)$ of users who satisfy the following requirements:

- 1) the users are famous persons;

2) they are active users of the social networks and have a large number of subscribers;

3) they'll have different primary professional interests;

4) all pairs of users a_i and a_r , $i, r = 1, 2, \dots, n$, do not maintain communication in the network (they are not friends, do not respond to each other's posts).

We call such users experts.

Suppose that over a certain period of time, experts have made m posts in a social network, and k of them are related to a certain cryptocurrency. We consider the context of the posts to be arbitrary. For the specified period of time, expert a_1 published m_1 posts, of which k_1 posts are related to a certain cryptocurrency, expert a_2 published m_2 posts, of which k_2 posts are related to a certain cryptocurrency, ..., expert a_n published m_n posts, of which k_n posts are related to a certain cryptocurrency:

$$m_1 + m_2 + \dots + m_n = m;$$

$$k_1 + k_2 + \dots + k_n = k,$$

where m_1, m_2, \dots, m_n are the frequencies of expert posts; k_1, k_2, \dots, k_n are the frequencies of expert posts related to a particular cryptocurrency, where i is the number of the expert, $i = 1, 2, \dots, n$.

It is necessary to calculate the frequencies of posts of all selected experts for an arbitrary time interval [10].

Rationale. This choice of experts is due to the need to form the set of such experts who will be independent of each other both in the space of the chosen social network and in the professional space.

Results. From the data presented in Table 2, it can be seen that the considered set of 3 experts $A = (\text{expert 1, expert 2, expert 3})$. According to the social networks data, it is known that $m = 20$ posts were published over a period of 10 hours, with $k = 13$ posts related to the selected cryptocurrency: $m_1 = 9$, $m_2 = 7$, $m_3 = 4$, and $k_1 = 6$, $k_2 = 5$, $k_3 = 2$.

Determining the expert whose posts will have the greatest impact on the rate of the selected cryptocurrency in a selected period of time

Task statement. Based on the list of experts $A = (a_1, a_2, \dots, a_n)$ obtained in section 1 and taking into account the frequencies of their posts in the selected social network for a specified small period of time — m_1, m_2, \dots, m_n , and k_1, k_2, \dots, k_n , it is necessary to determine the “main” expert.

The formulated problem can be easily interpreted as a classical probabilistic problem: m posts were written in a certain period of time. It is known that n experts published posts during this period, where m_1, m_2, \dots, m_n are the frequencies of expert posts, k_1, k_2, \dots, k_n are the frequencies of expert posts related to the chosen cryptocurrency, where i is the number of the expert, $i = 1, 2, \dots, n$. Action A is that in an arbitrary period of time someone wrote the post related to the se-

lected cryptocurrency. It is necessary to determine which expert is more likely to have made this post [11].

Rationale:

A = (the post related to the selected cryptocurrency was written at any time t from the interval $[0;T]$),

H_1 = (the post was written by expert 1),

H_2 = (the post was written by expert 2),

...

H_n = (the post was written by expert n).

We assume that actions H_i are pairwise independent, where i is the number of the expert, $i=1,2,\dots,n$. These assumptions can be made based on a list of requirements that experts must meet (see section 1).

According to the full probability formula:

$$P(A) = \sum_{i=1}^n P(H_i)P(A/H_i), \tag{1}$$

where

$$P(H_i) = \frac{m_i}{m}, \tag{2}$$

where m_i is the number of publications made by the expert i , and k_i is the number of publications made by the expert i related to the selected cryptocurrency, m is the total number of publications for the period $[0;T]$, $P(H_i)$ is the probability that the post was published by expert i , $P(A/H_i)$ is the probability that at any point in time t the post related to the selected cryptocurrency was written, provided that the post was written by expert i , $i=1,2,\dots,n$.

Then, using the Bayesian formula, we calculate the probability for each expert that he or she made the post, if it is known that the post was made during the period under consideration:

$$P(H_i/A) = \frac{P(H_i)P(A/H_i)}{P(A)}, \tag{3}$$

where $P(H_i)$ is the probability that the post was published by expert i , $P(A/H_i)$ is the probability that at any point in time t the post related to the selected cryptocurrency was written, provided that the post was written by expert i , $P(H_i/A)$ is the probability that the post was written by expert i , provided that it is known that at any point in time t the post related to the selected cryptocurrency was written.

Among the obtained a posteriori probabilities, the highest one is chosen. This means that this expert is most likely to have published a post in the time period under consideration and thus will have a greater impact on the rate of the selected cryptocurrency. It is this expert that will be considered the “main” expert for the forecasted time period $[T;T + \Delta t]$. This is due to the fact that the influence of the posts made during the time period $[0;T]$ also extends to a certain time period $[T;T + \Delta t]$.

It should also be noted that the obtained a posteriori probabilities can be further used to find average estimates of the effectiveness of predictive adaptive algorithms for changing the cryptocurrency rate under the influence of the sequential appearance of individual or group posts of experts over time. The creation of such algorithms and, as a result, the corresponding intelligent technologies is the subject of further research by the authors.

Result. As mentioned in section 1, according to the data presented in Table 2, we consider the set of 3 experts $A = (\text{expert 1, expert 2, expert 3})$. According to the social networks data, it is known that 20 posts were published during a period of 10 hours, 13 posts related to the selected cryptocurrency, with $m_1 = 9, m_2 = 7, m_3 = 4$ and $k_1 = 6, k_2 = 5, k_3 = 2$.

According to formula (2):

$$P(H_1) = \frac{9}{20}, \quad P(H_2) = \frac{7}{20}, \quad P(H_3) = \frac{4}{20},$$

$$P(A/H_1) = \frac{6}{9}, \quad P(A/H_2) = \frac{5}{7}, \quad P(A/H_3) = \frac{2}{4}.$$

According to formula (1)

$$P(A) = \frac{6}{20} + \frac{5}{20} + \frac{2}{20} = \frac{13}{20}.$$

According to formula (3):

$$P(H_1/A) = \frac{6}{13}, \quad P(H_2/A) = \frac{5}{13}, \quad P(H_3/A) = \frac{2}{13},$$

so, with probability of $\frac{6}{13}$ the post was most likely made by expert 1. Therefore, he was considered the “main” expert for this period of time. The next expert is expert 2 according to the probability of $\frac{5}{13}$, and the last one is expert 3 according to the probability of $\frac{2}{13}$.

Obtaining a forecast of the cryptocurrency rate taking into account the posts of the “main” expert

Task statement. Let the set of rates of the selected cryptocurrency for the time period $[0;T]$ be known as the set $X = \{x_t\}, t \in [0;T]$. You need to get the set of forecasts of the cryptocurrency rates taking into account the “main” expert chosen in section 1 for the period $[T;T + \Delta t]$ as the set $Y = \{y_t\}, t \in [T;T + \Delta t]$.

Rationale. To obtain forecasts, we will use the ATAPSN (algorithm for taking into account posts in social networks), taking into account the posts of the “main” expert [8].

The idea of the algorithm is to calculate the coefficient of significance of the posts of the “main” expert c_t , at time t from the interval $[T;T + \Delta t]$, which is calculated by the formula:

$$c_t = \delta_t \cdot ch_t, \tag{4}$$

where ch_t is the tone of the “main” expert’s post:

$$ch_t = \begin{cases} 1, & \text{if the post is positive,} \\ 0, & \text{if the post is neutral,} \\ -1, & \text{if the post is negative,} \end{cases} \quad (5)$$

δ_t is the accuracy of the previous forecast,

$$\delta_t = |y_t - x_t|, \quad (6)$$

where y_t is the predicted value of the cryptocurrency exchange rate obtained without taking into account posts on social networks, x_t is the actual value of the cryptocurrency exchange rate, where t is the point in time.

After determining the coefficient c_t from (4)–(6), a forecast of cryptocurrency exchange rate changes will be created based on the available data for the period of time.

$$y'_{t+1} = y_{t+1} + c_t.$$

Result. For the “main” expert identified in section 2, expert 1, a 10-hour forecast of the selected cryptocurrency rate was obtained (see Table 3).

Table 3. Forecast of the selected cryptocurrency rate for the expert 1

Hours	Forecast rates
1	466.19
2	473.44
3	513.78
4	534.78
5	508.31
6	508.6
7	525.67
8	510.85
9	512.55
10	515.43

Control of the accuracy of the obtained forecast

Task statement. For each of the experts selected in section 1, based on the ATAPSN algorithm (see section 3), we made a forecast of the rate of the selected cryptocurrency and indicated the actual rate for the period of time under consideration. The resulting forecasts, along with the actual cryptocurrency rates, are provided in the input dataset (see Table 1).

Based on the input dataset, the following statistical samples X, Y_j ($j = \overline{1,3}$) of size s each (where s is the number of forecasts made at the selected time point) should be formed:

X — the set of real cryptocurrency rates;

Y_1 — the set of predicted cryptocurrency rates obtained using the ATAPSN, taking into account expert 1 posts;

Y_2 — the set of predicted cryptocurrency rates obtained using the ATAPSN, taking into account expert 2 posts;

Y_3 — the set of predicted cryptocurrency rates obtained using the ATAPSN, taking into account expert 3 posts.

For each expert, it is necessary to calculate the accuracy of the MAPE forecast.

Rationale. The average relative sampling error is calculated by the formula:

$$MAPE_j = \frac{1}{s} \sum_{l=1}^s \frac{|x_l - y_{jl}|}{x_l} 100 .$$

where x_l are sample items X , y_{jl} are sample items Y_j , $j = (\overline{1,3})$, $l = (\overline{1,s})$ is the volume of samples X and Y_j [12].

Using this measure of forecast accuracy will allow us to control the quality of the dataset and rank experts in terms of the accuracy of forecasts obtained from their posts (see Table 4).

Table 4. Level of model adequacy

MAPE, %	Forecast accuracy
less than 10	High
10–20	Good
20–40	Satisfactory
40–50	Poor
more than 50	Unsatisfactory

Totally, these values are dependent on the purpose of the forecast. It is up to the researcher to set the limits of the accuracy indicator that satisfy him or her.

In case of low accuracy of the forecast, it is recommended to make changes to the significance factor to make the analysis of further changes more accurate (see section 3).

Result. For samples $Y_1 - Y_3$, we obtain the following values for the coefficients MAPE_j (see Table 5).

Table 5. Values of the coefficients MAPE

X	Y_1	Y_2	Y_3
467	466.19	497.8707174	502.2597
475	473.44	502.1439974	463.9021
516	513.78	506.4172773	490.0742
533	534.78	510.6905573	515.4008
508	508.31	514.9638372	523.135
510	508.6	519.2371172	528.5218
525	525.67	523.5103971	518.3018
512	510.85	527.7836771	515.8382
514	512.55	532.056957	529.6971
514	515.43	536.330237	521.8438
MAPE _j	0.23406855	3.27736587	3.1429464

Table 5 shows that sample Y_1 has the lowest MAPE (0.23%), the next sample Y_3 (3.14%), and the biggest error is in sample Y_2 (3.28%). It should be noted that the accuracy of all forecasts is high, according to Table 4, which indicates the quality of the built forecasting model.

According to the results obtained, it can be stated that in terms of forecast accuracy, expert 1 has the most significant posts, next expert 3, and finally expert 2. The results are fully consistent with the fact that expert 1 was chosen as the “main” expert, whose forecasts are the most significant.

CONCLUSIONS

In this article, we presented a modification of the ATAPSN algorithm [8], which allows taking into account the posts of a group of pre-selected experts and form a list of requirements for them.

This approach allows to increase the accuracy of forecasts of the selected cryptocurrency rates, which has been confirmed statistically.

This approach calculates the a posteriori probabilities that a post related to the selected cryptocurrency was written by a particular expert during the forecasting interval. They were used to determine the “main” expert.

The obtained a posteriori probabilities can be further used to find average estimates of the efficiency of predictive adaptive algorithms for changing cryptocurrency rates under the influence of the sequential appearance of individual or group posts by experts. The creation of such algorithms and, as a result, the corresponding intelligent technologies is the subject of further research by the authors.

It should be noted that there may be different “main” experts at different time intervals.

To use this approach, it is recommended to consider small time intervals, each of which allows you to more accurately determine your “main” expert. This increases the accuracy of forecasts of the selected cryptocurrency rates over the entire time interval.

Using the post frequencies in social networks as a parameter for determining the influence of experts allows us to apply the classical apparatus of probability theory, which guarantees the correctness of the results obtained.

The disadvantages include the fact that the accuracy of the forecast may be negatively affected by an unsuccessfully selected time interval for which the forecast was made, since it is not known in advance how long an expert’s post will be affecting the cryptocurrency rate. This indicates the need for the constant monitoring of both cryptocurrency rates and expert posts on social networks.

The proposed algorithm is an intermediate step towards the creation of a multi-expert model for forecasting cryptocurrency rates.

REFERENCES

1. P. Preethi, V. Uma, and A. Kumar, “Temporal Sentiment Analysis and Causal Rules Extraction from Tweets for Event Prediction,” *Procedia Computer Science*, no. 48, pp. 84–89, 2015. Available: <https://doi.org/10.1016/j.procs.2015.04.154>
2. Matthew S. Gerber, “Predicting crime using Twitter and kernel density estimation,” *Decision Support Systems*, no. 61, pp. 115–125, 2014. Available: <https://doi.org/10.1016/j.dss.2014.02.003>
3. O.V. Gavrilenko and I.V. Kuptsova, “Research of tasks of estimation of objects in computer social networks by means of reliable algorithms [Text],” *materials of the 8th international scientific and technical conference “Modern directions of development of information and communication technologies and means of management”*, Kharkiv, April 26–27, 2018, pp. 67–68.
4. A. Culotta and B. Huberman, “Towards detecting influenza epidemics by analyzing Twitter messages,” *Proceedings of the First Workshop on Social Media Analytics, ACM, 2010*, pp. 115–122.
5. M. Chau and J. Xu, “Mining communities and their relationships in blogs: A study of online hate groups,” *J. Hum.-Comput. Stud.*, 65(1), pp. 57–70, 2007.
6. O.V. Gavrilenko, “Review of methods of hybridization of intellectual information processing,” *Scientific Bulletin of Kherson State University, series “ECONOMIC SCIENCES”*, issue 8, part 3, p. 1, pp. 204–207, 2014.
7. O.V. Gavrylenko, K.D. Novakivska, and O.A. Shumeiko, “Selection of the most influential economic factors for forecasting the US dollar exchange rate,” *NTU Bulletin*, no. 54, pp. 26–35, 2022. doi: 10.33744/2308-6645-2022-4-54-026-035.
8. O. Gavrylenko, M. Miahkyi, and Y. Zhurakovskiy, “The task of analyzing publications to build a forecast for changes in cryptocurrency rates,” *Adaptive Automatic Control Systems*, vol. 2, no. 41, pp. 90–99, 2022. Available: <https://doi.org/10.20535/1560-8956.41.2022.271349>
9. *Binance cryptocurrency exchange*. [Online]. Available: <https://www.binance.com/en/strategy/spot/grid/DOGEUSDT>. Accessed on: 20 April 2023.
10. M.V. Kartashov, *Probability, processes, statistics*. Kyiv: VPC Kyiv University, 2007, 504 p.
11. V.P. Denysiuk, V.M. Bobkov, T.A. Pogrebetska, and V.K. Repeta, *Higher mathematics. P4. Probability theory and mathematical statistics*. Kyiv: NAU-druk, 2009, 256 p.
12. *Time series*. [Online]. Available: <https://kstat.pnu.edu.ua/wpcontent/uploads/sites/63/2018/04/%D0%A7%D0%B0%D1%81%D0%BE%D0%B2%D1%96-%D1%80%D1%8F%D0%B4%D0%B8.pdf>. Accessed on 20 April 2023.

Received 01.05.2023

INFORMATION ON THE ARTICLE

Petro I. Bidyuk, ORCID: 0000-0002-7421-3565, Educational and Research Institute for Applied System Analysis of the National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Ukraine, e-mail: pbidyuke_00@ukr.net

Olena V. Gavrilenko, ORCID: 0000-0003-0413-6274, National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Ukraine, e-mail: gelena1980@gmail.com

Mykhailo Yu. Myagkyi, ORCID: 0000-0002-8038-8839, National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Ukraine, e-mail: mishamyagkiy3@gmail.com

**АЛГОРИТМ ПРОГНОЗУВАННЯ КУРСУ КРИПТОВАЛЮТИ
З УРАХУВАННЯМ ВПЛИВУ ДОПИСІВ ГРУПИ ВІДОМИХ ЛЮДЕЙ
В СОЦІАЛЬНИХ МЕРЕЖАХ / П.І. Бідюк, О.В. Гавриленко, М.Ю. Мягкий**

Анотація. Наведено алгоритм прогнозування курсу обраної криптовалюти, з урахуванням дописів групи відомих особистостей в конкретній соціальній мережі. Експертами з-поміж них обирали ті, чії дописи в соціальних мережах досліджувалися, та які обізнані з галуззю фінансів, зокрема з ринком криптовалют, або так чи інакше з певною криптовалютою. Як датасет використано реальні курси криптовалюти за обраний період часу та статистику дописів експертів в обраній соціальній мережі. У межах дослідження застосовано такі методи, як формула повної ймовірності та формула Баєсса. З’ясовано, що дописи відомих людей в соціальних мережах по-різному впливають на курси криптовалют. Визначено «основного» експерта з урахуванням дописів якого отримано прогноз курсу обраної криптовалюти.

Ключові слова: курс криптовалюти, алгоритми прогнозування, пости в соціальних мережах, група експертів, «головний» експерт, інформаційна технологія інтелектуального аналізу.

**INFORMATION TECHNOLOGY FOR CREATING
INTELLIGENT COMPUTER PROGRAMS FOR TRAINING
IN ALGORITHMIC TASKS.
PART 2: RESEARCH AND IMPLEMENTATION**

A.S. KULIK, A.G. CHUKHRAY, O.V. HAVRYLENKO

Abstract. Information technologies, particularly artificial intelligence methods, involve more and more deeply into all spheres of human activity: science, technology, art, and education. Ukraine also has sufficient potential and needs to develop educational support, which is the subject of this paper. The work aims to demonstrate the components of information technology for the creation of Intelligent Tutoring Systems (ITS), which are involved in studying various engineering disciplines. The work includes methods of system analysis, mathematical and simulation modeling, technical diagnostics, and artificial intelligence. The proposed models and methods are implemented in ITS prototypes for teaching mathematics, programming, and the automatic control theory. The Intelligent Tutoring Systems were implemented in the educational process of KhAI University and other institutions in Ukraine, Great Britain, Austria, and China. Experimental studies have shown increased student learning success rates using ITS compared to traditional methods. Improved and adapted for computer training methods of technical diagnostics, Bayesian networks, and developed models of algorithmic tasks, the learning process and the learner are valuable from a scientific point of view. In a practical sense, the obtained results can be used to create new specialized ITSs and build an expandable common learning platform integrating the basic disciplines of a specific educational field.

Keywords: intelligent tutoring systems, experimental studies, results of implementation.

INTRODUCTION

The implementation of ITS is one of the highest priority directions of educational tools evolution. This fact is reasoned by numerous advantages of ITS usage over the classical approach: adoption for particular student, wide possibilities of virtual modeling of real objects and processes, decreasing of time and work efforts for completion and verification of tutoring courses, e-learning facilities, etc.

ITSs are characterized by supporting of inner and outer tutoring loops, minimal feedback (prompting hints and advices), nonlinear learning path, dynamic and customizable knowledge base, and self-learning support [1]. ITS usually include three main structural elements: a domain model, a student model and a pedagogical model, but researchers also often incorporate an interface model as the fourth element. Getting into account structure and functionality complexity of ITS, we can conclude that development process requires huge efforts and deep knowledge in programming and the subject area. One of the actual problems regards to providing of adaptive hints in the context of certain subject area. A hint

must be generated according to the place and type of student's mistake and assumes possibility of alpha and beta errors. Moreover, it should provide numerous rules (constraints) for each type of a student's mistake. Program implementation of such as laborious process usually requires a high degree of programming skills. In well-structured domains where solution of the tutoring task can be modeled as a tree, it is possible to skip manual rule generation (constraints) and automatically perform pedagogical decisions.

The goal of the presented work was to develop ITSs for different domains on the basis of system approach [2; 3], models, methods and tools, which were described in part 1 of the paper [4] and to carry out their experimental studies.

EXPERIMENTAL STUDIES

Experimental studies of the proposed method [4] were performed using m-ary trees for the metrics of K. Tai, K. Zhang, D. Shasha – the minimum editing distance between m-ary trees. The performance of the developed NearestHash method was compared with the speed of the “naive” method and the method of Bustos, Navarro, Chavez (BNC).

Fig. 1 shows a comparison of the performance of the NearestHash method with the performance of the BNC method (basic and modified) by input conditions [4]. In the experiment for 10 and 100 trees, the first 10 trees from the original list of trees were searched, and in the experiment with 1000 – the first 50. As can be concluded from the figures, at the first search the developed method surpasses the known and modified known for the first experiment from 10 trees 1.71 and 1.67 times; for the second experiment out of 100 trees – 1.66 and 1.69 times and for the third experiment out of 1000 trees – 3.8 and 2.3 times.

Consider the formation of the set $SSim2$ [4] based on bigrams, i.e. $q = 2$. Let it is given the strings $s0 = s0[1] \cdot s0[2] \cdot s0[3] \cdot \dots \cdot s0[length(s0) - 1] \cdot s0[length(s0)]$, and $snj = snj[1] \cdot snj[2] \cdot snj[3] \cdot \dots \cdot snj[length(snj) - 1] \cdot snj[length(snj)]$, where ‘ \cdot ’ is a concatenation symbol.

The third subtask, which is the set $SSim3$ [4] formation, is solved using a method developed in collaboration with A.Yu. Zavgorodny [5]. The ITS data model and SQL-queries constructed within the method allow to choose adaptively the next task for the student [6].

In the training mode the calculation task process model includes the pedagogical action choice, which is based on the information obtained from the student's model and the current step taken by the student in performing the task j . The various conditions that precede the choice of pedagogical action are:

1. When performing the task j , the student made the wrong step i . In this case, there are two options:
 - a) showing a tip for the student on the basis of activated DM;
 - b) transition to a simpler task, which contains needed KSC.
2. The task is performed correctly by the student. Then there are two possible scenarios:
 - a) transition to the next task of the same class;
 - b) transition to the task of another class.

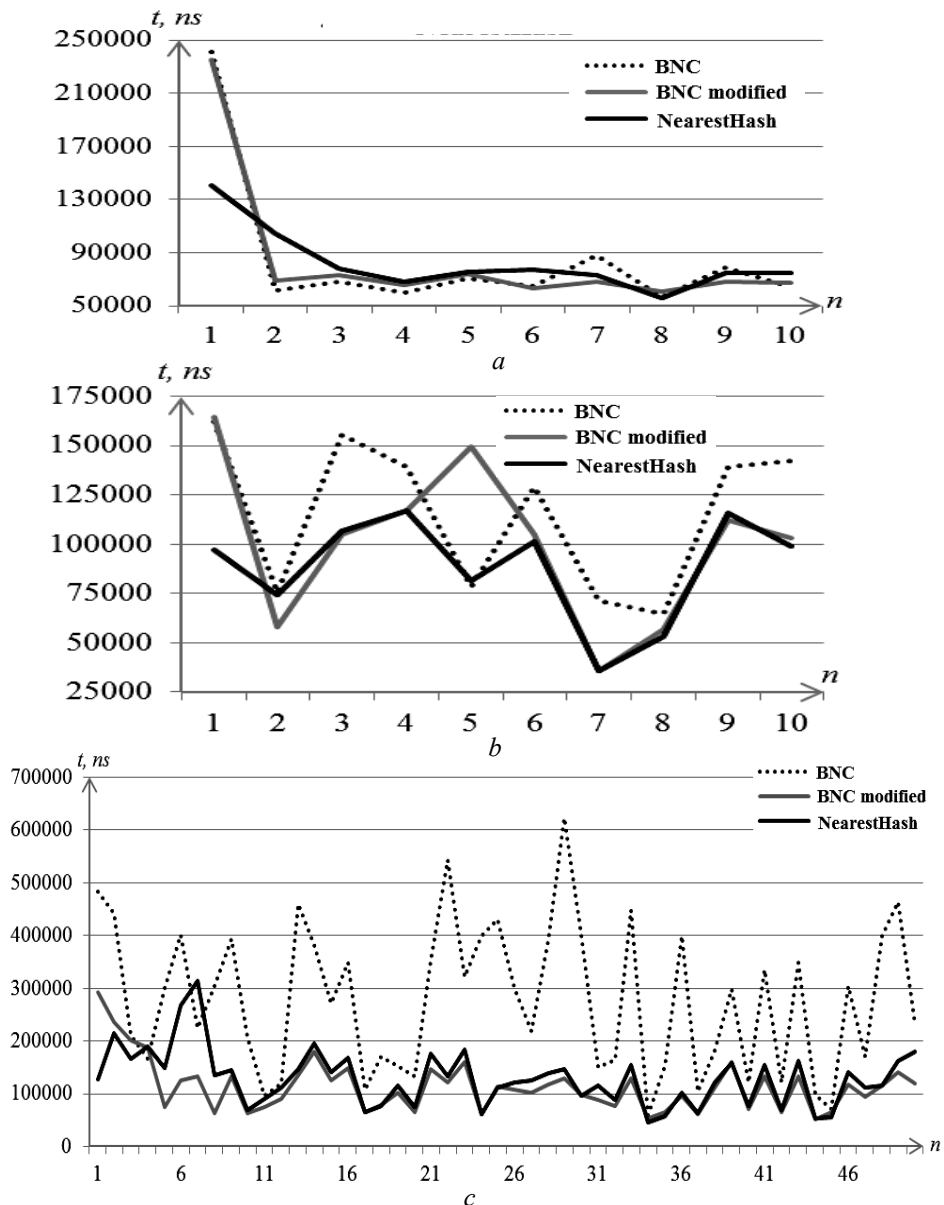


Fig. 1. Comparison of the three methods performance for set, consisting of a — 10, b — 100, c — 1000 randomly generated trees

All these transitions to the next task can be done in two ways:

- a) a task is selected by the program;
- b) a task is selected by the student.

In the case of student's incorrect step i while performing the task j hint on the basis of activated diagnostic models is shown. Then, in the model of the student after BN for the step i is inserted as many BN temporary layers, as incorrect steps have been made. In this case, if we represent the first layer on the left, such as the graph $G = (E, \Gamma)$, then all duplicate layers will be isomorphic to the original graph, and priori values of the probabilities of mastering KSC will be a posteriori values of probabilities for the previous layer.

An automatic transition from the current task with the several KSCs should be made to a simpler task that contains one of the KSCs deal with the appropriate SQL query. You can also jump to any task that contains fewer KSCs than the current task using a single SQL query. And the transition to tasks that contain lower number of KSC, than in the current task, can be done as a result of another SQL-query.

If a student performs the task correctly transition to the next task of the same class can be done using the method of generation and calculation of objects. As working with a class of tasks KSC are the same, then priori values of the mastering them probabilities will be a posteriori probabilities for the last completed task of this class. Thus, the values of the probabilities of mastering the relevant KSC will increase and reach a certain threshold of “mastery”.

If a student solves the task correctly the new task of another class should be chosen from two variants:

- a) a task with at least one KSC, which is new for the student;
- b) a task with KSCs, which already occupied by the student, but is combined in another way.

The above transitions could be performed using SQL queries.

An alternative learning scenario is a scenario using task clustering. Tasks are grouped into clusters so that each cluster has similar tasks, and clusters are sorted by complexity – by the average number of KSCs in cluster tasks or by the average complexity of tasks, where the complexity of the task is defined as the total complexity of the KSC. In this case, the student begins training with the easiest cluster and does not leave it until all tasks in this cluster are completed. A tuple (x_1, x_2, \dots, x_n) is used as a representation of the task for clustering, where n is the total number of KSCs in the problem domain, $x_i \in \{0, 1\}$ — a component of the tuple that indicates the presence or absence of the KSC with number i in the task. Hemming’s metric are usually used as a metric for comparing different tuples. One of the known methods, for example k -means, is used for clustering of tasks. In addition, each task can be matched with a hierarchy of KSC, and each hierarchy – with its string representation. Then NearestHash is used to cluster tasks with a special data structure – a system of disjoint sets and metrics K. Tai, K. Zhang, D. Shasha – the minimum editing distance between trees [7; 8].

Since solution of the considered ATs requires to perform calculations using formulas and algorithms that cover not one specific task, but a whole class of tasks, it is necessary to separately organize training in the knowledge of these formulas and algorithms. In such way the student ability to perform any task from a given class could be improved. Let’s draw the following analogy. Currently, neural networks are considered to be one of the most effective models of artificial intelligence. An artificial neural network is learned from examples, but there is no guarantee that it will work properly with any source data outside the training sample. Similarly, when teaching a person only by examples, there is no guarantee that he has mastered a general rule, algorithm or formula that covers all possible cases.

The program compiled by the student might be exactly the same as the reference program, for example in the case of small tasks. Then, if the texts of these two programs match, then correctness of the solution could be stated. If the texts

do not match, then program testing is required. If the student's program does not match any of the reference programs, then after its testing there are two possible cases: a) the program is incorrect, b) the program is similar to the correct one. If the student's program has passed all the tests and the same or exceeds the existing reference programs in productivity, the ITS can make this program as a new optimal. As criteria of optimality the memory capacity and conciseness should be used as well. Thus, the automatic self-learning of ITS has been realized.

If the student's program does not pass at least one test, the next diagnostic task is solved – finding a place of discrepancy at the lexical level. For this purpose it is used modified Levenstein metric — the minimum editing distance between arrays of lexems of two programs — and modified Wagner–Fisher method to determine this distance as well as the trace of the matrix [8], i.e. the minimum conversion sequence from one program to another.

For the experiment computer Intel Core I5-3210, 2.5 GHz, RAM – 4 GB has been used. The comparison program was written in Java and run in the Eclipse SDK, Version: 3.5.1. As a result of comparison of two programs consisting of 28 bytes, it took 386295 ns. For comparison, when calculating unmodified Levenstein distances by the Wagner–Fisher method for files with a size of 28 bytes, the average calculation time is 443109 ns, for files with a size of 2895 bytes — 135536867 ns \approx 0.1 s, for files of size 7251 bytes — 1250641880 ns \approx 1.25 s, and for files of size 10701 bytes there was an error due to memory overflow.

To solve the problem of diagnosing at the structural level, two abstract syntax trees (AST) are obtained as a result of parsing the reference program and the program compiled by the student. The first way to compare AST is to use the width search method. The second way to compare the AST is to calculate the distance between the trees by the method of K. Zhang, D. Shasha and the editing trace, i.e. the minimum conversion sequence at the structural level. Thus nodes of trees are operands and operators.

If the program compiled by the student does not pass at least one test and the ITS stores more than one reference program for this task, the NearestHash method is used to find the reference program closest (at the lexical or syntactic level) to the student's program.

ALGORITHMS AND SOFTWARE IMPLEMENTATION

The method and models for diagnosing algorithms built by the student in a graphical environment was developed. Game models are given by two models: a model like “biathlon” and a model like “sailing regatta”.

As for the first model, each student performs the calculated AT for a certain period of time $t_{finish} - t_{start}$. Since the task consists of n steps, in the terms of biathlon a step i could be considered as a shooting range i . In addition, for each wrong step a student should get penalty as additional Δt to the total time. Thus, a tutor is able:

- 1) to arrange competitions in the classroom with academic group;
- 2) to monitor students time at each shooting range;
- 3) to determine winners at the end of each championship;
- 4) to determine the absolute record holders after championships series.

The following simplifications were chosen for the second model:

- 1) the yacht moves along a straight line without course change;
- 2) no currents;
- 3) the wind speed is stable;
- 4) the density of environment is constant.

The student’s task is to synthesize the transfer function of the corrective device, which provides the best quality indicators when turning the sail and, as a consequence, when moving the yacht. To determine the effectiveness of the obtained student solution, a step-by-step calculation of the values of velocity over time, as well as the distance traveled. The regatta is visualized: students can compete with each other and with computer models.

The generalized scenario of the “internal cycle” of the ITS for the calculated AT is shown in Fig. 2. The scheme of the student’s requests analysis method is shown in Fig. 3.

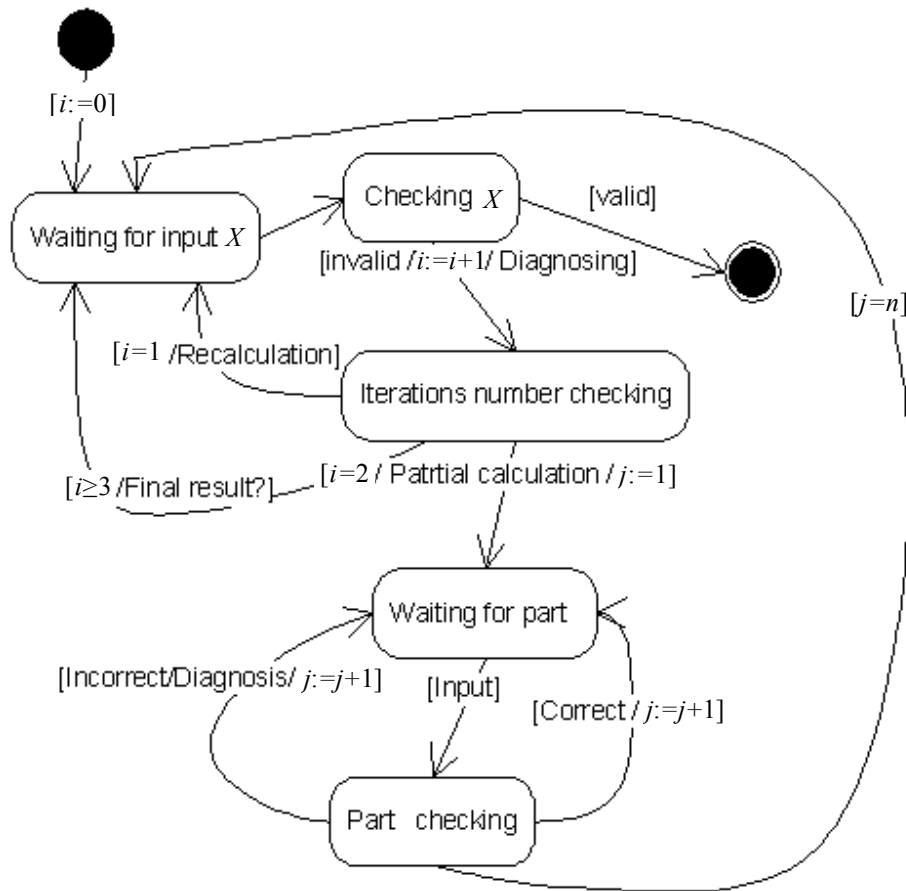


Fig. 2. Generalized diagram of the “internal cycle” of the ITS for the calculated AT

The methods of determining the similarity of SQL-queries are: $q1$ — the classical method of q-grams; $q2$ – the method of q-grams with the previous replacement of keywords in SQL; $q3$ – the method of q-grams derived from syntactic trees. The total similarity of the two SQL queries $qry1$ and $qry2$ is defined

$$\text{as } SIM_{qry1, qry2}^{total} = \sum_{i=1}^3 SIM_{qry1, qry2}^{qi} \text{ [9].}$$

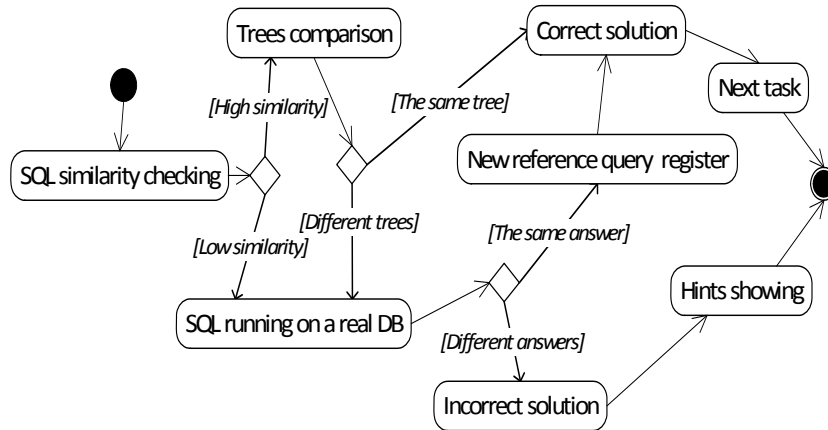


Fig. 3. Diagram of the student's SQL-queries analysis method

Fig. 4 shows the sequence of providing tips for a student. An indication of the error in the query is provided first. If the student is not able to correct the error on his own, then a more detailed indication of the error is provided, namely its location based on a comparison of trees. Next, there is a hint about the KSC that must be used while SQL query construction. Finally, a similar task is demonstrated with explanation of the necessary steps for its implementation.

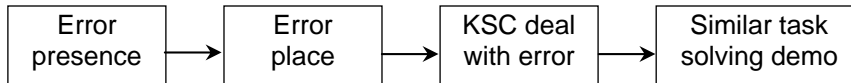


Fig. 4. The sequence of providing tips to the student

The corresponding software was created for the different disciplines of aircraft control systems department, particularly for automatic control theory (ACT). Fig. 5

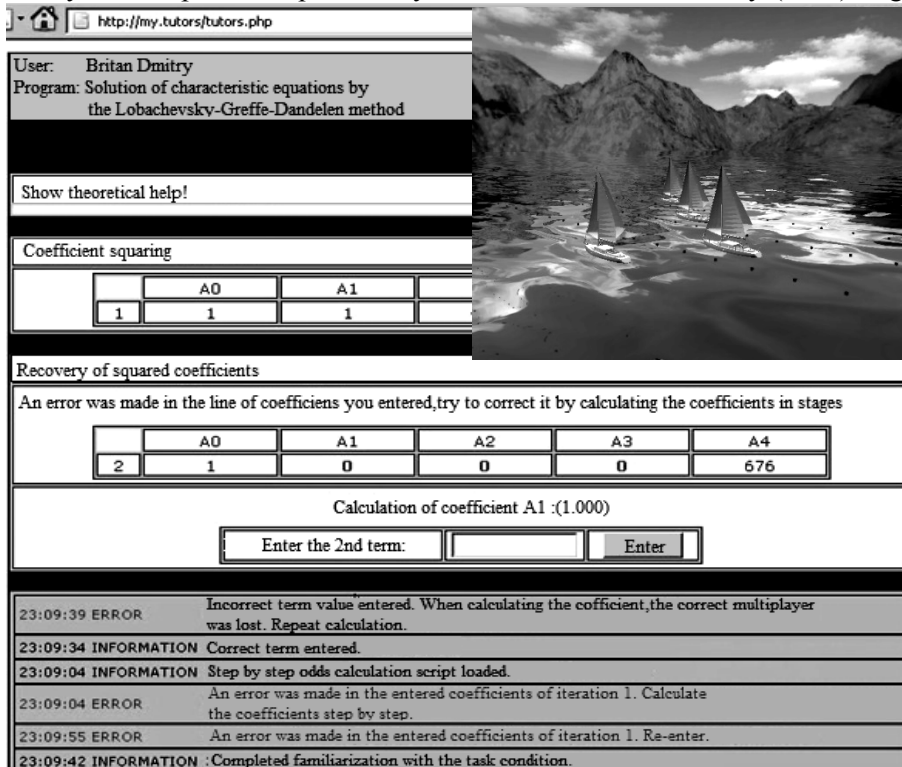


Fig. 5. ITS for the LGD method and ITS "Sailing Regatta"

shows the screen forms of ITS for the method of Lobachevsky-Greffe-Dandelen (LGD) and ITS “Sailing Regatta”.

The experiments showed that the minimum reduction in the time of manual completion of tasks by students due to training with ITS support for the LGD method was 23.3%, and the maximum – 50%. In this case, all students obtained proper solutions not only in the learning process with the help of ITS, but also manually during a post-test.

Figs. 6, 7 shows the screen forms of ITS for the construction of frequency characteristics of the of automatic control object and ITS for the construction of transient characteristics of the automatic control system.

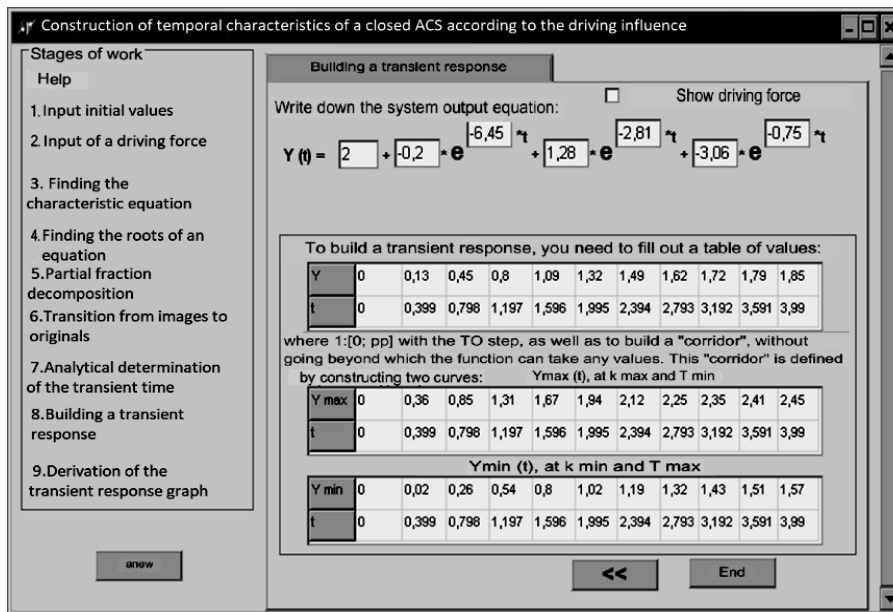


Fig. 6. ITS for object frequency characteristics construction

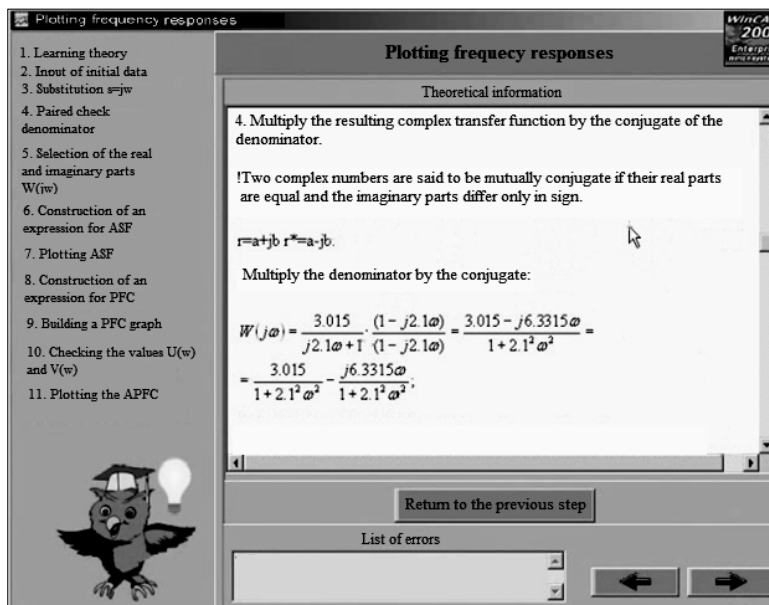


Fig. 7. ITS for the transient characteristics construction

As a result of experimental researches it is established that due to ITS for construction of frequency characteristics of OAU reduction of time of performance of a task by students made from 25 to 95.4%. As the conducted experiments showed, thanks to ITS for construction of transitional characteristics of ACS the minimum reduction of time at manual performance of a task by students made 15%, and the maximum – more than 36.6%.

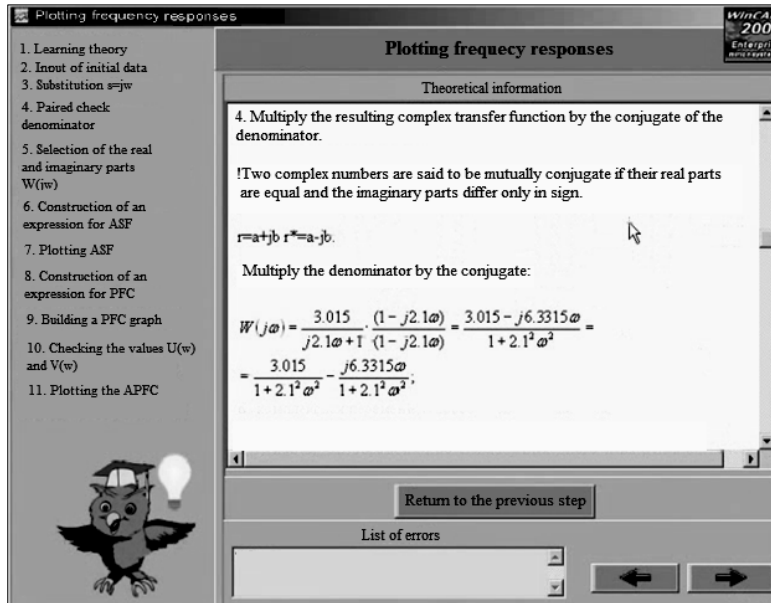


Fig. 7. ITS for the transient characteristics construction

Fig. 8 shows the screen forms of virtual laboratory work (VLW) “Experimental determination of the parameters of the transfer functions of the objects of automatic stabilization”.

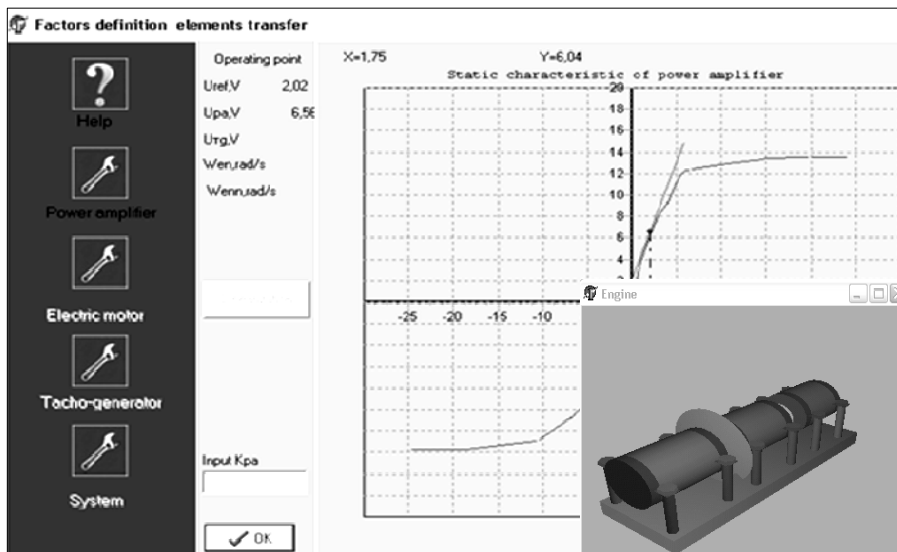


Fig. 8. Screen forms examples of VLW from automatic control theory

The minimum reduction in the time students performed this laboratory work due to ITS was 33.3%, and the maximum was 43.3%.

Fig. 9 shows screen forms of software for testing knowledge and skills in mathematics, created by order of Kharkiv Regional Centre of Education Quality Estimation, as well as Plymouth Inoovation Centre in Mathematics Teaching (Great Britain).

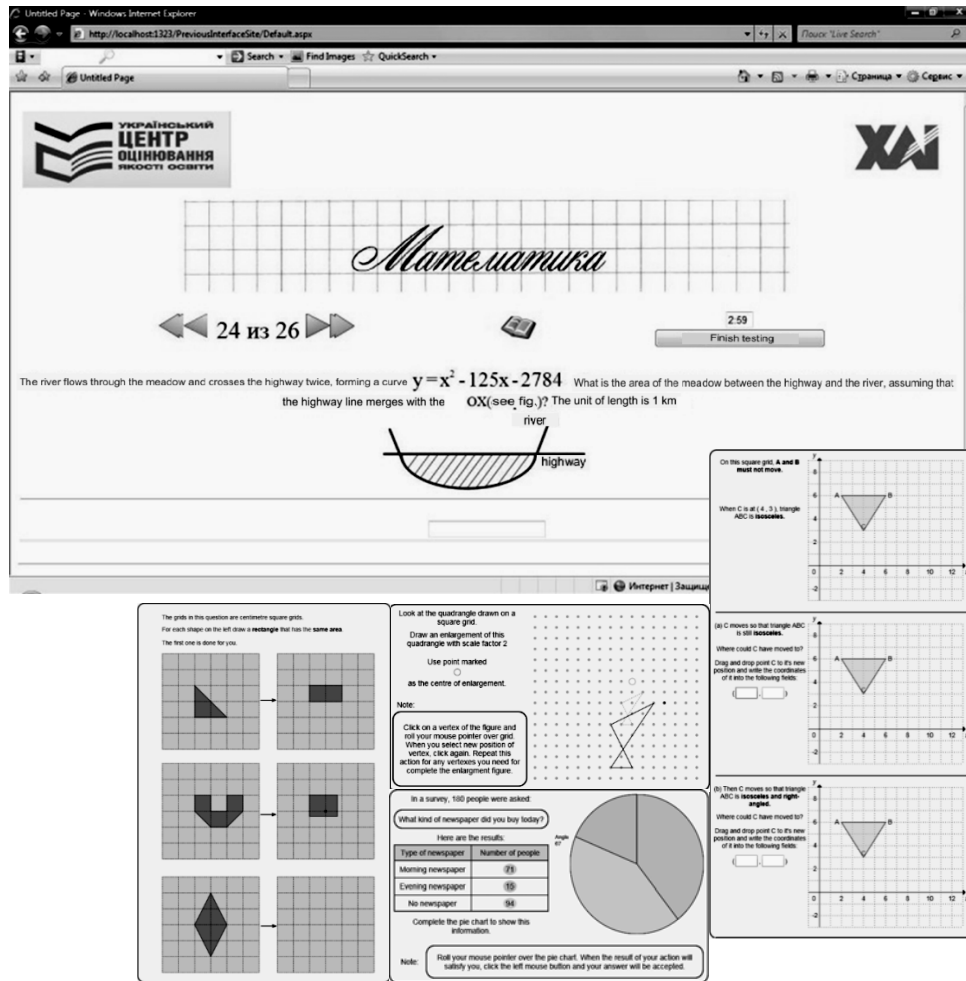


Fig. 9. Software for testing knowledge and skills in mathematics

The minimum time reduction for the development of the site for Centre of Education Quality Estimation was 553 hours, and the minimum savings – 8188 pounds, which is confirmed by the relevant act.

Fig. 10 shows the screen forms of ITS for algorithms [10] and ITS for SQL [11]. In the case of ITS for algorithms, each student out of twenty in the academic group should perform two tasks during the experiments. As a result of the experiment:

– in the first task: 10 from 20 persons obtained at once a conditional solution and 10 obtained at once a full solution. From those who obtained a conditional solution, 8 persons managed to obtain a complete solution due to ITS support later;

– in the second task: 7 from 18 persons obtained at once a conditional solution and 6 – a full solution. From those who obtained a conditional solution, 3 persons obtained a full solution with ITS support later.

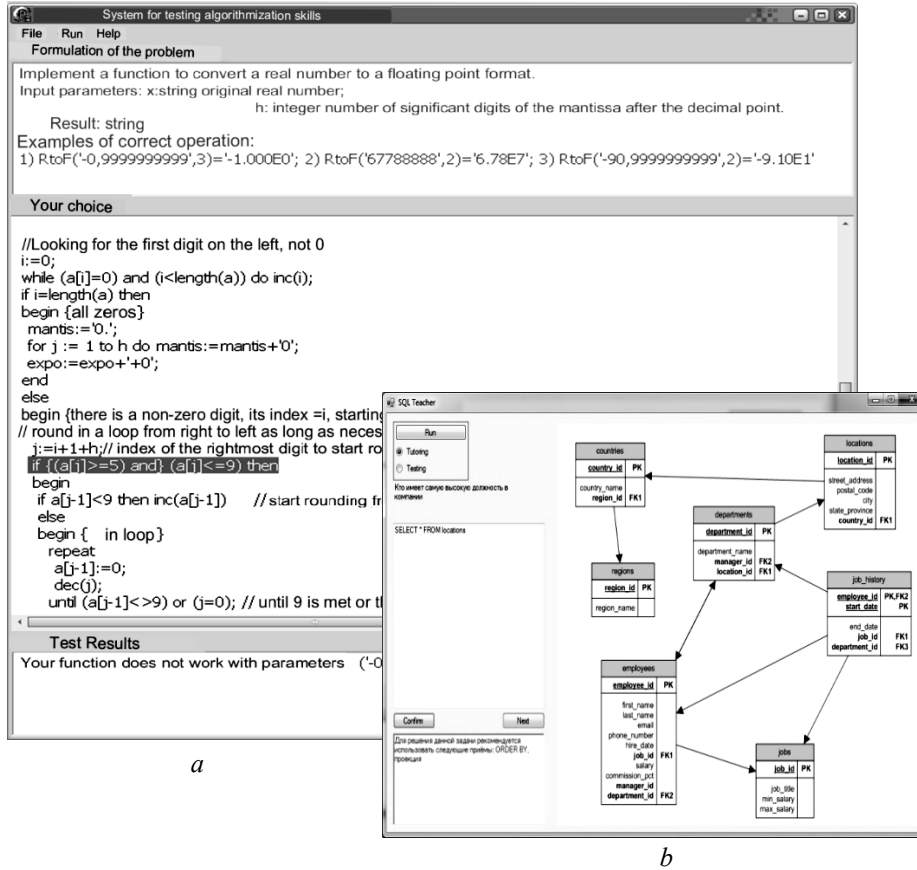


Fig. 10. ITS screen forms: *a* — for algorithms; *b* — for SQL

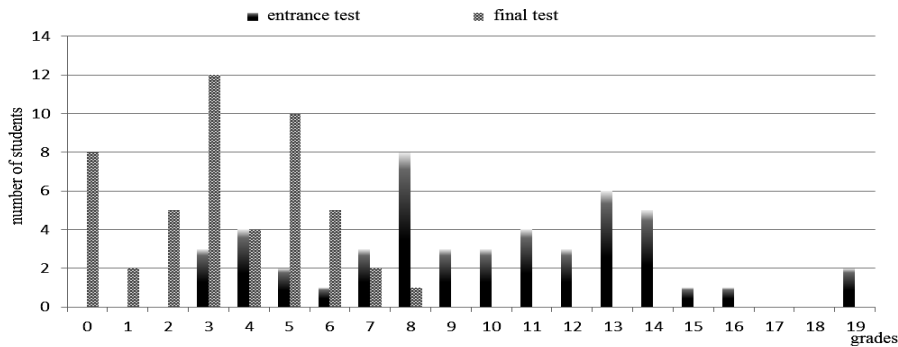


Fig. 11. Histogram for entrance and final students testing in ITS for SQL

Regarding the ITS for the SQL language, the experimental study of this ITS was conducted on 49 KhAI students. At the first stage, students performed 5 tasks of entrance testing, for which they were given a grades. Then students were solving 10 tasks in the studying mode and after that they passed a final test, which consists of the same tasks as the entrance one. As a result, for each student the score for the final test exceeds the score for the entrance test. The combined histogram for entrance and final assessments is shown in Fig. 11.

Experimental studies of the software for the ITS automated creation and run universal environment were also carried out [12]. The screen form of the environment and the histogram with results of students training by means of the created ITS for the discipline "Basics of systems modeling" are shown in Figs. 12, 13.

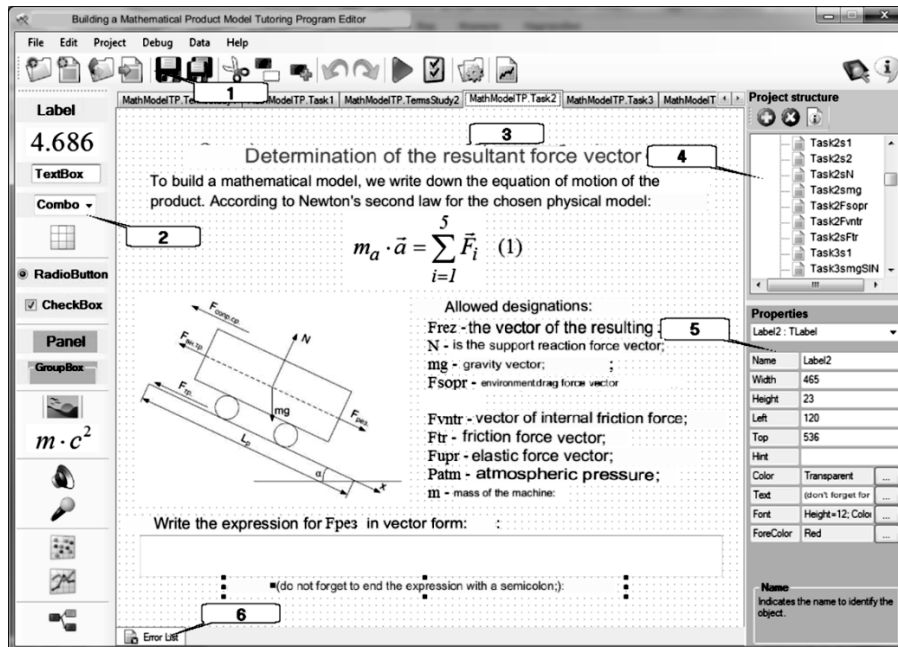


Fig. 12. Universal environment for the ITSs automated creation and translation

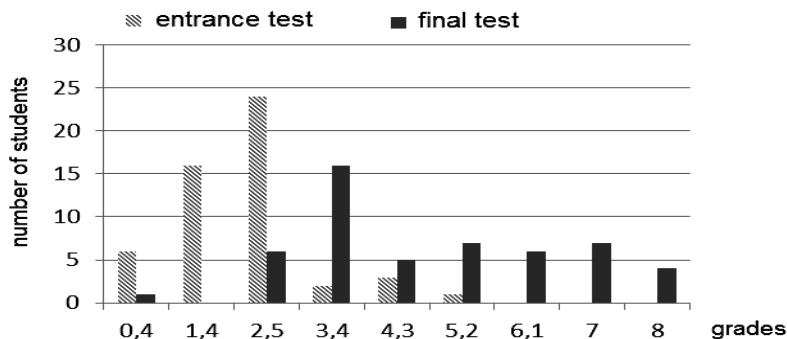


Fig. 13. Histogram of the students' learning results with ITS

The experiment participants were 52 students, who were asked to take an entrance test, training and final test from the discipline in the ITS environment. As a result, each student had increased his grades.

CONCLUSIONS

The second part of the paper presents practical results for the development and implementation of specific ITSs. There were developed different ITSs for specific domains. The experimental studies allow stating following:

1. Every student who studies by means of ITS has occupied necessary knowledge, skills and competencies while performing tasks, despite possible mistakes made during solving, ignorance or incompetence of a student.

2. Task solving time has been reduced at least twice on average.
3. Students got skills not only in certain tasks solving, but also in operating with whole task classes.
4. The tutor routine load has been significantly reduced.
5. Motivation of students to master new competencies has been increased.

The scientific and practical results of the work were shown in one doctoral and three candidate dissertations. There are over 100 scientific publications, indexed in the Scopus, reported and approved at scientific conferences in Ukraine, USA, Germany, Austria, Spain, Poland.

The developed ITSs have been implemented in a number of organizations, educational institutions and enterprises in Ukraine, Austria, Great Britain and China.

REFERENCES

1. A. Chukhray, *Methodology for learning algorithms: monograph*. National Aerospace University "KhAI", 2017, 336 p.
2. M.Z. Zgurovsky and N.D. Pankratova, *System analysis: problems, methodology, applications*. Kyiv: Nauk. opinion, 2011.
3. N.D. Pankratova, "System analysis in the dynamics of diagnosing complex technical systems," *System Research & Information Technology*, no. 4, pp. 33–49, 2008.
4. A. Kulik, A. Chukhray, and O. Havrylenko, "Information technology for creating intelligent computer programs for training in algorithmic tasks. Part 1: Mathematical foundations," *System Research & Information Technologies*, no. 4, pp. 27–41, 2021.
5. A. Kulik, A. Chukhray, and A. Zavgorodniy, "Similar strings detecting methods," in *Proceedings of the East-West Fuzzy Colloquium, Zittau, Germany, IPM, 2005*, pp. 183–191.
6. J.P. Martínez Bastida, O. Havrylenko, and A. Chukhray, "Developing a self-regulation environment in an open learning model with higher fidelity assessment," *Communications in Computer and Information Science*, vol. 826, pp. 112–131, 2018.
7. K. Zhang and D. Shasha, "Simple fast algorithms for the editing distance between trees and related problems," *Society for Industrial and Applied Mathematics. Journal on Computing*, no. 18 (6), pp. 1245–1262, 1989.
8. A. Chukhray and O. Havrylenko, "Proximate Objects Probabilistic Searching Method," *Advances in Intelligent Systems and Computing, 1113 AISC, 2020*, pp. 219–227.
9. A. Chukhray and O. Havrylenko, "The method of student's query analysis while intelligent computer tutoring in SQL," *Radioelectronic and Computer Systems*, no. 2 (98), pp. 78–86, 2021. doi: 10.32620/re`ks.2021.2.07.
10. Ie. Vagin, O. Havrylenko, J.P. Martínez Bastida, and A. Chukhray, "Computer Intelligent Tutoring System "SQLTOR,"" *ICTERI2019 ICT in Education, Research and Industrial Applications. Integration, Harmonization and Knowledge Transfer*, pp. 541–546. Available: <http://ceur-ws.org/Vol-2387/20190525.pdf>
11. D. Gaydachuk, O. Havrylenko, J.P. Martínez Bastida, and A. Chukhray, "Structural diagnosis method for computer programs developed by trainees," *Proceedings of 15th International Conference on ICT in Education, Research, and Industrial Applications, 2019, Kherson, Ukraine*, pp. 485–490.
12. A. Chukhray and O. Havrylenko, "The engineering skills training process modeling using dynamic Bayesian nets," *Radioelectronic and Computer Systems*, no 2 (98), pp. 87–96, 2021. doi: 10.32620/reks.2021.2.08.

Received 27.12.2022

INFORMATION ON THE ARTICLE

Anatoliy S. Kulik, ORCID: 0000-0001-8253-8784, National Aerospace University “Kharkiv Aviation Institute”, Ukraine, e-mail: anatelykulik@gmail.com

Andrey G. Chukhray, ORCID: 0000-0002-8075-3664, National Aerospace University “Kharkiv Aviation Institute”, Ukraine, e-mail: achukhray@gmail.com

Olena V. Havrylenko, ORCID: 0000-0001-5227-9742, National Aerospace University “Kharkiv Aviation Institute”, Ukraine, e-mail: lm77191220@gmail.com

ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ СТВОРЕННЯ ІНТЕЛЕКТУАЛЬНИХ КОМП’ЮТЕРНИХ ПРОГРАМ ДЛЯ НАВЧАННЯ АЛГОРИТМІЧНИМ ЗАВДАННЯМ. Частина 2: Дослідження та реалізація / А.С. Кулік, А.Г. Чухрай, О.В. Гавриленко

Анотація. Інформаційні технології і, зокрема, методи штучного інтелекту дедалі глибше проникають у всі сфери людської діяльності: науку, техніку, мистецтво та освіту. Україна також має достатній потенціал і надзвичайно потребує засобів підтримки навчання, розробці яких і присвячено цю статтю. Метою роботи є демонстрація компонентів інформаційної технології створення інтелектуальних системи навчання, які залучено до вивчення різних дисциплін технічного профілю. У роботі використано методи системного аналізу, математичного та імітаційного моделювання, технічного діагностування, штучного інтелекту. Запропоновані моделі та методи реалізовано в прототипах ITS для навчання математиці, програмуванню, теорії автоматичного управління. ITS було впроваджено в навчальному процесі університету ХАІ, інших установах України, Великобританії, Австрії, Китаю. Експериментальні дослідження показали підвищення показників успішності навчання студентів за допомогою ITS у порівнянні з традиційними методами. З наукової точки зору мають цінність покращені і адаптовані для комп’ютерного навчання методи технічної діагностики, Байєсових мереж та розроблені моделі алгоритмічних задач, процесу навчання та особи, що навчається. У практичному сенсі отримані результати можна використовувати для створення нових спеціалізованих ITS, а також для побудови єдиної розширюваної навчаючої платформи, яка об’єднує базові дисципліни певної галузі.

Ключові слова: інтелектуальні системи навчання, експериментальні дослідження, результати впровадження.

**ASSESSMENT OF THE ECONOMICAL DIMENSION OF
SUSTAINABLE DEVELOPMENT OF THE UKRAINE'S REGIONS
BASED ON THE BRIGHTNESS OF NIGHT LIGHTS**

M. ZGUROVSKY, K. YEFREMOV, S. GAPON, I. PYSHNOGRAIEV

Abstract. When assessing the level of development of territories, the problem of finding objective qualitative data that will characterize it arises. One of the possible sources of such data is the remote sensing of the Earth (RSE). The article is devoted to the analysis of the possibility of using the product of RSE – the map of night lights, for modeling the economical dimension of the sustainable development of the regions of Ukraine. Using the regression and correlation analysis and neural networks, appropriate models for assessing the level of economic development of the Kherson region, Donetsk region, and the AR of Crimea were obtained. The study was carried out by the team of the World Data Center for Geoinformatics and Sustainable Development of the Igor Sikorsky Kyiv Polytechnic Institute. It was part of research on the analysis of the behavior of complex socio-economic systems and processes of sustainable development in the context of the quality and safety of people's lives.

Keywords: sustainable development, spatial data analysis, economical development, night lights, mathematical modeling.

INTRODUCTION

The effectiveness and quality of management decisions regarding the behavior of complex socio-economic and security systems depends on the completeness and adequacy of the assessment of the management situation. For the construction of such an estimate, an important factor is the quality of the data used during the study.

In this study, a system is understood as a territory that is either an object of management or an environment for a certain management situation. During the assessment of the level of development of territories, the problem of finding objective qualitative data that will characterize it arises. One of the possible sources of such data is the application of remote sensing of the Earth (RSE).

The study is devoted to the analysis of the possibility of using one of the products of the RSE — the map of night lights, for modeling the economical dimension of sustainable development of the regions of Ukraine and other administrative-territorial entities.

The possibilities of using the map of night lights were considered by various scientists. In particular, at the global level, Jiansheng Wu et al. [1] investigated the relationship between GDP and night lights using 15 years of observations. In [2], the authors reviewed the indicators determining Human Well-Being for 2006 and showed the existing relationship with night lights. At the regional level, there were similar studies for the population of Japan [3] and Ukraine [4], the level of urbanization and GRP of the regions of China [5; 6], etc.

In a previous study [7], the spatial correlation of indicators of sustainable development and the brightness of night lights according to the data of 2011 was considered. The preservation of zonal trends with the components of the quality and safety of people's lives, the Index of the economical dimension, the Competitiveness Index and the Sustainable Development Index of Ukraine's regions was confirmed.

This study, in contrast to the results of foreign scientists, investigates the possibility of an integral assessment of the level of economic development based on maps of night lights at the regional level in dynamics, and also presents models for determining the level of economic development of the regions of Ukraine for the first time.

The study was carried out by the team of the World Data Center for Geoinformatics and Sustainable Development of the Igor Sikorsky Kyiv Polytechnic Institute and was part of research on the analysis of the behavior of complex socio-economic systems [8] and sustainable development processes in the context of the quality and safety of people's lives [9; 10; 11].

METHODOLOGY FOR CALCULATING THE INDEX OF BRIGHTNESS OF NIGHT LIGHTS OF THE UKRAINE'S REGIONS

The maps of night lights (Fig. 1) used in the study are the result of data processing by the Department of Geological and Atmospheric Sciences, Iowa State University [12] on artificial illumination of the Earth's surface at night from the sources of the Earth Observation Group, Colorado School of Mines [13].



Fig. 1. Map of night lights (raster image (composite) of artificial night luminosity of the Earth's surface) for 2013

The Night Light Brightness Index was calculated for the territory of Ukraine and each of its regions for the period 2004 — September 2022 (monthly available data is presented for 2022). For this, a vector mask of the borders of countries and regions and their urbanized territories was created on the night luminance maps. Geospatial calculation functions averaged indicators within each urbanized area of each region. The Night Light Brightness Index of an administrative-territorial formation is the sum of the values for the illumination of each cell of the map within the urbanized territories reduced to the area of the administrative-territorial formation:

$$I_i^j = \frac{\sum_{k=1}^{K_i} L_{i,k}^j}{S_i},$$

where i — number of the administrative-territorial formation; j — year of calculation; I_j^i — the value of the Night Lights Brightness Index; S_i — the area of the administrative-territorial formation in km^2 ; K_i — the number of cells, the brightness of which is counted to the i -th administrative-territorial formation; $L_{i,k}^j$ — the brightness of the k -th cell of the i -th administrative-territorial entity for the j -th year, which is measured in the range from 0 to 63.

Thus, the regions with the highest values of luminosity in their urbanized territories received the highest values of the Night Lights Brightness Index, and vice versa, the regions with the lowest values of luminosity in their urbanized territories — the lowest. The method of zonal statistics was used to calculate the sum of the values of the raster cells within the vector masks.

The result of calculations of the Night Lights Brightness Index for regions of Ukraine is presented in the Table 1 and in Fig. 2.

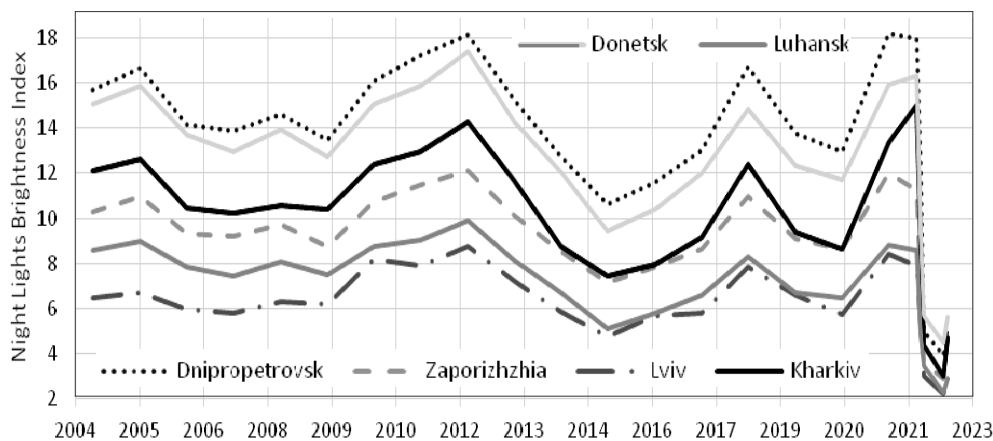


Fig. 2. Evaluation of the brightness of the urbanized territories of the regions of Ukraine from 2004 to September 2022 (monthly available data is presented for 2022)

THE NIGHT LIGHTS BRIGHTNESS INDEX ANALYSIS

The Night Lights Brightness Index had a decreasing trend several times:

- global economic crises (2006–2009);
- annexation of territories of Ukraine in 2014;
- the pandemic of COVID-19;
- the Russia's invasion on the territory of Ukraine in 2022.

Table 1. Evaluation of the brightness of the urbanized territories of the regions of Ukraine from 2007 to 2022*

Region	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022/1	2022/2	2022/3	2022/8	2022/9
AR of Crimea	5.15	5.41	4.78	5.99	6.16	6.88	5.74	4.60	3.87	4.22	4.90	6.01	5.46	5.22	7.40	7.44	5.73	4.31	3.33	4.14
Cherkasy	5.66	5.85	5.51	7.57	7.45	8.66	7.08	4.96	3.92	4.51	4.95	7.02	5.31	4.84	7.65	6.92	4.68	2.66	1.96	2.55
Chernihiv	6.94	7.36	6.97	8.87	9.12	9.90	8.18	6.69	5.43	6.08	6.65	8.65	7.14	6.52	9.36	8.67	5.35	2.57	1.89	2.97
Chernivtsi	4.15	4.56	4.83	5.73	5.64	6.46	5.43	3.14	2.30	2.96	2.96	4.51	3.55	3.04	4.40	3.64	2.80	1.38	1.12	1.39
city Sevastopol	10.55	10.78	10.16	11.39	12.07	13.13	11.81	10.97	10.30	10.51	11.38	12.93	11.97	11.55	13.92	25.52	27.37	20.63	16.91	18.28
Dnipropetrovsk	13.88	14.61	13.48	16.07	17.23	18.14	15.24	12.71	10.63	11.61	13.04	16.70	13.77	12.95	18.22	17.99	12.16	5.02	3.91	5.20
Donetsk	12.93	13.90	12.76	15.09	15.86	17.42	14.26	12.07	9.42	10.39	11.98	14.86	12.34	11.74	15.93	16.34	9.26	5.68	4.51	5.64
Ivano-Frankivsk	4.74	5.36	5.68	6.93	6.28	7.69	6.06	3.53	2.81	3.32	3.39	5.25	4.03	3.65	5.70	5.76	3.83	1.83	1.21	1.79
Kharkiv	10.25	10.60	10.40	12.38	12.97	14.26	11.64	8.75	7.44	7.96	9.17	12.37	9.40	8.64	13.38	15.00	8.74	4.35	3.01	4.71
Kherson	6.87	7.27	6.49	7.97	8.46	8.95	7.57	6.34	5.45	5.83	6.49	8.11	6.88	6.55	8.93	8.99	5.54	2.53	2.09	2.89
Khmelnitskiy	6.39	6.69	6.40	8.87	8.91	10.19	8.43	5.91	4.81	5.45	6.00	8.55	6.58	5.95	9.27	8.28	5.88	3.25	2.41	3.10
Kirovohrad	9.68	9.99	9.19	11.51	11.97	13.46	11.00	8.77	7.25	8.05	9.00	11.63	9.39	8.75	12.93	12.67	7.85	3.74	2.93	4.19
Kyiv	7.02	7.34	7.03	9.67	9.60	11.24	9.33	7.11	5.75	6.64	7.10	10.00	7.88	7.10	10.82	10.38	7.54	3.11	2.27	4.50
Kyiv city	77.62	76.17	72.72	84.49	84.41	91.10	85.63	86.35	79.17	85.33	84.16	96.83	91.22	87.62	97.84	127.43	110.01	41.59	36.05	35.65
Luhansk	7.46	8.05	7.48	8.78	9.05	9.91	8.13	6.68	5.09	5.80	6.61	8.30	6.71	6.45	8.83	8.61	4.77	3.45	2.25	2.90
Lviv	5.77	6.32	6.19	8.16	7.89	8.75	7.23	5.84	4.78	5.69	5.77	7.86	6.59	5.72	8.42	7.90	5.08	3.02	2.20	2.64
Mykolayiv	8.24	8.49	7.91	9.98	10.53	11.58	9.46	7.28	6.03	6.57	7.52	9.97	7.87	7.33	11.15	11.72	7.50	4.02	3.12	4.23
Odesa	6.74	6.85	6.46	8.17	8.62	9.52	7.91	5.86	4.92	5.28	6.01	8.20	6.55	6.13	9.09	10.03	6.72	3.53	2.84	3.97
Poltava	7.51	7.66	7.75	9.81	9.92	11.22	9.09	6.65	5.57	6.16	6.82	9.41	7.43	6.58	10.45	8.87	6.63	3.25	2.42	3.13
Rivne	4.21	4.41	4.02	5.88	5.80	6.89	5.48	3.87	2.99	3.64	3.83	5.90	4.40	3.97	6.31	5.51	3.90	2.21	1.63	2.14
Sumy	7.04	7.28	7.33	9.12	9.20	10.39	8.37	6.33	5.23	5.76	6.38	8.75	6.88	6.25	9.70	11.20	6.79	3.20	2.36	4.87
Temopyl	4.97	5.33	5.35	7.11	7.06	8.43	6.74	4.36	3.44	4.11	4.33	6.56	4.91	4.40	7.05	6.20	4.23	2.26	1.66	2.25
Vinnitsya	6.58	6.81	6.53	8.70	8.68	10.23	8.36	5.98	4.74	5.39	5.95	8.50	6.47	5.81	9.26	8.84	6.52	3.51	2.40	3.34
Volyn	3.89	4.15	3.80	5.50	5.36	6.16	5.00	3.58	2.83	3.45	3.63	5.26	4.20	3.78	5.85	5.24	3.51	2.07	1.55	1.94
Zakarpattya	3.60	3.95	3.92	4.55	4.72	5.32	4.41	3.06	2.53	3.13	3.22	4.09	3.72	3.24	4.66	4.84	3.64	2.33	1.82	1.93
Zaporizhzhya	9.19	9.75	8.75	10.73	11.47	12.14	10.12	8.52	7.17	7.84	8.62	10.97	9.08	8.66	12.02	11.23	6.93	3.56	2.92	3.92
Zhytomyr	5.27	5.41	5.15	7.04	7.11	8.39	6.65	4.89	3.95	4.53	4.94	7.20	5.36	4.94	7.76	8.13	5.41	2.50	1.79	2.90

* Monthly available data is presented for 2022

The Night Lights Brightness Index for the territory of Ukraine in 2018 (Fig. 3) grew by 64% compared to 2015 and amounted to 9.225 and 5.599, respectively. In 2021, the historical maximum of the index for the studied period was observed and amounted to 10.083. The minimum is 2.577 in August 2022 which was reached due to consequences of the Russia's military actions.



Fig. 3. Map of night lights (raster image (composite) of artificial night luminosity of the Earth's surface) of the regions of Ukraine for 2015

Only in four regions (Kyiv city, Sevastopol city, Kherson and Kyiv regions) the brightness of night lights was the lowest in 2009 as a result of the global economic crisis, and in all other regions — in 2015 (if we excluded 2022).

In Kyiv city, unlike other regions, had a growing trend till January 2022. The dynamics of the brightness index is shown in Fig. 4, 5.

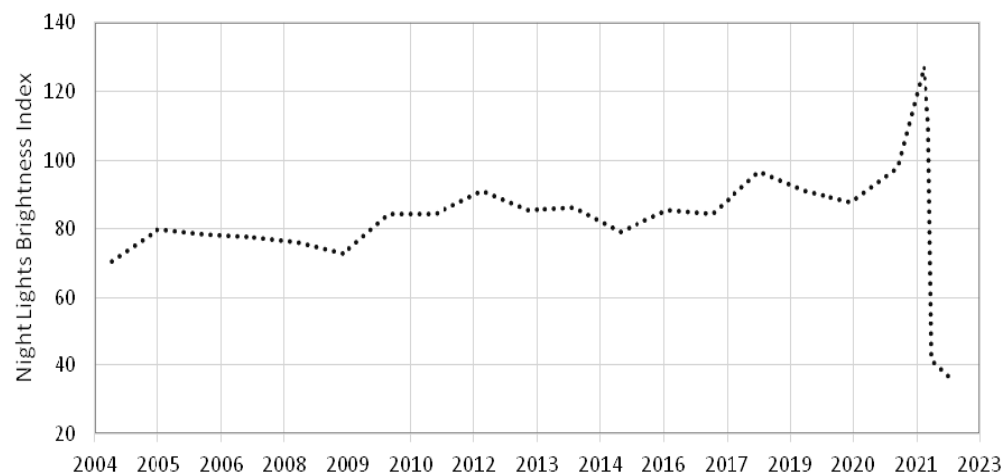


Fig. 4. Evaluation of the brightness of the territory of the Kyiv city from 2005 to 2018

Let's consider Donetsk and Luhansk regions. In 2015, the value of the Night Lights Brightness Index was affected greatly (Fig. 2). Fig. 6 shows the change in the intensity of night lights between 2013 and 2015 in the territory of these regions. Places of decrease in intensity are marked in dark grey on the map, and places of increase are marked in light grey. As can be seen from the figure, a decrease in the indicator is observed in most of the territories of the regions, which also indicates a decrease in economic activity.

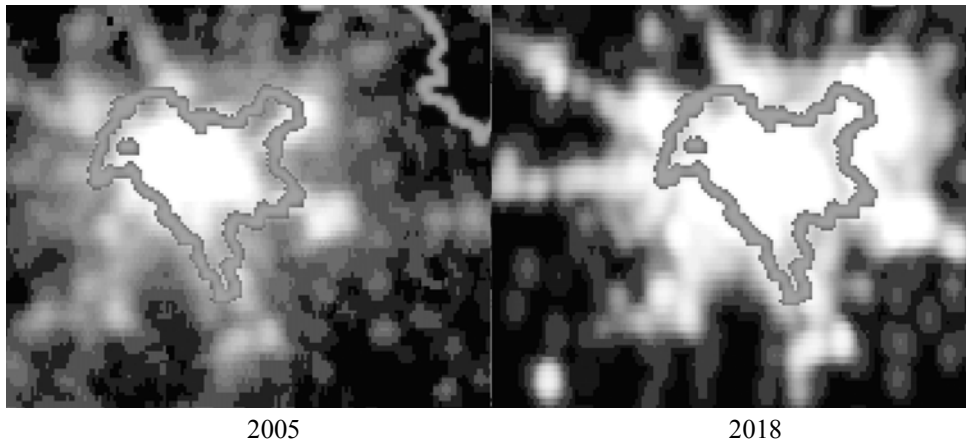


Fig. 5. Comparison of the brightness of the territory of the Kyiv city in 2005 and 2018.

As for August 2022, the most affected regions are Kharkiv, Sumy, Ivano-Frankivsk, Dnipropetrovsk. The decreasing was 78–80%. The less one is Sevastopol city (33%). It should be noted here that along with the unconditional impact of Russian aggression, this month is accompanied by a seasonal decrease in economic activity.

In order to verify the possibility of using night lights maps to assess the level of economic development of territories, a correlation-regression analysis of the Index of the economical dimension of sustainable development — an integral indicator calculated according to the methodology [9], and the Night Lights Brightness Index for the regions of Ukraine was conducted.



Fig. 6. Changes in the intensity of night lights between 2013 and 2015 in Donetsk and Luhansk regions

In the Table 2 the results of calculating the correlation coefficient of the studied indicators for each region are shown. The obtained results indicate that the relationship between economic development and the brightness of night lights is uneven for different regions of the country. The value of the correlation coefficient for the Kyiv city stands out from a number of data due to the specifics of the functioning of the capital. If we consider the entire data set (except for the Kyiv city), the correlation coefficient for the Index of the economical dimension of sustainable development and the Night Lights Brightness Index is 0.716 (Fig. 7), which indicates the presence of a connection.

Table 2. Correlation coefficients of the Index of the economical dimension of sustainable development of the regions of Ukraine and the level of brightness of night lights

Region	Correlation coefficient
AR of Crimea	0.744
city Sevastopol	0.732
Donetsk	0.721
Odesa	0.689
Poltava	0.660
Luhansk	0.636
Ivano-Frankivsk	0.616
Zhytomyr	0.605
Dnipropetrovsk	0.590
Kherson	0.563
Khmelnyskiy	0.530
Cherkasy	0.519
Rivne	0.478
Kharkiv	0.476
Ternopyl	0.464
Kirovohrad	0.457
Vinnytsya	0.432
Mykolayiv	0.397
Chernihiv	0.390
Chernivtsi	0.355
Volyn	0.279
Zaporizhzhya	0.267
Zakarpattia	0.252
Sumy	0.235
Kyiv	0.000
Lviv	-0.028
Kyiv city	-0.227

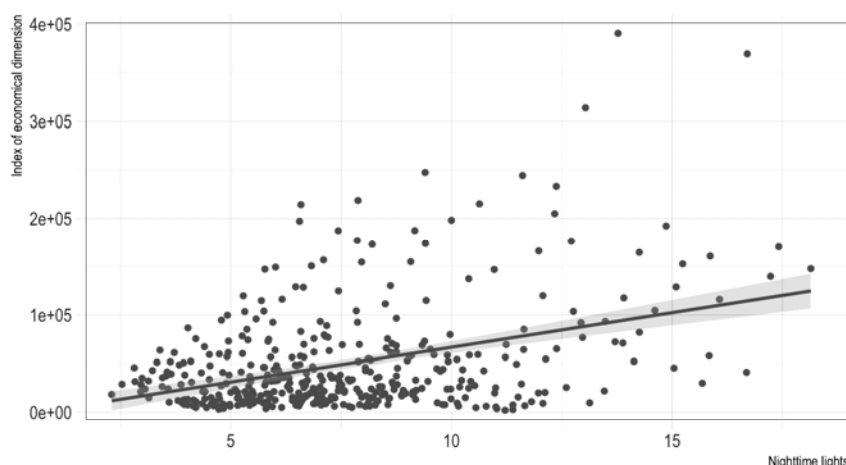


Fig. 7. Dependence of the Index of the economical dimension of sustainable development of the regions of Ukraine and the level of brightness of night lights

The presence of such a connection allows building models for assessing the level of the economic component of sustainable development based on satellite data for territories in the absence of the necessary statistical information.

ASSESSMENT OF THE LEVEL OF ECONOMIC DEVELOPMENT OF THE KHERSON REGION AND THE AUTONOMOUS REPUBLIC OF CRIMEA BASED ON THE NIGHT LIGHTS BRIGHTNESS INDEX

Three regions were selected for modeling:

- Kherson region (to check the methodology of building models);
- Autonomous Republic of Crimea (due to the absence of the objective statistical information since 2014);
- Donetsk region (due to the absence of the objective statistical information since 2014).

After the analysis of the relevant time series, the presence of a lag in 1 period was determined. The models are built in the form of nonlinear regression [14] and neural network [15] and have the following form:

1. Regressions:

a. Kherson region

$$I_{ec}(t) = 0,0763 * \ln(I_{nl}(t - 1)) + 0,00551,$$

where $I_{ec}(t)$. — Index of the economical dimension of sustainable development at a moment in time t , $I_{nl}(t - 1)$. — the Night Lights Brightness Index at a moment in time $t - 1$;

b. AR of Crimea

$$I_{ec}I_{nl}(t) = 0,30658 * LN(I_{nl}(t - 1)) - 0,2856,$$

where $I_{ec}(t)$ — Index of the economical dimension of sustainable development at a moment in time t , $I_{nl}(t - 1)$ — the Night Lights Brightness Index at a moment in time $t - 1$.

c. Donetsk region

$$I_{ec}(t) = 0,0019647232 * I_{nl}(t - 1)^{2,0489917756},$$

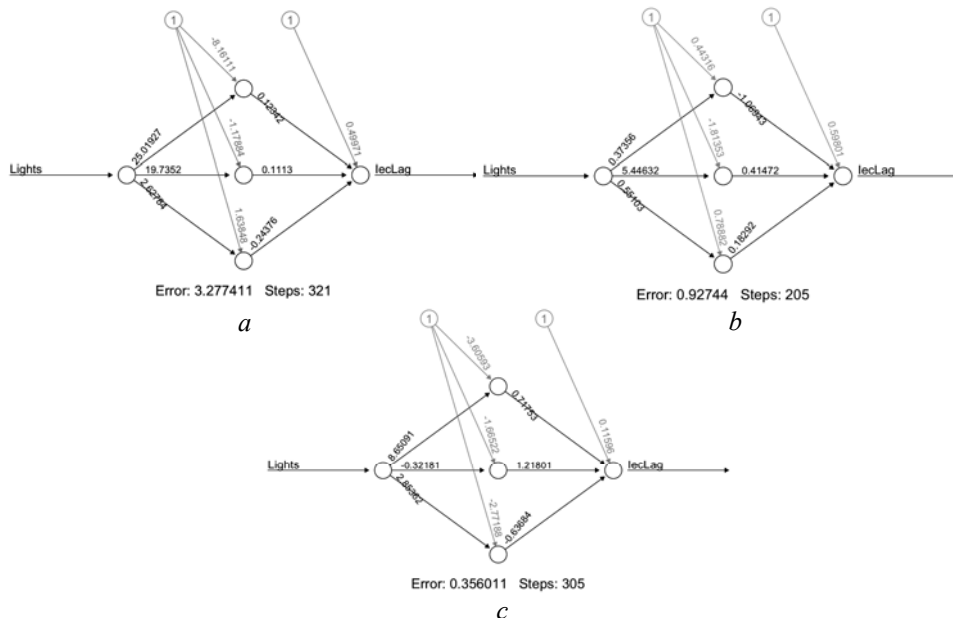


Fig. 8. Weights and structure of CNN for the: a — Kherson region; b — AR of Crimea; c — Donetsk region (I_{ec} — Index of the economical dimension of sustainable development at a moment in time t , $Lights$ — the Night Lights Brightness Index at a moment in time ($t-1$))

where $I_{ec}(t)$ — Index of the economical dimension of sustainable development at a moment in time t , $I_{nl}(t - 1)$ — the Night Lights Brightness Index at a moment in time $t - 1$.

2. Convolutional Neural Network (CNN):

- a. The weights and structure of the CNN for the Kherson region are presented in Fig. 8, *a*;
- b. The weights and structure of the CNN for AR of Crimea are presented in Fig. 8, *b*;
- c. The weights and structure of the CNN for the Donetsk region are presented in Fig. 8, *c*;

The model for the Kherson region has an average relative error of 7,6% for nonlinear regression and 6.4% for convolutional neural network. The results are presented in Fig. 9 and in Table 3. According to the built model, in 2021 the region showed growth and fall in 2022.

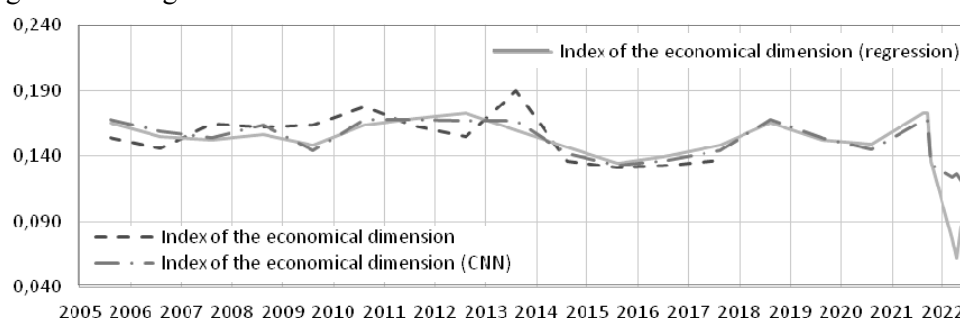


Fig. 9. Calculation of the Index of the economical dimension based on the brightness of night lights for the Kherson region

Table 3. Calculation of the Index of the economical dimension based on the brightness of night lights for the Kherson region

Year	The Index of the economical dimension	The Night Lights Brightness Index	The values of the Index of the economical dimension are modeled using non-linear regression	The values of the Index of the economical dimension are modeled using CNN
2009	0.162	6.487	0.157	0.164
2010	0.164	7.966	0.148	0.144
2011	0.178	8.455	0.164	0.168
2012	0.163	8.950	0.168	0.168
2013	0.155	7.570	0.173	0.167
2014	0.190	6.338	0.160	0.167
2015	0.136	5.447	0.146	0.141
2016	0.131	5.832	0.135	0.133
2017	0.133	6.490	0.140	0.137
2018	0.137	8.109	0.148	0.144
2019		6.883	0.165	0.168
2020		6.549	0.153	0.154
2021		8.930	0.149	0.145
01.2022		8.991	0.173	0.167
02.2022		5.542	0.173	0.167
03.2022		2.535	0.136	0.134
08.2022		2.094	0.076	0.124
09.2022		2.893	0.062	0.126
10.2022			0.087	0.122

The model for the Autonomous Republic of Crimea can be used to assess the level of economic development of the region after 2013. The average relative error of the obtained models is no more than 10.2%. The results are presented in Fig. 10 and in Table 4.

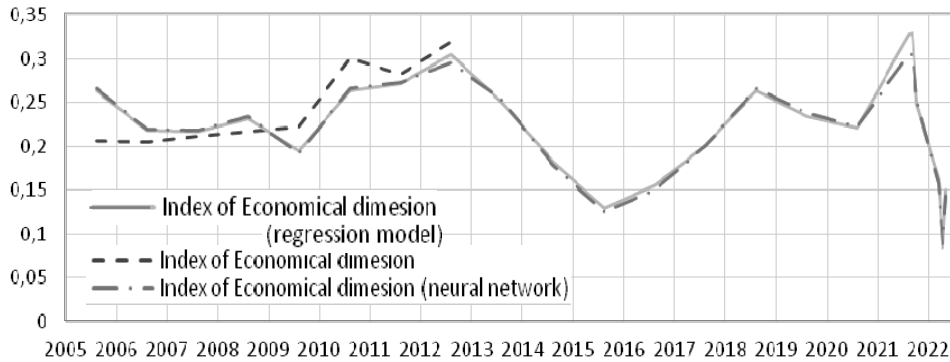


Fig. 10. Calculation of the Index of the economical dimension based on the brightness of night lights for AR of Crimea

Table 4. Calculation of the Index of the economical dimension based on the brightness of night lights for AR of Crimea

Year	The Index of the economical dimension	The Night Lights Brightness Index	The values of the Index of the economical dimension are modeled using non-linear regression	The values of the Index of the economical dimension are modeled using CNN
2009	0.216	4.778	0.194	0.192
2010	0.223	5.994	0.263	0.266
2011	0.301	6.161	0.272	0.273
2012	0.282	6.879	0.306	0.296
2013	0.320	5.739	0.250	0.254
2014		4.600	0.182	0.178
2015		3.872	0.129	0.124
2016		4.224	0.156	0.150
2017		4.900	0.202	0.201
2018		6.011	0.264	0.267
2019		5.461	0.235	0.238
2020		5.216	0.221	0.222
2021		7.398	0.328	0.305
01.2022		7.439	0.330	0.305
02.2022		5.731	0.250	0.253
03.2022		4.311	0.162	0.157
08.2022		3.335	0.084	0.089
09.2022		4.139	0.150	0.144
10.2022		4.778	0.194	0.192

According to the simulation results, after a certain economic growth in 2011–2013, the Autonomous Republic of Crimea fell in 2014–2016. In 2016–2021, there is an upward trend in the Index of the economical dimension of sustainable development.

The model for the Donetsk region has an average relative error of 19.4% for nonlinear regression and 18.2% for convolutional neural network. This is due to increased uncertainty in partial annexation and hostilities. The results are presented in Fig. 11 and in Table 5. According to the built model, the region has decreasing economic dimension trend after 2013 and after 2018. The models show even greater level of falling during 2022.

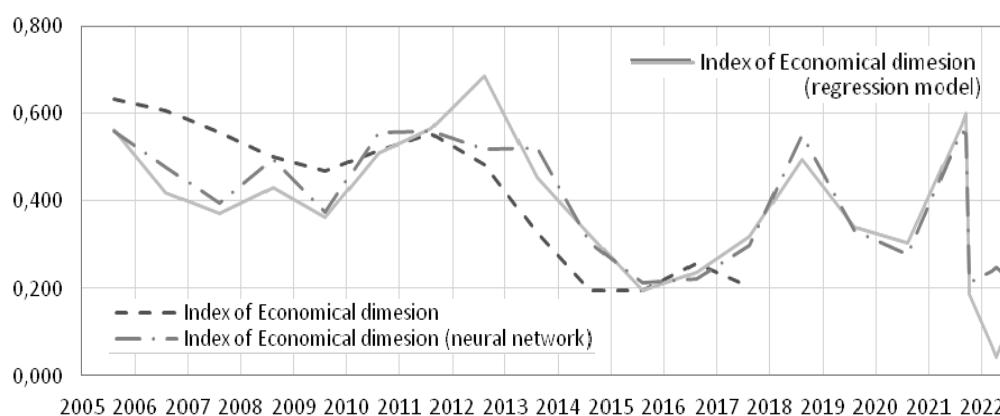


Fig. 11. Calculation of the Index of the economical dimension based on the brightness of night lights for the Donetsk region

Table 5. Calculation of the Index of the economical dimension based on the brightness of night lights for the Donetsk region

Year	The Index of the economical dimension	The Night Lights Brightness Index	The values of the Index of the economical dimension are modeled using non-linear regression	The values of the Index of the economical dimension are modeled using CNN
2009	0.503	12.762	0.432	0.495
2010	0.468	15.090	0.363	0.375
2011	0.517	15.862	0.511	0.558
2012	0.554	17.416	0.566	0.561
2013	0.482	14.255	0.685	0.520
2014		12.071	0.455	0.522
2015		9.425	0.323	0.305
2016		10.390	0.195	0.214
2017		11.977	0.238	0.222
2018		14.864	0.318	0.297
2019		12.337	0.495	0.552
2020		11.736	0.338	0.330
2021		15.934	0.305	0.279
01.2022		16.338	0.571	0.560
02.2022		9.257	0.601	0.553
03.2022		5.683	0.188	0.214
08.2022		4.510	0.069	0.238
09.2022		5.639	0.043	0.247
10.2022			0.068	0.239

CONCLUSIONS

In the course of the research, data on the luminosity of night lights on the territory of Ukraine was collected and processed. Based on the processed data, the Night Lights Brightness Index for the regions of Ukraine was formed and analyzed. The conducted correlation-regression analysis confirmed the possibility of using the specified index to determine the economic component of sustainable development on the territory of Ukraine.

Built for the Kherson region, the Autonomous Republic of Crimea and the Donetsk region, the assessment models of Index of the economical dimension of sustainable development showed an error of no more than 7.6%, 10.2% and 19.4%, which, among other things, allowed us to assess the level of economic development for the territory of the Autonomous Republic of Crimea and the Donetsk region for 2014–2022.

Using developed approach, we showed that the annexation of territories significantly affected the level of economic development of the territories of Crimea and Donbas. So, the Crimea fell by 0.174 points (57.8%), and the Donetsk region lost almost twice as much – 0.398 points (66.1%). At the same time, the pandemic period affected these regions in different ways. Crimea, after the restoration of its economic processes, lost only 0.044 points (16.6%), while the Donetsk region lost 0.232 points (44.3%).

And without having time to strengthen the trend towards economic recovery from the consequences of Covid-19, due to the full-scale invasion, these regions also suffered significantly. Not being a highly developed region, by the third decade of 2022 Crimea lowered its level of economic development by 0.230 points, when the Donetsk region lost much more – 0.432 points.

Thus, according to the results of the study, it was established that the degree of brightness of night lights can be used to assess the economic component of sustainable development of the regions of Ukraine, and in particular, the level of economic activity of the territories. This will help to make an assessment in the absence of ready-made qualitative statistical information, for example, for newly formed administrative units, or for regions for which, for various reasons, it is impossible to obtain sets of economic indicators.

Generally, with the application of the proposed approach, it was possible to identify and confirm the decrease in the economic activity of the regions of Ukraine during the pandemic and a full-scale invasion of Russia.

Further research is planned to be directed to the creation of a set of models for assessing the level of economic development of any territory within Ukraine with the aim of creating appropriate applications within the Information and Analytical Situational Center of the Igor Sikorsky Kyiv Polytechnic Institute.

REFERENCES

1. Jiansheng Wu, Zheng Wang, Weifeng Li, and Jian Peng, “Exploring factors affecting the relationship between light consumption and GDP based on DMSP/OLS nighttime satellite imagery,” *Remote Sensing of Environment*, vol. 134, pp. 111–119, 2013. Available: <https://doi.org/10.1016/j.rse.2013.03.001>.

2. T. Ghosh, S.J. Anderson, C.D. Elvidge, and P.C. Sutton, "Using Nighttime Satellite Imagery as a Proxy Measure of Human Well-Being," *Sustainability*, 5(12), pp. 4988–5019, 2013. Available: <https://doi.org/10.3390/su5124988>
3. Hasi Bagan and Yoshiki Yamagata, "Analysis of urban growth and estimating population density using satellite images of nighttime lights and land-use and population data," *GIScience & Remote Sensing*, 52:6, pp. 765–780, 2015. doi: 10.1080/15481603.2015.1072400.
4. V.I. Lialko, O.A. Apostolov, L.O. Yelistratova, and A.Ya. Khodorovskyi, "Otsinka sotsialno ekonomichnoho rozvytku oblastei Ukrainy za roky nezalezhnosti na pidstavi danykh suputnyka DMSP/OLS pro nichne osvittennia [Evaluation of the socio-economic development of the regions of Ukraine during the years of independence based on DMSP/OLS satellite data on night lighting]," *Ukrainskyi zhurnal dystantsiinoho zonduvannia Zemli*, no. 16, pp. 27–33, 2018. Available: http://nbuv.gov.ua/UJRN/ukjdz_2018_16_5.
5. Ting Ma, Chenghu Zhou, Tao Pei, Susan Haynie, and Junfu Fan, "Quantitative estimation of urbanization dynamics using time series of DMSP/OLS nighttime light data: A comparative case study from China's cities," *Remote Sensing of Environment*, vol. 124, pp. 99–107, 2012. Available: <https://doi.org/10.1016/j.rse.2012.04.018>.
6. Z. Dai, Y. Hu, and G. Zhao, "The Suitability of Different Nighttime Light Data for GDP Estimation at Different Spatial Scales and Regional Levels," *Sustainability*, 9:305, 2017. Available: <https://doi.org/10.3390/su9020305>
7. Michael Zgurovsky et al., "Parameterization of Sustainable Development Components Using Nightlight Indicators in Ukraine," *Conference proceedings of 2018 IEEE First International Conference on System Analysis & Intelligent Computing (SAIC), 2018*, pp. 1–5. doi: 10.1109/SAIC.2018.8516726.
8. Michael Zgurovsky, Andrii Boldak, Kostiantyn Yefremov, and Ivan Pyshnograiev, "Modeling and Investigating the Behavior of Complex Socio-economic Systems," *Conference proceedings of 2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON), 2017*, pp. 1113–1116. doi: 10.1109/UKRCON.2017.8100400.
9. M. Zgurovsky, K. Yefremov, I. Pyshnograiev, A. Boldak and I. Dzhygyrey, "Quality and Security of Life: A Cross-Country Analysis," *2022 IEEE 3rd International Conference on System Analysis & Intelligent Computing (SAIC), 2022*, pp. 1–5. doi: 10.1109/SAIC57818.2022.9923006.
10. *Sustainable Development Analysis: Global and Regional Contexts. P. 1. Global Analysis of Quality and Security of Life*. International Council for Science (ISC) and others; Scientific Supervisor of the Project M. Zgurovsky. K.: Igor Sikorsky Kyiv Polytechnic Institute, 2019, 328 p.
11. *Dashboard with the results of the assessment of the level of sustainable development of countries of the world and regions of Ukraine*. Available: <http://sdi.wdc.org.ua>
12. Xuecao Li, Yuyu Zhou, Min Zhao and Xia Zhao, "Harmonization of DMSP and VIIRS nighttime light data from 1992–2021 at the global scale," *Figshare. Scientific Data*, 7:168, 2020. Available: <https://doi.org/10.6084/m9.figshare.9828827.v7>
13. *Nighttime Lights Time Series*. Available: https://eogdata.mines.edu/nighttime_light
14. A. Kassambara, "Nonlinear Regression Essentials in R: Polynomial and Spline Regression Models," *Statistical tools for high-throughput data analysis*. Available: <http://www.sthda.com/english/articles/40-regression-analysis/162-nonlinear-regression-essentials-in-r-polynomial-and-spline-regression-models/>
15. David Kriesel, *A Brief Introduction to Neural Networks* [Text]. 2007, 244 p. Available: http://www.dkriesel.com/en/science/neural_networks

Received 12.02.2023

INFORMATION ON THE ARTICLE

Michael Z. Zgurovsky, ORCID: 0000-0001-5896-7466, National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Ukraine, e-mail: zgu-rovsm@hotmail.com

Kostiantyn V. Yefremov, ORCID: 0000-0003-3495-6417, National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Ukraine, e-mail: k.yefremov@wdc.org.ua

Sergii V. Gapon, ORCID: 0000-0002-8834-5825, National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Ukraine, e-mail: gapon@wdc.org.ua

Ivan O. Pyshnograiev, ORCID: 0000-0002-3346-8318, National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Ukraine, e-mail: pyshno-graiev@gmail.com

ОЦІНЮВАННЯ ЕКОНОМІЧНОГО ВИМІРЮВАННЯ СТАЛОГО РОЗВИТКУ РЕГІОНІВ УКРАЇНИ НА ОСНОВІ ЯСКРАВІСТІ НІЧНИХ ВОГНІВ /

М.З. Згуровський, К.В. Єфремов, С.В. Гапон, І.О. Пишнограєв

Анотація. Під час оцінювання рівня розвитку територій виникає проблема пошуку об'єктивних якісних даних, що будуть її характеризувати. Одним з можливих джерел таких даних є застосування методів дистанційного зондування Землі (ДЗЗ). Проаналізовано можливості використання одного з продуктів ДЗЗ – карти нічних вогнів, для моделювання економічного вимірювання сталого розвитку регіонів України. У результаті за допомогою регресійно-кореляційного аналізу та нейронних мереж отримано відповідні моделі оцінювання рівня економічного розвитку Херсонської області, Донецької області та АР Крим. Наведене дослідження виконано командою Світового центру даних «Геоінформатика та сталий розвиток» КПП ім. Ігоря Сікорського і є частиною досліджень з аналізу поведінки складних соціально-економічних систем та процесів сталого розвитку в контексті якості та безпеки життя людей.

Ключові слова: сталий розвиток, просторовий аналіз даних, економічний розвиток, нічні вогні, математичне моделювання.

УЗАГАЛЬНЮВАЛЬНА ІНТЕРВАЛЬНА НЕЧІТКА МОДЕЛЬ ТИПУ-2 ДЛЯ МОНІТОРИНГУ СТАНІВ СКЛАДНИХ СИСТЕМ З ВИКОРИСТАННЯМ ЕКСПЕРТНИХ ЗНАНЬ

Н.Р. КОНДРАТЕНКО, О.О. СНІГУР, Р.М. КОНДРАТЕНКО

Анотація. Запропоновано інтервальну узагальнювальну нечітку модель типу-2 в завданнях моніторингу станів складних систем з урахуванням експертних знань. Для узагальнення виходів нечітких моделей типу-2 з інтервальними функціями належності впроваджено теоретико-множинний підхід. Подано дослідження відповідності експертних оцінок вихідному значенню узагальнювальної нечіткої моделі протягом інтервалу спостережень. Наведено приклади використання узагальнювальної нечіткої моделі типу-2 в моніторингу станів артезіанської свердловини. Показано, що для підвищення якості прийняття рішень експерту необхідно звертати увагу на величину інтервального виходу узагальнювальної нечіткої моделі типу-2. Надано рекомендації експертам щодо поліпшення прийняття рішень стосовно оцінки вихідного інтервалу узагальнювальної моделі.

Ключові слова: нечітка модель типу-2, інтервальна функція належності, теоретико-множинний підхід, експертні знання.

ВСТУП

Відомо, що моделі ідентифікації складних систем, що будуються в умовах неповної вихідної інформації, успішно використовують апарат теорії нечітких множин [1]. Нечіткі множини типу-1 і типу-2 є достатньо потужним інструментом для розв'язання слабкоструктурованих проблем в умовах невизначеностей [2]. Неповнота вихідних даних є проблемою, що характерна для моделювання багатьох систем. Такими можна вважати соціальні, економічні, технічні системи різного характеру, але найбільш складними є завдання ідентифікації для екологічних та природничих систем, про характер роботи яких у багатьох випадках можна судити лише за вибіркою експериментальних даних.

Із слабкоструктурованих проблем актуальним є завдання моніторингу станів складних систем за умови неповноти вхідних даних. Упевнено можна вказати на необхідність моніторингу станів природничих систем, оскільки такі системи пов'язані з важкодоступними об'єктами; наприклад, це об'єкти геологічної розвідки. Для останніх характерними факторами є неточність вимірювань і часто неможливість вести безпосереднє спостереження об'єкта.

Особливість багатьох таких систем полягає в тому, що для отримання вхідних даних для подальшої експлуатації необхідно провести моніторинг станів системи зі спостереженнями протягом певного часу. Отримані дані можуть бути неоднорідними та суперечливими. Отже, для прийняття рішення розробники часто мають неповну інформацію про числові значення характеристик системи або є тільки інтервали можливих значень.

Використання інтервальних нечітких моделей як інструменту подання даних дозволяє виявити зв'язок між вхідними даними та виходом системи, стани якої досліджуються. Але досі залишається актуальною проблема дослідження функціонування інтервальних нечітких моделей в моніторингу станів складних систем в умовах неповної вхідної інформації з використанням знань експертів.

АНАЛІЗ ОСТАННІХ ДОСЛІДЖЕНЬ І ПУБЛІКАЦІЙ ТА ПОСТАНОВКА ПРОБЛЕМИ

Натепер обсяги даних, які нагромаджуються в різних галузях, різко збільшуються, що зумовлює пошук новітніх технологій. Це безумовно нечітке моделювання з використанням нечітких множин вищих порядків. Сучасний підхід до розв'язання завдань моніторингу станів складних систем за участю експертів реалізується за допомогою нечіткого моделювання. Нечіткі моделі використовують для свого функціонування основні поняття теорії нечітких множин, нечіткі правила та нечітке логічне виведення. Залежно від ступеня нечіткості нечітких множин, що враховується в побудові нечіткої моделі, розрізняють нечіткі моделі типу-1, загальні моделі типу-2 та інтервальні типу-2. Нечіткі моделі типу-1 будуються на основі нечітких множин з чіткими значеннями степенів належності, але такі нечіткі моделі дають на виході лише чітке значення. Інтервальні нечіткі моделі типу-2 будуються на основі нечітких множин з інтервальними значеннями степенів належності. Останні моделі дають на виході точкові та інтервальні значення. Звідси перед розробниками виникає задача вибору типу нечітких множин, за допомогою яких можна було б достатньо просто обробляти різні види невизначеностей. У праці [3] розглядаються питання, пов'язані з розвитком теорії нечітких множин типу-1 і типу-2 та наводиться мотивація адекватно подавати невизначеності, які існують в предметній області, за допомогою множин типу-2.

Множини типу-1 і типу-2 використовуються в методиках з подання знань експертів. У працях [4–6] розглядаються напрями, де в умовах неповноти вхідних даних експерту надається право визначати ту підмножину вхідних даних, яка, на його думку, є важливою. Також робиться висновок про доцільність використання нечітких моделей, побудованих на нечітких множинах типу-2 з інтервальними функціями належності.

Приклади практичного використання інтервальних нечітких моделей наведено у праці [7], де ставиться завдання зменшення помилок квантування інформації для нелінійних об'єктів. За рахунок використання інтервальних нечітких моделей вдається контролювати невизначеності, які виникають у нелінійних системах і надалі більш ефективно їх обробляти.

Як приклад успішного застосування інтервальних нечітких моделей, де експерти беруть участь у формуванні вхідного вектора параметрів, можна навести працю [8]. У ній показано особливості інтервального нечіткого моделювання на прикладі оцінювання стану артезіанської свердловини в довільний момент часу в межах від початку гідрогеологічної розвідки до її повного завершення.

Множину інтервальних нечітких моделей запропоновано для підвищення точності інтервального прогнозування часових послідовностей у праці [9]. Особливість цієї роботи полягає у викладеній методиці побудови

множини різновходових нечітких моделей типу-2. Кожна з таких моделей є звичайною інтервальною нечіткою моделі типу-2, яка містить базу правил, процедуру зведення до нечіткості, процедуру нечіткого логічного виведення та процедуру пониження типу.

У праці [10] запропоновано метод побудови нечітких баз знань типу-2 з інтервальними функціями належності, що буде використовуватись у медичній діагностиці та прогнозуванні станів захворювання. Показано, що побудована модель є адекватною медичній галузі і здатна в умовах якісного характеру вхідних даних враховувати як знання фахівця з медичної практики, так і дослідні медико-біологічні дані. У праці [11] пропонується метод побудови інтервальних нечітких логічних систем типу-2 у медичній галузі на основі експериментальних даних, що мають пропуски.

У праці [12] пропонується підхід для оцінювання якості виготовлення промислової продукції декількома моделями, у тому числі і моделями, які будуються на основі інтелектуальних технологій, а саме з використанням інтервальних нечітких моделей типу-2. Зазначено, що для оцінювання якості результатів, які отримують за допомогою кожної із запропонованих моделей, кінцеве рішення належить експерту.

Виходячи з наведеного, доцільно провести дослідження і з'ясувати, як для розв'язання завдання моніторингу станів складної системи в умовах невизначеності вхідних даних можна проаналізувати правильність рішень, що приймаються експертами, які надають вагому інформацію. Проте недостатньо досліджені інтервальні нечіткі моделі, що будуються з урахуванням знань експертів, та методи, які дозволяють на основі процедур агрегації виходів інтервальних нечітких моделей будувати нечіткі моделі, що є адекватними предметній області.

ВИКЛАД ОСНОВНОГО МАТЕРІАЛУ

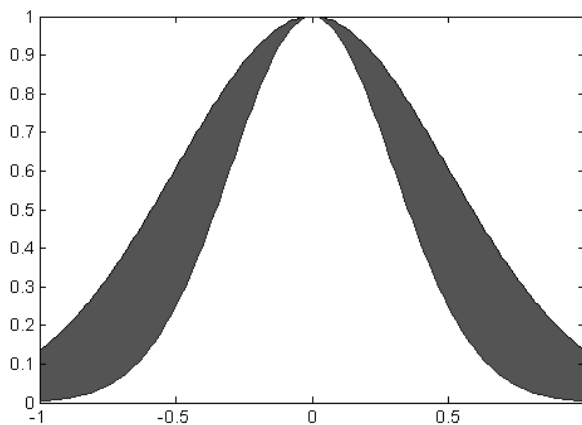


Рис. 1. Інтервальна функція належності

Вирішення поставленого завдання подамо в декілька етапів:

Етап 1. Ретельна робота з експертами. Для побудови нечітких моделей експертам пропонується працювати з нечіткими множинами типу-2 і будувати нечітку модель з інтервальними функціями належності. Тип функції належності задається. Це модифікована гаусова функція [1, 4]; типовий вигляд графіка функції зображено на рис. 1.

Для інтервальної нечіткої моделі експерти будують свій вхідний вектор, тобто можуть відкидати певні ознаки повного вхідного вектора, який пропонується розробником для досліджуваної системи.

Етап 2. Генерування множини інтервальних нечітких моделей. Наводиться генерація множини нечітких моделей з інтервальними функціями

належності із заданої експериментальної вибірки. Для побудови кожної моделі використовується метод генерації нечіткої моделі з експериментальних даних; ця процедура виконується кожним з експертів. Крім того, експерту надається право коригування вхідного вектора, де експерт залишає ті вхідні ознаки, які він вважає найважливішими. Оскільки експерту надається право коригувати вхідний вектор, то, як наслідок, у результаті такого коригування можлива поява малоінформативних вхідних даних. У такому випадку для моделі, що генерується, здійснюється перетворення експериментальної вибірки за допомогою алгоритму, який враховує неповноту експериментальних даних і забезпечує контрольоване розтягнення одного з параметрів функцій належності вхідних змінних з одночасним збереженням адекватності прийняття рішень нечіткою системою; це виконується за допомогою алгоритма, наведеного у праці [11].

Загальна частина роботи кожного експерта реалізується таким чином.

Задається вибірка експериментальних даних \mathbf{X} :

$$\mathbf{X} = \{X_1, X_2, \dots, X_n\}$$

де $X_i = \{x_{1i}, x_{2i}, \dots, x_{ki}, y_i\}$, $i = 1, \dots, n$; n — кількість експериментальних прикладів; k — кількість вхідних змінних; y — вихідна величина.

На основі вибірки відбувається генерація множини нечітких моделей, кожна з яких є інтервальною нечіткою моделлю. Структуру нечіткої моделі з інтервальними функціями належності, яка входить до множини, зображено на рис. 2.

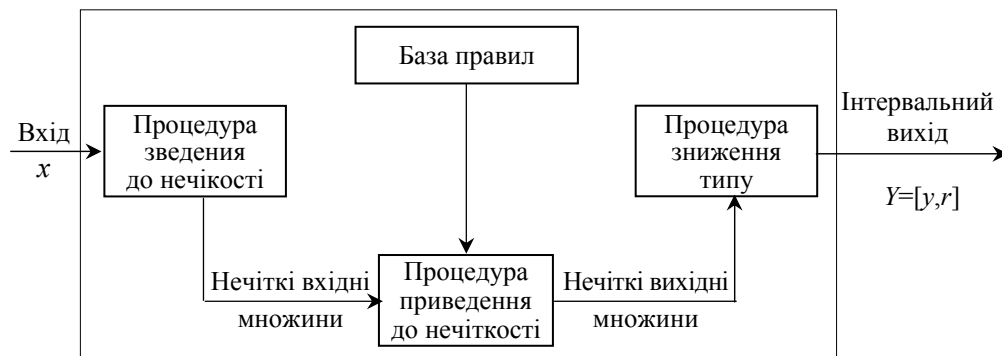


Рис. 2. Структура нечіткої моделі з інтервальними функціями належності

Модель відображає чіткі входи (x_1, x_2, \dots, x_n) в інтервальні та чіткі виходи: $\tilde{Y} = [y_l, y_r]$. Для опису нечітких термів лінгвістичних змінних будемо використовувати інтервальні нечіткі множини типу-2. Тоді математична модель буде являти собою інтервальну нечітку модель типу-2, що включає базу правил (нечітку базу знань), процедуру зведення до нечіткості, процедуру нечіткого логічного виведення, процедуру пониження типу та процедуру зведення до чіткості.

Для обчислення крайніх точок вихідного значення інтервальної нечіткої моделі типу-2 y_l і y_r використовуємо алгоритм Карніка–Менделя [1].

Таким чином, кожен експерт генерує свій набір ознак і, як наслідок, окрему модель. Для отримання кінцевого результату створюється узагальнювальна модель, яка може агрегувати виходи інтервальних типу-2 нечітких моделей (IT2FM1, IT2FM2, ..., IT2FMN). Структуру узагальнювальної інтервальної нечіткої моделі зображено на рис. 3.

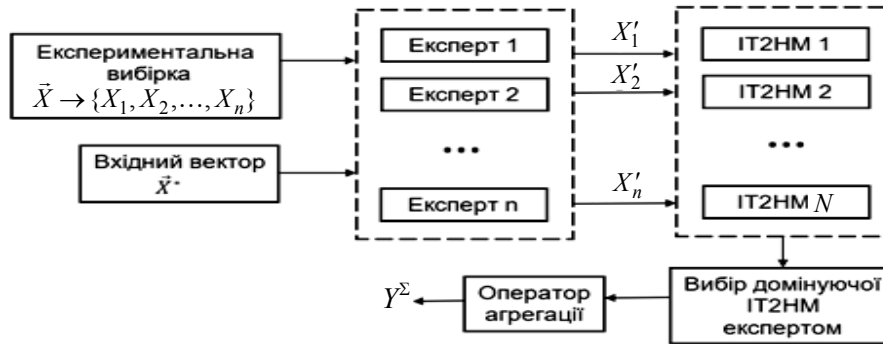


Рис. 3. Структура узагальнюючої інтервальної нечіткої моделі

Етап 3. Обчислення інтервального виходу Y нечіткої моделі, що є узагальнювальною. Y є результатом здійснення операцій теоретико множинного підходу. Реалізація агрегації та отримання кінцевого значення Y здійснюється на основі теоретико-множинного підходу (операції об'єднання \cup та перетину \cap теорії множин) таким чином. Спочатку виконується операція об'єднання окремих інтервальних виходів нечітких моделей, а потім відбувається перетин з інтервальним виходом моделі, яка є домінуючою або є моделлю за повним вектором, або вибирається експертом за певними рекомендаціями. Остання дія експерта є заключною.

Таким чином, кінцевий вихід узагальнювальної моделі створюється в результаті реалізації оператора агрегації у вигляді комбінації операцій теорії множин, а саме: U/\cap .

Можливий і інший підхід до отримання кінцевого результату. У крайньому випадку, коли під час побудови узагальнювальної нечіткої моделі, експерт не може прийняти рішення у виборі домінуючої моделі, використовується інша комбінація операцій об'єднання U та перетину \cap . Тоді узагальнювальна модель створюється автоматично з урахуванням розташування інтервальних виходів окремих моделей. У випадку наявності спільної частини інтервальних виходів нечітких моделей діє операція їх перетину, інакше — об'єднання. Тоді комбінація операцій така — \cap/U .

ЕКСПЕРИМЕНТАЛЬНІ ДОСЛІДЖЕННЯ ТА РЕЗУЛЬТАТИ МОНІТОРИНГУ

Розглянемо результати моніторингу станів складної системи за допомогою узагальнювальної моделі. За предметну область візьмемо артезіанську свердловину як складну гідрогеологічну систему. Задача моніторингу такої системи характеризується трудомісткістю отримання вихідних даних, що зумовлено важкодоступністю проведення експериментальних досліджень та є довготривалим періодом геологорозвідувальних робіт (ГРР) (від 6 місяців до декількох років) і має три етапи — від одного до трьох років. Система має багато вхідних параметрів, але беремо до уваги 24 вхідні змінні та одну вихідну величину; її назвемо «перспективність свердловини» і опишемо за допомогою лінгвістичної змінної (позначимо LV) зі шкалою від 0 до 10 [8]. Шкала розбивається на лінгвістичні терми: «недостатня перспективність» — від 1 до 5 балів; «достатня» — від 6 до 7 балів та «висока» — від 8 до 10 балів. Вихід моделі є інтервальним (позначимо через I), кінцеву оцінку для стану свердловини виставляє експерт. Результати роботи узагальнювальної нечіткої моделі родовища наведено в табл. 1.

Таблиця 1. Результати роботи узагальнювальної нечіткої моделі родо-вища (тестова вибірка)

№ експ.	Етапи	Експерт 1	Експерт 2	Вихід домінуючої моделі	Вихід моделі після U/∩.	Оцінка експерта
1	1	[0,03; 0,39]	[0,11; 0,24]	[0; 0,88]	[0,03; 0,39] Н	1 / недостатня
	2	[0,14; 0,92]	[0,3; 0,53]	[0,53; 4,11]	[0,53; 0,92] Н	
	3	[0; 0,02]	[0; 0]	[0; 2,64]	[0; 0,02] Н	
2	1	[0,04; 0,53]	[1,69; 3,12]	[0,01; 1,2]	[0,04; 1,2] Н	1 / недостатня
	2	[0,18; 1,09]	[1,51; 2,01]	[0,05; 2,36]	[0,18; 2,01] Н	
	3	[2,42; 3,44]	[3,11; 3,76]	[0,08; 2,81]	[1,68; 2,81] Н	
3	1	[7,21; 7,38]	[7,81; 7,82]	[0,63; 5,84]	[1,3; 5,48] Н, Д	2 / недостатня
	2	[6,95; 7,08]	[6,88; 6,88]	[0,63; 5,8]	[1,38; 5,48] Н, Д	
	3	[6,72; 6,99]	[6,8; 6,82]	[0,46; 5,81]	[1,25; 5,45] Н, Д	
4	1	[6,33; 6,66]	[6,66; 6,72]	[6,41; 6,96]	[6,41; 6,72] Д	5 / достатня
	2	[6,44; 6,93]	[7,59; 7,66]	[5,55; 7,48]	[5,82; 7,48] Д, В	
	3	[6,82; 7,3]	[7,75; 7,85]	[5,08; 7,82]	[6,62; 7,82] Д, В	
	2	[4,14; 4,2]	[3,84; 3,86]	[5,55; 6,53]	[5,71; 6,56] В	
	3	[5,56; 5,6]	[5,61; 5,62]	[4,91; 6,48]	[5,56; 5,62] Д	
6	1	[4,3; 5,98]	[6,35; 6,51]	[7,31; 7,62]	[7,32; 7,7] Д, В	6 / достатня
	2	[3,4; 4,45]	[4,39; 4,57]	[5,37; 6,15]	[5,37; 6,15] Д	
	3	[5,11; 5,42]	[4,78; 5,22]	[3,9; 5,71]	[4,78; 5,71] Д	
7	1	[7,64; 7,73]	[7,87; 7,87]	[9,03; 9,71]	[9,03; 9,71] В	8 / висока
	2	[8,27; 8,32]	[8,19; 8,2]	[9,87; 9,99]	[9,88; 10] В	
	3	[7,93; 8,12]	[8,17; 8,24]	[9,92; 10]	[9,92; 9,99] В	
8	1	[7,37; 7,42]	[7,92; 7,94]	[8,53; 8,75]	[8,54; 8,76] В	9 / висока
	2	[8,16; 8,21]	[7,5; 7,57]	[9,94; 9,99]	[9,95; 10] В	
	3	[7,94; 8,07]	[8,36; 8,39]	[9,89; 9,94]	[9,91; 10] В	
9	1	[7,89; 8,05]	[7,82; 7,82]	[8,61; 8,74]	[8,62; 8,74] В	9 / висока
	2	[8,9; 9,25]	[8,12; 8,12]	[9,64; 9,84]	[9,64; 9,77] В	
	3	[9,1; 9,54]	[9,21; 9,29]	[9,89; 9,98]	[9,9; 9,99] В	

Покажемо фрагмент реалізації теоретико-множинного підходу на підставі наведених даних.

Опишімо кожний експеримент та виконаємо основні обчислення. На першому етапі ГРР експертом 1 виділено 10 інформативних ознак, вихід моделі [0,03; 0,39]; експертом 2 виділено 11 ознак, вихід моделі [0,11; 0,24], модель з повним вектором 24 ознаки має інтервальний вихід — [0; 0,88].

Перший крок: реалізуємо операцію об'єднання інтервалів моделей для експертів 1 і 2:

$$[0,03; 0,39] \cup [0,11; 0,24] = [0,03; 0,39]$$

Другий крок: виконуємо операцію перетину з моделлю, що будується на повному векторі:

$$[0,03; 0,39] \cap [0; 0,88] = [0,03; 0,39].$$

Обчислення інтервального виходу узагальнювальної моделі для інших експериментів виконуються аналогічно.

У загальному випадку операція перетину може виконуватись не тільки з виходом моделі з повним вектором, але і моделі, яка, на думку незалежного експерта, є найбільш важливою, домінуючою, або базовою.

Як видно з табл. 1 для експерименту 1 значення лінгвістичної змінної «перспективність свердловини» за оцінкою експерта є недостатньою, оскільки крайні точки вихідного інтервального значення належать терму «недостатня», а ширина інтервалу, що відображає невизначеність, — значною.

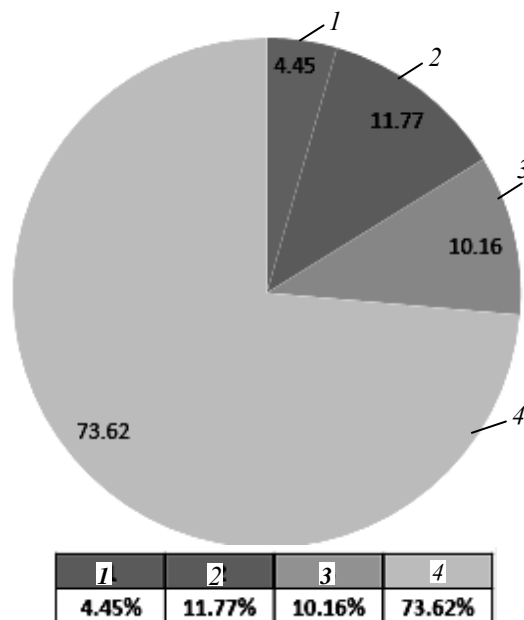
Експеримент 2 теж вказує на недостатню перспективність свердловини і має більшу невизначеність, ніж попередній.

Експеримент 3 має більшу невизначеність, ніж два попередні. Він показує, що лінгвістична змінна «перспективність свердловини», може бути як недостатньою, так і достатньою.

Експеримент 4, на відміну від попередніх, має значну невизначеність, але лінгвістична змінна «перспективність свердловини» може бути в діапазоні від терму «достатня» до терму «висока».

З експериментів 5 і 6 видно, що встановлюється значення лінгвістичної змінної «достатня», невизначеність незначна. Починаючи з експерименту 8 модель видає значення, близьке до терму «висока». Робимо висновок, що найбільш поширені ситуації є такими, коли межі інтервалу чітко вказують на належність лінгвістичної змінної певному стану, тоді експерт дуже швидко приймає рішення. Інший випадок, коли ширина вихідного інтервалу включає декілька значень лінгвістичної змінної, тоді експерту треба приймати кінцеве рішення, яке базується на його досвіді.

Рис. 4. Розташування областей невизначеності для експериментів 2 і 3



Розташування областей невизначеностей (рис. 4), які для наочності розраховані як добуток значення правої межі виходу моделі на ширину відповідного інтервалу для

кожного етапу експерименту 2 та аналогічно для експерименту 3. На діаграмі видно, що для експерименту 3, де лінгвістична оцінка подвійна, зона невизначеності набагато більша.

Таким чином, величина інтервального виходу узагальнювальної моделі вказує на існуючі в початкових даних невизначеності та вимагає від експерта уважного підходу до прийняття кінцевого рішення.

Для подальшого моніторингу стану родовища по етапах ГРП складемо таблицю, яка дасть можливість проаналізувати чи є зв'язок між шириною інтервалу та часу проведених робіт, а також проведемо дослідження чи існує відповідність між оцінками експерта та величиною інтервалу для кожного етапу всіх експериментів. Результати обчислення ширини інтервального виходу узагальнювальної моделі по етапах зі значенням лінгвістичної змінної наведено в табл. 2.

Таблиця 2. Результати обчислення ширини інтервального виходу узагальнювальної моделі з оцінкою експерта за лінгвістичною змінною

№	Етап 1		Етап 2		Етап 3		Оцінка експерта
	Вихід системи	Ширина інтервалу	Вихід системи	Ширина інтервалу	Вихід системи	Ширина інтервалу	
Результати роботи агрегованої моделі для оцінки 1/недостатня							
1	[0,03; 0,39]	0,36	[0, 53; 0,92]	0,39	[0,00; 0,02]	0,02	Недостатня
2	[0,04; 1,2]	1,16	[0,18; 2,01]	1,83	[1,68; 2,81]	1,13	Недостатня
Результати роботи узагальнювальної моделі для оцінки 2/недостатня							
3	[1,3; 5,48]	4,18	[1,38; 5,48]	4,1	[1,25; 5,25]	4,2	Недостатня
Результати роботи узагальнювальної моделі для оцінки 5/достатня							
4	[6,41; 6,72]	0,31	[5,82; 7,48]	1,66	[6,62; 7,82]	1,2	Достатня
Результати роботи узагальнювальної моделі для оцінки 6/достатня							
5	[5,66; 6,2]	0,54	[5,71; 6,56]	0,85	[5,56; 5,62]	0,06	Достатня
6	[7,32; 7,7]	0,38	[5,37; 6,15]	0,78	[4,78; 5,71]	0,93	Достатня
Результати роботи узагальнювальної моделі для оцінки 8/висока							
7	[9,03; 9,71]	0,68	[9,88; 10]	0,12	[9,92; 9,99]	0,07	Висока
Результати роботи узагальнювальної моделі для оцінки 9/висока							
8	[8,54; 8,76]	0,22	[9,95; 10]	0,05	[9,91; 10]	0,09	Висока
9	[8,62; 8,74]	0,12	[9,64; 9,77]	0,13	[9,9; 9,99]	0,09	Висока

Дані, що наведені в табл. 2, вказують на зміну значень вихідного інтервалу моделі залежно від терміну проведення ГРП. Отримані результати дають змогу порівняти величину інтервалу, що виникає на виході моделі,

з оцінкою, яку виставляє для кожного значення експерт. А саме, для експерименту 1 інтервальне вихідне значення на всіх етапах дуже мале, тому оцінка експертом лінгвістичної змінної «перспективність свердловини «недостатня» є очевидною. За експериментом 2 ширина інтервального виходу збільшується, але залишається стабільною на всіх етапах. Це дозволяє експерту, хоч перспективність і недостатня, все ж таки надати рекомендацію для продовження розвідувальних робіт. Для експериментів 3 і 4 тенденція стабільності ширини інтервалу зберігається. Надалі для експерименту 4 по експеримент 6 з'являється терм «достатня», ширина інтервалу починає зменшуватись і, нарешті, для експериментів 7, 8, 9 спостерігається її зменшення на всіх етапах, тобто інтервальна оцінка перетворюється в точкову. Це підтверджує факт, що зникнення невизначеності означає одночасно збільшення значення балів, які набирає лінгвістична змінна.

Розглянемо, чи впливає на оцінку експертів величина інтервального виходу моделі. Для своїх оцінок експерт, як правило, орієнтується на верхню границю інтервалу, що є поширеною практикою. Але правильність його суджень повинна підтверджуватись зменшенням або збільшенням інтервалу на виході моделі. Наведемо приклади наявності чи відсутності такої відповідності:

- оцінка «недостатня» / 1,2 бали — ширина інтервалу: від 0,36 до близько 4;
- оцінка «достатня» / — ширина інтервалу: від 0,31 до близько 1,16;
- оцінка «висока» / 8 балів — ширина інтервалу: від 0,68 до 0,07;
- оцінка «висока» / 9 балів — ширина інтервалу: від 0,22 до 0,09;
- оцінка «висока» / 9 балів — ширина інтервалу: від 0,12 до 0,09.

Характерною є оцінка «висока», коли інтервал фактично зникає і відповідно оцінка експерта повинна бути найбільшою, що і відбувається за шкалою. Це оцінка 9, яка відповідає фактично точковому значенню. Решта оцінок експерта є теж коректними, оскільки для низьких балів ширина інтервалу є найбільшою. Продемонструємо наведені судження графічно. Із рис. 5 видно як зменшується невизначеність результатів експериментів; процес відбувається послідовно з переходом від етапу 1 до етапу 3 (позначено світлими і темними відтінками).

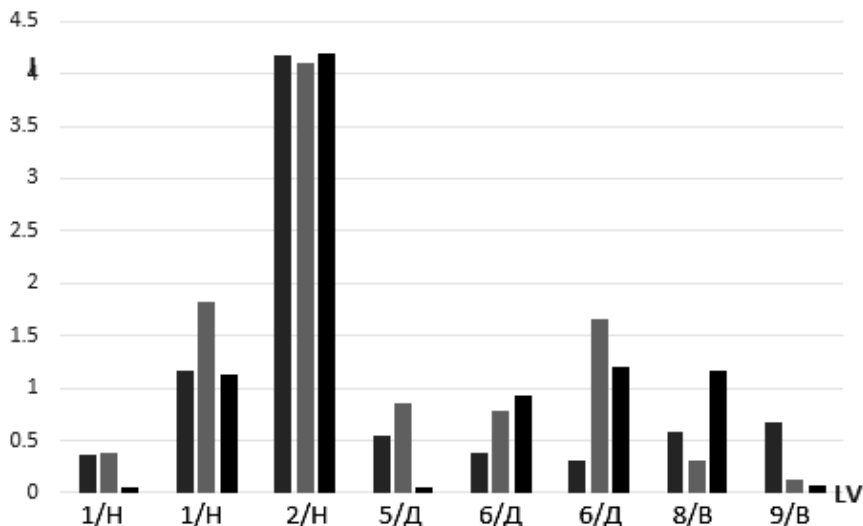


Рис. 5. Інтервальне вихідне значення моделі (I) та лінгвістичними змінними (LV); Н — недостатня; Д — достатня; В — висока

Аналіз отриманих результатів показує, що для підвищення якості прийняття рішень експерту необхідно звертати увагу на величину інтервального виходу узагальнювальної нечіткої моделі типу-2 та давати оцінку невизначеності, яка притаманна вихідному значенню. З огляду на складність проблеми розробнику варто скористатись такими практичними рекомендаціями:

- за умови рівності крайніх правих точок інтервальних значень для вихідної змінної перевагу необхідно надати тому значенню, яке має менший рівень невизначеності;
- якщо отримана в результаті роботи моделі ширина інтервалу включає декілька рівнів і не дозволяє чітко встановити єдине якісне значення з вихідної системи якісних термів, то необхідно провести додаткове експериментальне дослідження необхідних параметрів, або оновити експертні дані, з метою зменшення існуючої невизначеності;
- кожна окремо інтервальна модель, яка будується на основі знань експерта, може давати достатньо широкий діапазон лінгвістичних оцінок для стану системи. Це не означає, що рішення, отримане узагальнювальною моделлю, буде некоректним, швидше побудована модель вказує на значну невизначеність, тобто вхідних даних не вистачає для визначення вихідної величини. Це повинні враховувати фахівці, в обов'язки яких входять завдання моніторингу станів складних систем.

Отримані результати показують, що побудована узагальнювальна модель може виконувати функції помічника експерта з питань моніторингу станів складних систем в умовах неповноти даних. Результати дослідження ефективно можуть бути застосовані в довготривалих дослідженнях природних процесів задля скорочення терміну прийняття рішення та заощадження необхідних для цього ресурсів.

ВИСНОВОК

Запропоновано теоретико-множинний підхід до узагальнення виходів нечітких моделей типу-2 з інтервальними функціями належності, які входять до узагальнювальної моделі в завданні моніторингу станів складних систем. Наведено приклади використання узагальнювальної нечіткої моделі типу-2 в завданні моніторингу станів артезіанської свердловини в природному середовищі. Проведено дослідження відповідності інтервального виходу узагальнювальної моделі з лінгвістичними оцінками експертів та показано адекватність рішень, прийнятих експертом. Надано рекомендації експертам щодо покращення прийняття рішень стосовно оцінки вихідного інтервалу узагальнювальної моделі.

ЛІТЕРАТУРА

1. Q. Liang and J.M. Mendel, "Interval Type-2 fuzzy logic systems: theory and design," *IEEE Trans. on Fuzzy Syst.*, vol. 8, pp. 535–550, 2000.
2. Y.P. Zaychenko, *Fuzzy models and methods in intelligent systems*, (in Russian). Kyiv: Slovo, 2008.
3. Arnab Kumar De, Debjani Chakraborty, and Animesh Biswas, "Literatura review on type-2 fuzzy set theory," *Soft Computing*, 26 (18), pp. 9049–9068, 2022. doi: 10.1007/s00500-22-07304-4.
4. N.R. Kondratenko and O.O. Snihur, "Investigating adequacy of interval type-2 fuzzy models in complex objects identification problems," *System Research and Information Technologies*, no. 4, pp. 94–104, 2019.

5. N.R. Kondratenko, "Improving Adequacy of Type-2 Fuzzy Models by Using Type-2 Fuzzy Sets," *Naukovi visti NTUU KPI*, no. 6, pp. 56–61, 2014.
6. Xiaoming Tang, Li Deng, Jimin Yu, and Hongchun Qu, "Output Feedback Predictive Control of Interval Type-2 T-S Fuzzy Systems With Markovian Packet Loss," *IEEE Transactions on Fuzzy Systems*, vol. 26, no. 4, August 2018.
7. N.R. Kondratenko, O.V. Cheboraka, and O.A. Tkachuk, "Interval Type-2 Fuzzy Models Concerning Identification Problems of Multiple-input Multiple output Objects," *Information Processing Systems*, no. 3(93), pp. 48–52, 2011.
8. N. Kondratenko and O. Snihur, "Interval Fuzzy Modeling of Complex Systems under Conditions of Input Data Uncertainty," *Eastern-European Journal of Enterprise Technologies*, vol. 4/4 (82), pp. 20–28, 2016.
9. N.R. Kondratenko, O.V. Cheboraka, and S.M. Kuzemko, "Study of Aggregating Interval Type-2 Fuzzy Models Capabilities for Forecasting of Time Series," *Visnyk VPI*, no. 4, pp. 22–27, 2010.
10. N.R. Kondratenko, "Vykorystannya nechitkykh baz znanj z funktsiyamy nalezhnosti typu-2 u medychnyi diahnozytsi [Use of fuzzy knowledge bases with type-2 membership functions in medical diagnostics]," *Materials of the articles of the International scientific and practical conference "Actual problems of medical, biological physics and informatics"*, Vinnytsia, 2022. Available: <https://drive.google.com/file/d/1icajVT7OKyVxlfXZd1czwhS13EhtmUj8/view?usp=sharing>
11. N.R. Kondratenko, N.B. Zelinsjka, and S.M. Kuzemko, "Fuzzy Logic Systems with Allowance for the Blank in Experimental Data Taken," *Naukovi visti NTUU KPI*, no. 5, pp. 37–41, 2004.
12. P. Melin and O. Castillo, "An Intelleigent hybrid approach for industrial quality control combining neural networks, fuzzy logic and fractal theory," *Information Sciences*, vol. 147, no. 177, pp. 1543–1557, 2007.

Received 13.09.2022

INFORMATION ON THE ARTICLE

Nataliia R. Kondratenko, ORCID: 0000-0002-4450-1603, Vinnytsia National Technical University, Ukraine, e-mail: kondrn2014@gmail.com

Olga O. Snihur, ORCID: 0000-0002-9268-6876, PhD, private entrepreneur, Ukraine, e-mail: olha.snihur@gmail.com

Roman M. Kondratenko, ORCID: 0000-0002-7982-0786, BelISA, Belarus, e-mail: rkondratenko.niei@gmail.com

INTERVAL TYPE-2 GENERALIZING FUZZY MODEL FOR MONITORING THE STATES OF COMPLEX SYSTEMS USING EXPERT KNOWLEDGE / N.R. Kondratenko, O.O. Snihur, R.M. Kondratenko

Abstract. A type-2 interval generalizing fuzzy model is proposed for monitoring complex systems' states. A set-theoretic approach is proposed to generalize the results of type-2 fuzzy models with interval membership functions. The study of the correspondence of expert assessments to the output value of the generalizing fuzzy model over the observation interval is presented. Examples of the use of generalizing fuzzy model type-2 in the task of monitoring the conditions of an artesian well are given. It is shown that in order to improve the quality of decisions made, the expert needs to pay attention to the value of the interval output of the generalizing type-2 fuzzy model. Recommendations are presented to experts to improve decision-making regarding the estimation of the output interval of the generalizing model.

Keywords: type-2 fuzzy model, interval membership function, set-theoretic approach, expert knowledge.

НЕЙРОННІ МЕРЕЖІ: ДОСЛІДЖЕННЯ ПРАВИЛ
ПРИЙНЯТТЯ НИМИ РІШЕНЬ

А.І. ПЕТРЕНКО, І.А. ВОХРАНОВ

Анотація. Питання отримання більшої зрозумілості поведінки нейронних мереж є досить актуальним, особливо у галузях із високим рівнем ризиків. Для вирішення цієї задачі досліджено можливості нового алгоритму декомпозиції DeepRED, здатного витягувати правила прийняття рішень глибинними нейронними мережами з декількома прихованими шарами DNN (Deep Neural Networks). Дослідження алгоритму DeepRED проводилося на прикладі вилучення правил експериментальної нейронної мережі за виконання класифікації зображень бази даних MNIST рукописних цифр, що дозволило виявити ряд обмежень алгоритму DeepRED.

Ключові слова: вилучення правил, нейронні мережі, DeepRED, машинне навчання, дерева рішень, графи рішень.

ОПИС ЗАДАЧІ І СТАНУ ЇЇ РІШЕННЯ

Натепер, нейронні мережі (NN, Neural Networks) мають дуже широкий спектр застосувань. Вони здатні вирішувати такі задачі, як задачі класифікації, з дуже високою ефективністю. Проте, не зважаючи на всі їх переваги, все ще існує проблема з можливістю використання нейронних мереж у галузях із високим рівнем ризиків. У таких сферах, як медицина, фінанси та енергетика, розуміння та передбачуваність поведінки систем є дуже критичними. Будь-який неочікуваний сценарій може привести до ризиків для життя чи здоров'я людей, або до ризику великих фінансових втрат [1, 2]. Саме тому, питання отримання більшої зрозумілості поведінки нейронних мереж є досить актуальним.

Нейронні мережі формують свою логіку завдяки використанню шарів нейронів. Структуру всіх нейронних мереж, в загальному, можна звести до одного спільного вигляду: вхідний шар, вихідний шар та набір прихованих навчальних шарів (рис. 1).

Щоб подолати вказану незручність нейронних мереж, протягом останніх трьох десятиліть, детально розробляються різні способи пояснення “логіки” прийняття рішень нейронними мережами. Першим запровадженим та найбільш перспективним на сьогоднішній день підходом у цій галузі є **вилучення правил** (RE, Rule Extraction) із штучних нейронних мереж. Вилучення правил — це підхід, що зосереджується на розкритті прихованих в мережі правил, з метою допомогти пояснити, як саме нейронна мережа

приходить до остаточного рішення. Такі вилучені правила можуть використовуватись для зменшення небезпеки, відстеження стабільності та відмовостійкості системи, перевірки та підтвердження можливості використання нейронної мережі для конкретної задачі, тощо. Більшість дослідників зосереджуються на тому, що вилученні правила мають бути максимально зрозумілими, але в той же час повинні якомога точніше імітувати поведінку нейронної мережі.

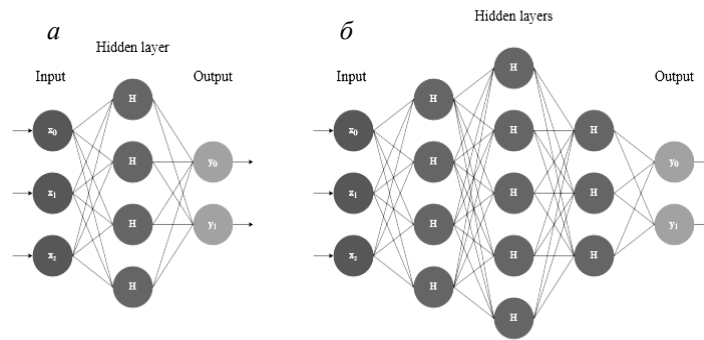


Рис. 1. Структура нейронних мереж: а — звичайної нейронної мережі, б — глибинної нейронної мережі

Існуючі методи вилучення правил поділяються на три групи: **декомпозиційні, педагогічні та еклектичні** [2]. Декомпозиційні методи повністю спираються на архітектуру мережі і використовують активації та ваги всіх нейронів, включаючи приховані шари. Зазвичай ці методи аналізують кожен нейрон, після чого, отримані описи поведінки цих нейронів певним чином об'єднуються та формують правила, що імітують поведінку всієї моделі [7].

Основними представниками декомпозиційних підходів є: метод КТ [10], метод на основі правил нечіткої логіки [6], поліноміальний алгоритм Цукімото [4], CRED [12], DeepRED [7].

Педагогічні методи, на відміну від декомпозиційних, не враховують внутрішню структуру нейронної мережі, а розглядають NN як єдину сутність (чорну скриньку). Їх принцип полягає в тому, щоб витягувати правила шляхом прямого відображення вхідних даних у вихідні. Іншим чином, це можна розглядати, як добре відому задачу апроксимації, де нейронна мережа виступає в ролі цільової функції, що приймає визначений набір вхідних параметрів та повертає певний результат класифікації. Маючи цю функцію, алгоритми намагаються знайти узгодженість між вхідними варіаціями та результатами. Найбільш популярними представниками педагогічних методів є: VIA [17] [18], TREPAN [13], BIO-RE [5], ANN-DT [3], STARE [20], KDRuleEx [8][9], RxREN [11].

Еклектичні методи представляють собою поєднання інших підходів і розглядають лише частину NN як чорний ящик. Оскільки визначення еклектичних підходів досить розпливчате, іноді дослідники мають різні погляди на те, які методи правильно відносити до еклектичних, а які ні. Двома, досить відомими методами, які можна віднести до еклектичних, є MofN [2] та FERNN [15].

Більшість алгоритмів вилучення правил запропоновані лише для невеликих нейронних мережах з *одним прихованим шаром*. Проте, нещодавно було запропоновано новий алгоритм — **DeepRED**, який здатен працювати з *глибинними нейронними мережами*. Цей алгоритм декомпозиції витягує проміжні правила для кожного шару NN, на основі чого, формує представлення того, як конкретна нейронна мережа приймає рішення [7]. Він вважа-

ється найбільш досконалим з точки зору зрозумілості вилучаємих правил і їх максимального наближення до поведінки НМ. Слід визнати, що існують модифікації DeepRED, які поліпшують його властивості [14].

Стаття організована таким чином: розділ I містить опис задачі і огляд стану її рішення; розділ II — огляд алгоритму; розділ III — опис покращення алгоритму; розділ IV — експериментальні результати дослідження алгоритмів, а розділ V містить висновки щодо задачі, яка розглядається.

ОПИС РОБОТИ АЛГОРИТМУ DEEPRED

Основною сутністю будь-якої нейронної мережі є модель нейрона (рис. 2). Ідея нейронної моделі полягає в тому, що вхід x разом із зміщенням b зважуються вагами w , а потім підсумовуються разом. Зміщення b (*bias*) є скалярним значенням, тоді як вхідні дані x та ваги w мають векторне значення, тобто $x \in \mathbb{R}^n$ та $w \in \mathbb{R}^n$, де $n \in \mathbb{N}$, що відповідає розмірності вхідних даних. Їх сума $z = x^T x + b$ виступає в якості аргумента **функції активації** ϕ , в результаті чого, формується вихід нейронної моделі:

$$y = \phi(z) = \phi(w^T x + b) \tag{1}$$

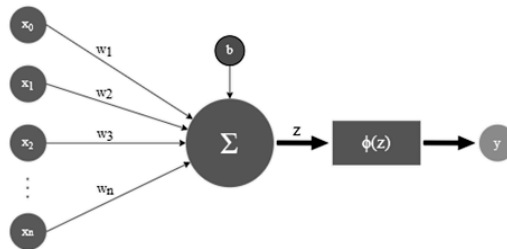


Рис. 2. Модель штучного нейрона

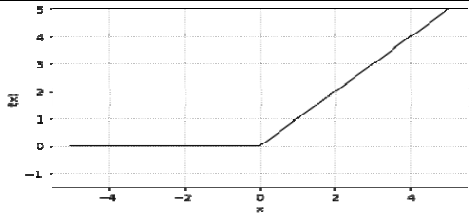
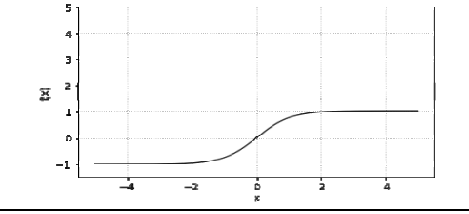
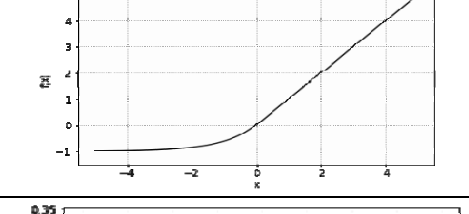
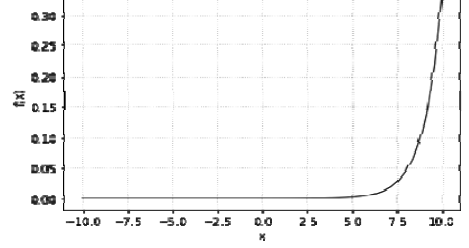
Деякі поширені функції активацій та їх графічні зображення приведені в табл. 1.

Вхідними даними алгоритму є навчальна вибірка даних і задані активації всіх шарів мережі. Найпопулярнішими функціями активації для глибоких нейронних мереж є функція активації ReLU і функція Softmax.

Таблиця 1. Деякі функції активацій та їх графічні зображення

Назва	Функція (ϕ)	Графічне зображення
Step (східчаста)	$\phi(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } x \geq 0 \end{cases}$	
Sigmoid (Logistic)	$\phi(x) = \frac{1}{1 + e^{-x}}$	

Продовження табл. 1

Назва	Функція (φ)	Графічне зображення
Rectified Linear Unit (ReLU)	$\varphi(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ x & \text{if } x > 0 \end{cases}$	
Hyperbolic Tangent (Tanh)	$\varphi(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$	
Exponential Linear Unit (ELU)	$\varphi(x) = \begin{cases} \alpha(e^x - 1) & \text{if } x \leq 0 \\ x & \text{if } x > 0 \end{cases}$	
Softmax	$\varphi(x) = \frac{e^{x_i}}{\sum_{j=1}^J e^{x_j}}$ for $i = 1, \dots, J$	

Алгоритм DeepRED було обрано в роботі, оскільки це метод декомпозиції, пристосований до аналізу глибоких нейронних мереж DNN (Deep Neural Networks), точніше до MLP (Multi-Layer Perceptron) з будь-якою глибиною, але без зворотних зв'язків (feedback loop). Багат шаровий перцептрон (MLP) відноситься до багаторівневої архітектури з усіма можливими зв'язками між шарами, які також позначаються як повністю зв'язані шари. Шари, у яких немає вхідних або вихідних нейронів, називають прихованими шарами (рис. 1.). Алгоритм використовує апроксимаційні моделі ієрархічно з глибиною, пропорційною загальному числу шарів NN. Апроксимуючі моделі правил прийняття рішень мають *структуру дерева*, вузли якого визначаються нейронами окремих шарів, а гілки якого мають значення, обчислені за формулою (1) для активацій нейронів, при цьому набір даних поділяється на менші підмножини. Для довільної мережі з k прихованими шарами заради зручності введемо позначення для кожного прихованого шару: h_1, \dots, h_k . Таким чином, в результаті першого кроку, отримується *дерево рішень (DT)*, що описує вихідний шар мережі через активації шару h_k . Наступним кроком алгоритму є обробка шару h_{k-1} . Для кожної умови із набору правил, отриманого на попередньому етапі, застосовується алгоритм C4.5 чи C5

[21, 22], щоб побудувати дерева рішень, які тепер будуть описувати шар h_k за допомогою h_{k-1} . Аналогічним чином, даний процес продовжується для кожного наступного шару, до тих пір, поки не буде отримано дерева рішень, що описують шар h_1 через входи нейронної мережі. При цьому, під час проходження алгоритму, застосовується механізм, що запобігає виконанню зайвих запусків C4.5. Якщо дерево рішень для опису певного випадку вже було побудоване раніше, наявні результати будуть просто скопійовані. Після отримання DT, що описують кожен шар мережі, виконується об'єднання отриманих дерев разом. Усі невідповідні та зайві правила в процесі викидаються. В результаті формується одне дерево рішень, що описує виходи NN на основі її входів, тобто описує поведінку самої нейронної мережі. На рис. 3 зображено псевдокод реалізації DeepRED, представлений автором алгоритму в його роботі [7].

ПОКРАЩЕННЯ РЕАЛІЗАЦІЇ АЛГОРИТМУ

Із псевдокоду, представленого на рис. 3, видно, що для ініціалізації дерев на першому кроці, в своїй первинній формі, алгоритм використовував набли-

```

Input: Neural network NN, training examples x
Output: Set of rules rules

activationValues = getHiddenActivationValues(NN,x)
activationValues(outputLayer) = NN(x) // one-hot encoded
foreach currentOutput ∈ outputNeurons do
    // intermediateTerms stores terms that describe higher-level terms
    intermediateTerms(outputLayer, 0) = currentOutput > 0.5
    // currentOutput is the class of interest, 0 used as dummy
    foreach currentLayer ∈ hiddenLayersDescending(NN) do
        foreach term ∈ intermediateTerms(currentLayer+1) do
            if treeAlreadyExtractedFor(term, currentLayer+1) then
                | intermediateTerms(currentLayer, term) = copyTermsFor(term, currentLayer+1)
            else
                | intermediateTerms(currentLayer, term) = C4.5(activationValues(currentLayer),
                | activationValues(currentLayer+1), term)
                // Describe term in next deeper layer by terms of current layer
            end
        end
    end
    while getNumberOfLayers(intermediateTerms) > 2 do
        | intermediateTerms = mergeIntermediateTerms(getNumberOfLayers(intermediateTerms),
        | getNumberOfLayers(intermediateTerms)-1)
        | intermediateTerms = deleteUnsatisfiableTerms(intermediateTerms)
        | intermediateTerms = deleteRedundantTerms(intermediateTerms)
    end
    rules[] = intermediateTerms2Rules(intermediateTerms)
    // Describes currentOutput by rules consisting of input neuron splits
    delete(intermediateTerms)
end

```

Рис. 3. Псевдокод оригінальної реалізації алгоритму DeepRED [7]

ження типу: “*IF output_i > 0,5 THEN class = i*”. Основною причиною цього є те, що оригінальна робота була зосереджена лише на проблемі бінарної класифікації, де використання таких правил доцільне. Проте, якщо потрібно застосовувати DeepRED у задачах з багатокласовою класифікацією, таке наближення частіше всього буде занадто грубим. Альтернативою в даному випадку є побудова DT на основі активацій останнього прихованого шару та результатів класифікації мережі в якості вхідних та вихідних векторів відповідно[14].

Оскільки взаємодії між шарами NN швидше мають форму графа, а не дерева, набагато точніший та більш виразний опис того, що відбувається в процесі злиття правил, забезпечує використання DDAG (Decision directed acyclic graph, або орієнтований ациклічний граф рішень) отриманих з DT, замість самих DT. Це, можна сказати, розширення до привичного дерева рішень. Єдиною відмінністю DDAG від DT є лише те, що він може мати структуру, відмінну від деревоподібної, тобто вузли можуть мати більше одного вхідного ребра (рис. 4). Завдяки такій структурі, DDAG має ряд пе-

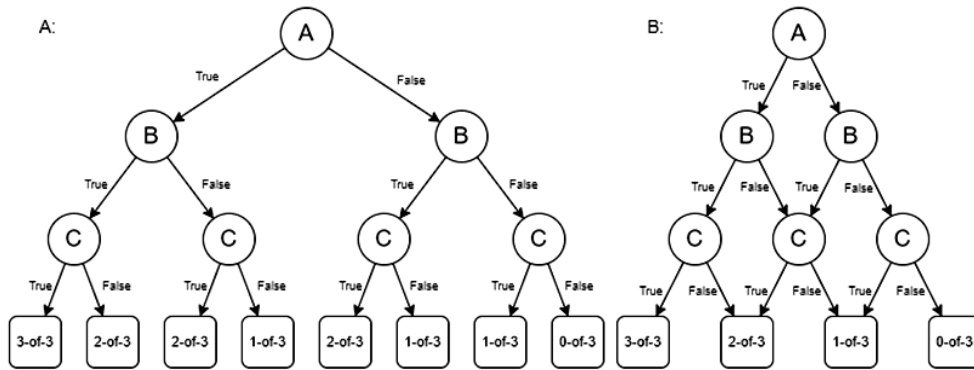


Рис. 4. Представлення правил M-of-{A,B,C} у вигляді: а — DT, б — DDAG

реваг для вирішення даного завдання, основними з яких є економніше використання пам'яті (що грає чималу роль при роботі в середовищі з обмеженими ресурсами) та складніша оцінка представлення правил.

```

Algorithm 3.1 DeepRED pseudocode using DDAG
function DEEPRED( $(h_l(X))_{l=0}^L$ )
     $g_L \leftarrow \text{INIT\_DT}(h_L(X))$ 
    for  $l = L - 1, L - 2, \dots, 0$  do
        Set  $T$  as all unique splits in  $ddag_{h_{l+1}}$ 
        for all  $t \in T$  do
             $\tilde{X} \leftarrow h_l$ 
             $\tilde{Y} \leftarrow t(h_l(X))$ 
             $DT_t \leftarrow \text{BUILD\_DT}(\tilde{X}, \tilde{Y})$ 
        end for
         $g_l \leftarrow \text{SUBSTITUTE}(dag_{l+1}, (DT_t)_{t \in T})$ 
         $g_l \leftarrow \text{REMOVE\_UNSATISFIABLE}(g_l)$ 
         $g_l \leftarrow \text{REMOVE\_REDUNDANT}(g_l)$ 
    end for
    return  $g_0$ 
end function
    
```

Рис. 5. Псевдокод модифікованої реалізації алгоритму DeepRED

Також, за рахунок меншої кількості вузлів, що потрібні для представлення одного і того ж набору правил, порівняно з DT, DDAG набагато зручніші при їх аналізі, оскільки мають більш зручну для читання форму.

Псевдокод модифікованої реалізації алгоритму DeepRED приведений на рис. 5. Спочатку будується DT для останньої активації h_L , тобто обчислюються аргументи максимуму $h_L(x)$ для кожного значення $x \in X$.

Після ініціалізації цикл *for* продовжує розглядати шари у зворотному напрямку. На кроці *l* створюються DT для кожного розділеного вузла в DDAG *gl+1*. Потім DTs замінюють вузли всередині DDAG, створюючи таким чином новий DDAG *gl* із вхідними даними нижнього рівня. Нарешті, виконується виключення незадовільних та надлишкових вузлів з умов відповідності всіх вхідних даних вузлів значенням їх функцій активації. Заміна вузла за допомогою DT виконується шляхом переключення вхідних ребір *кореневого вузла (root)* DT і підключення усіх ребер, що ведуть до *листових (leaf) вузлів True і False* (істинних та хибних), до істинних та хибних ребер початкового вузла відповідно.

РЕЗУЛЬТАТИ ДОСЛІДЖЕНЬ

Як нейронну мережу, на якій досліджувалася задача класифікації рукописних символів, взято MLP з двома прихованими шарами, розмірами в 100 та 30 нейронів. На всіх проміжних шарах мережі використовувалася функція активації ReLU, а як функція активації вихідного шару застосовувалася Softmax. Таким чином, враховуючи розмірності вхідних та вихідних векторів, використовувалася нейронна мережа з наступними розмірами всіх шарів:

- 784 — вхідний шар (відповідає вектору вхідних значень);
- 100 — перший прихований шар;
- 30 — другий прихований шар;
- 2 — вихідний шар (відповідає вектору результатів класифікації).

Для навчання цієї експериментальної NN, з якої в ході експериментів вилучались правила прийняття рішень, використовувалась база даних рукописних цифр MNIST (Modified National Institute of Standards and Technology dataset) [23]. Вона містить навчальну вибірку розміром 60000 зображень та тестову вибірку розміром 10000 зображень. Всі зображення бази монотонні (у відтінках сірого різної інтенсивності) розміром $28 \times 28 = 784$. Для того, щоб сформувані вхідну вибірку для бінарної класифікації, притаманної алгоритму DeepRED, із бази даних MNIST було відібрано лише зображення, що містять цифри 0 та 1 (рис. 6). Кожен елемент вектора вхідних значень приймає значення в діапазоні $[0, 1]$, де 0 — означає порожній піксель, а 1 — повністю зафарбований піксель.

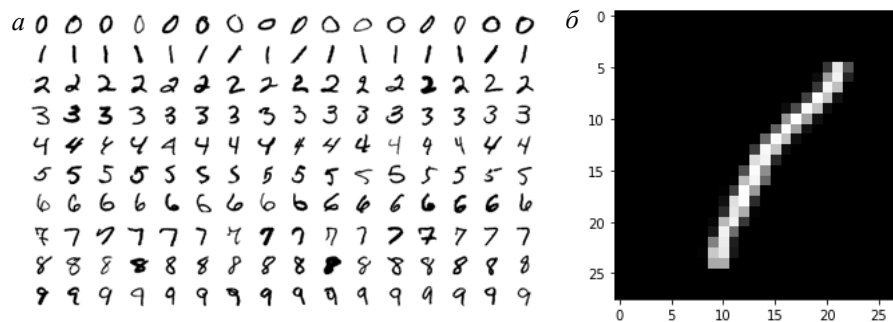


Рис. 6. Зображення бази даних MNIST: *a* — приклади зображень вибірки, *б* — перетворення зображення у вектор вхідних значень

Процес прийняття рішення щодо класифікації зображення з цифрою “1” шляхом проходження вузлами отриманого графа при вилученні правил ілюструється графом, який із-за своїх розмірів наданий окремо у форматі PDF-файла за посиланням: <https://drive.google.com/file/d/1-dzEd8GsNpLpAlHl-zP-sQZjafxFgbs5/view?usp=sharing>

На цьому графі в даному випадку, вузол {0} відповідає результату класифікації, який стверджує, що вхідне зображення містить цифру 0, а вузол {1}, відповідно, що вхідне зображення містить цифру 1. Інші вузли графа, що відповідають за відображення правил прийняття рішень нейронної мережі мають вигляд правил у форматі “ $x_i > \text{intensity}$ ”, де x_i — це значення i -го елемента вектора вхідних значень, що відповідає інтенсивності відповідного пікселя зображення (відповідно до рис 6, б), а intensity — це значення інтенсивності, вилучене з “логіки” нейронної мережі, з яким ведеться порівняння інтенсивності пікселя зображення.

Порівняльні дослідження алгоритму DeepRED проводилися для випадків: бінарної класифікації, класифікації з 3-ма класами та класифікації з 4 ма класами. Відповідні результати перших двох запусків алгоритму наведені у табл. 3 та табл. 4. У них використано такі показники для порівняльного аналізу:

- DDAG fidelity — точність класифікації DDAG відносно класифікації NN на тестовій вибірці;
- DDAG accuracy — точність класифікації DDAG відносно еталонних значень тестової вибірки;
- DDAG recall — точність класифікації DDAG відносно класифікації NN на навчальній вибірці;
- final DDAG size — розміри фінального графа вилучених правил;
- DDAG size before pruning — розмірами графа рішень, перед видаленням зайвих правил.

З таблиць видно, що складність алгоритму значно зростає зі збільшенням кількості класів, а також зі збільшенням кількості шарів NN та їх розмірів завдяки зростанню розмірів побудованих DDAG, що призводить до росту обчислювальної складності. Окрім того, кінцеві результати сильно залежать від того, як буде ініціалізовано дерево на першому кроці.

Таблиця 3. Результати запусків Deep RED у випадку бінарної класифікації (зображення з цифрами 1 та 3)

Конфігурація MLP	NN accuracy	DDAG fidelity	DDAG accuracy	DDAG recall	Final DDAG size	DDAG size before pruning
MLP(20)	0.9995	0.9674	0.9748	0.9669	26	27
MLP(50)	0.9995	0.9646	0.9753	0.9732	16	16
MLP(100)	0.9991	0.9655	0.9711	0.9698	20	20
MLP(300)	0.9995	0.9734	0.9739	0.9713	18	18
MLP(50,25)	0.9986	0.9347	0.9608	0.9546	46	51
MLP(100,30)	0.9995	0.9669	0.9776	0.9722	28	43
MLP(300, 150)	0.9995	0.9757	0.9795	0.9693	19	19
MLP(50, 25, 9)	0.9986	0.9618	0.9734	0.9650	11	14
MLP(100,50,30)	0.9936	0.9417	0.9548	0.9522	65	69

Таблиця 4. Результати запусків Deep RED у випадку класифікації з 3 класами (зображення з цифрами 0, 1, 2)

Конфігурація MLP	NN accuracy	DDAG fidelity	DDAG accuracy	DDAG recall	Final DDAG size	DDAG size before pruning
MLP(20)	0.9930	0.9005	0.8967	0.8888	88	113
MLP(50)	0.9946	0.8599	0.8590	0.8579	91	109
MLP(100)	0.9939	0.8815	0.8770	0.8778	109	132
MLP(300)	0.9936	0.8561	0.8513	0.8557	87	163
MLP(100,30)	0.9940	0.7353	0.7426	0.7252	214	451
MLP(300,150)	0.9952	0.5329	0.3943	0.3942	147	328
MLP(50, 25, 9)	0.9942	0.8780	0.6222	0.6143	123	332
MLP(100,50,30)	0.9959	0.7925	0.4559	0.4499	134	357

ВИСНОВКИ

У процесі дослідження алгоритму DeepRED виявлено, що один із найбільш перспективних способів його використання — це застосування в поєднанні з процесом навчання нейронної мережі. За рахунок того, що вилучення правил проводиться безпосередньо під час навчання мережі, з’являється краще розуміння того, як та чи інша зміна в архітектурі NN впливає на отримані результати. По-суті, при правильному використанні результати вилучення правил можуть бути застосовані у вигляді додаткового критерія оцінки якості навчання.

На сьогоднішній день, вимірювання точності результатів класифікації на тестовій вибірці — це ледь не єдиний критерій, яким керуються під час навчання нейронних мереж. Але насправді, ця точність не завжди в повній мірі відображає те, наскільки добре пройшов навчальний етап. Якщо в однакових умовах з однієї мережі вдається вилучати чіткі та лаконічні правила, а з іншої — дещо ускладнені та заплутані, і, при цьому, точність у них однакова, логічніше буде віддати перевагу першій, оскільки, в загальному випадку, вона буде більш передбачувана.

Опираючись на результати, що були отримані в рамках проведених практичних експериментів, можна сказати, що вилучення правил з нейронних мереж за допомогою декомпозиційного алгоритму DeepRED виглядає досить перспективно. Вилучені за допомогою алгоритму правила мають досить зрозумілу форму та відносно не складні при їх аналізі. Особливо вдалим рішенням, хотілося б відмітити ідею покращення алгоритму, за рахунок розширення дерев рішень до графів рішень (DDAG). Це дозволило сильно полегшити можливості читання отримуваних правил. Звісно, варто також зазначити, що проблеми зі зручністю аналізу вилучених правил все ще будуть актуальними при досить великих розмірах DDAG, але як уже відмічалось, такі алгоритми постійно балансують в рамках компромісу точності та зрозумілості результатів.

Підсумовуючи все, можна зробити висновок, що вирішувати повноцінно задачу вилучення правил з нейронних мереж, у тій формі, у якій DeepRED представлений на даний момент, все ще досить проблематично. Основною проблемою, яка сильно ускладнює можливість його масового використання, є проблема універсальності. За рахунок того, що алгоритм здатний працювати лише з MLP, відсіюється дуже велика частка можливостей, щодо застосування нейронних мереж у критичних сферах. Адже, на сьогоднішній день, найбільш перспективними нейронними мережами, для вирішення складних задач є мережі, що мають архітектуру, складнішу за MLP. Проте, в той же час, варто розуміти, що DeepRED — це найкращий декомпозиційний алгоритм для вилучення правил, серед тих, що існують на даний момент. При цьому, сама ідея, на якій засновано даний алгоритм, виглядає перспективною, і, схоже на те, що можливостей щодо її покращень ще досить багато. Зокрема чекає вирішення проблеми з сильною залежністю від етапу початкової ініціалізації дерева і ообота над подальшої оптимізацією розмірів графів рішень, наприклад, за рахунок використання вузлів M-of-N.

ЛІТЕРАТУРА

1. Frank Emmert-Streib, Zhen Yang, Han Feng, Shailesh Tripathi, and Matthias Dehmer, “An Introductory Review of Deep Learning for Prediction Models With Big Data,” *Front. Artif. Intell.*, 2020. Available: <https://www.frontiersin.org/articles/10.3389/frai.2020.00004/full>
2. H. Jacobsson, “Rule Extraction from Recurrent Neural Networks: A Taxonomy and Review,” *Neural Computation*, vol. 17, no. 6, pp. 1223–1263, 2005. Available: <https://doi.org/10.1162/0899766053630350>
3. A. Bondarenko, L. Aleksejeva, V. Jumuts, and A. Borisovs, “Classification Tree Extraction from Trained Artificial Neural Networks,” *Procedia Computer Science*, vol. 104, pp. 556–563, 2017. Available: <https://doi.org/10.1016/j.procs.2017.01.172>
4. Hiroshi Tsukimoto, “Extracting rules from trained neural networks,” *IEEE Transactions on Neural networks*, 2000. Available: <https://ieeexplore.ieee.org/document/839008>
5. Ismail Taha and Joydeep Ghosh, “Symbolic interpretation of artificial neural networks,” *Knowledge and Data Engineering, IEEE, 1999*. Available: <https://ieeexplore.ieee.org/document/774103>
6. J.M. Benitez, J.L. Castro, and I. Requena, “Are artificial neural networks black boxes?” *IEEE Transactions on Neural Networks*, 1997. Available: <https://ieeexplore.ieee.org/document/623216>
7. J.R. Zilke, E.L. Mencia, and F. Janssen, “DeepRED—Rule extraction from deep neural networks,” in *T. Calders, M. Ceci, D. Malerba (Eds.): Discovery Science 19th International Conference, DS 2016, Bari, Italy, Oct. 19–21, 2016, Proceedings, LNAI 9956*, pp. 457–473, 2016. Available: https://doi.org/10.1007/978-3-319-46307-0_29
8. Kamal Kumar Sethi, Durgesh Kumar Mishra, and Bharat Mishra, “Extended Taxonomy of Rule Extraction Techniques and Assessment of KDRuleEx,” *International Journal of Computer Applications*, 2012. Available: https://www.researchgate.net/publication/258652014_Extended_Taxonomy_of_Rule_Extraction_Techniques_and_Assessment_of_KDRuleEx

9. Kamal Kumar Sethi, Durgesh Kumar Mishra, and Bharat Mishra, “KDRuleEx: A Novel Approach for Enhancing User Comprehensibility Using Rule Extraction,” *Third International Conference on Intelligent Systems Modelling and Simulation, IEEE*, 2012. Available: <https://ieeexplore.ieee.org/document/6169675>
10. LiMin Fu, “Rule generation from neural networks,” *IEEE Transactions on Systems, Man, and Cybernetics*, 1994. Available: <https://ieeexplore.ieee.org/document/299696>
11. M. Gethsiyal Augasta and Thangairulappan Kathirvalavakumar, “Reverse engineering the neural networks for rule extraction in classification problems,” *Neural Processing Letters*, 2012. Available: https://www.researchgate.net/publication/216628594_Reverse_Engineering_the_Neural_Networks_for_Rule_Extraction_in_Classification_Problems
12. Makoto Sato and Hiroshi Tsukimoto, “Rule extraction from neural networks via decision tree induction,” *International Joint Conference on Neural Networks, IEEE*, 2001. Available: <https://ieeexplore.ieee.org/document/938448>
13. Mark W. Craven and Jude W. Shavlik, “Extracting tree-structured representations of trained networks,” *Advances in Neural Information Processing Systems*, 1996. Available: <https://proceedings.neurips.cc/paper/1995/file/45f31d16b1058d586fc3be7207b58053-Paper.pdf>
14. Matej Fanta, Petr Pulc, and Martin Holena, *Rules Extraction from Neural Networks Trained on Multimedia Data*. 2019. Available: <http://ceur-ws.org/Vol-2473/paper4.pdf>
15. Rudy Setiono and Wee Kheng Leow, “FERNN: An algorithm for fast extraction of rules from neural networks,” *Applied Intelligence*, 2000. Available: <https://link.springer.com/article/10.1023/A:1008307919726>
16. Rudy Setiono, *Extracting M-of-N Rules from Trained Neural Networks*. National University of Singapore. Available: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.37.4104&rep=rep1&type=pdf>
17. Sebastian Thrun, “Extracting provably correct rules from artificial neural networks,” *Technical Report*. University of Bonn, Institut für Informatik III, 1993. Available: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.31.9441>
18. Sebastian Thrun, “Extracting rules from artificial neural networks with distributed representations,” *Advances in Neural Information Processing Systems*, 1995. Available: <https://proceedings.neurips.cc/paper/1994/file/bea5955b308361a1b07bc55042e25e54-Paper.pdf>
19. Tameru Hailesilassie, “Rule Extraction Algorithm for Deep Neural Networks: A Review,” *International Journal of Computer Science and Information Security*, vol. 14, no. 7, 2016. Available: <https://arxiv.org/ftp/arxiv/papers/1610/1610.05267.pdf>
20. Zhi-Hua Zhou, Shi-Fu Chen, and Zhao-Qian Chen, “A statistics based approach for extracting priority rules from trained neural networks,” *International Joint Conference on Neural Networks, IEEE*, 2000. Available: <https://ieeexplore.ieee.org/document/861337>
21. S.L. Salzberg, *C4.5: Programs for Machine Learning by Quinlan J. Ross Quinlan*. Morgan Kaufmann Publishers, Inc., 1993, pp. 38–48. Available: <https://doi.org/10.1007/BF00993309>
22. Geoffrey E. Hinton, Simon Osindero, and Yee-Whye Teh, “A fast learning algorithm for deep belief nets,” *Neural Computation*, 2006.
23. Daniel Etzold, *MNIST — Dataset of Handwritten Digits*. 2015. Available: <https://medium.com/mllearning-ai/mnist-dataset-of-handwritten-digits-f8cf28edafe>

Надійшла 01.06.2023

INFORMATION ON THE ARTICLE

Anatolii I. Petrenko, ORCID: 0000-0001-6712-7792, Educational and Research Institute for Applied System Analysis of the National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Ukraine, e-mail: tolja.petrenko@gmail.com

Ilya A. Vokhranov, Educational and Research Institute for Applied System Analysis of the National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Ukraine.

NEURAL NETWORKS: STUDYING THEIR DECISION-MAKING RULES /
A.I. Petrenko, I.A. Vokhranov

Abstract. The question of a better understanding of the behavior of neural networks is quite relevant, especially in industries with a high level of risks. To solve this problem, the possibilities of the new DeepRED decomposition algorithm, capable of extracting decision-making rules by deep neural networks with several hidden layers, are explored in the paper. The study of the DeepRED algorithm was carried out on the example of extracting the rules of an experimental neural network during the classification of images of the MNIST database of handwritten digits, which made it possible to reveal a number of limitations of the DeepRED algorithm.

Keywords: rule extraction, neural networks, DeepRED, machine learning, decision trees, decision graphs.

**GUARANTEED ROOT-MEAN-SQUARE ESTIMATES OF THE
FORECAST OF MATRIX OBSERVATIONS UNDER CONDITIONS
OF STATISTICAL UNCERTAINTY**

O.G. NAKONECHNYI, G.I. KUDIN, P.M. ZINKO, T.P. ZINKO

Abstract. We investigate the problem of linear estimation of unknown mathematical expectations based on observations of realizations of random matrix sequences. Constructive mathematical methods have been developed for finding linear guaranteed RMS estimates of unknown non-stationary parameters of average values based on observations of realizations of random matrix sequences. It is shown that such guaranteed estimates are obtained either as solutions to boundary value problems for systems of linear differential equations or as solutions to the corresponding Cauchy problems. We establish the form and look for errors for the guaranteed RMS quasi-minimax estimates of the special forecast vector and parameters of unknown average values. In the presence of small perturbations of known matrices in the model of matrix observations, quasi-minimax RMS estimates are found, and their guaranteed RMS errors are obtained in the first approximation of the small parameter method. Two test examples for calculating the guaranteed root mean square estimates and their errors are given.

Keywords: matrix observations, linear estimations, guaranteed RMS estimates, guaranteed RMS estimate errors, quasi-minimax guaranteed vector estimates, difference equation, small parameter method, matrix perturbation.

INTRODUCTION

This article examines estimates of unknown mathematical expectations based on observations of realizations of random matrix sequences. Scientific publications [1–14], in which estimates of distribution parameters were studied, are devoted to the problems of matrix sequence statistics. We formulate and solve new problems of estimating the mean values of random matrix sequences. Under the condition that the mean values belong to sets of a special form, we have developed constructive algorithms for guaranteed root-mean-square estimates of the mean values. It is shown that such estimates can be obtained either as solutions of boundary value problems for a system of linear differential equations, or as solutions of the corresponding Cauchy problems. In the case of the dependence of the average values on a small parameter, asymptotic distributions were obtained both for the guaranteed estimates and for the guaranteed root mean square errors of such estimates.

STATEMENT OF THE PROBLEM

We consider matrix observations of the form:

$$Y_k = \rho_k(x(k)) + \eta_k, \quad k = \overline{0, N-1}, \quad (1)$$

where $\rho_k(x(k)) = \sum_{s=1}^m A_s(k)x_s(k)$, $k = \overline{0, N-1}$,

$A_s(k) \in H_{n \times p}$, $s = \overline{1, m}$, $k = \overline{0, N-1}$ are known matrices;

$H_{n \times p}$ is the space of matrices $n \times p$ dimensions;

$x(k) = (x_1(k), \dots, x_m(k))^T$, $k = \overline{0, N}$ are unknown vectors, belonging to a limited set

$$G = \{x(k), k = \overline{0, N} : \sum_{k=0}^{N-1} |f(k)|^2 q_k^2 \leq 1\},$$

$$f(k) = x(k+1) - x(k), \quad k = \overline{0, N-1},$$

(to simplify the calculations, we assume that $x(0)$ is known vector and, without limiting the generality, we put $x(0) = 0$);

q_k^2 , $k = \overline{0, N-1}$ are known positive real numbers;

T is a transposition symbol;

$\eta_k \in H_{n \times p}$, $k = \overline{0, N-1}$ is a sequence of random matrices.

It is assumed that the average value of the random matrices η_k , $k = \overline{0, N-1}$ is equal to the null matrix, i.e. $E\eta_k = 0$ (E is a symbol of mathematical expectation), and correlation matrices $R_k \in H_{n \times n}$, $k = \overline{0, N-1}$ are known and determined by relations

$$E\langle \eta_k, Z_1 \rangle \langle \eta_k, Z_2 \rangle = \langle R_k Z_1, Z_2 \rangle, \forall Z_i \in H_{n \times p}, \quad i = \overline{1, 2}, \quad k = \overline{0, N-1},$$

where $\langle \eta_k, Z_i \rangle = sp(\eta_k Z_i^T)$ is a scalar product of matrices.

Let's introduce linear operators that act from space R^l into space $H_{n \times p}$:

$$\bar{\rho}_k(U_k)(z) = \sum_{i=1}^l U_{ik} z_i, \quad k = \overline{0, N-1},$$

$$U_{ik} \in H_{n \times p}, \quad U_k = (U_{1k} : \dots : U_{lk}), \quad z = (z_1, \dots, z_l)^T, \quad i = \overline{1, l}, \quad k = \overline{0, N-1},$$

and operators conjugated to them $\bar{\rho}_k^*(U_k)(Y_k)$:

$$\bar{\rho}_k^*(U_k)(Y_k) = (\langle U_{1k}, Y_k \rangle, \dots, \langle U_{lk}, Y_k \rangle)^T,$$

$$U_{ik} \in H_{n \times p}, \quad Y_k \in H_{n \times p}, \quad i = \overline{1, l}, \quad k = \overline{0, N-1}.$$

It is necessary to evaluate the vector $Vx(N)$, where $V \in H_{l \times m}$.

Definition 1. A vector $\widehat{Vx}(N)$ of the form

$$\widehat{Vx}(N) = \sum_{k=0}^{N-1} \bar{\rho}_k^*(U_k)(Y_k) + c =$$

$$= \sum_{k=0}^{N-1} (\langle U_{1k}, Y_k \rangle, \dots, \langle U_{lk}, Y_k \rangle)^T + c, \quad c \in R^l$$

is called a linear estimate of a vector $Vx(N)$.

Definition 2. The value

$$\sigma^2(U_0, \dots, U_{N-1}) = \max_G E \left[Vx(N) - \widehat{Vx}(N) \right]^2$$

is called *the guaranteed root mean square (RMS) error of the linear estimate* $\widehat{Vx}(N)$.

SOLVING THE PROBLEMS OF LINEAR ESTIMATION OF THE FORECAST OF MATRIX OBSERVATIONS.

I. Let's introduce vectors $z(k) \in R^m$, $k = \overline{0, N}$, which are solutions of the difference equation:

$$z(k) = z(k+1) - \rho_k^* \overline{\rho_k}(U_k)(a), \quad k = \overline{N-1, 0}, \quad z(N) = V^T a, \quad a \in R^l, \quad (2)$$

where ρ_k^* , $k = \overline{0, N-1}$ are operators conjugated to ρ_k .

Denote by $z^i(k)$, $i = \overline{1, l}$ the solutions of the difference equation (2) at $a = e^i$, where e^i , $i = \overline{1, l}$ are the base vectors of space R^l , and also enter the matrix Z :

$$Z = (z_{ij})_{i,j=\overline{1,l}}; \quad z_{ij} = \sum_{k=0}^{N-1} (z^i(k+1), z^j(k+1)) q_k^{-2}; \quad i, j = \overline{1, l}. \quad (3)$$

The vectors $z^i(k+1)$, $i = \overline{1, l}$ finds from the difference equations:

$$z^i(k) = z^i(k+1) + b^i(k), \quad z^i(N) = V_{(i)}, \quad (4)$$

where $b^i(k) = (\langle A_1(k), U_{ik} \rangle, \dots, \langle A_m(k), U_{ik} \rangle)^T$,

$$V_{(i)} = (V_{i1}, \dots, V_{im})^T, \quad i = \overline{1, l}, \quad k = \overline{0, N-1}.$$

There is a formula

$$z^i(k+1) = V_{(i)} + \sum_{j=1}^{N-(k+1)} b^i(N-j), \quad i = \overline{1, l}, \quad k = \overline{0, N-1}. \quad (5)$$

Statement 1. Let $x(k)$, $k = \overline{0, N} \in G$, then the following equality holds:

$$\sigma^2(U_0, \dots, U_{N-1}) = \max_{|a|=1} ((Za, a)^{1/2} + |(c, a)|)^2 + \sum_{k=0}^{N-1} \sum_{i=1}^l \langle R_k U_{ik}, U_{ik} \rangle.$$

Proof. Fair equality:

$$E \left| Vx(N) - \widehat{Vx}(N) \right|^2 = \left| (Vx(N)) - \sum_{k=0}^{N-1} (\overline{\rho_k}^*(U_k)(\rho_k(x(k)))) - c \right|^2 +$$

$$\begin{aligned}
 & + E \left| \sum_{k=0}^{N-1} \bar{\rho}_k^*(U_k) \eta_k \right|^2 = \\
 & = \max_{|a|=1} \left((V^T a, x(N)) - \left(a, \sum_{k=0}^{N-1} \bar{\rho}_k^*(U_k) (\rho_k(\rho_k(x(k)))) - (a, c) \right) \right)^2 + \\
 & \quad + \sum_{k=0}^{N-1} \sum_{i=1}^l \langle R_k U_{ik}, U_{ik} \rangle = \\
 & = \max_{|a|=1} \left((V^T a, x(N)) - \sum_{k=0}^{N-1} (x(k), \rho_k^* \bar{\rho}_k(U_k)(a)) - (a, c) \right)^2 + \\
 & \quad + \sum_{k=0}^{N-1} \sum_{i=1}^l \langle R_k U_{ik}, U_{ik} \rangle.
 \end{aligned}$$

Since

$$\begin{aligned}
 (V^T a, x(N)) &= (z(N), x(N)) = \sum_{k=0}^{N-1} ((z(k+1), x(k+1)) - (z(k), x(k))) = \\
 &= \sum_{k=0}^{N-1} ((z(k+1) - z(k)), x(k)) + \sum_{k=0}^{N-1} (z(k+1), f(k)),
 \end{aligned}$$

then, considering that $z(k+1) - z(k) = \rho_k^* \bar{\rho}_k(U_k)(a)$ we get that

$$\begin{aligned}
 \sigma^2(U_0, \dots, U_{N-1}) &= \max_G E \left| Vx(N) - \widehat{Vx}(N) \right|^2 = \\
 &= \max_{|a|=1} \max_G (\sum_{k=0}^{N-1} (z(k+1), f(k)) - (a, c))^2 + \sum_{k=0}^{N-1} \sum_{i=1}^l \langle R_k U_{ik}, U_{ik} \rangle = \\
 &= \max_{|a|=1} ((\sum_{k=0}^{N-1} |z(k+1)|^2 q_k^{-2})^{1/2} + |(a, c)|)^2 + \sum_{k=0}^{N-1} \sum_{i=1}^l \langle R_k U_{ik}, U_{ik} \rangle.
 \end{aligned}$$

From the fact that equalities are fulfilled

$$(\sum_{k=0}^{N-1} |z(k+1)|^2 q_k^{-2}) = \sum_{i,j=1}^l (\sum_{k=0}^{N-1} (z^i(k+1), z^j(k+1))(a, e^i)(a, e^j) q_k^{-2}) = (Za, a),$$

we conclude that the statement 1 is correct.

Corollary 1. There is an equality:

$$\min_{U, c} \max_G E \left| Vx(N) - \widehat{Vx}(N) \right|^2 = \lambda_{\max}(\hat{Z}) + \sum_{k=0}^{N-1} \sum_{i=1}^l \langle R_k \hat{U}_{ik}, \hat{U}_{ik} \rangle, \quad \hat{c} = 0,$$

where $\hat{Z} = Z = (z_{ij})_{i,j=1,\overline{l}}$, $z_{ij} = z_{ij}(\hat{U})$, $\lambda_{\max}(\hat{Z})$ is the maximum eigenvalue of the matrix \hat{Z} , and $\hat{U}_{ik} \in H_{n \times p}$, $i = 1, \overline{l}$, $k = 0, \overline{N-1}$ are found from the condition:

$$(\hat{U}_{ik}, i = 1, \overline{l}, k = 0, \overline{N-1}; \hat{c}) \in \text{Arg} \min_{c; U_{ik}, i=1, \overline{l}, k=0, \overline{N-1}} \sigma^2(U_0, \dots, U_{N-1}).$$

Corollary 2. Let $l = 1$. The estimation error $(V, x(N))$, where $V \in R^m$ is an arbitrary vector, is as follows:

$$\sigma^2(U_0, \dots, U_{N-1}) = \max_G E \left| Vx(N) - \widehat{Vx}(N) \right|^2 =$$

$$= \sum_{k=0}^{N-1} |z(k+1)|^2 q_k^{-2} + \sum_{k=0}^{N-1} \langle R_k U_{1k}, U_{1k} \rangle + c^2,$$

where $z(k)$, $k = \overline{0, N}$ is a solution of the difference equation $z(k) = z(k+1) - \rho_k^*(U_{1k})$, $k = \overline{N-1, 0}$, $z(N) = V$, for this case $\hat{c} = 0$.

Statement 2. Let's put the parameter $l = 1$ in the statement 1, then:

1) guaranteed root mean square estimate for $\widehat{V_{(1)}}x(N)$ has the form:

$$\overline{\widehat{V_{(1)}}x(N)} = \sum_{k=0}^{N-1} sp(\hat{U}_{1k} Y_k^T);$$

2) the guaranteed root mean square error of the linear estimate $\widehat{V_{(1)}}x(N)$ has the form:

$$\sigma^2(\hat{U}_{10}, \dots, \hat{U}_{1, N-1}) = (p(N), V_{(1)}),$$

where $\hat{U}_{1k} = R_k^+ \rho_k(p(k))$, $k = \overline{0, N-1}$; R_k^+ is a pseudo-inverse operator;

$p(k)$, $k = \overline{0, N}$ are vectors that are determined from the system of equations:

$$\begin{cases} z(k) = z(k+1) - \rho_k^*(\hat{U}_{1k}), & k = \overline{N-1, 0}, & z(N) = V_{(1)}, \\ p(k+1) = p(k) + q_k^{-2} z(k+1), & k = \overline{0, N-1}, & p(0) = 0, \end{cases} \quad (6)$$

Proof. Let's define \hat{U}_{1k} $k = \overline{0, N-1}$ from conditions:

$$\frac{d}{d\tau} \sigma^2(U_0 + \tau \tilde{\mathfrak{G}}_0, \dots, U_{N-1} + \tau \tilde{\mathfrak{G}}_{N-1})_{\tau=0} = 0, \quad \text{for } \forall \tilde{\mathfrak{G}}_k, k = \overline{0, N-1}.$$

There is an equality:

$$\begin{aligned} & \frac{1}{2} \frac{d}{d\tau} \sigma^2(U_0 + \tau \tilde{\mathfrak{G}}_0, \dots, U_{N-1} + \tau \tilde{\mathfrak{G}}_{N-1})_{\tau=0} = \\ & = \sum_{k=0}^{N-1} (z(k+1), \tilde{z}(k+1)) q_k^{-2} + \sum_{k=0}^{N-1} \langle R_k U_k, \tilde{\mathfrak{G}}_k \rangle, \end{aligned}$$

where $\tilde{z}(k) = \tilde{z}(k+1) - \rho_k^*(\tilde{\mathfrak{G}}_k)$, $\tilde{z}(N) = 0$, $k = \overline{0, N-1}$.

If input the vectors $p(k) \in R^m$, $k = \overline{0, N}$, which are solutions of difference equations

$$p(k+1) = p(k) + q_k^{-2} z(k+1), \quad k = \overline{0, N-1}, \quad p(0) = 0,$$

then we will get:

$$\begin{aligned} \sum_{k=0}^{N-1} (z(k+1), \tilde{z}(k+1)) q_k^{-2} &= \sum_{k=0}^{N-1} ((\tilde{z}(k+1), (p(k+1) - p(k)))) = \\ &= \sum_{k=0}^{N-1} ((\tilde{z}(k) - \tilde{z}(k+1)), p(k)) = \\ &= -\sum_{k=0}^{N-1} (\rho_k^*(\tilde{\mathfrak{G}}_k), p(k)) = -\sum_{k=0}^{N-1} \langle \rho_k(p(k)), \tilde{\mathfrak{G}}_k \rangle. \end{aligned}$$

As a result, we get equality:

$$\sum_{k=0}^{N-1} \langle (-\rho_k(p(k)) + R_k U_{1k}), \tilde{\mathfrak{G}}_k \rangle = 0,$$

from which the representation derives

$$\hat{U}_{1k} = R_k^+ \rho_k(p(k)), \quad k = \overline{0, N-1},$$

which had to be shown.

Solution of the boundary value problem (6)

One of the options for solving the boundary value problem (6) is the possibility of reducing it to the solution of the Cauchy problem for vectors $p(k)$, $k = \overline{0, N}$. This requires solving the Cauchy problem for the first equation of system (6). Then, after substituting the result into the second equation of the system, solve the Cauchy problem for the required vectors $p(k)$, $k = \overline{0, N}$.

It is also possible to use the homogeneity of the considered problem for the required vectors $p(k)$, $k = \overline{0, N}$. This requires the use of a base e^i , $i = \overline{1, m}$ of the vector space R^m . Expansions of vectors $z(k)$, $p(k)$, $k = \overline{0, N}$ in this basis have the form:

$$z(k) = \sum_{i=1}^m x_i z_i(k), \quad p(k) = \sum_{i=1}^m x_i p_i(k), \quad k = \overline{0, N}, \quad (7)$$

where the vectors $p_i(k)$, $z_i(k)$, $k = \overline{0, N}$, $i = \overline{1, m}$ are defined as solutions of m boundary problems:

$$\begin{cases} z_i(k+1) = z_i(k) - F_k p_i(k), & k = \overline{0, N-1}, \quad z_i(0) = e^i; \\ p_i(k+1) = p_i(k) + q_k^{-2} z_i(k+1), & k = \overline{0, N-1}, \quad p_i(0) = 0, \quad i = \overline{1, m}, \end{cases}$$

where

$$F_k = \rho_k^* R_k^+ \rho_k = \begin{pmatrix} sp(A_1^T R_k^+ A_1) & \cdots & sp(A_1^T R_k^+ A_m) \\ \vdots & \ddots & \vdots \\ sp(A_m^T R_k^+ A_1) & \cdots & sp(A_m^T R_k^+ A_m) \end{pmatrix}.$$

Unknown coefficients x_i , $i = \overline{1, m}$ in the expansions (7) are found as solutions of the system of linear algebraic equations that ensure the fulfillment of the boundary condition $z(N) = V$:

$$\sum_{i=1}^m z_i(N) x_i = V.$$

According to the distribution of vectors $p(k)$, $k = \overline{0, N}$ (formula (7)), the expressions for the matrices $\hat{U}_{1k} = R_k^+ \rho_k(p(k))$, $k = \overline{0, N-1}$ of the required estimate $\widehat{Vx}(N)$ are obtained.

Another possibility of solving the boundary value problem (6) is to reduce it to a difference boundary value problem of the second order with respect to vectors $p(k)$, $k = \overline{0, N}$ and to find a general solution to the obtained problem. Arbitrary constants of the general solution are determined from the boundary conditions of problem (6).

II. Let's introduce vectors $\hat{p}(k)$, $\hat{x}(k)$, $k = \overline{0, N}$, that are the solutions of the system of difference equations:

$$\begin{cases} \hat{p}(k) = \hat{p}(k+1) + \rho_k^* R_k^+ (Y_k - \rho_k(\hat{x}(k))), & \hat{p}(N) = 0; \\ \hat{x}(k+1) = \hat{x}(k) + q_k^{-2} \hat{p}(k+1), & \hat{x}(0) = 0, \quad k = \overline{0, N-1}. \end{cases} \quad (8)$$

Remark 1. If we take into account that the equality holds

$$\rho_k^* R_k^+ (Y_k - \rho_k(\hat{x}(k))) = (sp(A_1^T R_k^+ Y_k), \dots, sp(A_m^T R_k^+ Y_k))^T - F_k \hat{x}(k),$$

then we can find the solution of linear differential equation system (8) according to the solution scheme of linear differential equation system (6).

Statement 3. The following equality holds

$$\overline{\overline{Vx(N)}} = (\hat{x}(N), V). \tag{9}$$

Proof. For a guaranteed estimate, the following relations are fulfilled:

$$\overline{\overline{Vx(N)}} = \sum_{k=0}^{N-1} \langle \hat{U}_k, Y_k \rangle = \sum_{k=0}^{N-1} \langle R_k^+ \rho_k(p(k)), Y_k \rangle = \sum_{k=0}^{N-1} (p(k), \rho_k^*(R_k^+ Y_k)). \tag{10}$$

Let's denote $\Delta_k = \hat{p}(k) - \hat{p}(k+1)$, $k = 0, N-1$. Then

$$\rho_k^*(R_k^+ Y_k) = \Delta_k + \rho_k^*(R_k^+ \rho_k(\hat{x}(k))).$$

Hence

$$(p(k), \rho_k^*(R_k^+ Y_k)) = (p(k), \Delta_k) + (p(k), \rho_k^*(R_k^+ \rho_k(\hat{x}(k)))).$$

Now we sum up both parts of the last equality:

$$\sum_{k=0}^{N-1} (p(k), \rho_k^*(R_k^+ Y_k)) = \sum_{k=0}^{N-1} (p(k), \Delta_k) + \sum_{k=0}^{N-1} (p(k), \rho_k^*(R_k^+ \rho_k(\hat{x}(k)))) \tag{11}$$

and calculate the first term on the right-hand side:

$$\begin{aligned} \sum_{k=0}^{N-1} (p(k), \Delta_k) &= \sum_{k=0}^{N-1} (p(k), \hat{p}(k) - \hat{p}(k+1)) = \\ &= \sum_{k=0}^{N-1} (\hat{p}(k+1), p(k+1) - p(k)) = \sum_{k=0}^{N-1} (\hat{x}(k+1) - \hat{x}(k), z(k+1)) = \\ &= \sum_{k=0}^{N-1} (\hat{x}(k), z(k) - z(k+1)) + (\hat{x}(N), V) = \\ &= (\hat{x}(N), V) - \sum_{k=0}^{N-1} (\rho_k^*(R_k^+ \rho_k(p(k))), \hat{x}(k)) = \\ &= (\hat{x}(N), V) - \sum_{k=0}^{N-1} (p(k), \rho_k^*(R_k^+ \rho_k(\hat{x}(k)))) . \end{aligned} \tag{12}$$

The required equality (9) follows from formulas (10)–(12).

Remark 2. The system of equations (8) can be obtained by solving the minimization problem of the function

$$\begin{aligned} J(f(0), \dots, f(N-1)) &= \\ &= \sum_{k=0}^{N-1} \langle R_k^+ (Y_k - \rho_k(x(k))), Y_k - \rho_k(x(k)) \rangle + \sum_{k=0}^{N-1} q_k^2 f^2(k). \end{aligned}$$

III. Below we consider the case when the set G is in the space of possible values $(x(0), \dots, x(N))$, $x(k) = (x_1(k), \dots, x_m(k))^T$, $k = 0, N$ is unbounded and has the form:

$$G = \{(x(0), \dots, x(N)) : \sum_{k=0}^{N-1} |x(k+1) - x(k)|^2 q_k^2 \leq 1, x(0) \in R^m\},$$

where q_k^2 , $k = \overline{0, N-1}$ are known positive real numbers.

It is necessary to determine the guaranteed root mean square error:

$$\sigma^2(U_0, \dots, U_{N-1}) = \max_G E \left| Vx(N) - \widehat{Vx}(N) \right|^2,$$

where $\widehat{Vx}(N) = \sum_{k=0}^{N-1} (\langle U_{1k}, Y_k \rangle, \dots, \langle U_{lk}, Y_k \rangle)^T + c$, $c \in R^l$.

Let's introduce vectors $z(k) \in R^l$, $k = \overline{0, N-1}$, that are solutions of difference equations

$$z(k) = z(k+1) - \overline{\rho}_k^* \overline{\rho}_k(U_k)(a), \quad z(N) = a, \quad k = \overline{0, N},$$

and the set $U = \{U_{ik}, i = \overline{1, l}, k = \overline{0, N-1} : z(0) = 0\}$.

Statement 4.

1) If $U_{ik}, i = \overline{1, l}, k = \overline{0, N-1} \in U$, then

$$\begin{aligned} \sigma^2(U) &= \max_G E \left| Vx(N) - \widehat{Vx}(N) \right|^2 = \\ &= \max_{|a|=1} ((Za, a)^{1/2} + |(a, c)|)^2 + \sum_{k=1}^{N-1} \sum_{i=1}^l \langle R_k U_{ik}, U_{ik} \rangle. \end{aligned} \quad (13)$$

2) If $U_{ik}, i = \overline{1, l}, k = \overline{0, N-1} \notin U$, then

$$\sigma^2(U) = \max_G E \left| Vx(N) - \widehat{Vx}(N) \right|^2 = \infty.$$

Proof.

1) If $U_{ik}, i = \overline{1, l}, k = \overline{0, N-1} \in U$, then we obtain the formula (13) similarly to the statement 1;

2) If $U_{ik} \notin U$, then there may exist \bar{a} such that

$$\bar{z}(0) = z(0)_{a=\bar{a}} \neq 0 \text{ by } \forall U_{ik}, i = \overline{1, l}, k = \overline{0, N-1}.$$

Therefore, given unbounded of the set G , we obtain the relation:

$$\begin{aligned} \sigma^2(U) &\geq \max_G (\sum_{k=0}^{N-1} (\bar{z}(k+1), f(k)) + (\bar{z}(0), x(0)) - (\bar{a}, c))^2 + \\ &+ \sum_{k=0}^{N-1} \sum_{i=1}^l \langle R_k U_{ik}, U_{ik} \rangle = \infty. \end{aligned}$$

IV. We present a guaranteed linear RMS estimate of the scalar product $(a, x(N))$ according to matrix observations of the form

$$Y_k = \rho_k(x(k)) + \eta_k, \quad k = \overline{0, N-1},$$

through the solutions of the Cauchy problem for linear differential equations.

Denote by V_k , $k = 0, 1, 2, \dots$ the sequence of linear operators of the form:

$$V_k = (R_k + \rho_k P_{k-1} \rho_k^*) \rho_k P_{k-1},$$

where matrices P_k are solutions of difference equations:

$$P_k = P_{k-1} + V_k^*(R_k + \rho_k P_{k-1} \rho_k^*)V_k + Q_k,$$

$$Q_k = q_k^{-2} I_m, \quad P_{-1} = 0, \quad k = \overline{0, 1, 2, \dots}$$

Statement 5. The equality holds

$$(a, \widehat{x(N)}) = \sum_{k=0}^{N-1} \langle \widehat{U}_k, Y_k \rangle, \quad \widehat{U}_k = V_k z(k+1),$$

where $z(k)$, $k = \overline{1, N}$ are the solutions of difference equations:

$$z(k) = z(k+1) - V_k z(k+1) \equiv (I_m - V_k)z(k+1), \quad k = \overline{N-1, 1}, \quad z(N) = a.$$

Moreover

$$\max_G E((a, x(N)) - (a, \widehat{x(N)}))^2 = (P_{N-1} a, a).$$

Proof. Let's solve the problem of optimal system control:

$$z(k) = z(k+1) - \rho_k^*(U_k), \quad z(N) = a, \quad k = \overline{N-1, 0}$$

with the criterion

$$\bar{J}(U_0, \dots, U_{N-1}) = \sum_{k=0}^{N-1} (Q_k z(k+1), z(k+1)) + \sum_{k=0}^{N-1} \langle R_k U_k, U_k \rangle$$

by the method of dynamic programming.

Let's introduce the Bellman function

$$B_k(x) = \min_{U_0, \dots, U_k} [\sum_{i=0}^k (Q_i z(i+1), z(i+1)) + \sum_{i=0}^k \langle R_i U_i, U_i \rangle], \quad z(k+1) = x,$$

for which the Bellman equation holds

$$B_k(x) = (Q_k x, x) + \min_{u \in H_{n \times p}} [B_k(x - \rho_k^*(u)) + \langle R_k u, u \rangle], \quad B_{-1}(x) = 0.$$

$$k = \overline{-1, N-1}$$

(by definition we consider that $\sum_{k=0}^{-1} = 0$).

We find the function $B_k(x)$ in the form $B_k(x) = (P_k x, x)$. Let's choose the matrices P_k , $k = \overline{-1, N-1}$ so that the Bellman equation holds true. After appropriate transformations (similarly as it is done, for example, in [15]) we obtain the expressions for \widehat{U}_k , V_k , P_k , $k = \overline{0, N-1}$.

It is obvious that $B_{N-1}(x) = (P_{N-1} a, a) = \max_G (E((a, x(N)) - (a, \widehat{x(N)}))^2$.

Statement 6. For a guaranteed linear RMS estimate of the scalar product the following representation takes place

$$(a, \widehat{x(N)}) = (a, \hat{x}(N)),$$

where the vector $\hat{x}(N)$ is a solution of difference equations

$$\hat{x}(k+1) = \hat{x}(k) + V_k^*(Y_k - \rho_k(\hat{x}(k))), \quad k = \overline{0, N-1}, \quad \hat{x}(0) = 0. \quad (14)$$

Proof. Since equalities are fulfilled:

$$\hat{U}_k = V_k z(k+1), \quad k = \overline{0, N-1},$$

then $\sum_{k=0}^{N-1} \langle \hat{U}_k, Y_k \rangle = \sum_{k=0}^{N-1} (z(k+1), V_k^* Y_k)$.

Expressions take place:

$$\begin{aligned} V_k^* Y_k &= \hat{x}(k+1) - \hat{x}(k) + V_k^* \rho_k(\hat{x}(k)); \\ \sum_{k=0}^{N-1} (z(k+1), V_k^* Y_k) &= \sum_{k=0}^{N-1} (z(k+1), \hat{x}(k+1) - \hat{x}(k)) + \\ &+ \sum_{k=0}^{N-1} (z(k+1), V_k^* \rho_k(\hat{x}(k))); \\ \sum_{k=0}^{N-1} (z(k+1), \hat{x}(k+1) - \hat{x}(k)) &= -\sum_{k=0}^{N-1} (z(k+1) - z(k), \hat{x}(k)) + (a, \hat{x}(N)) = \\ &= (a, \hat{x}(N)) - \sum_{k=0}^{N-1} (\rho_k^* (V_k z(k+1)), \hat{x}(k)) = \\ &= (a, \hat{x}(N)) - \sum_{k=0}^{N-1} (z(k+1), V_k^* \rho_k(\hat{x}(k))). \end{aligned}$$

From here we get the necessary equality.

Remark 3. The vector $\hat{x}(N)$ is found as a solution of the linear difference equation (14). It is possible to obtain the vector $\hat{x}(N)$ even if the vectors $f(k) = \hat{x}(k+1) - \hat{x}(k)$, $k = \overline{0, N-1}$ are random and uncorrelated ($Ef(k) = 0$, $Ef(k)f^T(k) = q_k^2$, $k = \overline{0, N-1}$). The given estimators are such that minimize the root mean square error in the category of linear estimators.

V. Definition 3. The vector

$$\widetilde{Vx}(N) = \sum_{k=0}^{N-1} (\langle \hat{U}_{1k}, Y_k \rangle, \dots, \langle \hat{U}_{lk}, Y_k \rangle)^T,$$

which components are calculated according to formulas

$$\hat{U}_{ik} = R_k^+ \rho_k(p_{(i)}(k)), \quad i = \overline{1, l}, \quad k = \overline{0, N-1},$$

and $p_{(i)}(k)$ are vectors that are determined from the systems of difference equations

$$\begin{cases} z_{(i)}(k) = z_{(i)}(k+1) - \rho_k^*(\hat{U}_{ik}), & z_{(i)}(N) = V_{(i)}, \\ p_{(i)}(k+1) = p_{(i)}(k) - q_k^{-2} z_{(i)}(k+1), & p_{(i)}(0) = 0, \\ i = \overline{1, l}, & k = \overline{0, N-1}. \end{cases}$$

is called the quasi-minimax guaranteed estimation of the vector

$$\widehat{Vx}(N) = \sum_{k=0}^{N-1} (\langle U_{1k}, Y_k \rangle, \dots, \langle U_{lk}, Y_k \rangle)^T.$$

Statement 7. For the guaranteed root mean square error of quasi-minimax estimates there is equality:

$$\sigma^2(\hat{U}_0, \dots, \hat{U}_{N-1}) = \lambda_{\max}(\hat{Z}) + \sum_{k=0}^{N-1} \sum_{i=1}^l \langle R_k \hat{U}_{ik}, \hat{U}_{ik} \rangle,$$

where $\hat{Z} = (\hat{z}_{ij})_{i,j=\overline{1,m}}$, $\hat{z}_{ij} = \sum_{k=0}^{N-1} (\hat{z}_i(k+1), \hat{z}_i(k+1)) q_k^{-2}$, $i, j = \overline{1, m}$, and elements $\hat{z}_i(k+1)$, $i = \overline{1, m}$ are found as solutions of difference equations:

$$\begin{aligned} \hat{z}_i(k) &= \hat{z}_i(k+1) + \hat{b}_i(k), \quad k = \overline{N-1, 1}, \quad \hat{z}_i(N) = V_{(i)}, \\ \hat{b}_i(k) &= (\langle A_1(k), \hat{U}_{ik} \rangle, \dots, \langle A_m(k), \hat{U}_{ik} \rangle)^T, \\ V_{(i)} &= (V_{i1}, \dots, V_{im})^T, \quad i = \overline{1, l}, \quad k = \overline{1, N-1}. \end{aligned}$$

Finding elements \hat{z}_{ij} , $i, j = \overline{1, m}$ of the matrix \hat{Z} is carried out according to the algorithm for calculating elements of matrix Z (formulas (3)–(5)).

Quasi-minimax RMS estimates for small matrix perturbations.

Assume that the known matrices of model (1) have the form:

$$A_s(k) = A_s^{(0)}(k) + \varepsilon A_s^{(1)}(k) \in H_{n \times p}, \quad s = \overline{1, m}, \quad k = \overline{0, N-1},$$

where $\varepsilon \in R^1$ is small parameter; and the operators are as follows:

$$\rho_k^{(\varepsilon)}(x(k)) = \rho_k^{(0)}(x(k)) + \varepsilon \rho_k^{(1)}(x(k)),$$

where $\rho_k^{(0)}(x(k)) = \sum_{s=1}^m A_s^{(0)}(k)x_s(k)$, $\rho_k^{(1)}(x(k)) = \sum_{s=1}^m A_s^{(1)}(k)x_s(k)$, $k = \overline{0, N-1}$.

We determine the effect of small perturbations of the matrices on the estimates, as well as on their errors, using the results presented in statement 3.

Statement 8. Quasi-minimax guaranteed estimation of the vector $Vx(N)$ within the framework of the first approximation of the small parameter method has the form:

$$\widetilde{Vx}(N) = \sum_{k=0}^{N-1} (\langle \hat{U}_{1k}^{(\varepsilon)}, Y_k \rangle, \dots, \langle \hat{U}_{lk}^{(\varepsilon)}, Y_k \rangle)^T,$$

where $\hat{U}_{ik}^{(\varepsilon)} = \hat{U}_{ik}^{(0)} + \varepsilon \hat{U}_{ik}^{(1)} + o(\varepsilon)I_{n \times p}$,

$I_{n \times p} \in H_{n \times p}$ is the matrix, all elements of which are equal to one,

$$\hat{U}_{ik}^{(0)} = R_k^+ \rho_k^{(0)}(p_{(i)}^{(0)}(k)),$$

$$\hat{U}_{ik}^{(1)} = R_k^+ \rho_k^{(0)}(p_{(i)}^{(1)}(k)) + R_k^+ \rho_k^{(1)}(p_{(i)}^{(0)}(k)), \quad i = \overline{1, l}, \quad k = \overline{0, N-1},$$

and vectors $p_{(i)}^{(0)}(k)$, $i = \overline{1, l}$, $k = \overline{0, N}$ are defined as solutions of boundary value problems:

$$z_{(i)}^{(0)}(k) = z_{(i)}^{(0)}(k+1) - \rho_k^{(0)*} (R_k^+ \rho_k^{(0)}(p_{(i)}^{(0)}(k))), \quad z_{(i)}^{(0)}(N) = V_{(i)},$$

$$p_{(i)}^{(0)}(k+1) = p_{(i)}^{(0)}(k) + q_k^{-2} z_{(i)}^{(0)}(k+1), \quad p_{(i)}^{(0)}(0) = 0, \quad i = \overline{1, l}, \quad k = \overline{0, N-1},$$

$$z_{(i)}^{(1)}(k) = z_{(i)}^{(1)}(k+1) - \rho_k^{(0)*} (R_k^+ \rho_k^{(0)}(p_{(i)}^{(1)}(k))) -$$

$$- \rho_k^{(1)*} (R_k^+ \rho_k^{(0)}(p_{(i)}^{(0)}(k))) - \rho_k^{(0)*} (R_k^+ \rho_k^{(1)}(p_{(i)}^{(0)}(k))), \quad z_{(i)}^{(1)}(N) = 0,$$

$$p_{(i)}^{(1)}(k+1) = p_{(i)}^{(1)}(k) + q_k^{-2} z_{(i)}^{(1)}(k+1), \quad p_{(i)}^{(1)}(0) = 0, \quad i = \overline{1, l}, \quad k = \overline{0, N-1}.$$

Statement 9. There is the equality for the guaranteed root mean square error of quasi-minimax estimates within the framework of the first approximation of the small parameter method:

$$\sigma^2(\hat{U}_0^{(\varepsilon)}, \dots, \hat{U}_{N-1}^{(\varepsilon)}) = \lambda_{\max}(\hat{Z}^{(\varepsilon)}) + \sum_{k=0}^{N-1} \sum_{i=1}^l \langle R_k \hat{U}_{ik}^{(\varepsilon)}, \hat{U}_{ik}^{(\varepsilon)} \rangle.$$

Here $\hat{U}_{ik}^{(\varepsilon)} = \hat{U}_{ik}^{(0)} + \varepsilon \hat{U}_{ik}^{(1)} + o(\varepsilon) I_{n \times p}$, $i = \overline{1, l}$, $k = \overline{0, N-1}$, and the expression for the matrix $\hat{Z}^{(\varepsilon)}$ has the form

$$\hat{Z}^{(\varepsilon)} = \hat{Z}^{(0)} + \varepsilon \hat{Z}^{(1)} + o(\varepsilon) I_{m \times m},$$

where $\hat{Z}^{(0)} = (\hat{z}_{i,j}^{(0)})_{i,j=\overline{1,m}}$, $\hat{z}_{i,j}^{(0)} = \sum_{k=0}^{N-1} (\hat{z}_i^{(0)}(k+1), \hat{z}_j^{(0)}(k+1)) q_k^{-2}$;

$$\hat{Z}^{(1)} = (\hat{z}_{i,j}^{(1)})_{i,j=\overline{1,m}}, \quad \hat{z}_{i,j}^{(1)} = 2 \sum_{k=0}^{N-1} (\hat{z}_i^{(0)}(k+1), \hat{z}_j^{(1)}(k+1)) q_k^{-2}.$$

The vectors $\hat{z}_i^{(0)}(k+1)$, $i = \overline{1, m}$, $k = \overline{0, N-1}$ are found as solutions of zero-approximation difference equations:

$$\hat{z}_i^{(0)}(k) = \hat{z}_i^{(0)}(k+1) + \hat{b}_i^{(0)}(k), \quad z_i^{(0)}(N) = V_{(i)}; \quad (15)$$

$$\hat{b}_i^{(0)}(k) = \left(\langle A_1^{(0)}(k), \widehat{U}_{ik}^{(0)} \rangle, \dots, \langle A_m^{(0)}(k), \widehat{U}_{ik}^{(0)} \rangle \right)^T;$$

$$V_{(i)} = (V_{i1}, \dots, V_{im})^T, \quad i = \overline{1, l}, \quad k = \overline{0, N-1},$$

and vectors $\hat{z}_j^{(1)}(k+1)$, $j = \overline{1, m}$, $k = \overline{0, N-1}$ are found as solutions of first approximation difference equations:

$$\hat{z}_j^{(1)}(k) = \hat{z}_j^{(1)}(k+1) + \hat{b}_j^{(1)}(k), \quad z_j^{(1)}(N) = 0; \quad (16)$$

$$\begin{aligned} \hat{b}_j^{(1)}(k) = & (\langle A_1^{(1)}(k), \widehat{U}_{jk}^{(0)} \rangle, \dots, \langle A_m^{(1)}(k), \widehat{U}_{jk}^{(0)} \rangle)^T + \\ & + (\langle A_1^{(0)}(k), \widehat{U}_{jk}^{(1)} \rangle, \dots, \langle A_m^{(0)}(k), \widehat{U}_{jk}^{(1)} \rangle)^T, \quad j = \overline{1, l}, \quad k = \overline{0, N-1}. \end{aligned}$$

Finding the solutions of differential equations (15), (16) is carried out according to the algorithm for calculating elements of matrix $Z = (z_{ij})_{i,j=\overline{1,l}}$ (formulas (3)–(5)).

Example 1. Let the matrix observations have the form:

$$Y_k = \rho_k^{(\varepsilon)}(x(k)) + \eta_k, \quad k = \overline{0, N}; \quad (17)$$

$$\rho_k^{(\varepsilon)}(x(k)) = \rho_k^{(0)}(x(k)) + \varepsilon \rho_k^{(1)}(x(k)), \quad (18)$$

where $\rho_k^{(0)}(x(k)) = A^{(0)}x(k)$, $\rho_k^{(1)}(x(k)) = A^{(1)}(k)x(k)$,

$$A^{(0)} = I_2, \quad A^{(1)}(k) = \begin{pmatrix} k & 0 \\ 0 & 0 \end{pmatrix}, \quad k = \overline{0, N}, \quad \varepsilon \in R^1 \text{ is a small parameter;}$$

$x(k) \in R^1$, $k = \overline{0, N-1}$ and belong to a bounded set G :

$$G = \{x(k), k = \overline{0, N} : \sum_{k=0}^{N-1} |x(k+1) - x(k)|^2 \leq q_1^{-2}\};$$

$$\eta_k \in H_{2 \times 2}, k = \overline{0, N} \text{ is a sequence of random matrices.}$$

It is assumed that the average value of the random matrices $\eta_k, k = \overline{0, N-1}$ are equal to the zero matrix, and is the correlation matrices $R_k = q_0^2 I_2, k = \overline{0, N-1}$.

The guaranteed RMS estimate $\hat{x}(N)$ has the form:

$$\hat{x}(N) = \sum_{k=0}^{N-1} \langle \hat{U}_k^{(\varepsilon)}, Y_k \rangle, \quad (19)$$

where $\hat{U}_k^{(\varepsilon)} = q_0^{-2} \rho_k^{(\varepsilon)}(p^{(\varepsilon)}(k)), k = \overline{0, N-1}$;

$p^{(\varepsilon)}(k), k = \overline{0, N}$ are values that are determined from the system of difference equations:

$$\begin{cases} z^{(\varepsilon)}(k) = z^{(\varepsilon)}(k+1) - q_0^2 \langle A^{(\varepsilon)}(k), A^{(\varepsilon)}(k) \rangle p^{(\varepsilon)}(k), k = \overline{N-1, 1}, z^{(\varepsilon)}(N) = 1; \\ p^{(\varepsilon)}(k+1) = p^{(\varepsilon)}(k) + q_1^2 z^{(\varepsilon)}(k+1), k = \overline{0, N-1}, p^{(\varepsilon)}(0) = 0. \end{cases} \quad (20)$$

Applying the small parameter method for solving problems (19), (20), we introduce the following notation:

$$\hat{U}_k^{(\varepsilon)} = \widehat{U}_k(0) + \varepsilon \widehat{U}_k(1) + o(\varepsilon) I_{2 \times 2}, k = \overline{0, N-1},$$

$$\widehat{U}_k(0) = q_0^{-2} \rho_k^{(0)}(p^{(0)}(k)), \widehat{U}_k(1) = q_0^{-2} (\rho_k^{(0)}(p^{(1)}(k)) + \rho_k^{(1)}(p^{(0)}(k))),$$

$$p^{(\varepsilon)}(k) = p^{(0)}(k) + \varepsilon p^{(1)}(k) + o(\varepsilon), z^{(\varepsilon)}(k) = z^{(0)}(k) + \varepsilon z^{(1)}(k) + o(\varepsilon),$$

where $p^{(0)}(k), k = \overline{0, N}$ are the values of the zero approximation of the small parameter method, which are defined as solutions of the boundary value problem:

$$\begin{cases} z^{(0)}(k) = z^{(0)}(k+1) - \beta p^{(0)}(k), z^{(0)}(N) = 1, \\ p^{(0)}(k+1) = z^{(0)}(k) + q_1^2 z^{(0)}(k+1), p^{(0)}(0) = 0, \\ \beta = 2q_0^{-2}, k = \overline{0, N-1}, \end{cases} \quad (21)$$

and $p^{(1)}(k), k = \overline{0, N}$ are the values of the first approximation of the small parameter method, which are defined as solutions of the boundary value problem:

$$\begin{cases} z^{(1)}(k) = z^{(1)}(k+1) - p^{(1)}(k) - \beta k p^{(0)}(k), z^{(1)}(N) = 0, \\ p^{(1)}(k+1) = p^{(1)}(k) + q_1^2 z^{(1)}(k+1), p^{(1)}(0) = 0, k = \overline{0, N-1}. \end{cases} \quad (22)$$

Solution of the boundary value problem (21) of zero approximation

The boundary value problem (21) reduces to a boundary value problem for values $p^{(0)}(k), k = \overline{0, N}$:

$$\begin{aligned} p^{(0)}(k+2) - (2+q)p^{(0)}(k+1) + p^{(0)}(k) &= 0, & p^{(0)}(0) &= 0, \\ p^{(0)}(N) - p^{(0)}(N-1) &= q_1^2, \end{aligned} \quad (23)$$

$$k = \overline{0, N-2}, \quad q = 2q_1^2 q_0^{-2}.$$

The general solution of the corresponding homogeneous system has the form:

$$p_0^{(0)}(k) = C_1 \lambda^k + C_2 \lambda^{-k}, \quad \lambda = \left(2 + q + \sqrt{q^2 + 4q} \right) / 2.$$

Taking into account the boundary conditions, the solution of the boundary value problem (23) is as follows:

$$p_0^{(0)}(k) = C_1 (\lambda^k - \lambda^{-k}), \quad C_1 = q_1^2 \lambda^N / ((\lambda - 1)(\lambda^{2N-1} + 1)).$$

Therefore, in the zero approximation of the small parameter method, expressions are obtained:

$$\widehat{U}_k(0) = q_0^{-2} p^{(0)}(k), \quad k = \overline{0, N-1}. \quad (24)$$

The guaranteed RMS estimate in the zero approximation of the small parameter method has the form:

$$\hat{x}^{(0)}(N) = q_0^{-2} C_1 \sum_{k=0}^{N-1} \langle A^{(0)}, Y_k \rangle (\lambda^k - \lambda^{-k}), \quad (25)$$

and the error of this estimate is as follows:

$$\max_G \{E[x(N) - \hat{x}(N)]^2\}^{1/2} = \{p^{(0)}(N)\}^{1/2}, \quad (26)$$

where
$$p^{(0)}(N) = \frac{q_1^2 (\lambda - 1 / \lambda^{2N-1})}{(\lambda - 1)(\lambda + 1 / \lambda^{2N-1})}.$$

The representation of the estimation error in the zero approximation by formula (26) allows one to notice a decrease in its value with an increase in the quantity of observations, as well as to establish a limit value $p^{(0)}(N)$:

$$\lim_{N \rightarrow \infty} p^{(0)}(N) = q_1^2 \frac{(-1 + \sqrt{1 + 2q_1^{-2} q_0^2})}{2}.$$

Solving the boundary value problem (22) of the first approximation

The boundary value problem (22) is reduced to a boundary value problem for values $p^{(1)}(k)$, $k = \overline{0, N}$:

$$p^{(1)}(k+2) - p^{(1)}(k+1) + p^{(1)}(k) = \beta k (\lambda^k - \lambda^{-k}), \quad (27)$$

$$p^{(1)}(0) = 0, \quad p^{(1)}(N) - p^{(1)}(N-1) = 0, \quad k = \overline{0, N-2}.$$

The partial solution of the inhomogeneous equation (27) is represented by formulas with undefined coefficients:

$$p_{\text{part}}^{(1)}(k) = (B_1 k^2 + B_2 k + B_3) \lambda^k + (D_1 k^2 + D_2 k + D_3) \lambda^{-k}, \quad k = \overline{0, N},$$

which have the form:

$$B_1 = \beta / (\lambda^2 - 1), \quad B_2 = -B_1 (3\lambda^2 - 1) / (\lambda^2 - 1), \quad B_3 = 0, \\ D_1 = -B_1, \quad D_2 = B_2, \quad D_3 = 0.$$

The general solution of the inhomogeneous equation (27) has the form:

$$p^{(1)}(k) = F_1 \lambda^k + F_2 \lambda^{-k} + (B_1 k^2 + B_2 k)(\lambda^k - \lambda^{-k}), \quad k = \overline{0, N},$$

where arbitrary constants are determined from boundary conditions (27).

The solution of the boundary value problem (27) is as follows:

$$p^{(1)}(k) = [F_1 + (B_1 k^2 + B_2 k)](\lambda^k - \lambda^{-k}), \quad k = \overline{0, N},$$

where $F_1 = \frac{1}{(\lambda - 1)} \left\{ [B_1(N - 1) + B_2](N - 1) - (B_1 N + B_2 \lambda) N \frac{\lambda - (1/\lambda^{2N-1})}{1 - (1/\lambda^{2N-1})} \right\}$.

Thus, expressions are obtained for the corrections of the first approximation of the small parameter method:

$$\hat{U}_k(1) = q_0^{-2} (A^{(0)} p^{(1)}(k) + A^{(1)}(k) p^{(0)}(k)), \quad k = \overline{0, N-1},$$

$$\hat{x}^{(1)}(N) = \sum_{k=0}^{N-1} \langle \hat{U}_k(1), Y_k \rangle,$$

$$p^{(1)}(N) = [F_1 + (B_1 N^2 + B_2 N)](\lambda^N - \lambda^{-N}). \quad (28)$$

The guaranteed root mean square error in the first approximation of the small parameter method is represented by the formula:

$$\sigma^2(\varepsilon) = \sigma^2(0) + \varepsilon \sigma^2(1) + o(\varepsilon),$$

where

$$\sigma^2(0) = \frac{q_1^2 (\lambda - (1/\lambda^{2N-1}))}{(\lambda - 1) (1 + (1/\lambda^{2N-1}))},$$

$$\sigma^2(1) = p^{(1)}(N) = (F_1 + (B_1 N^2 + B_2 N))(\lambda^N - \lambda^{-N}).$$

Remark 4. It is worth noting that when using formula (28), it is necessary to take into account the specific values of the model parameters of the observation problem q_0, q_1, q, λ , as well as the number of observations N , namely: order of magnitude $\varepsilon \sigma^2(1)$ a smaller than order of magnitude $\sigma^2(0)$.

The extended possibilities of applying the small parameter method can be seen in the following example for other small perturbations of the known matrices in the model of the observation problem.

Example 2. Let the matrix observations have the form represented by formulas (17), (18), but with other matrices of small perturbation:

$$A^{(1)}(k) = a^{-k} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad k = \overline{0, N-1}. \quad (29)$$

Let's find the decomposition for the small parameter of the guaranteed value estimate $x(N)$, as well as its errors.

As in example 1, the guaranteed RMS estimate $\hat{x}(N)$ has the form:

$$\hat{x}(N) = \sum_{k=0}^{N-1} \langle \hat{U}_k^{(\varepsilon)}, Y_k \rangle, \quad (30)$$

where $\hat{U}_k^{(\varepsilon)} = q_0^{-2} \rho_k^{(\varepsilon)}(p^{(\varepsilon)}(k))$, $k = \overline{0, N-1}$; $p^{(\varepsilon)}(k)$, $k = \overline{0, N}$ are the values that are determined from the system of difference equations:

$$\begin{cases} z^{(\varepsilon)}(k) = z^{(\varepsilon)}(k+1) - q_0^2 \langle A^{(\varepsilon)}(k), A^{(\varepsilon)}(k) \rangle p^{(\varepsilon)}(k), & k = \overline{N-1, 1}, \quad z^{(\varepsilon)}(N) = 1, \\ p^{(\varepsilon)}(k+1) = p^{(\varepsilon)}(k) + q_1^2 z^{(\varepsilon)}(k+1), & k = \overline{0, N-1}, \quad p^{(\varepsilon)}(0) = 0. \end{cases} \quad (31)$$

Applying the small parameter method to solve problems (30), (31) we obtain a guaranteed root mean square estimate in the zero approximation and its error in the formulas form (24)–(26).

First approximation corrections of the small parameter method for the guaranteed RMS estimate require the definition of matrices $\hat{U}_k(1) = q_0^{-2} (\rho_k^{(0)}(p^{(1)}(k)) + \rho_k^{(1)}(p^{(0)}(k)))$, where $p^{(0)}(k)$, $k = \overline{0, N}$ are the values of the zero approximation of the small parameter method, and $p^{(1)}(k)$, $k = \overline{0, N}$ are the values of the first approximation of the small parameter method, which are defined as solutions of boundary value problems

$$p^{(1)}(k+2) - p^{(1)}(k+1) + p^{(1)}(k) = \beta a^{-k} (\lambda^k - \lambda^{-k}), \quad (32)$$

$$p^{(1)}(0) = 0, \quad p^{(1)}(N) - p^{(1)}(N-1) = 0, \quad k = \overline{0, N-2}.$$

The partial solution of the inhomogeneous equation (27) is represented by formulas with undefined coefficients:

$$p_{\text{part}}^{(1)}(k) = B_1 \left(\frac{\lambda}{a} \right)^k + B_2 (a\lambda)^{-k}, \quad k = \overline{0, N},$$

where these coefficients are calculated by formulas:

$$B_1 = \beta a^2 / (\lambda^2 - \lambda a + a^2), \quad B_2 = -\beta a^2 \lambda^2 / (a^2 \lambda^2 - \lambda a + 1).$$

The general solution of the inhomogeneous equation (27) has the form:

$$p^{(1)}(k) = F_1 \lambda^k + F_2 \lambda^{-k} + B_1 \left(\frac{\lambda}{a} \right)^k + B_2 (a\lambda)^{-k}, \quad k = \overline{0, N},$$

and arbitrary constants F_1, F_2 are determined from the boundary conditions (32):

$$F_1 = -(F_2 + B_1 + B_2),$$

$$F_2 = \frac{1}{(1 + 1/\lambda^{-2N+1})} \left[B_1 \left(a^{-N} \frac{(\lambda - a)}{(\lambda - 1)} - 1 \right) + B_2 \left(a^{-N} \lambda^{-2N+1} \frac{(\lambda a - 1)}{(\lambda - 1)} + 1 \right) \right].$$

Thus, expressions are obtained for the corrections of the first approximation of the small parameter method:

$$\hat{U}_k(1) = q_0^{-2} (A^{(0)} p^{(1)}(k) + A^{(1)}(k) p^{(0)}(k)), \quad k = \overline{0, N-1},$$

$$\hat{x}^{(1)}(N) = \sum_{k=0}^{N-1} \langle \hat{U}_k(1), Y_k \rangle,$$

$$p^{(1)}(N) = F_1 \lambda^N + F_2 \lambda^{-N} + B_1 \left(\frac{\lambda}{a} \right)^N + B_2 (a\lambda)^{-N}.$$

The guaranteed root mean square error in the first approximation of the small parameter method is represented by the formula:

$$\sigma^2(\varepsilon) = \sigma^2(0) + \varepsilon \sigma^2(1) + o(\varepsilon),$$

where $\sigma^2(0) = \frac{q_1^2 (\lambda - (1/\lambda^{2N-1}))}{(\lambda - 1) (1 + (1/\lambda^{2N-1}))}$,

$$\sigma^2(1) = p^{(1)}(N) = F_1 \lambda^N + F_2 \lambda^{-N} + B_1 \left(\frac{\lambda}{a}\right)^N + B_2 (\lambda a)^{-N}.$$

Obviously, that at certain values of the parameter a in model (29), the desired accuracy of the small parameter method can be achieved with a larger number of observations.

CONCLUSIONS

The article develops constructive mathematical methods for finding linear guaranteed root mean square estimates of unknown non-stationary parameters of average values based on observations of realizations of a sequence of random matrices. It is shown that, under certain conditions, such estimates are expressed in terms of solutions of the boundary value problem for the system of difference equations. Formulas are presented that allow obtaining recurrent estimates of unknown parameters. In the case of the dependence of the average values on a small parameter, the corresponding asymptotic formulas are given. Asymptotic distributions of linear parameter estimates and their root mean square errors are given for partial cases.

REFERENCES

1. Sheng Yue, *Application of random matrix theory in Statistics and machine learning*. Publicly Accessible Penn Dissertations, 4146, 2021, 240 p. Available: <https://repository.upenn.edu/edissertations/4146>
2. Yuan Ke, Stanislav Minsker, Zhao Ren, Qiang Sun, and Wen-Xin Zhou, "Uzer friendly covariance estimation for heavy-tailed distributions," *Statistical Science*, 34 (3), pp. 454–471, 2019.
3. Stanislav Minsker, "Sub-gaussian estimators of mean of a random matrix with heavy-tailed entries," *The Annals of Statistics*, 46 (6A), pp. 2871–2903, 2018.
4. Jun Tong, Rui Hu, Jiangtao Xi, Zhitao Xiao, Qinghua Guo, and Yu. Yanguang, "Linear shrinkage estimation of covariance matrices using complexity cross-validation," *Signal Processing*, 148, pp. 223–233, 2018.
5. Roberto Cabal Lopes, *Robust estimation of the mean a random matrix: a non-asymptotic study*. Centro de Investigacion en Matematicas, A.C., 2020, 187 p. Available: <https://cimat.repositorioinstitucional.mx/jspui/bitstream/1008/1082/1/TE%20785.pdf>
6. H. Battey, J. Fan, J. Lu, and Z. Zhu, "Distributed testing and estimation under sparse high dimensional models," *The Annals of Statistics*, 46 (3), pp. 1352–1382, 2018.
7. T.T. Cai and H. Wei, "Distributed Gaussian mean estimation under communication constraints: Optimal rates and communication-efficient algorithms," *arXiv preprint*, arXiv: 2001.08877, 2020.
8. T. Ke, Y. Ma, and X. Lin, "Estimation of the number of spiked eigenvalues in a covariance matrix by bulk eigenvalue matching analysis," *arXiv preprint*, arXiv: 2006.00436, 2020.
9. C. McKenna, "Factor analysis in high dimensional biological data with dependent observations," *arXiv preprint*, arXiv: 2009.11134, 2020.
10. Sourav Chatterjee, "Matrix Estimation by Universal Singular Value Thresholding," *The Annals of Statistics*, vol. 43, no. 1, 2015, pp. 177–214.
11. O.G. Nakonechnyi, G.I. Kudin, P.M. Zinko, and T.P. Zinko, "Perturbation Method in Problems of Linear Matrix Regression," *Problems of Control and Informatics*, no. 1, pp. 38–47, 2020.

12. O.G. Nakonechnyi, G.I. Kudin, P.M. Zinko, and T.P. Zinko, "Approximate guaranteed estimates of matrices in linear regression problems with a small parameter," *System Research & Information Technologies*, no. 4, pp. 88–102, 2020.
13. O.G. Nakonechnyi, G.I. Kudin, P.M. Zinko, and T.P. Zinko, "Guaranteed Root-Mean-Square Estimates of Linear Transformations of Matrices under Statistical Uncertainty," *Problems of Control and Informatics*, no. 2, pp. 24–37, 2021.
14. O.G. Nakonechnyi, G.I. Kudin, P.M. Zinko, and T.P. Zinko, "Minimax Root-Mean-Square Estimates of Matrix Parameters in Linear Regression Problems under Uncertainty," *Problems of Control and Informatics*, no. 4, pp. 28–37, 2021.
15. K. Ostrem, *Introduction to Stochastic Control Theory*. M.: Mir, 1973, 324 p.

Received 20.01.2023

INFORMATION ON THE ARTICLE

Oleksandr G. Nakonechnyi, ORCID: 0000-0002-8705-3070, Taras Shevchenko National University of Kyiv, Ukraine, e-mail: a.nakonechniy@gmail.com

Grygoriy I. Kudin, ORCID: 0000-0002-1322-4551, Taras Shevchenko National University of Kyiv, Ukraine, e-mail: gkudin@ukr.net

Petro N. Zinko, ORCID: 0000-0002-5111-4417, Taras Shevchenko National University of Kyiv, Ukraine, e-mail: petro.zinko@gmail.com

Taras P. Zinko, ORCID: 0000-0003-1263-9293, Taras Shevchenko National University of Kyiv, Ukraine, e-mail: taras.zinko@gmail.com

ГАРАНТОВАНІ СЕРЕДНЬОКВАДРАТИЧНІ ОЦІНКИ ПРОГНОЗУ МАТРИЧНИХ СПОСТЕРЕЖЕНЬ В УМОВАХ СТАТИСТИЧНОЇ НЕВИЗНАЧЕНОСТІ / О.Г. Наконечний, Г.І. Кудін, П.М. Зінько, Т.П. Зінько

Анотація. Досліджено задачу лінійного оцінювання невідомих математичних сподівань за спостереженнями реалізацій випадкових матричних послідовностей. Розроблено конструктивні математичні методи для знаходження лінійних гарантованих середньоквадратичних оцінок невідомих нестационарних параметрів середніх значень за спостереженнями реалізацій послідовності випадкових матриць. Показано, що такі гарантовані оцінки одержуються або як розв'язки крайових задач для систем лінійних різницевих рівнянь, або як розв'язки відповідних задач Коші. Установлено вигляд похибок для гарантованих середньоквадратичних квазімінімаксних оцінок спеціального вектора прогнозу та параметрів невідомих середніх значень. За наявності малих збурень відомих матриць у моделі матричних спостережень знайдено квазімінімаксні середньоквадратичні оцінки і в першому наближенні методу малого параметра отримано їх гарантовані середньоквадратичні похибки. Наведено два тестові приклади обчислення гарантованих середньоквадратичних оцінок та їх похибок.

Ключові слова: матричні спостереження, лінійне оцінювання, гарантована середньоквадратична оцінка, похибка гарантованої середньоквадратичної оцінки, квазімінімаксна гарантована оцінка вектора, різницеве рівняння, метод малого параметра, збурення матриць.

A GENETIC ALGORITHM IMPROVEMENT BY TOUR CONSTRAINT VIOLATION PENALTY DISCOUNT FOR MARITIME CARGO DELIVERY

V.V. ROMANUKE, A.Y. ROMANOV, M.O. MALAKSIANO

Abstract. The problem of minimizing the cost of maritime cargo delivery is considered. The cost is equivalent to the sum of the tour lengths of feeders used for the delivery. The problem is formulated as a multiple traveling salesman problem. In order to find its solution as the shortest route of the tours of feeders, a genetic algorithm is used where we present two inequalities constraining the tour length of every feeder to lie between the shortest and longest lengths. Apart from the constant tour constraint violation penalty in the genetic algorithm, we suggest a changeable penalty as an exponential function of the algorithm iteration, where we maintain the possibility of the penalty rate to be either increasing or decreasing, whose steepness is controlled by a positive parameter. Our tests show that the changeable penalty algorithm may return shorter routes, although the constant penalty algorithms cannot be neglected. As the longest possible tour of the feeder is shortened, the changeable penalty becomes more useful owing to a penalty discount required either at the beginning or at the end of the algorithm run to improve the selectivity of the best feeder tours. In optimizing maritime cargo delivery, we propose to run the genetic algorithm by the low and constant penalties along with the increasing and decreasing penalties. The solution is the minimal value of the four route lengths. In addition, we recommend that four algorithm versions be initialized by four different pseudorandom number generator states. The expected gain is a few percent, by which the route length is shortened, but it substantially reduces expenses for maritime cargo delivery.

Keywords: maritime cargo delivery, tour length, genetic algorithm, tour constraint violation penalty, penalty discount.

INTRODUCTION

The up-to-date market of cargo delivery is divided into three branches of transportation: ground-surface, water, and air. Among them the water transportation has been the most used. In general, this is the maritime transportation which is the basis of the world trading and commerce. Roughly about 80% of all goods are transported by river, sea and ocean. The amount of maritime cargo has been dramatically growing since 1980. In 2020, there were about 1.85 billion metric tons shipped all over the world, whereas it was only 0.1 billion metric tons in 1980. Quite naturally, the world fleet of containers has expanded. The gross tonnage of container carriers since 1980 has increased from 11 up to 275 million metric tons [1, 2].

The main advantages of maritime transportation over competitors are the cost and reliability, and also the possibility to deliver any cargo. The main drawbacks are relatively low speed of delivery and dependence on weather. It is impossible to influence weather, when a delivery is scheduled, but it is possible to increase the delivery speed by routing the most efficient tours. The efficient tour implies its minimally possible length, expressed in units of either distance or time.

Tour length minimization is a transportation optimization problem [3]. In particular, this is a version of the assignment problem or the traveling salesman problem [4, 5].

In fact, the traveling salesman problem solves the task of routing efficient tours, using which optimizes the cost of the delivery. It is an NP-hard problem in combinatorial optimization, whose exact solution usually takes too long to be obtained because exact algorithms perform reasonably fast only for small-sized problems [6]. Heuristic algorithms perform far much faster producing approximated solutions and saving computational resources (which are equivalent to time and budget) [7, 8].

One of the best heuristics is the genetic algorithm allowing to find tours whose length is practically close to the minimal length of the delivery [9, 10]. Sometimes the length found heuristically coincides with the length in the exact solution. For maritime cargo delivery with using multiple tours, the genetic algorithm requires such input parameters as follows: a map of ports, a number of feeders (in maritime transportation, a cargo boat is called the feeder), a population size, and a series of additional inputs including mutation operators. In detail, the map of ports is the two-coordinate location of ports which should be visited en route. The number of feeders defines the maximal number of tours by which the cargo can be delivered. The population size is the number of randomly generated tours to be processed by the algorithm.

To obtain the best approximated solution, the adjustable inputs (like the population size, mutation operators, and others) should be optimally configured. The optimal configuration is a very tough task being itself an optimization problem (similar, e. g., to the optimization in AutoML [11, 12]). In this way, rules of thumb are widely accepted based on recent experience [13, 14]. Another way to optimize the algorithm performance is to use penalty when a tour length exceeds an upper length. The upper length is determined by the capability of the feeder which can cover only the upper length distance and after that it will require fuel refill. This is a tour constraint whose violation imposes a penalty that expunges too lengthy tour from the processing. However, the tour constraint penalty is taken by rules of thumb as well [15, 16]. Therefore, a proper rationalization of the penalty would improve the genetic algorithm performance.

PROBLEM STATEMENT

The goal is to algorithmize the tour constraint penalty in order to optimize the genetic algorithm itself. Moreover, the penalty algorithmization is expected to make possible further minimization of the tour cost. For achieving the goal, the following five tasks are to be fulfilled:

1. To formalize variables used in the genetic algorithm for a maritime cargo delivery model. The model is to be formulated based on [15].
2. To substantiate the inclusion of the tour constraint penalty into the algorithm. The penalty must be changeable depending on the state of the algorithm convergence.
3. To show the advantage of the algorithm using the changeable penalty compared to the algorithm using the constant penalty.
4. To discuss the significance and practical applicability of the suggested improvement in the genetic algorithm.

5. To make an unbiased conclusion on the contribution to the field of genetic algorithms used, in particular, to optimize maritime cargo delivery. An outlook of how the research should be extended and advanced is to be made as well.

MARITIME CARGO DELIVERY MODEL

We have N ports, from one of which every feeder starts its tour and ends up by returning to that port. By default, the port is assigned number 1 and is called the hub. Let p_{k1} and p_{k2} be coordinates of port k . Coordinates of all the ports are gathered in matrix

$$\mathbf{P} = [p_{kl}]_{N \times 2}. \quad (1)$$

The distance between port k and port j is

$$\rho(k, j) = \sqrt{(p_{k1} - p_{j1})^2 + (p_{k2} - p_{j2})^2} \text{ by } k = \overline{1, N} \text{ and } j = \overline{1, N}.$$

All the distances are gathered in matrix

$$\mathbf{D} = [\rho(k, j)]_{N \times N}. \quad (2)$$

Obviously,

$$\rho(k, j) = \rho(j, k) \quad \forall k = \overline{1, N} \text{ and } \forall j = \overline{1, N}$$

and

$$\rho(k, k) = 0 \quad \forall k = \overline{1, N}.$$

So, matrix (2) is symmetric:

$$\mathbf{D} = \mathbf{D}^T.$$

Matrix of distances (2) is directly associated with durations of the maritime cargo delivery. The durations, in their turn, can be treated as the costs of the delivery.

The maximally possible number of feeders is denoted by M_{\max} , where $M_{\max} \in \mathbb{N} \setminus \{1\}$. We consider binary variable x_{kjm} associating ports k and j and feeder m , where $m = \overline{1, M}$ and M is a current total number of feeders:

$$M \leq M_{\max}. \quad (3)$$

Thus, $x_{kjm} = 1$ if ports k and j are included into the tour of feeder m , where the feeder visits either port j after port k or port k after port j : if $x_{kjm} = 1$ then $x_{jkm} = 0$ and if $x_{jkm} = 1$ then $x_{kjm} = 0$ for non-two-port tours. If a tour of feeder m is of just ports 1 and k , then $x_{1km} = x_{k1m} = 1$ because the feeder must return to the hub. Otherwise, if feeder m does not visit port j after port k nor port k after port j , $x_{kjm} = 0$ (although ports k and j still can be included into the tour of feeder m). So,

$$x_{kjm} \in \{0, 1\} \text{ by } k = \overline{1, N} \text{ and } j = \overline{1, N} \text{ and } m = \overline{1, M} \quad (4)$$

by

$$\sum_{j=2}^N \sum_{m=1}^M x_{1jm} = M \quad (5)$$

and

$$\sum_{k=2}^N \sum_{m=1}^M x_{k1m} = M, \quad (6)$$

where equality (5) means that each of M feeders only once departs from the hub, and equality (6) means that each of M feeders only once arrives to the hub.

Meanwhile, a feeder may not cover the distance greater than d_{\max} . Therefore, inequality

$$\sum_{k=1}^N \sum_{j=1}^N \rho(k, j) \cdot x_{kjm} \leq d_{\max} \quad \forall m = \overline{1, M} \quad (7)$$

constrains the tour of every feeder. Moreover, the feeder must not be charged for the delivery if its tour is too short. If d_{\min} is the shortest possible tour of the feeder, inequality

$$\sum_{k=1}^N \sum_{j=1}^N \rho(k, j) \cdot x_{kjm} \geq d_{\min} \quad \forall m = \overline{1, M} \quad (8)$$

also constrains the tour of every feeder.

Only one feeder can arrive at port j , being not the hub, from only one port (which can be the hub). This is expressed by equality

$$\sum_{k=1}^N \sum_{m=1}^M x_{kjm} = 1 \quad \forall j = \overline{2, N}. \quad (9)$$

Symmetrically, only one feeder can depart from port k , being not the hub, towards only one following port (which can be the hub). This is expressed by equality

$$\sum_{j=1}^N \sum_{m=1}^M x_{kjm} = 1 \quad \forall k = \overline{2, N}. \quad (10)$$

Every feeder must depart from the hub and arrive at it, so its tour is a closed loop. This is ensured by the following requirement:

$$\sum_{k \in Q_m} \sum_{j \in Q_m \setminus \{k\}} x_{kjm} \leq |Q_m| - 1$$

$$\forall Q_m \subset T_m = \{1, \{q_l^{(m)}\}_{l=2}^{A_m}\} \subset \{1, N\} \text{ by } 2 \leq |Q_m| < A_m \text{ and } \forall m = \overline{1, M} \quad (11)$$

with tour

$$T_m = \{1, \{q_l^{(m)}\}_{l=2}^{A_m}\} \subset \{1, N\} \quad (12)$$

of feeder m . Constraint (11) eliminates any subtours of every feeder. This ensures that a feasible route of delivering maritime cargo is of closed loops only, where every loop is a feeder tour starting off the hub and ending up by returning to the hub.

To optimize the maritime cargo delivery, we minimize the sum of all the tours of the feeders: objective function

$$\rho_{\Sigma}\left(N, M, \left\{\left\{\left\{x_{kjm}\right\}_{k=1}^N\right\}_{j=1}^N\right\}_{m=1}^M, d_{\min}, d_{\max}\right) = \sum_{k=1}^N \sum_{j=1}^N \sum_{m=1}^M x_{kjm} \cdot \rho(k, j)$$

is to be minimized subject to constraints (3)–(12). The minimization is implied to be done over binary variables (4) along with trying to minimize the total number of feeders used in the tours. That is, the minimization goal is to find such

$$M^* \in \overline{\{1, M_{\max}\}}$$

and

$$x_{kjm}^* \in \{0, 1\} \text{ for } k = \overline{1, N} \text{ and } j = \overline{1, N} \text{ by } m = \overline{1, M^*}$$

at which

$$\begin{aligned} & \sum_{k=1}^N \sum_{j=1}^N \sum_{m=1}^{M^*} x_{kjm}^* \cdot \rho(k, j) = \\ & = \rho_{\Sigma}\left(N, M^*, \left\{\left\{\left\{x_{kjm}^*\right\}_{k=1}^N\right\}_{j=1}^N\right\}_{m=1}^{M^*}, d_{\min}, d_{\max}\right) = \\ & = \min_{\substack{\left\{\left\{x_{kjm}\right\}_{k=1}^N\right\}_{j=1}^N, \\ m=1, M, M=1, M_{\max}}} \sum_{k=1}^N \sum_{j=1}^N \sum_{m=1}^M x_{kjm} \cdot \rho(k, j). \end{aligned} \quad (13)$$

The solution given formally as

$$\left\{\left\{\left\{x_{kjm}^*\right\}_{k=1}^N\right\}_{j=1}^N\right\}_{m=1}^{M^*} \quad (14)$$

allows to build a set of M^* the most rational tours of M^* feeders. Sum (13) of these tours is the shortest route to deliver maritime cargo and return to the hub.

THE TOUR CONSTRAINT PENALTY

Even for a few tens of ports, it is an intractably time-consuming computational task to find an exact solution of problem (13) subject to constraints (3)–(12). A solution whose route length is quite close to the shortest route length is obtained by genetic algorithms. One of the best genetic algorithms designed for solving problem (13) subject to constraints (3)–(7) and (9), (10) was presented in [15]. Herein, we add constraint (8) cutting off too short feeder tours, and add constraint (11) with tour (12) of feeder m eliminating subtours.

The genetic algorithm uses four forms of chromosome mutations: flip, swap, slide, and crossover. The crossover operation takes two chromosomes, cuts each chromosome in two parts in random places, and interchanges those parts. For generating a random place of the chromosome cut, the minimal number of ports every feeder should visit without counting the hub after starting off port 1 (hub) is used. This number is

$$H_{\min} = \psi\left(\frac{N-1}{M_{\max}}\right), \quad (15)$$

where function $\psi(x)$ returns the integer part of number x [17]. In addition, within the crossover operation, two chromosomes as tours of two different feeders may be merged into a single tour allowing to decrease the number of feeders used to deliver maritime cargo. This is done with using a merging probability β given at the input of the genetic algorithm.

Let H_m be the number of ports which feeder m should visit after starting off port 1 (hub). Thus, we denote the vector of the tour of feeder m (vector of ports which feeder m should visit in the order of the sequence of the vector elements) by

$$\mathbf{F}_m = [f_h^{(m)}]_{1 \times H_m}. \quad (16)$$

So,

$$\bigcup_{m=1}^M \{f_h^{(m)}\}_{h=1}^{H_m} = \overline{\{2, N\}}. \quad (17)$$

Initially, tours $\{\mathbf{F}_m\}_{m=1}^M$ of feeders are randomly generated by breaking the set of non-hub ports $\overline{\{2, N\}}$ with using integers (15) and M . Each feeder has a series of such tours called population.

For every element of the population, the following routine is executed during an iteration of the algorithm. First,

$$d_m = 0 \text{ for } m = \overline{1, M}.$$

The distance to the port following the hub is calculated as

$$d_m = \rho(1, f_1^{(m)}).$$

Then, the remaining distances except the last one are accumulated into d_m :

$$d_m^{(\text{obs})} = d_m, \quad d_m = d_m^{(\text{obs})} + \rho(f_k^{(m)}, f_{k+1}^{(m)}) \text{ for } k = \overline{1, H_m - 1}.$$

Finally, the distance of returning to the hub is:

$$d_m^{(\text{obs})} = d_m, \quad d_m = d_m^{(\text{obs})} + \rho(f_{H_m}^{(m)}, 1). \quad (18)$$

To improve selectivity of the best feeder tours to solution (14), and to expunge tours which violate conditions (7), (8), the tour constraint violation penalty is applied. Thus, as it was shown in [15], if $d_m > d_{\max}$ then a current accumulated distance d_m after (18) can be increased with a factor $\lambda > 0$:

$$d_m^{(\text{obs})} = d_m, \quad d_m = d_m^{(\text{obs})} + (d_m^{(\text{obs})} - d_{\max}) \cdot \lambda.$$

The increment of d_m in the case of $d_m < d_{\min}$ can be done in the same way.

However, we introduce a more flexible tour constraint violation penalty. In our version, the penalty rate depends on the iteration (denoted by i) of the genetic algorithm. It is either increasing or decreasing that is controlled by a positive parameter α :

$$r(i) = 1 + \frac{1 + \text{sign}(\alpha - 1)}{2} + \text{sign}(1 - \alpha)e^{-i \cdot |\alpha|},$$

where the case $\alpha = 1$ is excluded. The penalty rate is increasing if $\alpha > 1$, and it is decreasing if $\alpha < 1$. For instance, if $\alpha = 1.01$ then the penalty rate exponentially increases from a number close to α (because $r(1) \approx \alpha$ in this case) up to 2 (Fig. 1). If, say, $\alpha = 0.995$ then the penalty rate exponentially decreases from a number close to $1 + \alpha$ (in this case $r(1)$ is slightly greater than $1 + \alpha$) down to 1 (Fig. 2).

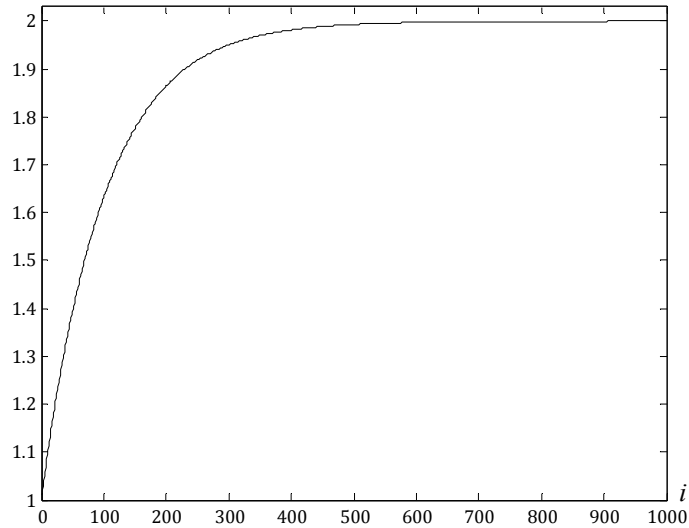


Fig. 1. The increasing tour constraint violation penalty by $\alpha = 1.01$

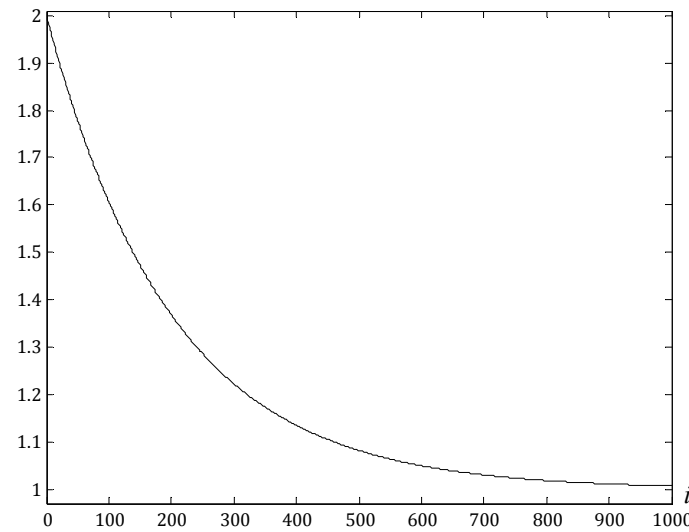


Fig. 2. The decreasing tour constraint violation penalty by $\alpha = 0.995$

Therefore, upon obtaining accumulated distance d_m by (18), if $d_m > d_{\max}$ then

$$d_m^{(\text{obs})} = d_m, \quad d_m = d_m^{(\text{obs})} + (d_m^{(\text{obs})} - d_{\max}) \cdot r(i).$$

If $d_m < d_{\min}$ then

$$d_m^{(\text{obs})} = d_m, \quad d_m = d_m^{(\text{obs})} + (d_{\min} - d_m^{(\text{obs})}) \cdot r(i).$$

Finally, sum

$$\tilde{\rho}_{\Sigma}(N, M, \{\mathbf{F}_m\}_{m=1}^M, d_{\min}, d_{\max}; \alpha) = \sum_{m=1}^M d_m$$

is calculated and minimized over the population. Obviously,

$$\tilde{\rho}_{\Sigma}(N, M, \{\mathbf{F}_m\}_{m=1}^M, d_{\min}, d_{\max}; \alpha) \geq \rho_{\Sigma} \left(N, M^*, \left\{ \left\{ \{x_{kjm}^*\}_{k=1}^N \right\}_{j=1}^N \right\}_{m=1}^{M^*}, d_{\min}, d_{\max} \right).$$

Herein, the question is which α is to be selected. The matter is that at different values of α the output of the genetic algorithm varies. This is so due to the state of the algorithm convergence varies depending on how tours violating requirements (7), (8) are expunged. Therefore, it is better to run through a set of the values and to select such a value at which the route is the shortest. The first run of the algorithm is done at $\alpha = 1.01$, whereupon the value is decreased by a factor slightly less than 1:

$$\alpha^{(\text{obs})} = \alpha, \quad \alpha = 0.999\alpha^{(\text{obs})}. \quad (19)$$

After 10 runs, the penalty rate is still an exponentially increasing curve because $\alpha \approx 1.000946$. Since the 11-th run, the penalty rate decreases because then $\alpha \approx 0.999945$. In fact, $\alpha > 0.95$ for the first 62 runs, whereas $\alpha < 0.95$ after the 63-rd run. So, we re-run the algorithm until $\alpha > 0.95$ starting with $\alpha = 1.01$ and proceeding by (19). Besides, we use an early stop condition imposed on the re-running. Denote by

$$\tilde{\rho}_{\Sigma}^* = \tilde{\rho}_{\Sigma} \left(N, M^*, \{\mathbf{F}_m\}_{m=1}^{M^*}, d_{\min}, d_{\max}; \alpha^* \right)$$

the shortest route length found so far. Denote by s_{fail} the counter of fails to improve the route (i. e., to shorten its length), and denote the maximal number of such fails by $s_{\text{fail}}^{(\text{max})}$. If

$$\tilde{\rho}_{\Sigma} \left(N, M^*, \{\mathbf{F}_m\}_{m=1}^{M^*}, d_{\min}, d_{\max}; \tilde{\alpha} \right) < \tilde{\rho}_{\Sigma}^* \quad (20)$$

for a next value of $\alpha = \tilde{\alpha}$, then

$$\tilde{\rho}_{\Sigma}^* = \tilde{\rho}_{\Sigma} \left(N, M^*, \{\mathbf{F}_m\}_{m=1}^{M^*}, d_{\min}, d_{\max}; \tilde{\alpha} \right)$$

and

$$\alpha^* = \tilde{\alpha},$$

whereupon the counter of fails is set at 0:

$$s_{\text{fail}} = 0.$$

Otherwise, if (20) is false,

$$s_{\text{fail}}^{(\text{obs})} = s_{\text{fail}}, \quad s_{\text{fail}} = s_{\text{fail}}^{(\text{obs})} + 1.$$

So, the algorithm is re-run while $\alpha > 0.95$ and $s_{\text{fail}} < s_{\text{fail}}^{(\text{max})}$.

To see whether the algorithm performs better with a decreasing tour constraint violation penalty, we need $s_{\text{fail}}^{(\text{max})} \geq 10$. In any way, the last value of α in ascertaining the performance must be less than 1. Therefore, in short, the described flexible penalty may be called the tour constraint penalty discount, although an increasing tour constraint violation penalty can give the shortest route as well (for simplicity, the route returned by the algorithm we will further call the shortest, although a shorter route may exist). In this case, it can be said that a penalty discount is given at the start of the algorithm run; as the run advances (the number of passed iterations increases), the discount decays.

TESTING

First, we test the algorithm for 10 to 50 ports randomly scattered. In this case, all ports coordinates (1) are in matrix

$$\mathbf{P} = 50 \cdot \Theta(N, 2)$$

by

$$N = 5 + 5n, \quad n = \overline{1, 9}$$

and an operator $\Theta(N, 2)$ returning a pseudorandom $N \times 2$ matrix whose entries are drawn from the standard uniform distribution on the open interval $(0; 1)$. The remaining parameters are:

$$M_{\text{max}} = 2, \quad s_{\text{fail}}^{(\text{max})} = 15, \quad \beta = 0.05,$$

$$d_{\text{max}} = \psi \left(1.25 \cdot \psi \left(0.5 \cdot \max_{k=1, N} \left\{ \sum_{j=1}^N \rho(k, j) \right\} \right) \right),$$

$$d_{\text{min}} = \zeta(0.1 \cdot d_{\text{max}}), \tag{21}$$

where function $\zeta(x)$ rounds number x to the nearest integer towards infinity. The maximal number of iterations is 3600, whereas the algorithm early stop condition is used, by which (a run of) the algorithm is stopped if the shortest route length does not change for 720 iterations (a one fifth of the maximal number of iterations). The test is repeated for 100 times for the algorithm used in three versions: with the tour constraint penalty discount, with the constant penalty by

$$r(i) = 1 \quad \forall i = \overline{1, 3600}, \tag{22}$$

and with the augmented constant penalty by

$$r(i) = 100 \quad \forall i = \overline{1, 3600}. \tag{23}$$

Overall, there are 900 route lengths (for instances of randomly generated ports) returned by each of the three versions, where $M^* = 1$ (the shortest route is of a single tour in every solution). We will refer to them as v_α, v_1, v_{100} , respectively. Whereas the pseudorandom number generator outputs the same instance for each of the three algorithm versions, re-runs for v_α are not initialized with the

same pseudorandom number generator seed. The advantage of the algorithm using the changeable penalty compared to the algorithm using constant penalties by (22), (23) can be seen in Table 1, where “better than” implies producing a shorter route. Table 2 shows that v_α has been worse than v_{100} in less than 1.5 %. Compared v_α to v_1 , this percentage is even far smaller – just 0.1111 % (only one of those 900 route lengths produced by v_α has appeared to be longer than the respective route length produced by v_1 ; this is an instance of 50 ports whose v_α -route length is 274.8312 and v_1 -route length is 274.6006). Table 3 shows that v_1 and v_{100} are more likely to produce the same result (strictly speaking, the equal lengths of the routes, whereas the routes themselves may differ in particular regions). The percentage of instances where v_α performs identically to v_1 or v_{100} is not that small. At least, this is 20% on average.

Table 1. The percentage of instances where one algorithm version produces a shorter route than the other version

		v_α better than v_1	v_α better than v_{100}	v_1 better than v_{100}
Overall average		75.7778	75.8889	19.3333
<i>N</i>	10	10	17	11
	15	55	51	8
	20	78	71	9
	25	83	84	14
	30	88	88	19
	35	88	89	22
	40	95	95	29
	45	90	90	28
	50	95	98	34

Table 2. The percentage of instances where one algorithm version produces a longer route than the other version

		v_α worse than v_1	v_α worse than v_{100}	v_1 worse than v_{100}
Overall average		0.1111	1.4444	21.5556
<i>N</i>	10	0	0	2
	15	0	0	15
	20	0	0	17
	25	0	0	17
	30	0	0	19
	35	0	3	28
	40	0	3	24
	45	0	6	40
	50	1	1	32

Table 3. The percentage of instances where two algorithm versions produce the same route lengths

		v_α and v_1	v_α and v_{100}	v_1 and v_{100}
Overall average		24.1111	22.6667	59.1111
N	10	90	83	87
	15	45	49	77
	20	22	29	74
	25	17	16	69
	30	12	12	62
	35	12	8	50
	40	5	2	47
	45	10	4	32
	50	4	1	34

In the test, v_α has produced 653 route lengths (72.5556 %) by $\alpha > 1$ (i. e., by an increasing tour constraint violation penalty). All the 100 instances of 10 ports are solved by $\alpha > 1$. Then, however, the percentage of instances where v_α has produced the shortest route by an increasing tour constraint penalty starts decreasing (Fig. 3). The way how the best values of α are distributed shown in Fig. 4 delusively hints at that the increasing penalty is better than the decreasing one. Meanwhile, it is worth noting that 224 of 900 instances has been v_α -solved by starting with $\alpha = 1.01$. The distributions of the best values of α for the number of ports in Fig. 5 confirm that the decreasing penalty becomes more influential as the number increases.

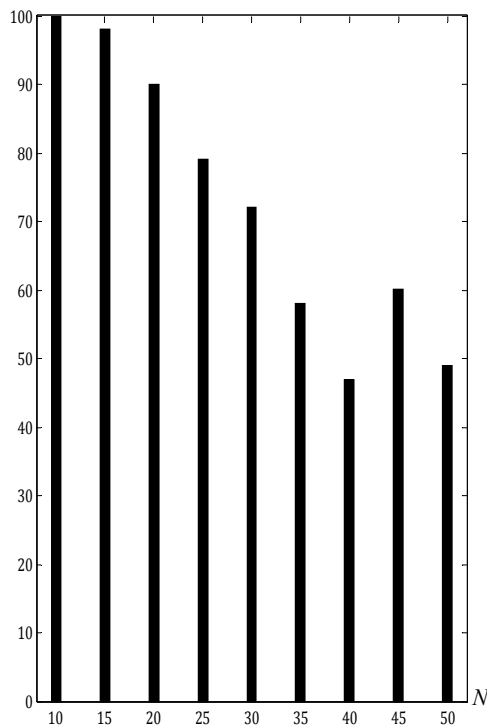


Fig. 3. The percentage of instances per number of ports where v_α has produced the shortest route by $\alpha > 1$

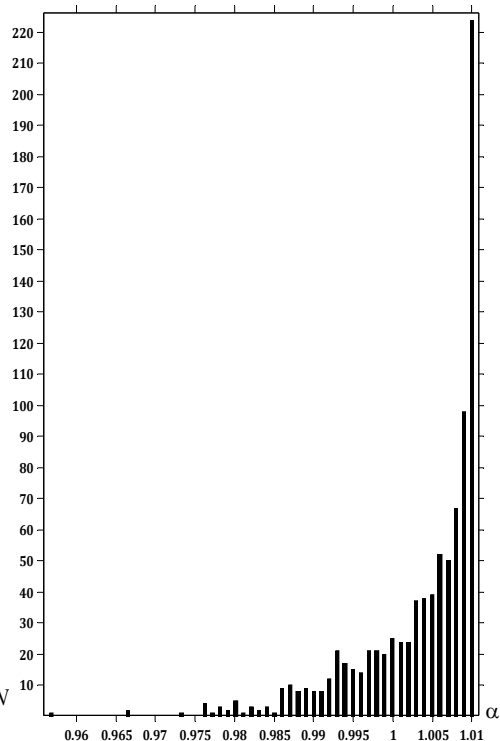


Fig. 4. The distribution of instances per α at which the shortest route is obtained

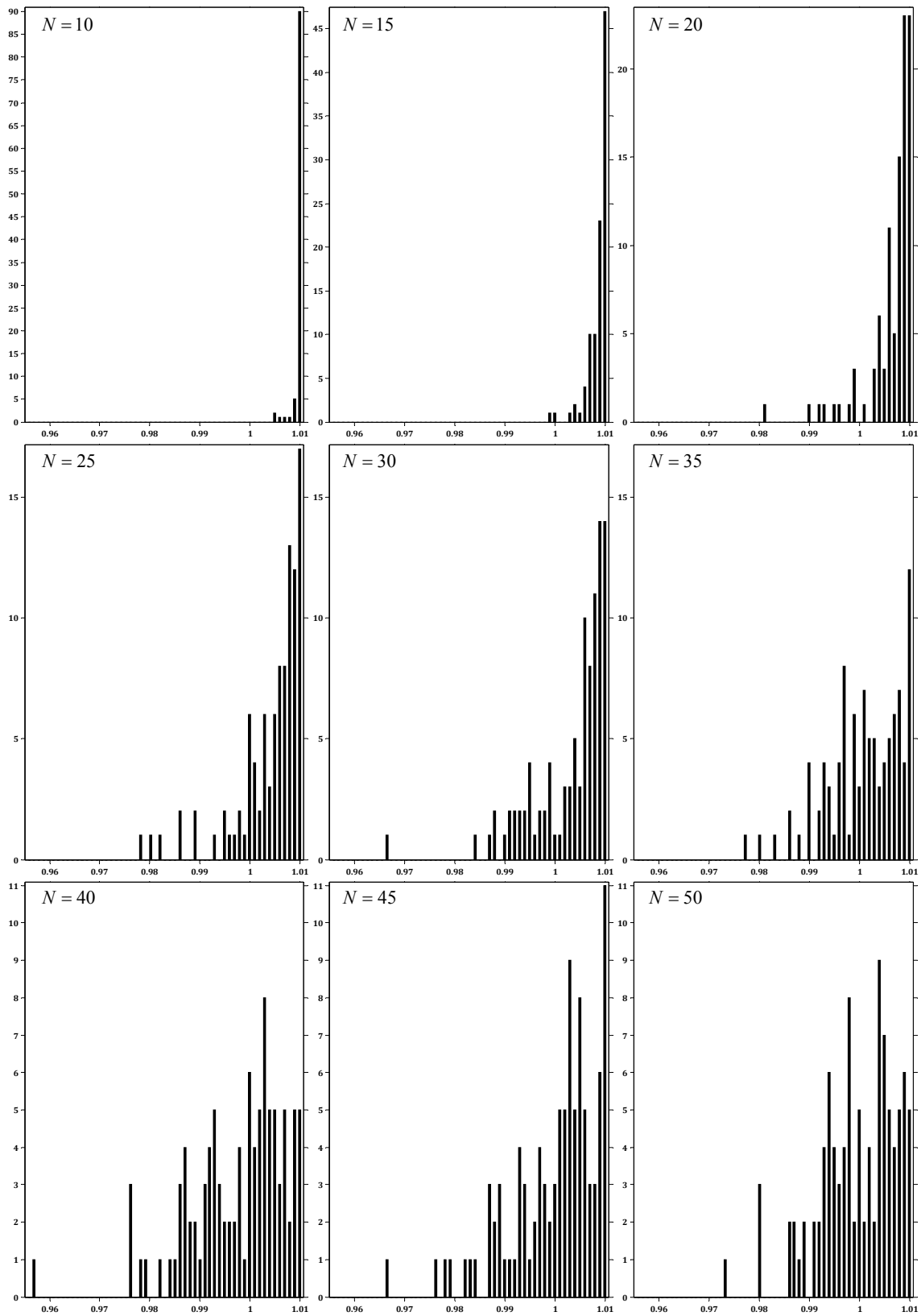


Fig. 5. The distributions of instances per α for the number of ports

A simple example of how the changeable penalty algorithm outperforms the two constant penalty algorithm versions is presented in Fig. 6. Compared to Fig. 7, where both the constant penalty algorithm versions produce the same route, the changeable penalty algorithm shortens the route by 3.7837 %, which is quite considerable and significant improvement. A more intricate example is presented in Fig. 8, where the shortest route through 50 ports is found at an increasing tour constraint violation penalty as well. Compared to Fig. 9, showing the shortest route found by v_1 , the changeable penalty algorithm shortens the route by 11.1689 %. Moreover, algorithm version v_{100} seeming to be a slightly more robust than v_1 produces a longer route (Fig. 10). Its length is 15.5299 % greater than that in Fig. 8.

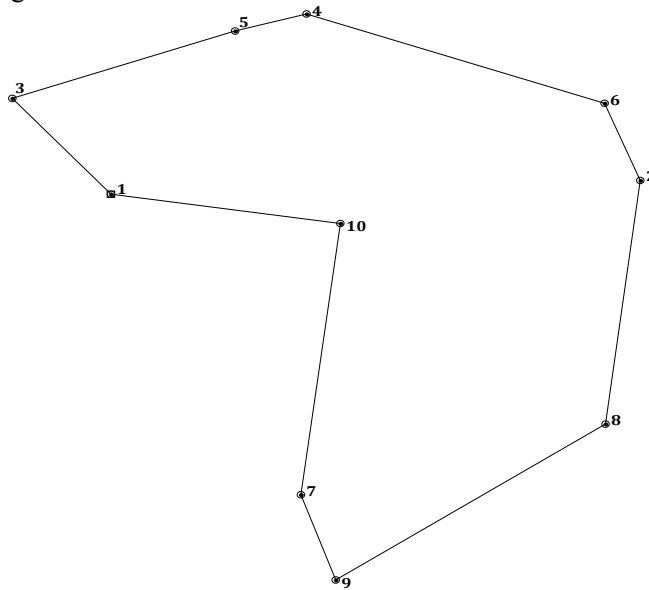


Fig. 6. The v_α -solution of an instance with 10 ports by $\alpha = 1.008$, where $\tilde{\rho}_\Sigma^* = 139.5179$

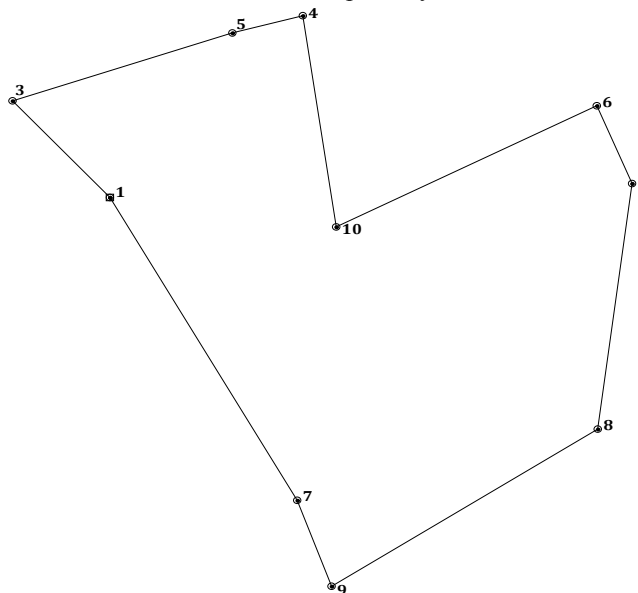


Fig. 7. The worse solution of the instance with 10 ports in Fig. 6 by v_1 and v_{100} , where $\tilde{\rho}_\Sigma^* = 145.0045$

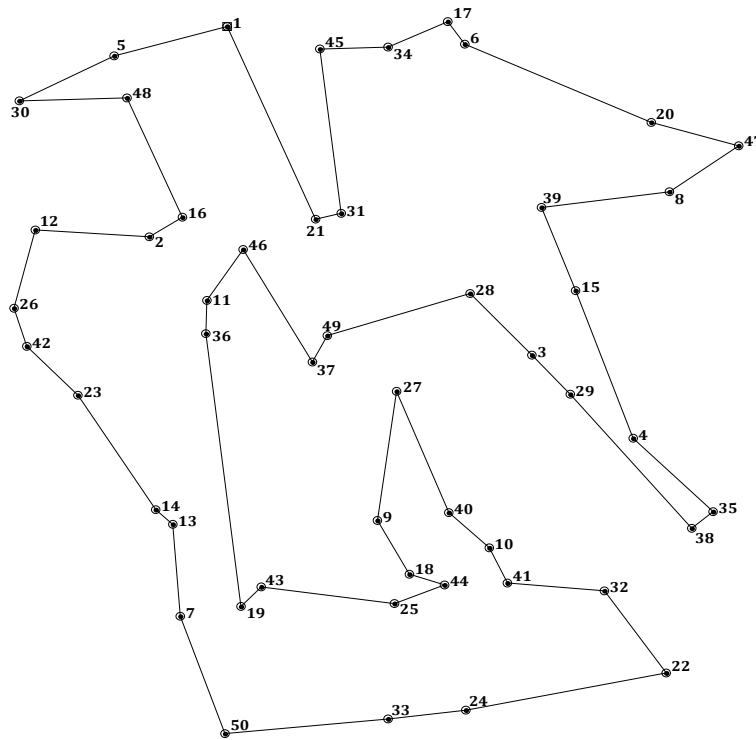


Fig. 10. The v_{100} -solution of the instance with 50 ports in Fig. 8, where $\tilde{\rho}_{\Sigma}^* = 337.1495$

It is noteworthy that the results reported in Tables 1–3 and Figs. 3–5 are statistically reliable, i. e. they are approximately repeatable for other pseudorandom number generator seeds. Thus, in another series of 900 instances, the overall average percentages from Table 1 first row are now 73.6667 %, 75.7778 %, 21 % (there have been obtained 189 routes by v_1 whose lengths are shorter than lengths by v_{100}), respectively. So, v_{α} is indeed “better than” v_1 and v_{100} in about 75 % of the modeled instances. The overall average percentages from Table 2 first row are now 0.1111 % (once again only one of those 900 route lengths produced by v_{α} has appeared to be longer than the respective route length produced by v_1), 0.8889 % (the difference is so big due to a few instances where v_{α} is “worse than” v_{100}), 16.8889 %, respectively. Just like in the first series, the only instance whose v_{α} -route is longer than v_1 -route has appeared to be of 50 ports, where the v_{α} -route length is 272.0435 and the v_1 -route length is 271.7023 (the difference in the first series is even smaller). Eventually, the overall average percentages from Table 3 first row are now 26.2222 %, 23.3333 %, and 62.1111 %, respectively, being really close to those ones in Table 3.

A very specific property of the genetic algorithm is that its result as the shortest route length depends on the pseudorandom number generator state seeded at the beginning of a test. Hence, we try re-running v_{α} initialized with the same pseudorandom number generator seed. By this set-up of the test, the shortest route length does depend on whether $\alpha > 1$ or $\alpha < 1$. Nevertheless, this dependence is weak: 769 instances have been v_{α} -solved by $\alpha = 1.01$ (the starting value for the increasing penalty rate), whereas just 99 instances have been v_{α} -solved by

$\alpha = 0.999945$ (the starting value for the decreasing penalty rate). Similarly to Tables 1–3, the comparison to v_1 and v_{100} is presented in Tables 4–6. These tables clearly show that the advantage of v_α is not that big, if any.

Table 4. The “better than” percentages for the same pseudorandom number generator seed test

		v_α better than v_1	v_α better than v_{100}	v_1 better than v_{100}
Overall average		42.5556	45.6667	21
<i>N</i>	10	11	15	10
	15	32	32	6
	20	37	37	8
	25	48	49	17
	30	46	48	25
	35	50	50	24
	40	52	63	28
	45	50	62	35
50	57	55	36	

Table 5. The “worse than” percentages for the same pseudorandom number generator seed test

		v_α worse than v_1	v_α worse than v_{100}	v_1 worse than v_{100}
Overall average		40.5556	37.3333	16.8889
<i>N</i>	10	6	4	4
	15	27	28	9
	20	47	45	9
	25	46	44	12
	30	51	47	15
	35	47	48	17
	40	48	37	26
	45	50	38	25
50	43	45	35	

Table 6. The “equal route lengths” percentages for the same pseudorandom number generator seed test

		v_α and v_1	v_α and v_{100}	v_1 and v_{100}
Overall average		16.8889	17	62.1111
<i>N</i>	10	83	81	86
	15	41	40	85
	20	16	18	83
	25	6	7	71
	30	3	5	60
	35	3	2	59
	40	0	0	46
	45	0	0	40
50	0	0	29	

Table 7 shows what the pure advantage is (when the algorithm version is simultaneously compared to the two other versions). Moreover, the average route length by v_α is 233.5604 (averaged over 900 route lengths), whereas the average route lengths by v_1 and v_{100} are 234.0213 and 234.41296, respectively. Therefore, at least a tiny advantage of v_α does exist, i. e. using the tour constraint penalty discount may indeed shorten the route.

Table 7. The percentages of performance comparison for the same pseudorandom number generator seed test

	better than the two other versions	worse than the two other versions	not worse than the two other versions
v_α	36	31.5556	53.6667
v_1	14.1111	12.5556	53.1111
v_{100}	9.6667	15.2222	48.5556

The final test is done for each of algorithm versions v_1 , v_{100} , v_α by $s_{fail}^{(max)} = 19$ and resetting every re-run with a new pseudorandom number generator seed being the same for v_1 , v_{100} , v_α . This is the purest experiment, where every instance is solved at least 20 times (and there are 19 attempts to shorten the very first route length). It is v_α -solved for 10 times by the increasing penalty and 10 times by the decreasing penalty, unless a shorter route is found by a decreased α . Now, the performance comparison similar to Table 7 is presented in Table 8. It is clearly seen that the purest advantage of v_α does exist (because there have been v_α -found 33 routes, which is 3.6667 %, each shorter than any of 20 routes by v_1 and any of 20 routes by v_{100} in the respective 33 instances). However, the purest advantage of v_{100} seems to be stronger. Amazingly enough, there have been v_α -found 134 routes (14.8889 %) each shorter than any of 20 routes by v_{100} in the respective 134 instances. Contrariwise, there have been v_{100} -found 143 routes (15.8889 %) each shorter than any route by v_α in the respective 143 instances. In other words, the algorithm version using the tour constraint penalty discount has an efficiency comparable (roughly speaking, almost the same) to the efficiency of the algorithm version using the high constant penalty. The respective percentages for v_α and v_1 are 8.3333 % (v_α outperforms v_1) and 11.2222 % (v_1 outperforms v_α).

Table 8. The percentages of performance comparison for the purest experiment

	better than the two other versions	worse than the two other versions	not worse than the two other versions
v_α	3.6667	6.1111	79
v_1	7.4444	5.8889	81.8889
v_{100}	12.2222	12.5556	80.2222

Consequently, the algorithm using the changeable penalty sometimes outperforms the algorithm using the constant penalty. Although this occasion is not very likely, there are expectedly about one problem of 25, when the changeable penalty algorithm will produce a shorter route than routes by the constant penalty algorithm versions. The latter can outperform as well, with slightly higher likelihoods.

The final test has revealed another interesting peculiarity. Among those 900 instances, 215 shortest routes by v_α have been found by $\alpha = 1.01$, 88 shortest v_α -routes have been found by $\alpha = 1.009$, and 68 shortest v_α -routes have been found by $\alpha = 1.008$ (the two nearest values to 1.01). The distribution resembles that one in Fig. 4. The ratio of the number of instances solved by $\alpha > 1$ to the number of instances solved by $\alpha < 1$ is about 2. Consequently, the increasing penalty has its one advantage over the decreasing penalty, but still the latter is “needed” roughly in every third problem solved by the changeable penalty algorithm.

DISCUSSION OF THE CONTRIBUTION

The experiment with controllable seed for generating maritime cargo delivery problem instances and for randomly generating tours (16) for (17) has shown that the changeable penalty, along with the constant penalty, is an important parameter of the genetic algorithm. It has been also revealed that the algorithm output depends on the seed. It is unclear how the best value of α could be selected. Moreover, it is impossible to foresee that the constant penalty algorithm version will not be outperformed by the changeable penalty algorithm. Therefore, the best decision is to use both the constant and changeable penalty versions (say, by running them on parallel processor cores), whereupon the shortest route length is trivially selected. In our case of study, we propose to run simultaneously four versions: v_1 , v_{100} , v_α by $\alpha = 1.01$, and v_α by $\alpha = 0.999945$ (here the penalty has the slowest descent; during the starting few thousand iterations, it decreases almost linearly).

An example of how the suggested changeable penalty improves the genetic algorithm is presented in Figs. 11–13. A maritime cargo delivery problem with 45 ports is solved, starting with the same pseudorandom number generator seed, by using the low constant penalty by (22), the high constant penalty by (23), and the increasing penalty with $\alpha = 1.01$. As we can see, the high constant penalty has improved the low constant penalty route length by just 0.08 % ($\tilde{\rho}_\Sigma^* = 300.90902$ in Fig. 12 against $\tilde{\rho}_\Sigma^* = 301.1532$ in Fig. 11). The improvement by the increasing penalty is far more significant: it is 3.8194 % ($\tilde{\rho}_\Sigma^* = 289.4161$ in Fig. 13) compared to $\tilde{\rho}_\Sigma^* = 300.90902$ in Fig. 12, and it is 3.8974 % compared to $\tilde{\rho}_\Sigma^* = 301.1532$ in Fig. 11.

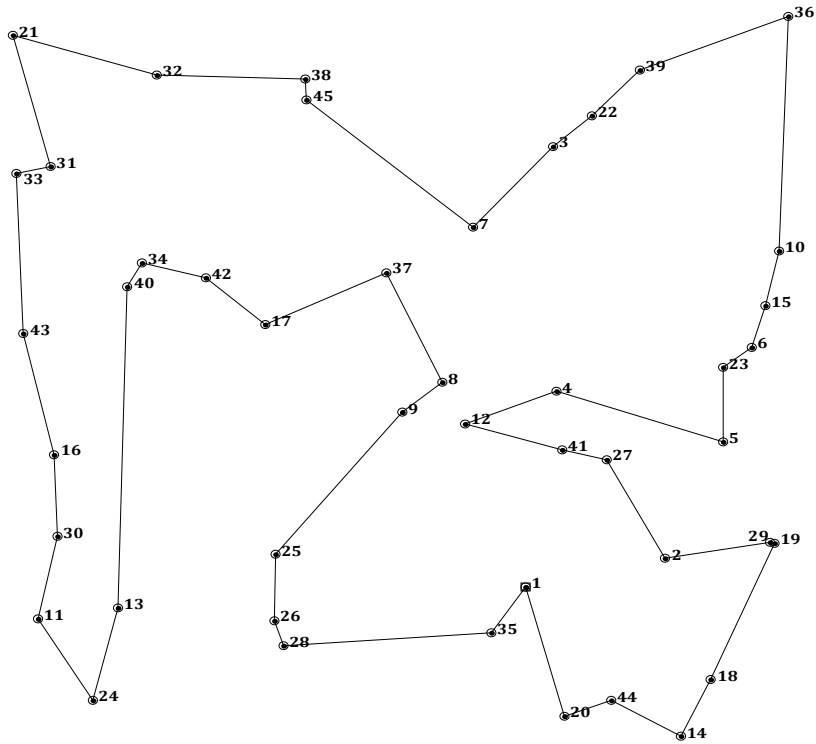


Fig. 11. The v_1 -solution of an instance with 45 ports, where $\tilde{\rho}_\Sigma^* = 301.1532$

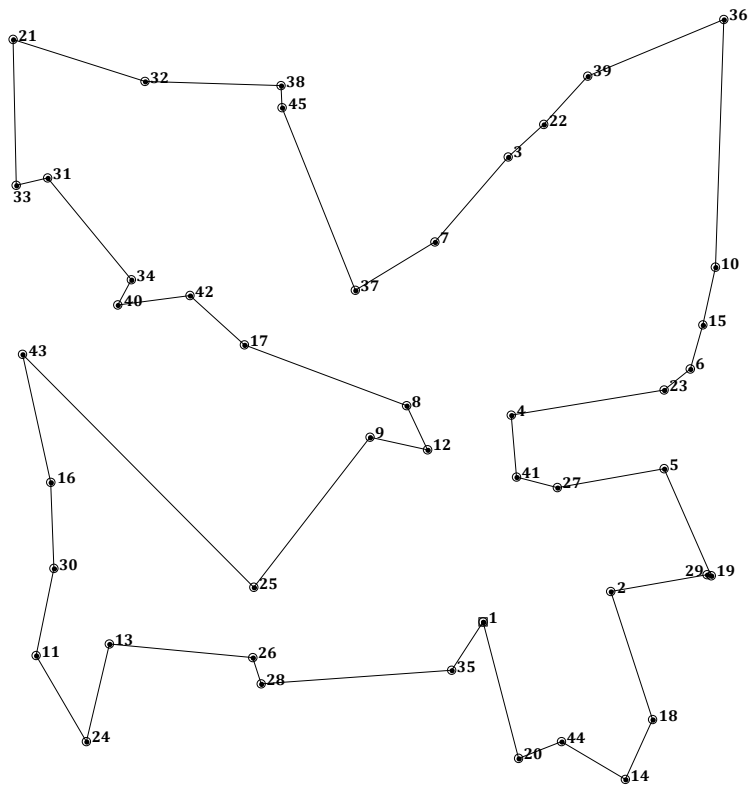


Fig. 12. The v_{100} -solution of the instance with 45 ports in Fig. 11, where $\tilde{\rho}_\Sigma^* = 300.90902$

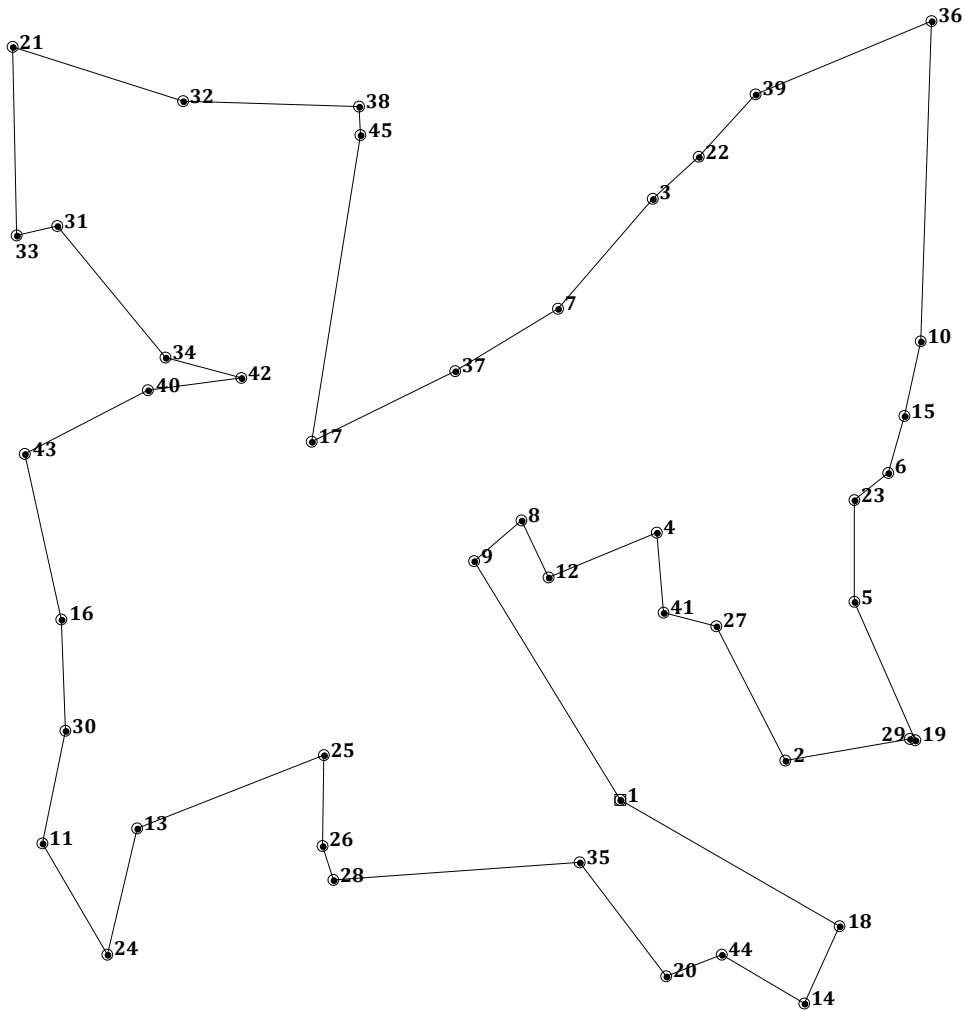


Fig. 13. The v_α -solution ($\alpha=1.01$) of the instance with 45 ports in Fig. 11, where $\tilde{\rho}_\Sigma^* = 289.4161$

Despite the example in Figs. 11–13 is a good demonstration of that the changeable penalty is a real improvement of the genetic algorithm, a counterexample is easily generated just on the same maritime cargo delivery problem by re-running the constant penalty algorithm versions and v_α (a re-run implies that the pseudorandom number generator state is changed). Thus, in a series of 302 re-runs, the shortest route length by v_1 varies between 272.8407 and 323.0998, and the shortest route length by v_α varies between the same lower and upper boundaries. The shortest route length by v_{100} varies between 272.8407 (the same lower boundary) and 317.6751, so its upper boundary is less than that for v_1 and v_α .

Nevertheless, as the maximally possible number of feeders is increased, the suggested changeable penalty further improves the genetic algorithm. In this case, we have

$$d_{\max} = \psi \left(1.25 \cdot \psi \left(\frac{1}{M_{\max}} \cdot \max_{k=1, N} \left\{ \sum_{j=1}^N \rho(k, j) \right\} \right) \right)$$

and (21), i. e. the longest and shortest possible tours of the feeder are shortened. For the same instance with $M_{\max} = 6$, the performance of v_{α} in a series of 738 re-runs is significantly better. The lower and upper boundaries in this case are 274.1295 and 321.0341 for v_1 , 274.70303 and 317.0557 for v_{100} , 272.8407 and 323.7964 for v_{α} , where the shortest route length $\tilde{p}_{\Sigma}^* = 272.8407$ is found in a re-run with $\tilde{p}_{\Sigma}^* = 288.4775$ (it is even shorter than that in Fig. 13) by v_1 and $\tilde{p}_{\Sigma}^* = 297.49702$ by v_{100} . Furthermore, in this “local” test series v_{α} is better than the two other versions in 27.9133 % and is worse than the two other versions in 23.9837 % of all re-runs. These rates are 3.9735 % and 1.9868 % for the series with $M_{\max} = 2$. By the way, the performance of the high constant penalty algorithm with $M_{\max} = 6$ significantly drops: the average route length is 290.4455 within the 302 re-runs with $M_{\max} = 2$, and it is 295.3313 with $M_{\max} = 6$.

At last, it is important to note that all the generated instances in our testings have been such that, in the solution returned by the algorithm, only one feeder was required to cover the shortest route ($M^* = 1$). This is practically possible and applicable due to we set the longest possible tour of the feeder at a relatively high value. Even though the shortened route length by the suggested improvement in the genetic algorithm is small, it is a significant decrement of the maritime cargo delivery cost.

CONCLUSION

We have presented a tour constraint violation penalty to the genetic algorithm for solving a maritime cargo delivery problem formulated as a multiple traveling salesman problem. The penalty is an exponential function of the iteration, where we maintain the possibility of the penalty rate to be either increasing or decreasing whose steepness is controlled by a positive parameter α . Our tests have shown that the changeable penalty algorithm may return shorter routes, although the constant penalty algorithms cannot be neglected. Therefore, our contribution to the field of genetic algorithms is the monotonous flexibility of the tour constraint violation penalty. The usefulness of this flexibility grows as the longest possible tour of the feeder is shortened. It is so due to a penalty discount is required either at the beginning or at the end of the algorithm run to improve selectivity of the best feeder tours. In optimizing maritime cargo delivery, we propose to run the genetic algorithm by the low and constant penalties along with the increasing and decreasing penalties, whereupon the solution is the minimal value of the four route lengths. In addition, we recommend the four algorithm versions to be initialized by four different pseudorandom number generator states. Although the gain is just a few percent (by which the route length is shortened) or less, it is a substantial reduction of expenses for maritime cargo delivery.

The research should be extended and advanced in the way of studying the pseudorandom number generator state influence. As we have revealed that the state influences the algorithm output, it is to ascertain lower and upper boundaries of the route length. Another open question is how many re-runs should be made (by changing the state) to obtain the shortest possible route in a maritime cargo delivery problem.

REFERENCES

1. T.R. Walker et al., “Chapter 27 — Environmental Effects of Marine Transportation,” in *World Seas: an Environmental Evaluation*. Cambridge, Massachusetts, USA: Academic Press, 2019, pp. 505–530. Available: <https://doi.org/10.1016/B978-0-12-805052-1.00030-9>
2. W. Li, R. Pundt, and E. Miller-Hooks, “An updatable and comprehensive global cargo maritime network and strategic seaborne cargo routing model for global containerized and bulk vessel flow estimation,” *Maritime Transport Research*, vol. 2, 100038, 2021. Available: <https://doi.org/10.1016/j.martra.2021.100038>
3. C. Archetti, L. Peirano, and M. G. Speranza, “Optimization in multimodal freight transportation problems: A Survey,” *European Journal of Operational Research*, vol. 299, iss. 1, pp. 1–20, 2022. Available: <https://doi.org/10.1016/j.ejor.2021.07.031>
4. X. Wu, J. Lu, S. Wu, and X. Zhou, “Synchronizing time-dependent transportation services: Reformulation and solution algorithm using quadratic assignment problem,” *Transportation Research Part B: Methodological*, vol. 152, pp. 140–179, 2021. Available: <https://doi.org/10.1016/j.trb.2021.08.008>
5. P.A. Miranda, C.A. Blazquez, C. Obreque, J. Maturana-Ross, and G. Gutierrez-Jarpa, “The bi-objective insular traveling salesman problem with maritime and ground transportation costs,” *European Journal of Operational Research*, vol. 271, iss. 3, pp. 1014–1036, 2018. Available: <https://doi.org/10.1016/j.ejor.2018.05.009>
6. D.-Z. Du and P.M. Pardalos, *Handbook of Combinatorial Optimization*. New York, NY, USA: Springer, 1998, 2406 p. Available: <https://doi.org/10.1007/978-1-4613-0303-9>
7. A. Hertz and M. Widmer, “Guidelines for the use of meta-heuristics in combinatorial optimization,” *European Journal of Operational Research*, vol. 151, iss. 2, pp. 247–252, 2003. Available: [https://doi.org/10.1016/S0377-2217\(02\)00823-8](https://doi.org/10.1016/S0377-2217(02)00823-8)
8. A. Colomi, M. Dorigo, F. Maffioli, V. Maniezzo, G. Righini, and M. Trubian, “Heuristics from nature for hard combinatorial optimization problems,” *International Transactions in Operational Research*, vol. 3, iss. 1, pp. 1–21, 1996. Available: [https://doi.org/10.1016/0969-6016\(96\)00004-4](https://doi.org/10.1016/0969-6016(96)00004-4)
9. L.D. Chambers, *The Practical Handbook of Genetic Algorithms*. Chapman and Hall/CRC, 2000, 544 p.
10. R. L. Haupt and S.E. Haupt, *Practical Genetic Algorithms*. John Wiley & Sons, 2003, 253 p. Available: <https://doi.org/10.1002/0471671746>
11. C. Thornton, F. Hutter, H.H. Hoos, and K. Leyton-Brown, “Auto-WEKA: Combined selection and hyperparameter optimization of classification algorithms,” in *KDD’13: Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 847–855, 2013.
12. V.V. Romanuke, “Optimal training parameters and hidden layer neurons number of two-layer perceptron for generalized scaled objects classification problem,” *Information Technology and Management Science*, vol. 18, pp. 42–48, 2015. doi: 10.1515/itms-2015-0007
13. E. Merhej, S. Schockaert, and M. De Cock, “Repairing inconsistent answer set programs using rules of thumb: A gene regulatory networks case study,” *International Journal of Approximate Reasoning*, vol. 83, pp. 243–264, 2017. Available: <https://doi.org/10.1016/j.ijar.2017.01.012>
14. A. Santini, A. Viana, X. Klimentova, and J.P. Pedroso, “The probabilistic travelling salesman problem with crowdsourcing,” *Computers & Operations Research*, vol. 142, 105722, 2022. Available: <https://doi.org/10.1016/j.cor.2022.105722>
15. A. Király and J. Abonyi, “Redesign of the supply of mobile mechanics based on a novel genetic optimization algorithm using Google Maps API,” *Engineering Applica-*

- tions *Of Artificial Intelligence*, vol. 38, pp. 122–130, 2015. Available: <https://doi.org/10.1016/j.engappai.2014.10.015>
16. L. Kota and K. Jarmai, “Mathematical modeling of multiple tour multiple traveling salesman problem using evolutionary programming,” *Applied Mathematical Modelling*, vol. 39, iss. 12, pp. 3410–3433, 2015. Available: <https://doi.org/10.1016/j.apm.2014.11.043>
17. V. V. Romanuke, “Job order input for efficient exact minimization of total tardiness in tight-tardy progressive single machine scheduling with idling-free preemptions,” *Scientific Papers of O. S. Popov Odessa National Academy of Telecommunications*, no. 1, pp. 19–36, 2020. Available: <https://doi.org/10.33243/2518-7139-2020-1-1-19-36>

Received 04.07.2022

INFORMATION ON THE ARTICLE

Vadim V. Romanuke, ORCID: 0000-0001-9638-9572, Vinnytsia Institute of Trade and Economics of State University of Trade and Economics, Ukraine, e-mail: romanukevadimv@gmail.com

Andriy Y. Romanov, ORCID: 0000-0002-1714-3310, Odessa National Maritime University, Ukraine, e-mail: andreygorogogo@gmail.com

Mykola O. Malaksiano, ORCID: 0000-0002-4075-5112, Odessa National Maritime University, Ukraine, e-mail: malaksiano@gmail.com

ПОКРАЩЕННЯ ГЕНЕТИЧНОГО АЛГОРИТМУ НА ОСНОВІ ДИСКОНТУ ШТРАФУ ЗА ПОРУШЕННЯ ОБМЕЖЕНЬ РЕЙСУ ДЛЯ МОРСЬКОЇ ДОСТАВКИ ВАНТАЖІВ / В.В. Романюк, А.Ю. Романов, М.О. Малаксіано

Анотація. Розглянуто задачу мінімізації вартості морської доставки вантажів. Ця вартість еквівалентна сумі довжин рейсів фідерів, що використовуються для доставки. Задача формулюється у формі задачі декількох комівояжерів. Для знаходження розв’язку у формі найкоротшого маршруту, що складається з рейсів фідерів, використовується генетичний алгоритм, у якому дві нерівності, котрі обмежують довжину рейсу кожного фідера до інтервалу між найкоротшою та найбільшою довжинами. Окрім сталого штрафу за порушення обмежень рейсу у генетичному алгоритмі запропоновано змінюваний штраф у формі експоненціальної функції ітерації алгоритму, де залишається можливість як зростаючого, так і спадного штрафу, чия крутизна контролюється деяким додатним параметром. Тести показують, що алгоритм зі змінюваним штрафом може повертати коротші маршрути, хоча алгоритми зі сталими штрафами не можуть бути відкинуті. Зі скороченням найдовшого рейсу фідера змінюваний штраф стає більш корисним завдяки тому, що деякий дисконт штрафу потрібний на початку або наприкінці прогону алгоритму задля покращення селективності найкращих рейсів фідерів. Для оптимізації морської доставки вантажів запропоновано запускати даний генетичний алгоритм за низького та високого штрафів разом зі зростаючим та спадаючим штрафами, після чого розв’язком є мінімальне значення з чотирьох відповідних довжин маршрутів. Рекомендовано ініціалізувати ці чотири версії алгоритму чотирма різними станами генератора псевдовипадкових чисел. Очікуваний виграш складає декілька відсотків скорочення довжини маршруту, але для морської доставки вантажів це є значним скороченням витрат.

Ключові слова: морська доставка вантажів, довжина рейсу, генетичний алгоритм, штраф за порушення обмежень рейсу, дисконт штрафу.

ЗАСТОСУВАННЯ КРИВИХ БЕЗЬЄ ДЛЯ ОПИСУ ФОРМИ КОНСТРУКЦІЇ ПРИ ОПТИМІЗАЦІЇ КЛЕЙОВОГО З'ЄДНАННЯ

С.С. КУРЕННОВ, К.П. БАРАХОВ, Г.С. БАРАХОВА

Анотація. Розв'язано задачу топологічної оптимізації симетричного двозрізного з'єднання. Запропонована математична модель з'єднання змінної товщини є узагальненням класичної моделі Голанда–Рейсснера. Форма накладки описується за допомогою кривої Безьє. Шуканими величинами в задачі оптимізації є координати контрольних точок кривої Безьє. Цільовою функцією може бути як довжина з'єднання, так і площа поперечного перерізу накладки. Обмеження накладаються на напруження в клейовому шарі та в накладці. Пряму задачу зі знаходження напруженого стану з'єднання за заданих геометричних параметрів розв'язано за допомогою методу скінченних різниць. Для розв'язання задачі оптимізації використано генетичний алгоритм. З метою покращення збіжності генетичного алгоритму запропоновано острівну модель еволюції, яка забезпечує швидкість еволюційного відбору і стабільність досягнутих результатів. Розв'язано модельну задачу.

Ключові слова: тришарова конструкція, топологічна оптимізація, генетичний алгоритм.

ВСТУП

Клейові з'єднання внапуск є невід'ємною частиною конструкцій із композиційних матеріалів. Поширення клейових з'єднань у композитних конструкціях зумовлено їх високою технологічністю, герметичністю, малою масою, високою аеродинамічною ефективністю. Клейові з'єднання внапуск не порушують структуру композитів і дозволяють реалізувати у конструкції їх високі міцнісні та механічні властивості. Однак відомим недоліком з'єднань внапуск є концентрація напружень у клейовому шарі на краях ділянки склеювання [1; 2]. Для зниження концентрації напружень і підвищення міцності з'єднань використовуються різні конструктивні рішення, такі як збільшення товщини клейового шару на краях з'єднання [3], зменшення товщини пластин біля краю з'єднання [4], використання декількох різних типів клеїв [5], інші способи [6; 7]. Застосування симетричних двосторонніх з'єднань дозволяє усунути вигин конструкції і зменшити відривні напруження у клейовому шарі [8–10].

Задача топологічної оптимізації є якісно складнішою задачею, ніж задача класичної параметричної оптимізації. Топологічна оптимізація з'єднання внапуск, як правило, полягає в знаходженні оптимальної довжини з'єднання і залежності зміни товщини пластин, що з'єднуються по довжині ділянки склеювання. Одним із можливих способів розв'язання цієї задачі є дискретизація шуканої функції. У цьому випадку задача зводиться до знаходження товщини елементів конструкції в системі точок [11]. Перехід від неперервних функцій до дискретних дозволяє оптимізувати з'єднання з покровою зміною товщини [12]. Якщо ж шукана функція є неперервною, то то

її можна описати за відомими значеннями в системі точок за допомогою сплайнів [13], функцій Безье [14, 15] або рядів Фур'є [16; 17]. Оскільки кількість параметрів досить велика, а розв'язання оберненої задачі оптимізації навіть числовими методами є неможливим, то для розв'язання задачі оптимізації використовуються наближені методи, наприклад, генетичні алгоритми. Цей метод полягає в знаходженні оптимальних параметрів задачі шляхом розв'язання послідовності прямих задач. Зазвичай напружений стан конструкції за заданих параметрів задачі знаходять за допомогою методу скінченних елементів [13–15]. Оптимізації також може підлягати товщина клейового шару на краях ділянки склеювання [18], розмір і форма видавлених надлишків клею на краю з'єднання [19], а також структура композиту [20; 21]. Загальним недоліком застосування методу скінченних елементів для розв'язання задач топологічної оптимізації є відносно повільна швидкість роботи алгоритму.

Метою цієї роботи є розв'язання задачі топологічної оптимізації клейового з'єднання в одновимірній постановці. Застосування математичних моделей з'єднання [1], які добре себе зарекомендували, і які використовуються для опису напруженого стану з'єднань в аналітичній формі, дає змогу зменшити розмірність задачі без істотної втрати точності і тим самим підвищити швидкість виконання розрахунків.

Генетичні алгоритми дозволяють розв'язувати задачі оптимізації за наявності обмежень [22]. У цьому випадку є обмеження на максимальні напруження, розміри конструкції тощо. Відомим недоліком генетичних алгоритмів є складність їх налаштування. Для підвищення швидкості збіжності генетичного алгоритму у роботі використовується удосконалена острівна модель генетичного алгоритму (Island Model GA) [23; 24]. Острівні моделі можна класифікувати за декількома ознаками, такими як можливі напрямки міграцій та умови еволюційного відбору на островах. Якщо для всіх островів умови однакові, модель називають гомогенною [25; 26]. Відповідно, якщо умови різні, то модель називають гетерогенною [27]. У запропонованому у цій роботі варіанті еволюційного алгоритму на одному з трьох островів мутагенез відбувається частіше і з більшою дисперсією, ніж на двох інших островах. Така комбінація високої варіабельності на одному острові та стабільності на двох інших у сукупності з регулярною міграцією кращих особин між островами забезпечує задовільну швидкість роботи еволюційного алгоритму та стабільність досягнутих результатів [17].

ПОСТАНОВКА ЗАДАЧІ

Розглянемо конструкцію, що складається із двох пластин, з'єднаних за допомогою симетричних накладок (рис. 1, *а*). Така конструкція не відчуває вигину за розтягування-стискання і тому часто використовується в машинобудуванні. З огляду на симетрію конструкції розглядаємо лише її четверту частину. Поперечні переміщення центрального шару (основної пластини) дорівнюють нулю. Якщо розглядати деформування даної конструкції у межах теорії стрижнів, можна розглянути лише ділянку склеювання. Конструкцію навантажено поздовжніми зусиллями $2F$. Товщину клейового шару вважаємо постійною по довжині з'єднання і однаковою на всіх ділянках. Довжина ділянки склеювання L . Диференціальний елемент області склеювання і діючі на нього сили чинники наведено на рис. 1, *б*.

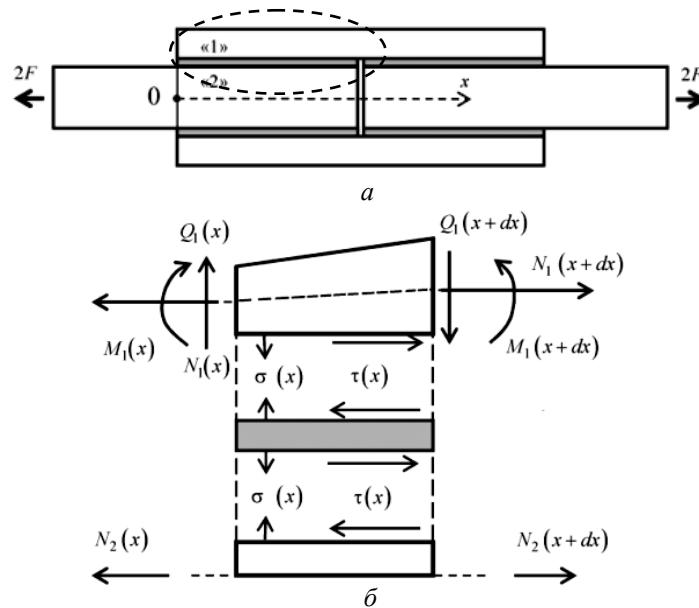


Рис. 1. Клейове з'єднання: а — схема конструкції, б — диференціальний елемент

Рівняння рівноваги накладки та основної пластини мають вигляд:

$$\frac{dN_1}{dx} = -\tau; \quad \frac{dN_2}{dx} = \tau; \quad \frac{dQ_1}{dx} = \sigma,$$

$$\frac{dM_1}{dx} - s_1(x)\tau - N_1 \frac{ds_1}{dx} + Q_1 = 0, \quad (1)$$

де N_1, N_2 — поздовжні зусилля в несних шарах; Q_1, M_1 — зрізне зусилля і згинальний момент у накладці; τ, σ — дотичні і нормальні напруження в клейовому шарі; s_1 — відстань від нейтральної осі накладки до клейового шару у разі симетричної структури накладки $s_1(x) = 0.5 \delta_1(x)$, де $\delta_1(x)$ — товщина накладки.

Рівняння деформації стрижнів мають вигляд:

$$N_1 = B_1 \frac{dU_1}{dx}; \quad N_2 = B_2 \frac{dU_2}{dx}; \quad D_1 \frac{d^2 W_1}{dx^2} = M_1, \quad (2)$$

де U_1, U_2 — поздовжні переміщення несних шарів; W_1 — поперечні переміщення накладки; $B_1(x)$ і B_2 — жорсткості шарів на розтягування-стиснення, якщо шари однорідні за товщиною, то $B_1(x) = \delta_1(x)E_1$, $B_2 = \delta_2 E_2$, де E_1, E_2 — модуль пружності відповідного шару; $D_1(x)$ — згинальна жорсткість накладки; $D_1(x) = \frac{\delta_1^3(x)E_1}{12}$.

Напруження в клейовому шарі можна записати у вигляді [1; 9]

$$\sigma = K \cdot W_1; \quad \tau = P \left(U_1 - U_2 + s_1(x) \frac{dW_1}{dx} \right), \quad (3)$$

де K , P — жорсткості клейового шару на розтягування-стиснення і на зсув, які можуть бути обчислені, наприклад як $K = \frac{E_0}{\delta_0}$, $P = \frac{G_0}{\delta_0}$, де δ_0 — товщина клейового шару; E_0 , G_0 — модуль пружності і модуль зсуву клею.

Крайові умови мають такий вигляд:

$$\begin{aligned} N_2(0) = F, \quad N_2(L) = 0, \quad N_1(0) = 0, \quad Q_1(0) = 0, \\ M_1(0) = 0, \quad U_1(L) = 0, \quad Q_1(L) = 0, \quad dW_1/dx|_{x=L} = 0. \end{aligned}$$

Систему рівнянь (1)–(3) можна звести до системи диференціальних рівнянь відносно U_1 , U_2 і W_1 :

$$\begin{aligned} \frac{B_1}{P} \frac{d^2 U_1}{dx^2} + \frac{1}{P} \frac{dB_1}{dx} \frac{dU_1}{dx} - U_1 + U_2 - s_1 \frac{dW_1}{dx} = 0, \quad U_1 + \frac{B_2}{P} \frac{d^2 U_2}{dx^2} - U_2 + s_1 \frac{dW_1}{dx} = 0 \\ \frac{D_1}{P} \frac{d^4 W_1}{dx^4} + \frac{2}{P} \frac{dD_1}{dx} \frac{d^3 W_1}{dx^3} + \left(\frac{1}{P} \frac{d^2 D_1}{dx^2} - s_1^2 \right) \frac{d^2 W_1}{dx^2} - 2s_1 \frac{ds_1}{dx} \frac{dW_1}{dx} + \frac{K}{P} W_1 - \\ - \frac{B_1}{P} \frac{ds_1}{dx} \frac{d^2 U_1}{dx^2} + \frac{ds_1}{dx} U_2 - \left(\frac{1}{P} \frac{ds_1}{dx} \frac{dB_1}{dx} + s_1 + \frac{B_1}{P} \frac{d^2 s_1}{dx^2} \right) \frac{dU_1}{dx} - \\ - \frac{ds_1}{dx} U_1 + s_1 \frac{dU_2}{dx} = 0. \end{aligned} \quad (4)$$

До цих самих змінних U_1 , U_2 і W_1 можна звести і крайові умови.

ЗАДАЧА ОПТИМІЗАЦІЇ

Задачу оптимізації сформулюємо так. Необхідно знайти довжину з'єднання L і залежність товщини накладки від поздовжньої координати $\delta_1(x)$, які забезпечують екстремальне значення деякого критерію оптимальності за умови виконання обмежень за міцністю. Як критерій оптимальності візьмемо масу накладки, яка з точністю до довільного множника пропорційна площі поперечного перерізу накладки:

$$M = \int_0^L \delta_1(x) dx \rightarrow \min. \quad (5)$$

Обмеження можуть бути накладені на максимальні напруження в клейовому шарі. При цьому можна використовувати різні критерії міцності, наприклад, критерій максимальних головних напружень [28]:

$$\sigma_g^*(x) = \frac{|\sigma(x)|}{2} + \frac{\sqrt{\sigma^2(x) + 4\tau^2(x)}}{2} \leq \sigma_g^{\max}, \quad (6)$$

де $x \in [0; L]$; $\sigma_g^*(x)$ — модуль першого головного напруження в клейовому шарі; σ_g^{\max} — границя міцності клею.

Руйнування з'єднання може відбуватися у формі розриву накладки. Тому необхідно ввести обмеження на максимальні напруження в накладці:

$$\sigma_p^* = \frac{N_1(x)}{\delta_1(x)} + \frac{\delta_1(x)}{2D_1(x)} M_1(x) \leq \sigma_p^{\max}, \quad (7)$$

де $x \in [0; L]$; $\sigma_p^*(x)$ — максимальні напруження у поперечному перерізі накладки; σ_p^{\max} — границя міцності матеріалу накладки.

На функцію $\delta_1(x)$ можуть бути накладені обмеження знизу і зверху. Товщина накладки не повинна бути меншою за деяке задане значення

$$\delta_1(x) \geq \delta_{\min}, \quad (8)$$

де δ_{\min} — деяка технологічно мінімально можлива товщина накладки.

Обмеження на максимальну товщину накладки може бути продиктовано міркуваннями аеродинамічної ефективності, можливостями контролю якості, технології виробництва тощо.

ГЕНЕТИЧНИЙ АЛГОРИТМ ОПТИМІЗАЦІЇ

Для розв'язання задачі запропоновано використовувати генетичний алгоритм. Для цього беремо за шукані змінні довжину з'єднання L і товщину накладки у вузлових точках $\delta_i^{(1)}$ і знаходимо такі їх оптимальні значення, які забезпечують мінімум маси накладки (5) під час виконання обмежень за міцністю (6) і (7). Однак, на відміну від задачі про знаходження оптимального розподілу матеріалу вздовж балки [11], якщо значення товщини $\delta_i^{(1)}$ в сусідніх точках значно відрізняються (що може статися внаслідок схрещування або мутацій у процесі виконання генетичного алгоритму), то напруження в клейовому шарі (3), які обчислено з використанням скінченно-різницевого методу, матимуть неправдоподібні стрибки. Тому шукатимемо оптимальну залежність $\delta_i^{(1)}$ серед гладких функцій. Це також впливає з інтуїтивних міркувань про те, що ймовірно шукана функція $\delta_1(x)$ є гладкою, не має розривів, кутових точок і стрибків. Шукаємо функцію $\delta_1(x)$ у вигляді кривої Безьє, яка має досить багато контрольних точок. У цьому випадку задача оптимізації зводиться до пошуку оптимальних координат опорних точок. Тобто форма і довжина накладки визначаються впорядкованим набором координат контрольних точок кривої Безьє $[(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)]$, де n — кількість контрольних точок. Тоді довжина з'єднання $L = x_n$.

Генетичний алгоритм потребує розв'язання прямої задачі зі знаходження напруженого стану з'єднання за відомої довжини з'єднання та функції $\delta_1(x)$. Для числового розв'язання системи (4) використовується прямий метод скінченних різниць.

Якщо функція $\delta_1(x)$ задана, то відомі і функції $s_1(x)$, $B_1(x)$ і $D_1(x)$. Для реалізації методу скінченних різниць розіб'ємо ділянку склеювання $x \in [0; L]$ на систему вузлових точок з номерами від 0 до N . Інтервал розбиття $h = L/N$. Переміщення несних шарів у точках позначимо через $U_1(x_i) = u_i^{(1)}$, $U_2(x_i) = u_i^{(2)}$ і $W_1(x_i) = w_i^{(1)}$.

Розв’язання системи (4) зводиться до знаходження величин $u_i^{(1)}$, $u_i^{(2)}$ і $w_i^{(1)}$. Наявність крайових умов дозволяє ввести як невідомі переміщення у зовнішніх вузлах за межами ділянки склеювання. Функція $\delta_1(x)$ та пов’язані з нею функції $s_1(x)$, $B_1(x)$ і $D_1(x)$ задано тільки на ділянці склеювання $x \in [0; L]$. Тобто задано лише $\delta_0^{(1)}, \dots, \delta_N^{(1)}$, $B_0^{(1)}, \dots, B_N^{(1)}$, $D_0^{(1)}, \dots, D_N^{(1)}$. Тому в крайніх точках x_0 і x_N похідні від переміщень у виразі (4) і в крайових умовах можна записати в різницевій формі за симетричним скінченно-різницеvim шаблоном. Однак для запису похідних від $s_1(x)$, $B_1(x)$ і $D_1(x)$ у різницевій формі в межових точках необхідно використовувати односторонні праві та ліві шаблони. Записавши у різницевій формі систему (4) для точок $0, 1, \dots, N$, а також крайові умови, отримаємо систему лінійних рівнянь відносно невідомих $u_{-1}^{(1)}, \dots, u_{N+1}^{(1)}$, $u_{-1}^{(2)}, \dots, u_{N+1}^{(2)}$, і $w_{-2}^{(1)}, \dots, w_{N+2}^{(1)}$, яка містить $3N + 11$ рівнянь. Розв’язавши отриману систему рівнянь, знайдемо переміщення несних шарів у вузлових точках. Це дозволяє знайти напруження у клейовому шарі (3) та інші силові чинники в з’єднанні.

Для реалізації генетичного алгоритму необхідно ввести фітнес-функцію, яка давала б можливість ранжувати за якістю різні набори шуканих параметрів $[(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)]$. Якщо за критерій оптимальності взяти масу конструкції (5) і накласти обмеження на максимальні перші головні напруження у клейовому шарі (6) і мінімальну товщину накладки (8), то фітнес-функцію можна подати у вигляді

$$\Phi = \int_0^L \delta_1(x) dx + \sum_j f_j, \quad (9)$$

де f_j — штрафні функції, які більші нуля, якщо обмеження порушені, і дорівнюють нулю, якщо обмеження виконується. Залежність штрафних функцій від відхилення відповідного параметра від обмеження може бути лінійною або ступеневою, наприклад

$$f_1 = \begin{cases} Z_1 (\max(\sigma_g^*) / \sigma_g^{\max} - 1)^2, & \max(\sigma_g^*) > \sigma_g^{\max}; \\ 0, & \max(\sigma_g^*) \leq \sigma_g^{\max}, \end{cases}$$

де Z_1 — деяке велике число. Аналогічним чином уводяться функції f_2 , f_3 тощо, які відповідають обмеженням (7), (8) та іншим обмеженням.

Генетичні алгоритми мають деякі недоліки, найістотнішим з яких є складність налаштування. За високої мінливості порушується збіжність і навіть знайдені прийнятні значення шуканих параметрів ризикують бути втраченими в результаті мутацій. За низької мінливості наблизений розв’язок знаходиться швидко, але потім уповільнюється збіжність і вироджується популяція (стабілізація біля деяких неоптимальних значень параметрів). Одним з можливих виходів з цієї суперечності є острівна модель еволюційного алгоритму. Згідно з цією моделлю загальна популяція розбивається на кілька ізольованих субпопуляцій (островів), і на кожному з островів еволюційний процес відбувається незалежно. Але із заданою періоди-

чністю найкращі особини з популяції можуть мігрувати з острова на острів. У роботі пропонується модель з трьома островами, на одному з яких ймовірність мутацій і дисперсія мутацій вищі ніж на двох інших островах. Таке поєднання двох відносно стабільних островів з одним островом, на якому рівень мутагенезу більш високий, дозволяє поєднувати швидкість пошуку розв'язків зі стабільністю і збереженням кращих розв'язків у загальній популяції. Робота еволюційного алгоритму на одному острові складається з таких етапів:

- 1) створення стартової популяції;
- 2) селекція — відбір найкращих особин для схрещування;
- 3) поділ відібраних особин на пари;
- 4) схрещування та мутації;
- 5) повернення нових особин у популяцію;
- 6) вимирання найгірших особин;
- 7) перевірка критерію зупинки. Якщо критерій виконується, то зупинка алгоритму, якщо ні — повернення до пункту 2.

Усі елементи алгоритму припускають різні способи реалізації. Наприклад, селекція може виконуватися у формі відбору заданої кількості кращих особин або у формі ймовірнісного відбору. Причому ймовірність може залежати як від значень цільової функції (9), так і від місця особи в ранжованому списку. Поділ особин на пари може бути абсолютно випадковим, а може залежати від ступеня близькості або відмінності особин. Такий підхід потребує введення міри близькості особин. У наведеній роботі з цією метою використовується сума квадратів відстаней між контрольними точками кривих Безье кожних двох особин. Тобто як критерія схожості особин з номерами i та j використовується сума

$$\Delta_{i,j} = \sum_{k=1}^n [(x_k^{(i)} - x_k^{(j)})^2 + (y_k^{(i)} - y_k^{(j)})^2].$$

Чим особини i та j більше схожі, тим величина $\Delta_{i,j}$ буде меншою.

У цій роботі використано стратегію аутбридингу, тобто чим більше особини відрізняються одна від одної, тим з більшою ймовірністю вони утворюють пару для схрещування. Схрещування реалізовано у формі випадкового обміну контрольних точок двома батьківськими особинами. Тобто контрольна точка з номером k нащадка особин i та j є результатом випадкового вибору з двох відповідних контрольних точок батьківських особин $(x_k^{(i)}, y_k^{(i)})$ і $(x_k^{(j)}, y_k^{(j)})$. Мутації в даному випадку полягають у зміні координат деяких контрольних точок нащадка на деяку випадкову величину, яка має нульове середнє значення. Кількість контрольних точок, що мутують, є також випадковою величиною. Вимирання найгірших особин, як і відбір особин для схрещування, також може бути реалізовано кількома способами. Критерієм зупинки може бути виконання заданої кількості ітерацій або досягнення популяцією заданого рівня гомогенності.

Наведений алгоритм застосовується для кожного з островів окремо. Загальний алгоритм оптимізації складається з таких етапів:

- 1) створення стартових популяцій на кожному з островів;
- 2) виконання K циклів еволюційного відбору всіх островах;

3) міграція з острова на острів деякої кількості найкращих особин;

4) перевірка критерію зупинки. Якщо критерій досягнуто — зупинка алгоритму. Якщо критерію не досягнуто — повернення до пункту 2.

Після зупинки роботи алгоритму необхідно виділити оптимальний розв'язок із всієї сукупності особин. Середнє ж значення параметрів у популяції є стійкішим до випадкових відхилень, ніж параметри окремої особини. Тому як розв'язок задачі оптимізації пропонується брати зрізну вибірккову середню параметрів кращих особин усіх (або однієї з трьох) популяцій. Для обчислення зрізного вибіркового середнього можна використовувати, наприклад, половину особин популяції.

ЧИСЛОВА РЕАЛІЗАЦІЯ ТА РЕЗУЛЬТАТИ

Розглянемо результати розв'язку запропонованого в роботі алгоритму топологічної оптимізації клейового з'єднання на конкретному прикладі. Розглянемо з'єднання, яке має такі параметри: $E_1 = 100$ ГПа, $E_2 = 70$ ГПа, $\delta_2 = 3$ мм, $\delta_0 = 0,1$ мм, $E_0 = 2,274$ ГПа, $G_0 = 0,54$ ГПа, $\sigma_g^{\max} = 30$ МПа, $\delta_{\min} = 0,5$ мм. З'єднання навантажено поздовжнім зусиллям $F = 300$ кН/м. Максимальні напруження в накладці обмежимо величиною $\sigma_p^{\max} = 115$ МПа.

Використовується 6 контрольних точок кривої Безьє. Розрахунки показали, що довжина оптимального з'єднання $L = 133,85$ мм. Графік зміни товщини накладки по довжині ділянки склеювання і положення контрольних точок кривої Безьє зображено на рис 2.

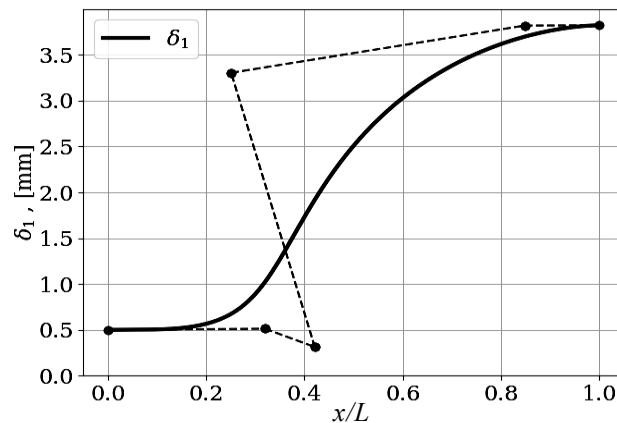


Рис. 2. Товщина накладки та контрольні точки кривої Безьє

Як бачимо, на лівому краю з'єднання товщина накладки дорівнює мінімально допустимій 0,5 мм. На рис. 3 показано напруження в з'єднанні: напруження в клейовому шарі — на рис. 3, а, (у безрозмірній формі); максимальні напруження в накладці та основі — на рис. 3, б.

Як бачимо, на обох кінцях з'єднання максимальні головні напруження дорівнюють максимально допустимим. Максимально допустимі напруження в накладці σ_p^{\max} позначено на рис. 3, б чорною пунктирною лінією. Обмеження (7) також виконується як рівності.

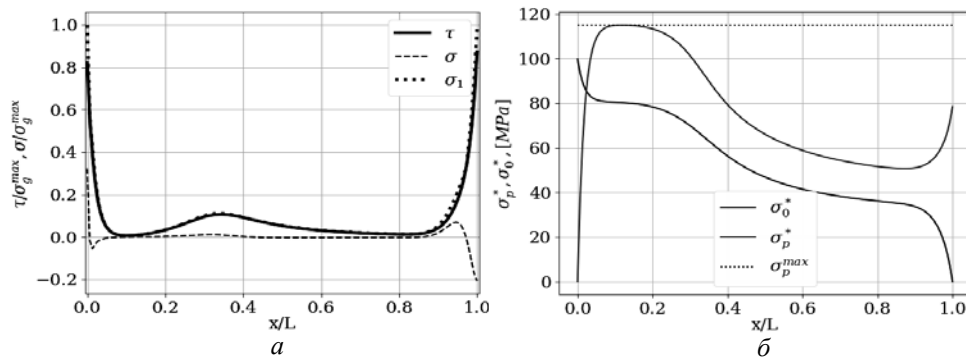


Рис. 3. Напруження в клейовому шарі

Для верифікації запропонованої моделі виконано розрахунок напруженого стану з'єднання за допомогою методу скінченних елементів. Використано двовимірну скінченно-елементну модель. Клейовий шар розбивався на елементи з максимальним розміром $0,1\delta_0$. На рис. 4 зображено графіки компонентів напружень (3) у клейовому шарі в околі кінців з'єднання (тобто в найбільш напружених ділянках), які обчислені за допомогою запропонованої моделі та скінченно-елементного моделювання (наведено напруження в серединній площині клейового шару).

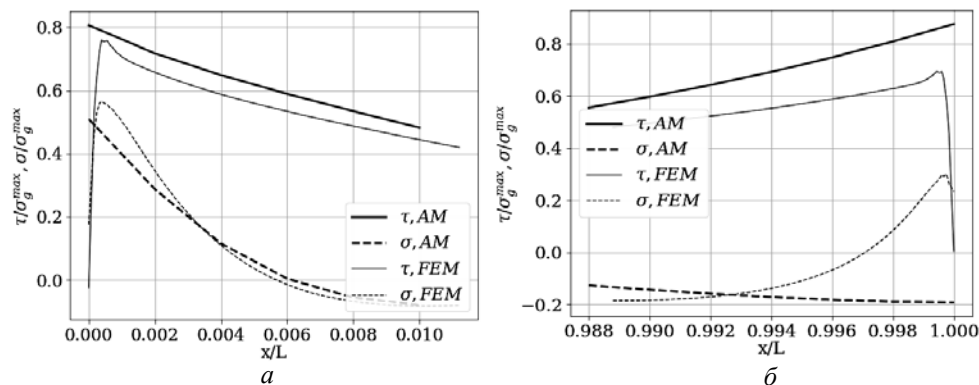


Рис. 4. Напруження в клейовому шарі на кінцях з'єднання: АМ — аналітична модель; FEM — скінченно-елементна модель

ВИСНОВКИ

Результати розв'язання низки задач та аналізу результатів:

1. Залежність довжини та форми накладки від навантаження має нелінійний характер.
2. Наявність у задачі оптимізації обмеження на мінімально допустиму товщину накладки призводить до того, що знайдена оптимальна форма містить на ненавантаженому краю накладки горизонтальну площину мінімально допустимої товщини. Таке конструктивне рішення, наскільки відомо авторам роботи, раніше ніким не пропонувалося.
3. Домогтися рівномірного розподілу напружень у з'єднанні за заданих умов задачі неможливо. Ключовим обмеженням, мабуть, є стала по довжині з'єднання товщина основної пластини. Як наслідок несна здатність з'єднання обмежена.

4. Використання кривих Безье порівняно з розвиненням профілю накладки у тригонометричний ряд має ряд переваг — менша кількість шуканих параметрів, можливість автоматично враховувати деякі відомі ефекти (наприклад, горизонтальність дотичної до графіка товщини на початку координат).

Запропонований підхід може бути розвинений та узагальнений у таких напрямках:

1. Застосування запропонованого генетичного алгоритму оптимізації для розв'язання задач топологічної оптимізації з'єднань коаксіальних циліндричних труб та з'єднань з коловою симетрією [9; 10].

2. Розвиток та ускладнення острівної моделі генетичного алгоритму. Використання, наприклад, різних цільових функцій на кожному з островів, і навіть комбінацій генетичних алгоритмів з іншими сучасними методами оптимізації.

3. Оптимізація з'єднань з урахуванням температурних та технологічних напружень у конструкції [29].

ЛІТЕРАТУРА

1. L.F.M. da Silva, P.J.C. Das Neves, R.D. Adams, and J.K. Spelt, "Analytical models of adhesively bonded joints. Part I: Literature survey," *International Journal of Adhesion and Adhesives*, vol. 29, pp. 319–330, 2009. Available: <https://doi.org/10.1016/j.ijadhadh.2008.06.005>
2. M. Shishesaz and M. Hosseini, "Effects of joint geometry and material on stress distribution, strength and failure of bonded composite joints: an overview," *The Journal of Adhesion*, vol. 96, iss. 12, pp. 1053–1121, 2018. doi: 10.1080/00218464.2018.1554483.
3. K.P. Barakhov and I.M. Taranenko, "Influence of Joint Edge Shape on Stress Distribution in Adhesive Film," in *Nechyporuk M., Pavlikov V., Kritskiy D. (eds) Integrated Computer Technologies in Mechanical Engineering – 2021. ICTM 2021. Lecture Notes in Networks and Systems*, vol 367, pp. 123–132. Springer, Cham, 2022. Available: https://doi.org/10.1007/978-3-030-94259-5_12
4. R. Haghani, M. Al-Emrani, and R. Kliger, "Effect of Laminate Tapering on Strain Distribution in Adhesive Joints: Experimental Investigation," *Journal of Reinforced Plastics and Composites*, vol. 29, iss. 7, pp. 972–985, 2009. doi: 10.1177/0731684408102698.
5. F. Ramezani, M.R. Ayatollahi, A. Akhavan-Safar, and L.F.M. Da Silva, "A comprehensive experimental study on bi-adhesive single lap joints using DIC technique," *International Journal of Adhesion and Adhesives*, vol. 120, 102674, 2020. Available: <https://doi.org/10.1016/j.ijadhadh.2020.102674>
6. X. Shang, E.A.S. Marques, J.J.M. Machado, R.J.C. Carbas, D. Jiang, and L.F.M. da Silva, "Review on techniques to improve the strength of adhesive joints with composite adherends," *Composites Part B: Engineering*, vol. 177, 107363, 2019. doi: 10.1016/j.compositesb.2019.107363.
7. J. Kupski and S. Teixeira de Freitas, "Design of adhesively bonded lap joints with laminated CFRP adherends: Review, challenges and new opportunities for aerospace structures," *Composite Structures*, vol. 268, 113923, 2021. Available: <https://doi.org/10.1016/j.compstruct.2021.113923>
8. S. Amidi and J. Wang "An analytical model for interfacial stresses in double-lap bonded joints," *The Journal of Adhesion*, vol. 95, iss. 11, pp. 1031–1055, 2019. Available: <https://doi.org/10.1080/00218464.2018.1464917>
9. S. Kurennov and N. Smetankina, "Stress-Strain State of a Double Lap Joint of Circular Form. Axisymmetric Model," in *Nechyporuk M., Pavlikov V., Kritskiy D. (eds) Integrated Computer Technologies in Mechanical Engineering – 2021. ICTM 2021*.

- Lecture Notes in Networks and Systems*, vol 367, pp. 36–46. Springer, Cham, 2022. Available: https://doi.org/10.1007/978-3-030-94259-5_4
10. S. Kurennov, K. Barakhov, D. Dvoretzka, and O. Poliakov, “Stress State of Two Glued Coaxial Tubes Under Nonuniform Axial Load,” in *Nechyporuk M., Pavlikov V., Kritskiy D. (eds) Integrated Computer Technologies in Mechanical Engineering – 2020. ICTM 2020. Lecture Notes in Networks and Systems*, vol. 188, pp. 389–400. Springer, Cham, 2021. Available: https://doi.org/10.1007/978-3-030-66717-7_33
 11. S. Kurennov, K. Barakhov, I. Taranenko, and V. Stepanenko, “A genetic algorithm of optimal design of beam at restricted sagging,” *Radioelectronic and Computer Systems*, no. 2, pp. 83–91, 2022. Available: <https://doi.org/10.32620/reks.2022.1.06>
 12. Mehdi Veisytabar, Arash Reza, and Younes Shekari, “Stress analysis of adhesively-bonded single stepped-lap joints based on three-parameter fractional viscoelastic foundation model,” *Proceedings of the Institution of Mechanical Engineers, Part L: Journal of Materials: Design and Applications*, no. 5, pp. 933–949, 2022. Available: <https://doi.org/10.1177/14644207211062497>
 13. H.L. Groth and P. Nordlund, “Shape optimization of bonded joints,” *International Journal of Adhesion and Adhesives*, vol. 11, no. 4, pp. 204–212, 1991. doi: 10.1016/0143-7496(91)90002-y.
 14. H.C. Ayaz Ümütlü and Z. Kiral, “Airfoil shape optimization using Bézier curve and genetic algorithm,” *Aviation*, vol. 26(1), pp. 32–40, 2022. Available: <https://doi.org/10.3846/aviation.2022.16471>
 15. P. Zou, J. Bricker, and W. Uijttewaal, “Optimization of submerged floating tunnel cross section based on parametric Bézier curves and hybrid backpropagation – genetic algorithm,” *Marine Structures*, vol. 74, 102807, 2020. doi: 10.1016/j.marstruc.2020.102807.
 16. S.S. Kurennov, Topological optimization of a symmetric single-lap adhesive joint, *System Research & Information Technologies*, no. 2, pp. 75–84, 2022. Available: <https://doi.org/10.20535/SRIT.2308-8893.2022.2.05>
 17. S. Kurennov, K. Barakhov, and O. Vambol, “Topological optimization of a symmetrical adhesive joint. Island model of genetic algorithm,” *Radioelectronic and Computer Systems*, no. 3, pp. 67–83, 2022. doi: 10.32620/reks.2022.3.05.
 18. A.R. Rispler, L. Tong, G.P. Steven, and M. R. Wisnom, “Shape optimisation of adhesive fillets,” *International Journal of Adhesion and Adhesives*, 20(3), pp. 221–231, 2000. doi: 10.1016/s0143-7496(99)00047-0.
 19. G. Belingardi, L. Goglio, and A. Tarditi, “On the Optimization of Single Lap Metal/Plastics Adhesive Joints,” *Key Engineering Materials*, vol. 221–222, pp. 161–172, 2001. doi: 10.4028/www.scientific.net/kem.221-222.161.
 20. V.S. Symonov, I.S. Karpov, and J. Juračka, “Optimization of a Panelled Smooth Composite Shell with a Closed Cross-Sectional Contour by Using a Genetic Algorithm,” *Mechanics of Composite Materials*, 49(5), pp. 563–570, 2013. Available: <https://doi.org/10.1007/s11029-013-9372-0>
 21. M. HassanVand, H. Abbaszadeh, and M. Shishesaz, “Optimization of adhesive single-lap joints under bending moment,” *The Journal of Adhesion*, vol. 98, iss. 11, pp. 1687–1712, 2021. Available: <https://doi.org/10.1080/00218464.2021.1932485>
 22. V. Sineglazov, K. Riazanovskiy, and O. Chumachenko, “Multicriteria conditional optimization based on genetic algorithms,” *System Research & Information Technologies*, no. 3, pp. 89–104, 2020. doi: 10.20535/SRIT.2308-8893.2020.3.07.
 23. A.A. Gozali and S. Fujimura, “DM-LIMGA: Dual Migration Localized Island Model Genetic Algorithm—a better diversity preserver island model,” *Evol. Intel.*, 12, pp. 527–539, 2019. Available: <https://doi.org/10.1007/s12065-019-00253-2>
 24. J.M. Palomo-Romero, L. Salas-Morera, and L. García-Hernández, “An island model genetic algorithm for unequal area facility layout problems,” *Expert Systems with Applications*, 68, pp. 151–162, 2017. Available: <https://doi.org/10.1016/j.eswa.2016.10.004>
 25. X. Sun, P. Chou, C.-C. Wu, and L.-R. Chen, “Quality-Oriented Study on Mapping Island Model Genetic Algorithm onto CUDA GPU,” *Symmetry*, 11(3), 318, 2019. doi: 10.3390/sym11030318

26. Y. Niu, X. Xu, and S. Guo, "Structural Optimization Design of a Typical Adhesive Bonded Honeycomb-Core Sandwich T-joint in Side Bending Using Multi-Island Genetic Algorithm," *Applied Composite Materials*, 2021. doi: 10.1007/s10443-021-09882-2
27. Lucas Angelo da Silveira, Thaynara Arielly de Lima, and Mauricio Ayala-Rincón, *Reconfigurable Hetero-geneous Parallel Island Models*. 2022. Available: <https://doi.org/10.48550/arXiv.2205.02916>
28. S.S. Kurennov, "Refined Mathematical Model of the Stress State of Adhesive Lap Joint: Experimental Determination of the Adhesive Layer Strength Criterion," *Strength of Mater.*, 52, pp. 779–789, 2020. Available: <https://doi.org/10.1007/s11223-020-00231-5>
29. P. Tsokanas, T. Loutas, G. Kotsinis, V. Kostopoulos, W. M. van der Brink, and F. Martin de la Escalera, "On the fracture toughness of metal-composite adhesive joints with bending-extension coupling and residual thermal stresses effect," *Composites Part B: Engineering*, vol. 185, 107694, 2019. doi: 10.1016/j.compositesb.2019.107694

Надійшла 29.11.2022

INFORMATION ON THE ARTICLE

Sergey S. Kurennov, ORCID: 0000-0002-3835-3288, National Aerospace University "Kharkiv Aviation Institute", Ukraine, e-mail: kurennov.ss@gmail.com

Kostiantyn P. Barakhov, ORCID: 0000-0003-1714-7917, National Aerospace University "Kharkiv Aviation Institute", Ukraine, e-mail: kpbarakhov@gmail.com

Hanna S. Barakhova, ORCID: 0000-0001-5209-836X, National Aerospace University "Kharkiv Aviation Institute", Ukraine, e-mail: h.s.barakhova@khai.edu

APPLICATION OF BEZIER CURVES FOR DESCRIPTION OF STRUCTURE SHAPE IN OPTIMIZATION OF ADHESIVE JOINTS / S.S. Kurennov, K.P. Barakhov, H.S. Barakhova

Abstract. The problem of topological optimization of a symmetrical double-shear adhesive joint has been solved. The suggested mathematical model of a joint with variable thickness generalizes the classic Holland–Reissner model. The shape of the doubler is described by means of the Bezier curve. Seeking parameters in the optimization problem are coordinates of reference points of the Bezier curve. Both joint length and doubler cross-section area can be considered an objective function. The restriction is applied on stress in adhesive film and doubler. The direct problem of finding the joint stress state at given geometric parameters was solved using the finite difference method. The genetic algorithm was used to solve the optimization problem. In order to improve the convergence of the genetic algorithm, the island model of evolution is suggested, which ensures quick evolution selection and stability of obtained results. The model problem is solved.

Keywords: three layer construction, topological optimization, genetic algorithm.

ESTIMATION OF THE PARAMETERS OF GENERALIZED LINEAR MODELS IN THE ANALYSIS OF ACTUARIAL RISKS

R.S. PANIBRATOV, P.I. BIDYUK

Abstract. Methods of estimating the parameters of generalized linear models for the case of paying insurance premiums to clients are considered. The iterative-recursive weighted least squares method, the Adam optimization algorithm, and the Monte Carlo method for Markov chains were implemented. Insurance indicators and the target variable were randomly generated due to the problem of public access to insurance data. For the latter, the normal and exponential law of distribution and the Pareto distribution with the corresponding link functions were used. Based on the quality metrics of model learning, conclusions were made regarding their construction quality.

Keywords: actuarial risk, generalized linear models, simulation modeling, exponential family of distributions, iterative-recursive weighted least squares method, Adam method, Monte Carlo method for Markov chains.

INTRODUCTION

Actuarial risk is classified as the risk that the assumptions implemented by actuaries in the model for estimating the prices of insurance policies may be inaccurate or incorrect. This term is also identified as “insurance risk”. The level of actuarial risk is directly proportional to the reliability of the assumptions implemented in the pricing models used by insurance companies to set premiums. Probability estimates are used to set the price of insurance policies, allowing insurers to implement payments subject to normal business operations. If the proposed hypotheses are wrong, the events that were not considered become the reasons for increasing the frequency of payments, which, in turn, leads to serious financial consequences for the insurer.

Monograph [1] provides useful methodology for system analysis of complex processes. The methodology was applied by the authors to analysis of actuarial processes. In studies [2; 3], mathematical modeling issues related to complicated non-stationary processes and systems are discussed.

Generalized Linear Models (GLM) allow making explicit assumptions about the nature of the insurance data and their relationship with the predicted variables. Furthermore, GLM provides statistical diagnostics that helps in selecting only significant variables and testing model assumptions. This approach is widely

recognized as a standard way of pricing for different types of insurance in different markets of countries.

The GLM consists of a wide variety of models, including the linear regression model as a special case. Assumptions for the latter, usually including normal distribution, constant variance, and additivity of effects, are rejected. For example, the target variable can be taken from an exponential family of distributions [4].

The exponential family of distributions has the following general form:

$$f_i(y_i, \theta_i, \varphi) = \exp \left\{ \frac{y_i \theta_i - b(\theta_i)}{a_i(\varphi)} + c(y_i, \varphi) \right\},$$

where $a_i(\varphi)$, $b(\theta)$, and $c(y_i, \varphi)$ are functions, that are defined at the beginning; θ_i is a parameter related to the mean value; φ is a scale parameter related to variance.

The variance can vary together with the mean of the distribution. The influence of explanatory variables is assumed to be additive on an another scale. The following assumptions are made for GLM:

Stochastic component: each component of Y is independent and is taken from the one distribution of exponential family.

Systematic component: p covariates (explanatory variables) form linear predictor η :

$$\eta = X\beta.$$

Link function: the relationship between the random and systematic components is established through a link function that is differentiable and monotonic:

$$E[Y] = \mu = g^{-1}(\eta).$$

PROBLEM STATEMENT

The purpose of the study is to implement methods for estimating GLM parameters for the analysis of actuarial risks, using different distribution laws and link functions for the predicted variable.

METHODS OF ESTIMATING PARAMETERS OF GENERALIZED LINEAR MODELS

GLM parameter estimation is a rather important issue and deserves appropriate attention. The following algorithms were used to evaluate parameters for the purpose of comparing methods: iterative-recursive weighted least squares method; Adam optimization algorithm, Monte Carlo method for Markov chains.

Iterative-recursive weighted least squares method

The heteroskedastic model can be adjusted using the weighted least squares method (WLS),

$$\tilde{\beta} = (X^T W X)^{-1} X^T W y,$$

where y is a target variable; $W = \text{Diag} \left[\text{Var}(y_i) \left(\frac{\partial \eta_i}{\partial \mu_i} \right)^2 \right]^{-1}$ is a diagonal matrix of weights; X is a matrix of covariates.

We will use Fisher's scoring [5]:

$$\beta^{(t+1)} = (X^T W X)^{-1} (X^T W \beta^{(t)} + X^T A (y - \mu)),$$

where $A = \text{Diag} \left[\text{Var}(y_i) \left(\frac{\partial \eta_i}{\partial \mu_i} \right) \right]^{-1}$.

The matrices A and W are related in a next way

$$A = W \left(\frac{\partial \eta}{\partial \mu} \right) = W \text{Diag} \left(\frac{\partial \eta_i}{\partial \mu_i} \right).$$

Then equation for estimating parameters of GLM can be rewritten as follows:

$$\beta^{(t+1)} = (X^T W X)^{-1} (X^T W \beta^{(t)} + X^T W z),$$

where $z = \eta + \left(\frac{\partial \eta}{\partial \mu} \right) (y - \mu)$.

At each step of algorithm:

1. The current estimate of β is used to calculate a new working variable z and a set of weights W .
2. Update (regress) z using X and W to obtain new values.

Iteratively reweighted least squares with random effects (IRWLSR), which was proposed in [6], successfully overcomes the challenge posed by high dimensional intractable integrals in fitting GLMMs for non-normal data. IRWLSR is used for Maximum Likelihood estimations of generalized linear mixed effects models (GLMMs). The benefit is that a working linear mixed effects model (WWLMM) for normal data performs the bulk of the calculation and prediction. Only the working responses and weights in the WWLMM are updated using the GLMM distribution and link function. The complete algorithm can be applied with ease even if high-dimensional intractable integrals are present in the likelihood function because it does not require any numerical evaluations of intractable integrals.

Adam (Adaptive moment estimation)

The adaptive movement estimation algorithm or Adam is an extension of the gradient optimization algorithm. It was created to speed up the optimization process to improve the ability of the optimization algorithm. This is achieved by increasing the step size for each input parameter being optimized. Each step size is automatically adapted through a search process based on partial derivatives or gradients for each input variable. This involves computing the first and second moments of the gradient as an exponentially decreasing first moment and second moment for the input variables.

First, the moment vector and exponentially weighted norm at infinity are adjusted for each optimization parameter, which is denoted as m and v . Their initial values are equal to zero.

The algorithm consists of the following steps [7]:

The gradient of the target function for the current step is calculated:

$$g(k) = f(x(k-1)).$$

1. Next, the first moment is updated using the gradient and the hyperparameter β_1 :

$$m(k) = \beta_1 m(k-1) + (1-\beta_1)g(k).$$

2. The second moment is then updated using the gradient square and the hyperparameter β_2 :

$$v(k) = \beta_2 * v(k-1) + (1-\beta_2) * g^2(k).$$

The last two values are biased because their initial values are equal to zero.

3. The first and second moments are adjusted according to the following formulas:

$$\tilde{m}(k) = \frac{m(k)}{1-\beta_1^k}; \quad \tilde{v}(k) = \frac{v(k)}{1-\beta_2^k}.$$

4. The optimum point is calculated by the expression:

$$x(k) = x(k-1) - \alpha * \frac{\tilde{m}(k)}{\sqrt{\tilde{v}(k) + \varepsilon}},$$

where α is the hyperparameter of step size; ε is a small value that ensures there will not be zero division error.

According to [8], the Adam optimization algorithm combining adaptive coefficients and composite gradients based on randomized block coordinate descent is proposed, which improves the algorithm's performance. Starting with the improvement of the Adam algorithm to accelerate the convergence speed, accelerate the search for the global optimal solution, and enhance the high-dimensional data processing ability. The suggested approach improves the Adam optimization algorithm's performance to some degree, but it ignores the effects of second-order momentum and various learning rates on the performance of the original algorithm.

Quasi-Hyperbolic Momentum (QHM) and Quasi-Hyperbolic Adam (QHAdam) are computationally affordable, easily understood, and straightforward to use. They were implemented in [9] by the authors. In many situations, they can be great substitutes for momentum and the Adam algorithm. They specifically make use of high exponential discount factors possible by employing rapid discounting.

As objective functions, the log-likelihood functions of the distributions of the predicted variables were used with their link functions.

Bayesian approach for parameter estimation

The flexibility of the Bayesian technique to adapt, or the capacity to estimate model parameters iteratively in the situation where new measurements are con-

tinually coming in, is one of its benefits. The Monte Carlo technique for Markov chains (MCMC) may be used to estimate the model's parameters to the data at a particular step i , and in general, the process can be expressed as follows:

$$\tilde{\theta}_i = \tilde{\theta}_{i-1} + \alpha \Delta \tilde{\theta}(i),$$

where $\tilde{\theta}_i$ is an estimation of some parameter θ of the model at step i ; α is a weight factor; $\Delta \tilde{\theta}(i)$ is the increment in the value of the estimate to be evaluated.

The Markov condition for representing stochastic processes using Markov chains is satisfied since the estimate of the parameter at the subsequent stage solely depends on the estimate at the preceding phase. The Monte Carlo approach is used to generate data from the suitable distribution of the parameter, which is required in order to estimate the rise in the estimate. Therefore, the Monte Carlo approach for Markov chains gets its broad name.

This architecture has concerns with re-estimating the parameter distribution in response to fresh observations and producing data from the proper distribution for estimating unknown parameters. The MCMC method may be used to answer these queries.

Typically, the Bayesian method for creating a regression model looks like follows:

$$y_k \sim N(a_1 x_{k1} + \dots + a_n x_{kn}, \sigma), \quad k=1, \dots, m,$$

where N is the Normal distribution with appropriate parameters.

In general, the distribution type can be chosen based on how the data are actually distributed. It can be rewritten in the matrix form:

$$y \sim N(X\alpha, \sigma_e I), \alpha \sim N(\alpha_0, K_{\alpha,0}),$$

where α_0 is a vector of mean values at zero step; $K_{\alpha,0}$ is the corresponding covariance matrix.

The zero step refers to the initial estimation that was made and will be improved as new data and slow adaptation become available.

Ordinary Least Squares (OLS) can be used when the measured factors and the dependent variable are both normally distributed random variables. The convenience of MCMC, however, comes from the ability to update the values of a few distribution parameters without performing a full recalculation of all matrix operations. This is made possible by the so-called property of conjugate prior distributions from the family of normal, gamma, and their inverse distributions.

The regression model has the following form:

$$y = a + bx + \varepsilon.$$

Primary estimation of vector α_0 can be written in the next form:

$$\alpha_0 = \begin{bmatrix} a_0 \\ b_0 \end{bmatrix}$$

and covariance matrix:

$$K_{\alpha,0} = \begin{bmatrix} \sigma_{a,0}^2 & 0 \\ 0 & \sigma_{b,0}^2 \end{bmatrix}.$$

Applying the aforementioned property to each new pair of observations, it is possible to arrive at a formula for the updated values (t_1, s_1) , (t_2, s_2) of the distribution parameters:

$$\alpha_1 = K_{\alpha,1}(K_{\alpha,0}^{-1}\alpha_0 + K_y^{-1}\mu_y);$$

$$K_{\alpha,1} = (K_{\alpha,0}^{-1} + K_y^{-1})^{-1};$$

$$\mu_y = \left[\frac{t_1 s_2 - t_2 s_1}{t_1 - t_2}, \frac{s_1 - s_2}{t_1 - t_2} \right]^T;$$

$$K_y = \begin{bmatrix} \left[\left(\frac{t_1}{t_1 - t_2} \right)^2 + \left(\frac{t_2}{t_1 - t_2} \right)^2 \right] \sigma_e^2 & 0 \\ 0 & 2 \left(\frac{1}{t_1 - t_2} \right)^2 \sigma_e^2 \end{bmatrix}.$$

However, in the majority cases, when working with non-stationary processes, there is a case of probabilistic uncertainty, or uncertainty about the distribution, which means it is impossible to guarantee an unchanged normal distribution of variables and parameters, making it impossible to always use the above property. It is also theoretically incorrect to continue using OLS because the conditions required for its use are broken. Even in situations when the kind of distribution sought is unknown, MCMC offers a method for approximating the unknown distribution by producing data for the posterior distribution.

Let's say that an unknown parameter's distribution has to be calculated. To do this, it is recommended that a large number of points be produced from this distribution, from which the required distributional parameters may then be estimated using the law of large numbers. Nevertheless, as the posterior distribution of the parameters is unknown, a tool for such creation is required. The Metropolis-Hastings method is a potent tool for resolving the given issue. The algorithm's primary steps may be summarized as follows:

1. The method begins by working with an initial value α_0 , which can initially be any random integer.
2. The next step is to choose a new value for α^* that is suggested as being produced from the chosen distribution. Some propositional function $g(\alpha^*, \alpha_0)$ serves as the foundation for the production of α^* . The function g is often selected to be Gaussian and must be symmetric. The Bayes theorem is used to account for the effect of the newly acquired observations in this process.
3. Then calculations are made to determine the resulting value's acceptance rate β :

$$\beta = \frac{g(y | \alpha^*) g(\alpha^*)}{g(y | \alpha_0) g(\alpha_0)}.$$

4. If $u \leq \beta$, or the newly created point α^* , is accepted and becomes α_1 , a random number $0 \leq u \leq 1$ is chosen; if not, the value is rejected, and the procedure stays at the previous point.

5. Until there are sufficient observations, steps 2, 3, 4 are repeated.

The Metropolis-Hastings method asserts that the remaining points originate from the posterior distribution $f(\alpha|y, X)$ after certain initial values are discarded since the values created in the first stages are not yet stable.

The computation of new posterior distributions is repeated using the considered technique when new data is introduced since the prior posterior distributions, according to the Bayes scheme, become a priori for the following iteration.

In [10], authors analyzed insurance claim data and represented the total claims during a specific time frame as a random sum of individual claims' individual positive random variables. According to the data, the variances of both the random total and the random claim size are as big as cubic powers of their respective means. They used natural exponential family modeling to match distributions with cubic variance functions to the insurance data. The authors took into account three discrete counting factors for the random sum and three positive continuous distributions for the claim size, both of which were derived from natural exponential families. In order to create sampling algorithms, they provided a comprehensive study of the nontrivial discrete counting variables. They ran Monte Carlo simulations to compute tail probabilities, particularly for big losses, using samples from the overall claim distribution. Two methods increased these models' effectiveness. The first is that convolutions belong to the same family as the individual distribution because the claim size distributions fulfill the reproducibility condition. The use of importance sampling is the second enhancement.

In [11], authors presented a variety of machine learning regression approaches ranging from multivariate adaptive regression splines and kernel regression to ordinary and generalized least-squares regression variants over GLM and generalized additive model (GAM) approaches for high-dimensional variable selection applications, such as the calibration step in the least-squares Monte Carlo (LSMC) framework. Regression algorithms proved to be suitable machine learning methods for proxy modeling of life insurance companies in their slightly disguised real-world example and given LSMC setting, with potential for both performance and computational efficiency gains by fine-tuning model hyperparameters and implementation designs. After all, none of the author's tried and true methods could entirely remove the bias seen in the validation figures and produce findings that were consistent throughout the validation sets. The range of suggested regression techniques can be further reduced by looking at whether these findings are systematic for the approaches, the consequence of the Monte Carlo error, or a mix of the two. Although such assessments would be quite expensive computationally, they would be extremely helpful in revealing new ways to improve the quality of approximations.

NUMERICAL EXPERIMENT

Insurance data is not always publicly available, so it was decided to generate insurance indicators and target variables randomly. The data consisted of the following variables:

- age (range from 18 to 64 years);
- sex;
- body mass index (normal distribution was used);

- number of children (range from 0 to 5);
- smoker status;
- region (generated from sample view [‘north’, ‘south’, ‘east’, ‘west’, ‘center’]);
- charges.

The last variable is the target and the following distribution laws with the corresponding link functions were used for it:

- normal distribution with a known variance σ and a logarithmic link function;
- exponential distribution with the identity link function;
- pareto distribution with a known scale parameter x_m and a link function of the form $f(x) = -1 - x$.

Gaussian noise with variable variance, which is a linear function, was added to the predicted variable.

RESULTS OF THE EXPERIMENTS

The following metrics were used to estimate the quality of the models:

Mean squared error:

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \widehat{Y}_i)^2.$$

Root mean squared error:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - \widehat{Y}_i)^2}.$$

Mean absolute error:

$$MAE = \frac{1}{n} \sum_{i=1}^n |Y_i - \widehat{Y}_i|.$$

The results of the estimation of GLM parameters using three methods are presented in Tables 1, 2, 3.

Table 1. Results of GLM construction for a target variable with a Gaussian distribution with known variance and a logarithmic link function

Metric	MCMC	ADAM	IRWLS
MSE	4410.78	686.65	2458.21
RMSE	66.41	26.20	49.58
MAE	51.88	25.81	49.58

Table 2. Results of GLM construction for the target variable with a Pareto distribution with a known scale parameter and a negative linear link function

Metric	MCMC	ADAM	IRWLS
MSE	52725.56	82188.07	638121.57
RMSE	229.62	286.69	798.83
MAE	154.02	205.31	773.33

Table 3. Results of GLM construction for the target variable with an exponential distribution and an identity link function

Metric	MCMC	ADAM	IRWLS
MSE	213.52	148.95	288.53
RMSE	14.61	12.20	16.98
MAE	11.17	3.64	15.66

According to the results of GLM construction for three cases, it can be seen that in most cases, the Adam method showed quite good results. The MCMC method also showed good results for the case of the Pareto distribution. The final choice of parameter estimation method is made after practical application of models to solve the problem of predicting possible losses.

CONCLUSIONS

The methods of estimating GLM parameters for the analysis of actuarial risks in case of payment of insurance premiums to clients are considered. The following three methods are implemented: the iterative-recursive weighted least squares method, the Adam optimization algorithm, and the Monte Carlo method for Markov chains. Since insurance data are often not publicly available, the insurance indicators and the target variable were randomly generated: age, gender, body mass index, number of children, smoking status, region, and charges. The latter was generated using: a normal distribution with a known variance and a logarithmic link function; exponential distribution with the identity link function; of the Pareto distribution with a known scale parameter and a link function of the form of a negative linear function. According to the results of the experiments, it can be concluded that the Adam optimization algorithm demonstrated good results in most cases. The Monte Carlo method for Markov chains also showed good results for the case of the Pareto distribution.

REFERENCES

1. M.Z. Zgurovsky and N.D. Pankratova, *System Analysis: Problems, Methodology, Applications*. Kyiv: Naukova Dumka, 2011, 727 p.
2. O.G. Nakonechny and S.V. Demydenko, "Minimax mean square estimates of trend in the regression problems," *Tavria Visnyk of Informatics and Mathematics*, no. 1, pp. 23–30, 2009.
3. O.G. Nakonechny, M.M. Veres, and B.V. Stashchuk, "Guaranteed estimates of linear equations solutions in Hilbert space with random parameters," *Journal of Computational and Applied Mathematics*, vol. 104, no. 1, pp. 85–96, 2011.
4. D. Anderson et al., *A Practitioner's Guide to Generalized Linear Models – a foundation for theory, interpretation and application*; 3rd edition. Towers Watson, 2007, 122 p.
5. P. McCullagh and J. Nelder, *Generalized Linear Models*; 2nd edition. Chapman & Hall, 1989, 532 p.
6. T. Zhang, "Iteratively reweighted least squares with random effects for maximum likelihood in generalized linear mixed effects models," *Journal of Statistical Computation and Simulation*, vol. 91, no. 16, pp. 3404–3425, 2021. doi: 10.1080/00949655.2021.1928127.

7. D. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014. doi: 10.48550/arXiv.1412.6980.
8. M. Liu et al., "An Improved Adam Optimization Algorithm Combining Adaptive Coefficients and Composite Gradients Based on Randomized Block Coordinate Descent," *Computational Intelligence and Neuroscience*, vol. 2023, 14 p., 2023. doi: 10.1155/2023/4765891.
9. J. Ma and D. Yarats, "Quasi-hyperbolic momentum and Adam for deep learning," *arXiv preprint arXiv:1810.06801*, 2018. doi: 10.48550/arXiv.1810.06801.
10. S. Bar-Lev and A. Ridder, "Monte Carlo methods for insurance risk computation," *arXiv preprint arXiv:1807.02201*, 2018. doi: 10.48550/arXiv.1807.02201.
11. A.S. Krah, Z. Nikolić, and R. Korn, "Machine learning in least-squares Monte Carlo proxy modeling of life insurance companies," *Risks*, vol. 8, no. 1, 21 p., 2020. doi: 10.3390/risks8010021.

Received 26.01.2023

INFORMATION ON THE ARTICLE

Roman S. Panibratov, ORCID: 0000-0002-8604-4420, Educational and Research Institute for Applied System Analysis of the National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine, e-mail: roman.panibratov@gmail.com

Petro I. Bidyuk, ORCID: 0000-0002-7421-3565, Educational and Research Institute for Applied System Analysis of the National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine, e-mail: pbidyuke_00@ukr.net

ОЦІНЮВАННЯ ПАРАМЕТРІВ УЗАГАЛЬНЕНИХ ЛІНІЙНИХ МОДЕЛЕЙ В АНАЛІЗІ АКТУАРНИХ РИЗИКІВ / П.І. Бідюк, Р.С. Панібратов

Анотація. Розглянуто методи оцінювання параметрів узагальнених лінійних моделей для аналізу актуарних ризиків у випадку виплат страхових премій клієнтам. Було реалізовано ітеративно-рекурентно зважуваний метод найменших квадратів, алгоритм оптимізації Adam та метод Монте-Карло для ланцюгів Маркова. Страхові показники та цільова змінна генерувалися випадковим чином у зв'язку з проблемою публічного доступу страхових даних. Для останньої використовувався нормальний та експоненціальний закон розподілу і розподіл Парето з відповідними функціями зв'язку. На основі метрик якості навчання моделей були зроблені висновки щодо їх якості побудови.

Ключові слова: актуарний ризик, узагальнені лінійні моделі, імітаційне моделювання, експоненційна множина розподілів, ітеративно-рекурентно зважуваний метод найменших квадратів, метод Adam, метод Монте-Карло для марківських ланцюгів.

ВІДОМОСТІ ПРО АВТОРІВ

Барахов Костянтин Петрович,

кандидат технічних наук, доцент Національного аерокосмічного університету імені М.Є. Жуковського «Харківський авіаційний інститут», Україна, Харків

Барахова Ганна Сергіївна,

аспірант Національного аерокосмічного університету імені М.Є. Жуковського «Харківський авіаційний інститут», Україна, Харків

Барілко Веста Євгеніївна,

студентка ІПСА КПІ ім. Ігоря Сікорського, Україна, Київ

Бідюк Петро Іванович,

професор, доктор технічних наук, професор кафедри математичних методів системного аналізу ІПСА КПІ ім. Ігоря Сікорського, Україна, Київ

Вохранов Ілля Анатолійович,

аспірант ІПСА КПІ ім. Ігоря Сікорського, Україна, Київ

Гавриленко Олена Валеріївна,

доцент, кандидат фізико-математичних наук, доцент кафедри інформаційних систем та технологій факультету інформатики та обчислювальної техніки КПІ ім. Ігоря Сікорського, Україна, Київ

Гавриленко Олена Володимирівна,

кандидат технічних наук, доцент кафедри систем управління літальних апаратів Національного аерокосмічного університету ім. М.Є. Жуковського «Харківський авіаційний інститут», Україна, Харків

Гапон Сергій Вікторович,

завідувач лабораторії географічних інформаційних систем Світового центру даних з геоінформатики та сталого розвитку КПІ ім. Ігоря Сікорського, Україна, Київ

Грішин Костянтин Дмитрович,

студент ІПСА КПІ ім. Ігоря Сікорського, Україна, Київ

Єфремов Костянтин Вікторович,

кандидат технічних наук, доцент кафедри штучного інтелекту КПІ ім. Ігоря Сікорського, директор Світового центру даних з геоінформатики та сталого розвитку КПІ ім. Ігоря Сікорського, Україна, Київ

Згуровський Михайло Захарович,

академік НАН України, професор, доктор технічних наук, ректор КПІ ім. Ігоря Сікорського, Україна, Київ

Зінько Петро Миколайович,

доцент, кандидат фізико-математичних наук, доцент кафедри системного аналізу та теорії прийняття рішень Київського національного університету імені Тараса Шевченка, Україна, Київ

Зінько Тарас Петрович,

кандидат технічних наук, молодший науковий співробітник НДС «Проблем системного аналізу» Київського національного університету імені Тараса Шевченка, Україна, Київ

Кондратенко Наталія Романівна,

професор, кандидат технічних наук, професор кафедри захисту інформації Вінницького національного технічного університету, Україна, Вінниця

Кондратенко Роман Михайлович,

аспірант Білоруського інституту системного аналізу та інформаційного забезпечення науково-технічної сфери, Білорусь, Мінськ

Кудін Григорій Іванович,

доцент, кандидат фізико-математичних наук, молодший науковий співробітник НДС «Проблем системного аналізу» Київського національного університету імені Тараса Шевченка, Україна, Київ

Кулік Анатолій Степанович,

професор, доктор технічних наук, професор кафедри систем управління літальних апаратів Національного аерокосмічного університету ім. М.С. Жуковського «Харківський авіаційний інститут», Україна, Харків

Куреннов Сергій Сергійович,

доцент, доктор технічних наук, професор Національного аерокосмічного університету імені М.С. Жуковського «Харківський авіаційний інститут», Україна, Харків

Малаксіано Микола Олександрович,

доцент, доктор технічних наук, завідувач кафедри технічної кібернетики й інформаційних технологій ім. професора Р.В. Меркта Одеського національного морського університету, Україна, Одеса

Мягкий Михайло Юрійович,

аспірант КПІ ім. Ігоря Сікорського, Україна, Київ

Наконечний Олександр Григорович,

професор, доктор фізико-математичних наук, завідувач кафедри системного аналізу та теорії прийняття рішень Київського національного університету імені Тараса Шевченка, Україна, Київ

Панібратов Роман Сергійович,

аспірант ІПСА КПІ ім. Ігоря Сікорського, Україна, Київ

Панкратова Наталія Дмитрівна,

член-кореспондент НАН України, професор, доктор технічних наук, заступник директора з наукової роботи ІПСА КПІ ім. Ігоря Сікорського, Україна, Київ

Петренко Анатолій Іванович,

професор, доктор технічних наук, професор кафедри системного проектування ІПСА КПІ ім. Ігоря Сікорського, Україна, Київ

Пишнограсв Іван Олександрович,

доцент, кандидат фізико-математичних наук, доцент кафедри штучного інтелекту КПІ ім. Ігоря Сікорського, Україна, Київ

Романов Андрій Юрійович,

аспірант кафедри технічної кібернетики й інформаційних технологій ім. професора Р.В. Меркта Одеського національного морського університету, Україна, Одеса

Романюк Вадим Васильович,

професор, доктор технічних наук, професор кафедри технічної кібернетики й інформаційних технологій ім. професора Р.В. Меркта Одеського національного морського університету, Україна, Одеса

Снігур Ольга Олексіївна,

кандидат технічних наук, приватний підприємець, Україна, Вінниця

Чухрай Андрій Григорович,

професор, доктор технічних наук, завідувач кафедри математичного моделювання та штучного інтелекту Національного аерокосмічного університету ім. М.С. Жуковського «Харківський авіаційний інститут», Україна, Харків