

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

№ 1

2022

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

Г.-В. ВЕБЕР, проф., Турція

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

В.М. КУНЦЕВИЧ, акад. НАН України

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**

• **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

№ 1

2022

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

V.M. KUNTHEVICH, Academician of NASU

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

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 intellectual systems for decision making support**
- **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, російською та англійською мовами.

Журнал можна 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 25.04.2022. Формат 70x108 1/16. Папір офс. Гарнітура Times.

Спосіб друку – цифровий. Ум. друк. арк. 14,411. Обл.-вид. арк. 28,56. Наклад 114 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, Russian and English languages are accepted. Information printed in the Journal is included in the Catalogue of periodicals of Ukraine.

СИСТЕМНІ ДОСЛІДЖЕННЯ ТА ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ

1 • 2022

ЗМІСТ

ТЕОРЕТИЧНІ ТА ПРИКЛАДНІ ПРОБЛЕМИ І МЕТОДИ СИСТЕМНОГО АНАЛІЗУ	
<i>Pankratova N.D., Pankratov V.A.</i> System approach to the underground construction objects planning based on foresight and cognitive modelling methodologies	7
<i>Петренко А.І.</i> Неминучі зміни в ІТ індустрії. Підготовка кадрів в умовах п'ятої промислової революції (Індустрія 5.0)	26
ПРОГРЕСИВНІ ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ, ВИСОКОПРОДУКТИВНІ КОМП'ЮТЕРНІ СИСТЕМИ	
<i>Matsuki Y., Bidyuk P.I.</i> The proof of hypothesis regarding distortion of time and space using the nuclear fusion model	37
<i>Martynenko A.A., Tevyashev A.D., Kulishova N.E., Moroz B.I.</i> System analysis of the problem of establishing the authenticity and authority of painting works	50
ПРОБЛЕМИ ПРИЙНЯТТЯ РІШЕНЬ ТА УПРАВЛІННЯ В ЕКОНОМІЧНИХ, ТЕХНІЧНИХ, ЕКОЛОГІЧНИХ І СОЦІАЛЬНИХ СИСТЕМАХ	
<i>Панібратов Р.С.</i> , Система підтримання прийняття рішень для оцінювання та прогнозування стану страхової компанії	61
ТЕОРЕТИЧНІ ТА ПРИКЛАДНІ ПРОБЛЕМИ ІНТЕЛЕКТУАЛЬНИХ СИСТЕМ ПІДТРИМАННЯ ПРИЙНЯТТЯ РІШЕНЬ	
<i>Zaychenko Yu., Zaichenko He., Hamidov G.</i> Hybrid GMDH deep learning networks – analysis, optimization and applications in forecasting at financial sphere	73
<i>Bodyanskiy Ye., Kostiuk S.</i> Adaptive hybrid activation function for deep neural networks	87
<i>Nedashkovskaya N.I., Androsov D.V.</i> Generative time series model based on encoder-decoder architecture	97
МАТЕМАТИЧНІ МЕТОДИ, МОДЕЛІ, ПРОБЛЕМИ І ТЕХНОЛОГІЇ ДОСЛІДЖЕННЯ СКЛАДНИХ СИСТЕМ	
<i>Кирилюк В.С., Левчук О.І., Гавриленко В.В.</i> Математичне моделювання контактної взаємодії двох пружних трансверсально-ізотропних півпросторів, один з яких містить приповерхневу виїмку еліптичного перерізу	110
<i>Shvandt M.A., Moroz V.V.</i> Overview of the detection and tracking methods of the lab animals	124
Відомості про авторів	149

SYSTEM RESEARCH AND INFORMATION TECHNOLOGIES

1 • 2022

CONTENT

THEORETICAL AND APPLIED PROBLEMS AND METHODS OF SYSTEM ANALYSIS	
<i>Pankratova N.D., Pankratov V.A.</i> System approach to the underground construction objects planning based on foresight and cognitive modelling methodologies	7
<i>Petrenko A.I.</i> Inevitable changes in the IT industry. Training in the conditions of the fifth industrial revolution (Industry 5.0)	26
PROGRESSIVE INFORMATION TECHNOLOGIES, HIGH-EFFICIENCY COMPUTER SYSTEMS	
<i>Matsuki Y., Bidyuk P.I.</i> The proof of hypothesis regarding distortion of time and space using the nuclear fusion model	37
<i>Martynenko A.A., Tevyashev A.D., Kulishova N.E., Moroz B.I.</i> System analysis of the problem of establishing the authenticity and authority of painting works	50
DECISION MAKING AND CONTROL IN ECONOMIC, TECHNICAL, ECOLOGICAL AND SOCIAL SYSTEMS	
<i>Panibratov R.S.</i> Decision support system for estimating and forecasting state of insurance company	61
THEORETICAL AND APPLIED PROBLEMS OF INTELLECTUAL SYSTEMS FOR DECISION MAKING SUPPORT	
<i>Zaychenko Yu., Zaichenko He., Hamidov G.</i> Hybrid GMDH deep learning networks – analysis, optimization and applications in forecasting at financial sphere	73
<i>Bodyanskiy Ye., Kostiuk S.</i> Adaptive hybrid activation function for deep neural networks	87
<i>Nedashkovskaya N.I., Androsov D.V.</i> Generative time series model based on encoder-decoder architecture	97
MATHEMATICAL METHODS, MODELS, PROBLEMS AND TECHNOLOGIES FOR COMPLEX SYSTEMS RESEARCH	
<i>Kirilyuk V.S., Levchuk O.I., Gavrilenko V.V.</i> Mathematical modeling of the contact interaction of two elastic transversely isotropic half-spaces, one of which contains a near-surface groove of an elliptical section	110
<i>Shvandt M.A., Moroz V.V.</i> Overview of the detection and tracking methods of the lab animals	124
Information about the authors	149

UDC 303.732.4, 519.816
DOI: 10.20535/SRIT.2308-8893.2022.1.01

**SYSTEM APPROACH TO THE UNDERGROUND
CONSTRUCTION OBJECTS PLANNING BASED ON FORESIGHT
AND COGNITIVE MODELLING METHODOLOGIES¹**

N.D. PANKRATOVA, V.A. PANKRATOV

Abstract. The system approach to the underground construction objects planning based on foresight and cognitive modeling methodologies is proposed. Using the foresight methodology allows with the help of expert estimation procedures to identify critical technologies and build alternatives of scenarios with quantitative characteristics. For the justified implementation of a particular scenario the cognitive modelling is used, which allows to build causal relationships based on knowledge and experience, understand and analyze the behaviour of a complex system for a strategic perspective with a large number of interconnections and interdependencies. The suggested system approach allows planning of underground objects on the basis of reasonable scenarios selection and justification of their creation priority.

Keywords: foresight, cognitive, impulse modelling, planning, scenarios, underground construction.

INTRODUCTION

The growth of megalopolises, their populations, expansion of infrastructure are of the characteristic features of the modern world. Regulation of urban planning in order to improve environmental standards and safety of life in constantly growing megacities is one of the most pressing, but insufficiently studied and complex world problems [1]. It leads to the search of new places to production facilities, social and other objects of human activity [2–4]. The space of megacities created by man in the process of underground construction becomes a new, underground habitat, which should be comfortable and safe for humans. Risks in underground space development which is characterized by space-temporal variability are considered in [5–8].

Various directions of implementing system approach for planning urban surface construction in megacities are known [9, 10]. Analyzing the trends of the future development of the underground space of megacities are considered the individual projects of underground urban studies that characterize the directions

¹This material is based upon work supported in part by the National Research Foundation of Ukraine under Grant 2020.01/0247

of urban underground construction in the near and medium perspective [11]. So, In Chicago, the second largest economic center in the United States, it is planned to build an underground city with a vertical layout, which will have 100 underground floors. As for the underground development, the studies went no further than general task setting and analysis of research methods [12].

Underground urban planning is a complex system in many aspects. Firstly, this system consists of many interconnected subsystems and objects. Secondly, the processes occurring in this system during construction and during operation are also complex and in some cases poorly predictable, because they are largely associated with various geological processes. The problems that accompany underground urban development can be attributed to poorly structured problems.

A system approach to the planning of underground urban studies, based on the methodology of foresight, as a tool for the concept of sustainable development of megacities was proposed in [13].

The goal of the system approach presented in this paper is to study some problems of the underground object's viability in extreme and emergency situations.

SYSTEM APPROACH TO THE UNDERGROUND CONSTRUCTION OBJECTS PLANNING BASED ON FORESIGHT AND COGNITIVE MODELLING METHODOLOGIES

In this paper the system approach to the underground construction objects planning based on the mathematical support of foresight methodology with the aim of scenarios alternatives creating and cognitive modeling to build scenarios for the development of the desired future and ways of their implementation is proposed. For realization of this system approach the totality of the properties and characteristics of the studied objects, as well as the features of the methods and procedures used to create them taking into account. Based on a comparison of the characteristics of the qualitative analysis methods, the requirements for their application, the disadvantages and advantages of each of them, researchers of foresight problems should choose the rational combination of methods, establish the correct sequence for their use, account the totality of requirements for systems and the features of the tasks to be solved.

The system approach in the form of a two-stage model based on a combination of foresight and cognitive modelling methodologies is developed and its scheme presented in Fig. 1 [14]. The involvement of scanning methods, STEEP analysis, brainstorming, SWOT analysis, TOPSIS (Technique for Order Preference by Similarity to an Ideal Solution) method and VIKOR method at the initial level of the first stage allows using expert assessment to identify critical technologies in economic, social, environmental, technical, technological, information and other directions [15–18]. The basis of this level is the analysis subsystems, which are connected by direct and feedback links to the monitoring system and field tests. The quantitative data obtained after analysis and processing are the initial ones for solving of foresight tasks. The construction of rational alternatives of scenarios for the development of strategically important underground objects are expedient to be performed on the basis of a collection of foresight activities. For this goal, in the process of creating alternatives of scenarios it becomes necessary to involve expert assessment methods, among which are the most commonly used methods of analytic hierarchy, Delphi methods and morphological analysis [19–22].

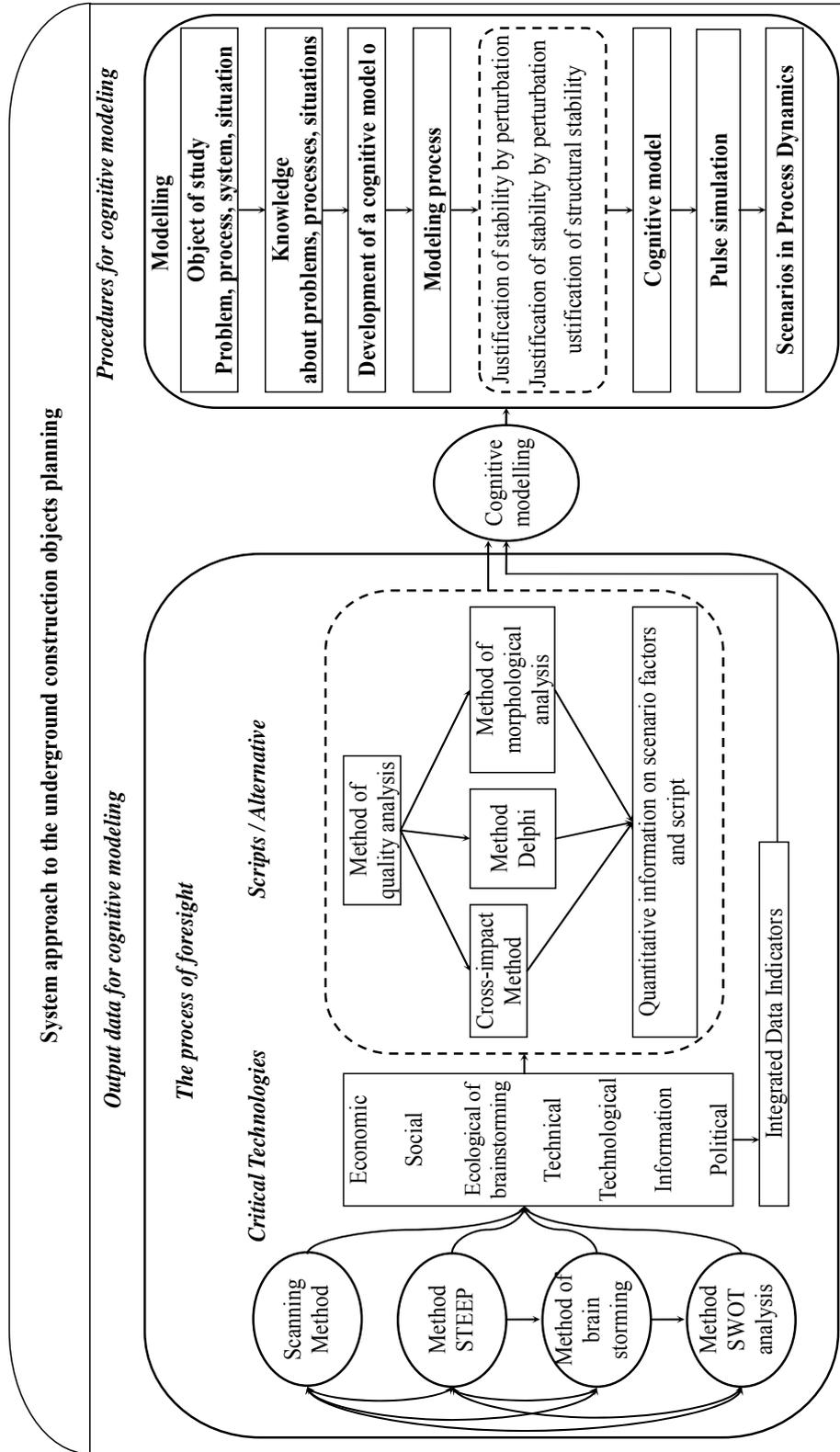


Fig. 1. Scheme of the system approach to the underground construction objects planning

In this paper to identify critical technologies the SWOT analysis method is used. For the purpose of ranking the obtained critical technologies and identifying the most topical ones, the TOPSIS method is applied [16,17]. The method TOPSIS of multicriterial analysis (ranking) of alternatives in addition to estimating the distance from the considered alternative to the ideal solution allows to take into account the distance to the worst solution. The trade-off in choosing the best alternative is based on the fact that the chosen solution must be at the same time as close to the ideal as possible and most remote from the worst solution. The obtained rating makes it possible to take into account the weight characteristics of critical technologies that are the vertices of the cognitive map when constructing a cognitive model. According to the VIKOR method, a compromise solution to the problem should be an alternative that is closest to the ideal solution. Moreover, to assess the degree of the alternative proximity to the ideal solution, a multicriteria measure is used [18]. As soon as the critical technologies are identified we cross to the system approach second level, using the qualitative methods for creation alternatives of socio-economic systems scenarios [21].

In some cases, when the output information for cognitive modeling is given in statistical form as separate logical groups, the method of constructing an integrated indicator data is proposed [23]. This enables all groups to aggregate in integrated indicator data used the proposed method of recovery of functional dependences for discrete preset samples or carry out decomposition integrated indicator to individual subject groups, followed by decomposition of the logical sequence characteristics. That is the construction of cognitive maps reasonably add or remove her vertex, vertex to break a sequence of interconnected nodes.

Formalization the method of constructing an integrated indicator data implies the use of this sequence of procedures:

- the selection of indicator which will characterize the specific area of one of the directions of sustainable development (economic, environmental, social);
- grouping by specific characteristics of the data sets, which influence the dynamics, selected at the stage 1 of the indicator formation;
- forming a database for a specific period on the basis of the discrete samples;
- recovery of functional dependencies by the discrete samples;
- analysis of the results based on the recovered dependence.

At the second stage of system approach for the construction of scenarios correspond to selected alternatives, cognitive modelling is involved [24]. which makes it possible to obtain a valid scenario for decision making. According to the developed methodology of cognitive modeling for complex systems [24–27], modeling is carried out in some steps. At the first step of the cognitive modelling, using the results of foresight methodology, theoretical and practical data on underground urban planning, the cognitive model as cognitive map is developed. The cognitive map in the form of a sign oriented graph and a functional graph in the form of a weighted sign digraph are created. At the second step of cognitive modeling the investigation of the cognitive model properties, methods of analysis of structural stability and resistance to disturbances, methods of analysis of model connectivity (simplicial analysis), and graph theory methods are used. The proofs of numerical stability of cognitive maps based on the representation of values and perturbations at the vertices of the graph in matrix form are presented [14]. At the third step of cognitive modeling, to determine the possible development of proc-

esses in a complex system and to create the scenario development, an impulse process model (simulation of disturbance propagation on cognitive models) is used, which allows to create the scenarios of development in the process of dynamics and to propose a scientifically based strategy for implementing the priority scenario [28].

So, at the first step of cognitive modelling the cognitive models such as a cognitive map – a sign oriented graph (1) and a functional graph in the form of a weighted sign digraph is created [24–29]

$$G = \langle V, E \rangle,$$

where G is a cognitive map in which V are concepts, a finite set of vertices of the cognitive map $V_i \in V, i = 1, 2, \dots, k; E = \{e_{ij}\}$ is the set of arcs e_{ij} of the graph, $i, j = 1, 2, \dots, m$, reflect the relationship between the vertices V_i and V_j ; the influence of V_i on V_j in the situation under study can be positive (+1) when an increase (decrease) in one factor leads to an increase (decrease) in another, negative (–1) when an increase (decrease) in one factor leads to a decrease (increase) in another, or absent (0). The cognitive map G corresponds to the square matrix of relations A_G :

$$A_G = \{a_{ij}\} = \begin{cases} 1, & \text{if } V_i \text{ is connected with } V_j, \\ 0, & \text{otherwise.} \end{cases}$$

The ratio a_{ij} can take the value “+1” or “–1”. The relation between variables (interaction of factors) is a quantitative or qualitative description of the effect of changes in one variable on others at the corresponding vertices.

Vector Functional Graph

$$\Phi = \langle G, X, F(X, E), \theta \rangle,$$

where G is a cognitive map; X is the set of vertex parameters,, is the space of vertex parameters; $F(X, E)$ is the arc transformation functional.

At the second step of cognitive modeling, to study the properties of the cognitive model, is used methods of structural stability and perturbation resistance analysis [24–26], methods for analyzing model connectivity (simplicial analysis [27, 28]), and graph theory methods [29]. The results of the analysis were compared with the available information on underground construction.

At the third step of cognitive modeling, to determine the possible development of processes in a complex system and develop development scenarios, is using the impulse process model (modeling the propagation of disturbances in cognitive models) [25, 29]:

$$x_{v_i}(n+1) = x_{v_i}(n) + \sum_{v_j: e=e_{ij} \in E}^{k-1} f(x_i, x_j, e_{ij})P_j(n) + Q_{v_i}(n), \quad (1)$$

where $x(n), x(n+1)$ are the values of the indicator at the vertex V_i at the simulation steps at time $t = n$ and the next $t = n+1$; $P_j(n)$ is the momentum that ex-

isted at the vertex V_j at the moment $t = n$; $Q_{V_i}(n) = \{q_1, q_2, \dots, q_k\}$ is the vector of external pulses (disturbing or controlling actions) introduced to the vertices V_i at time moment n . It allows to consider of modeling process in dynamic.

MODELLING OF UNDERGROUND CONSTRUCTION

In the framework of the foregoing, let us call the studied complex system “Natural-technical geosystem”.

The first step. Cognitive Model Development. Table 1 presents data on the vertices (concepts) of the hierarchical cognitive model without reference to a specific territory, in a generalized form. The generalizing concepts (indicators, factors), independent of the specifics, which can be disclosed and taken into account in the future when developing the lower levels of the hierarchical model, are using. Fig. 2 shows a hierarchical cognitive map I_G : “Natural-technical geosystem”. In Table 1 and Fig. 2, the vertices of the upper (first level) are denoted as $I - V_i$, $i = 5, 11, 13, 15, 16$.

Table 1. The vertices of the hierarchical cognitive map “Natural-technical geosystem”

Code	Vertex explanation	Vertex assignment
$I - V_{11}$	The viability of the underground urban development	Indicative
$I - V_{13}$	Disasters, extreme and emergency situations	Perturbing
$I - V_{15}$	Environmental risks	Perturbing
$I - V_{16}$	Economic risks	Perturbing
$I - V_5$	Genetic type and lithological composition of soils	Basic
V_1	Mountain and hydrostatic pressure, seismic impact	Basic
V_2	Surface Load Static Load Index	Basic
V_3	The indicator of the static load of the surrounding soil massif	Basic
V_4	Existing underground facilities	Disturbing
V_6	Estimated soil resistance	Basic
V_7	Aquifers and High Water	Disturbing
V_8	Relief Type and Morphometry	Basic
V_9	Engineering and geological processes	Disturbing
V_{10}	Mining construction technologies	Regulating
V_{12}	The level of comfort of work and rest during the construction and operation of underground structures	Indicative
V_{14}	Construction, operational, management risks	Disturbing
V_{17}	Staff qualifications	Regulating
V_{18}	Industrial Safety	Basic
V_{19}	Quality and construction time	Regulating

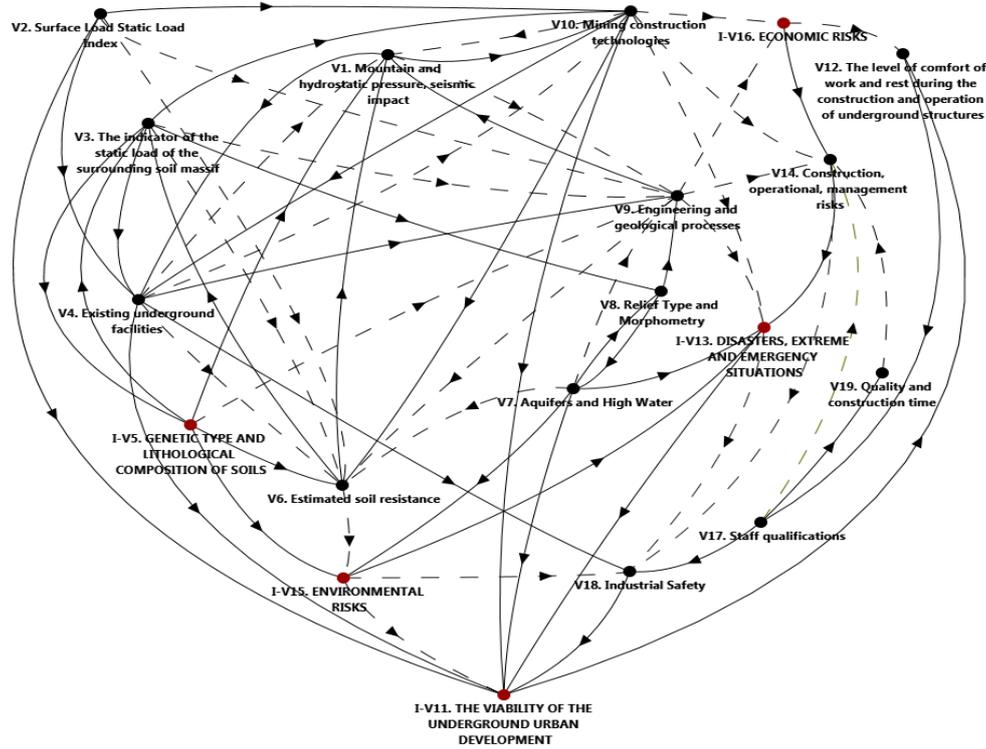


Fig. 2. Hierarchical cognitive map I_G “Natural-technical geosystem”

The cognitive model is a simulation model that makes it possible not to conduct an experiment on a “living” system, but to simulate its behavior and possible future development under the influence of various factors, generating new knowledge about the system. This allows to justify management decisions in a given situation.

The second step of modeling. Before using the cognitive model to determine its possible behavior, the second step of modeling analyzes the various properties of the model are fulfilled. In this case, the stability properties of the model must be analyzed.

The results of the analysis of the model properties obtained using the CMLS software system [30]. Figs 3 and 4 shows an example of determining the cycles of the cognitive model I_G .

The fig. 3 shows one of the positive feedback cycles, a sign of which is an even number of negative arcs in it. Fig. 4 one of the negative cycles

Impulse sustainability. The cognitive model I_G was not resistant to perturbations according to the accepted criterion [24]: the maximum modulo M root of the characteristic equation of the matrix of relations of the graph I_G is $|M| = 1,82 > 1$ (must be less than 1).

Structural stability. An analysis of the ratio of the number of stabilizing cycles (35 negative feedbacks) and process accelerator cycles (33 positive feedbacks) indicates the structural stability of such a system [24].

The given example of the analysis of the cycles of the cognitive model showed the variety of cycles of cause and effect relationships that exist in complex systems. There are 68 of them in the analyzed system. Without an appropri-

ate theoretical analysis, there is a great risk of the human factor in making managerial decisions, because its consequences may not be obvious due to the complexity of interactions in the system.

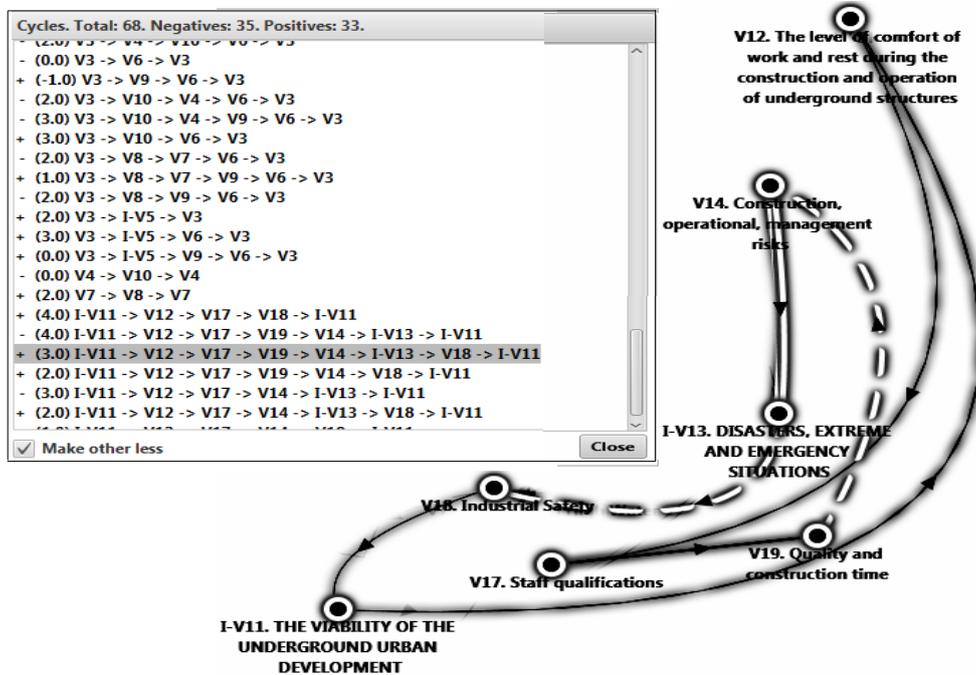


Fig. 3. Cognitive map cycles, one of the positive cycles is highlighted

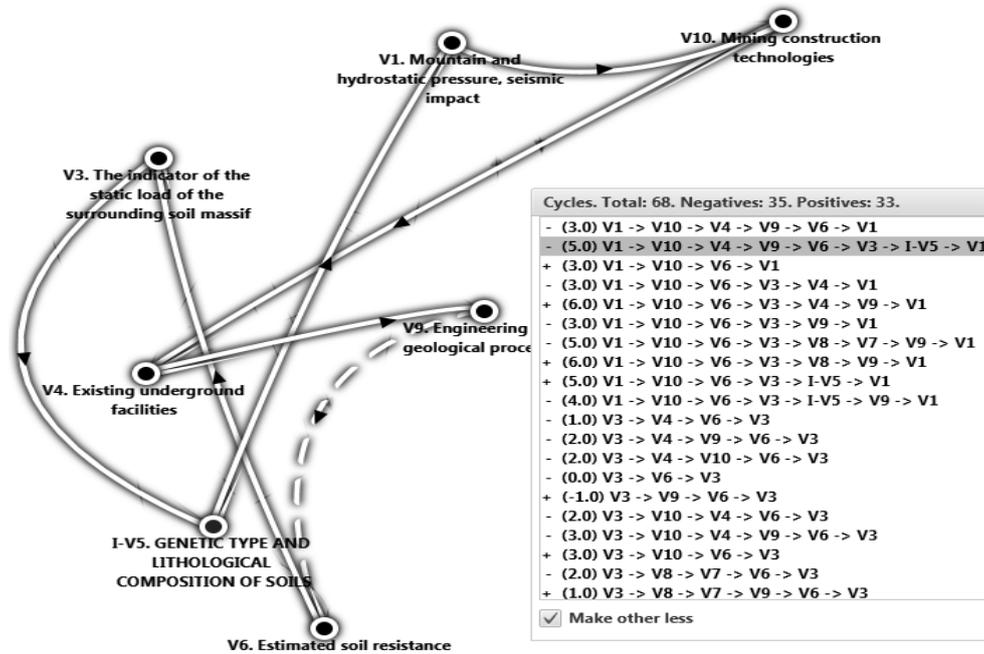


Fig. 4. Cognitive map cycles, one of the negative cycles is highlighted

Analysis of system connectivity, simplicial analysis. Immersed in the study of the structure of the cognitive model, it is desirable to conduct a simplicial

analysis of the properties of its connectivity. Such an analysis is carried out in order to study and understand the topological properties of the model and, accordingly, other connectivity faces of the complex system under study that are not detected in the above algebraic analysis. According to R.H. Atkin and J. Casti, connectedness is the essence of the concept of a large system [27, 28].

The connectivity properties of blocks (simplexes) characterize the “deep” connections of the cognitive model, the connections of its simplexes, and not just the vertices, as in the cognitive map. A simplex is formed by each vertex, which is the reason that some other vertices interact with each other. Figs 5, 6 and 7 show the results of a simplicial analysis of the I_G model. Figure 5 shows the transformed matrix of relations of the graph I_G with the dimensions ρ of simplexes $\sigma_{\rho}^{V_i}$ of rows (x) and columns (y) indicated in it in the decreasing order; $\rho = k - 1$, k is the number of elements in the corresponding row / column, the dimension of the simplex shows the number of edges connecting the vertices. In Fig. 6, the simplices $\sigma_{\rho=3}^{(V_{10})}$ ($\rho = 3$ means that three edges go out from each vertex) of the simplex are highlighted for one vertex V_{10} . This vertex V_{10} is the reason for the connection of the vertices V_1, V_2, V_3, V_4 . Those, vertex Mining

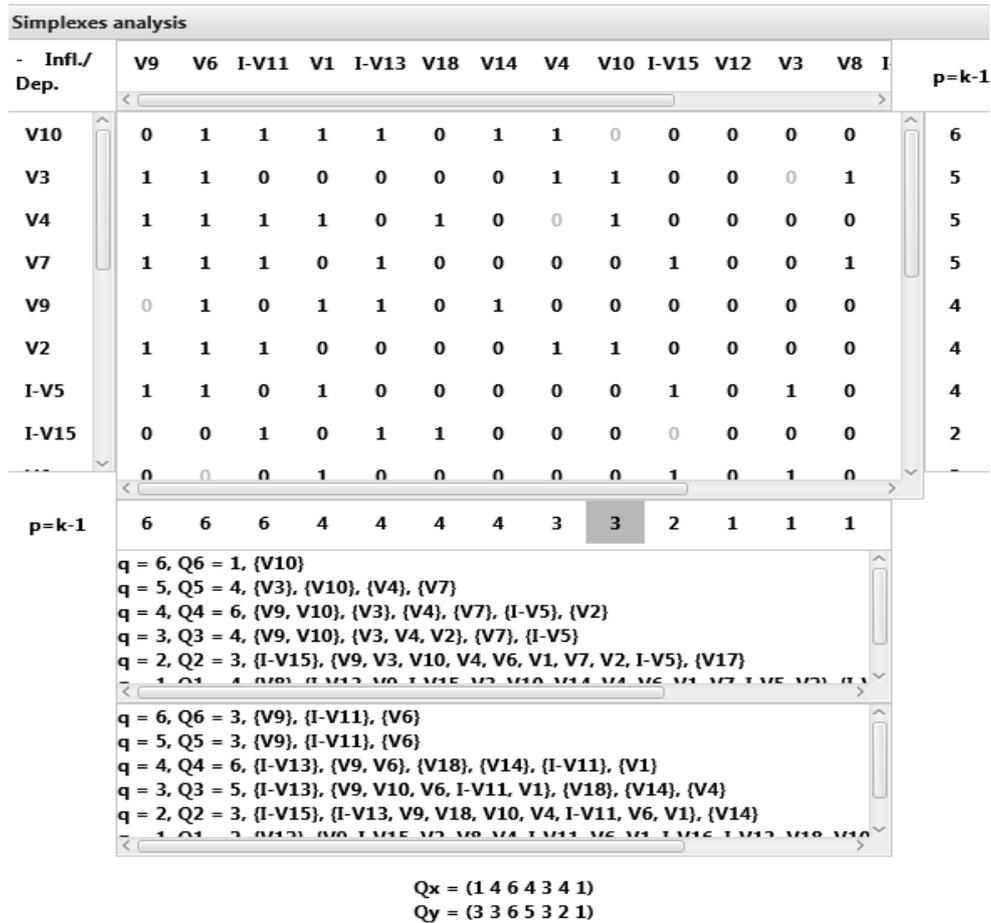


Fig. 5. Results of simplicial analysis (calculation)

construction technologies (V_{10}) is the reason that vertices Mountain and hydrostatic pressure, seismic impact (V_1), Surface Load Static Load Index (V_2), The indicator of the static load of the surrounding soil massif (V_3), Existing underground facilities (V_4) forming one block are interconnected.

Thus, these vertices are the cause of the simplex in the form of a tetrahedron.

Note that simplexes of higher dimension are not depicted on the plane; only their “projection” can be conditionally drawn – Fig. 7.

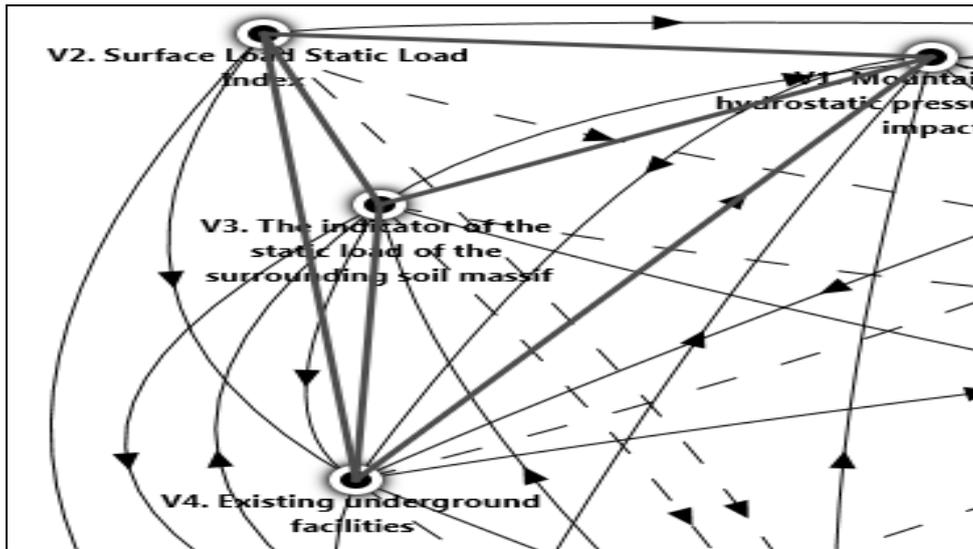


Fig. 6. Image of one of the simplexes of dimension $\rho = 3$

Simplexes form q -connected chains $q = 1$ (connection through a vertex), $q = 2$ (connection through an edge), $q = 3$ (connection through a plane), etc., thus uniting into simplicial complexes K_x (along the lines – “inputs”) and K_y (columns – “outputs”). Simplexes are q -connected or not in simplexes (they lack or have connections of simplexes along vertices, edges, planes, m -dimensional volumes). Simplicial complexes are characterized by the structural vectors Q_x and Q_y formed by vertex groups common to different simplexes.

The third stage of modeling. Scenario analysis is designed to anticipate possible trends in the development of situations on the model. To generate scenarios of the development of the system, impacts are introduced into the vertices of the cognitive map in the form of a set of impulses. The impulse process formula has the form (1).

It is possible to introduce perturbations Q of different sizes (normalized) to any of the vertices, as well as to their combination. In connection with a large number of theoretically possible variants of introduced disturbances, it is necessary to develop a plan for a computational experiment before excluding pulse simulation, eliminating at least almost impossible variants.

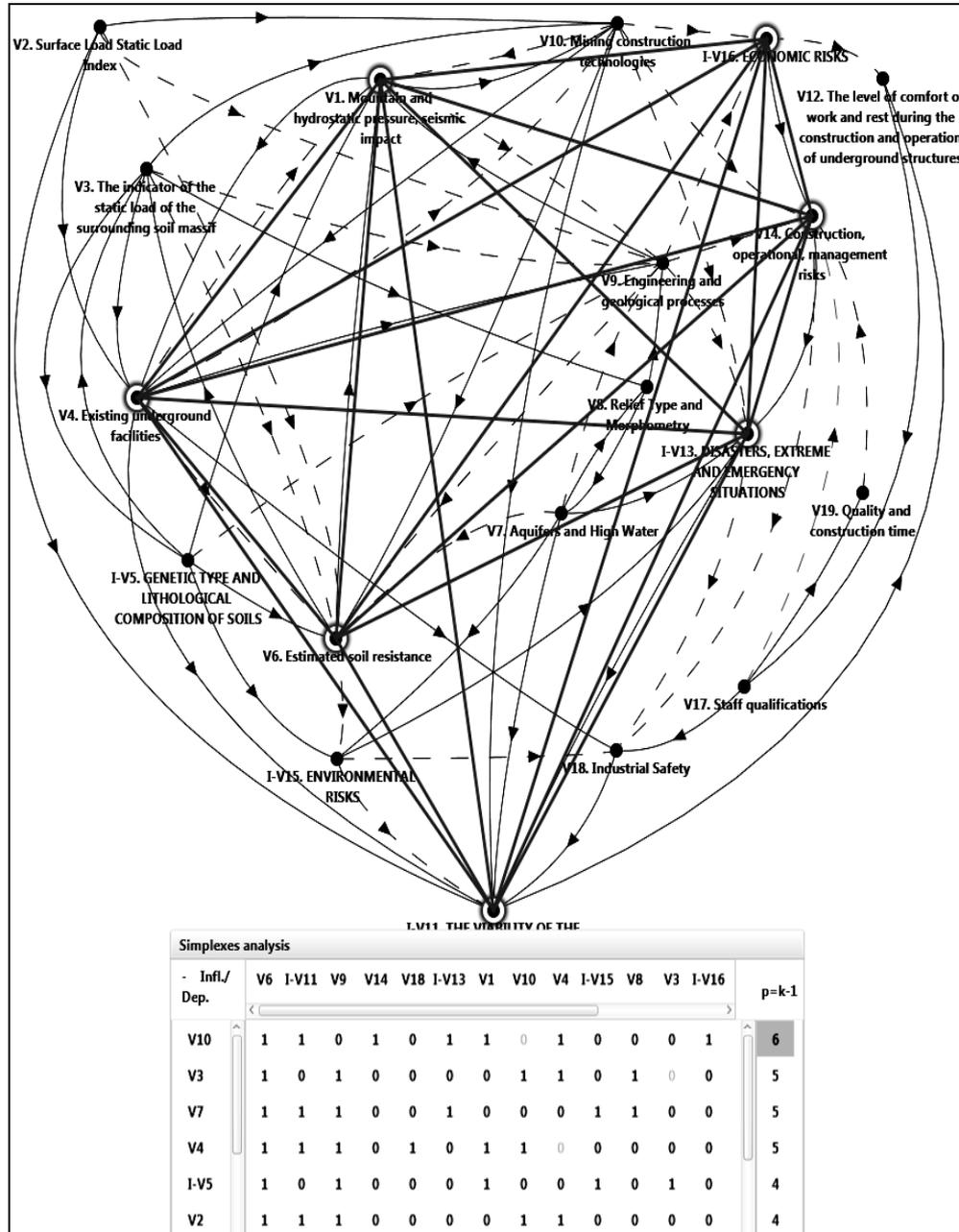


Fig. 7. Image of the projection of one of the simplexes of dimension $\rho = 6$

Introducing disturbances to the vertices, the decision-maker is looking for the answer to the question: “What will happen if ...?”

The CMLS software system [30] allows, in the process of pulse modeling and analysis of the obtained results, to introduce control or disturbing influences at any modeling step. This allows to change (correct) scenarios in model dynamics, to determine the effects that bring the processes closer to the desired.

The results of pulse modeling in four scenarios are presented.

Scenario No. 1. Assume good technology is used in underground construction. To the vertex V_{10} , the control action is introduced $q_{10} = +1$, the perturbation vector $Q = \{q_1 = 0, \dots, q_{10} = +1, \dots, q_{19} = 0\}$.

Fig. 8 shows graphs of pulsed processes. For the convenience of visual analysis of the image, the graphs of pulsed processes in the vertices V_{10} , $I-V_{13}$, $I-V_{15}$, $I-V_{16}$, $I-V_{11}$, $I-V_5$ are represented by two figures: Fig. 8,a from the first to the sixth step of modeling and Fig. 8,b from the sixth to tenth step of modelling. The image of pulsed processes at a larger number of simulation steps is not necessary, because system behavior trends under these conditions are already evident.

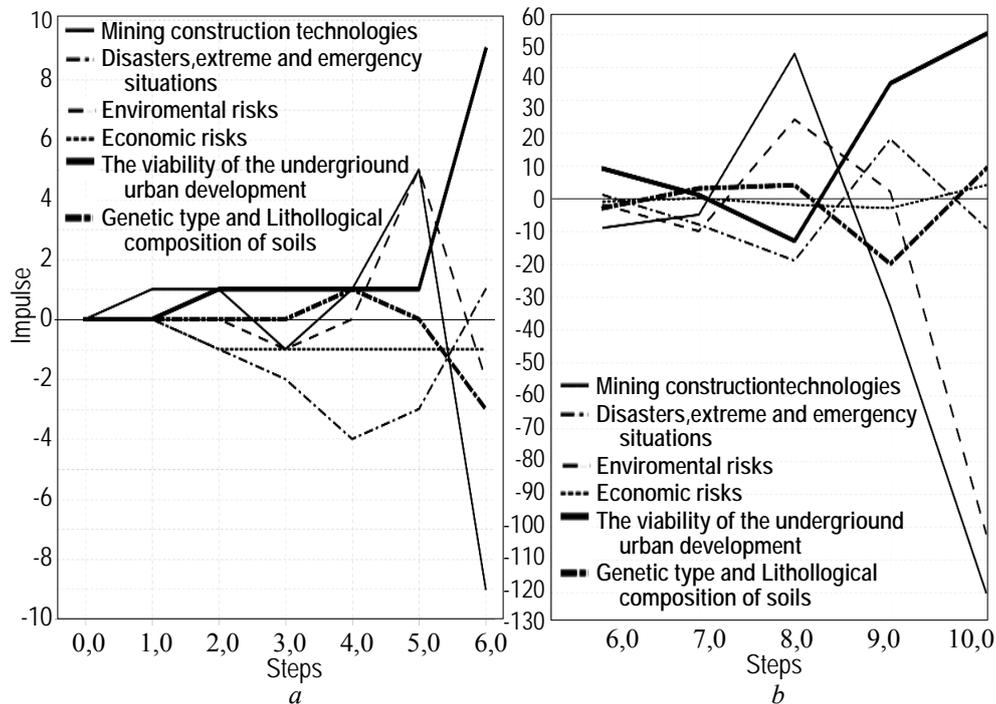


Fig. 8. Graphs of pulsed processes, from the first to the sixth step of modeling (a) (scenario No. 1); from the first to the sixth step of modeling (b) (scenario No. 1)

Modeling scenario No. 1, it is advisable to analyze whether changes in Mining construction technologies (V_{10}) can and in what way affect other vertices of the cognitive model. As can be seen from the graphs in Fig. 8, positive changes in V_{10} can contribute to positive trends in the development of vertices at the top hierarchical level: up to the 5th and 6th steps of the modeling, the declining trends of Disasters, extreme and emergency situations ($I-V_{13}$), Environmental risks ($I-V_{15}$), Economic risks ($I-V_{16}$), Genetic type and lithological composition of soils ($I-V_5$), The viability of the underground urban development ($I-V_{11}$) is growing.

All this may indicate that a single positive change in one of the vertices of the system model may not be enough to exclude the negative impact of risks and other negative influences.

Scenario No. 2. Suppose that the possibility of the simultaneous occurrence of all risks is increasing in the system. Disturbing effects are appearing $q_{14} = +1, q_{15} = +1, q_{16} = +1$, there is perturbation vector $Q = \{q_1 = 0, \dots, q_{14} = +1, q_{15} = +1, q_{16} = +1, \dots, q_{19} = 0\}$.

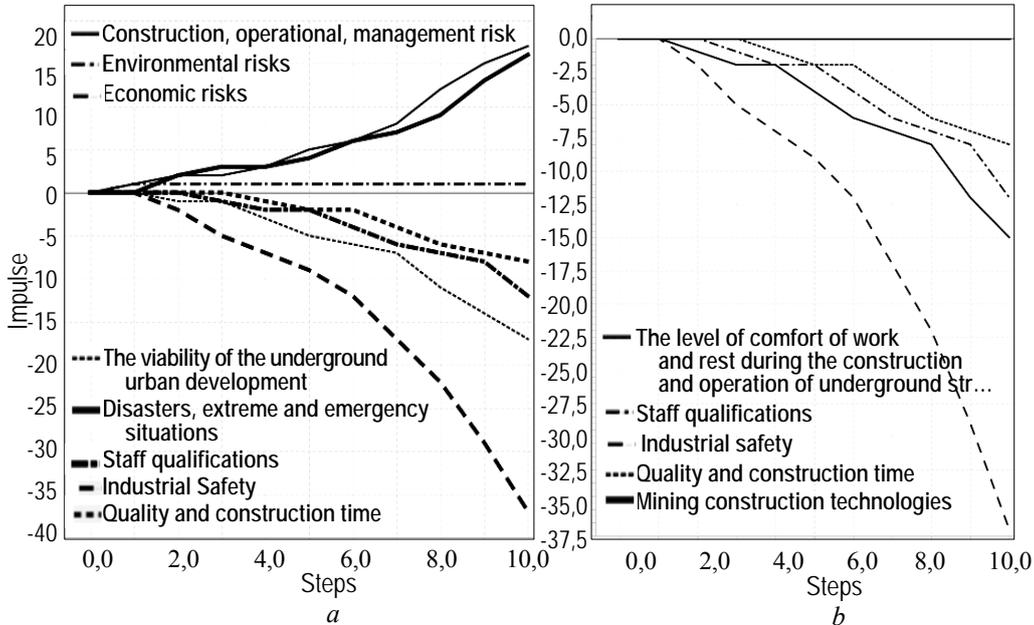


Fig. 9. Graphs of pulsed processes: 9,a and 9,b are scenarios No. 2

Pulse simulation results are presented in Fig. 9,a for vertices $V_{14}, I-V_{15}, I-V_{16}, I-V_{11}, I-V_{13}, V_{17}, V_{18}, V_{19}$ and Fig. 9,b for vertices $V_{12}, V_{17}, V_{18}, V_{19}, V_{10}$.

The simulation results of the second scenario show an extremely unfavorable option for the development of situations in the system. With increasing risks all indicators of the system fall at both the first and second levels of the hierarchy. This observation forces one to make a decision on the search for the necessary counteraction to the situations that have arisen.

Consider the third scenario. Suppose improving Engineering and geological processes (V_9), Mining construction technologies (V_{10}), Staff qualifications (V_{17}), Quality and construction time (V_{19}), but there are Disasters, extreme and emergency situations ($I-V_{13}$).

Scenario No. 3. Control actions $q_9 = +1, q_{10} = +1, q_{17} = +1, q_{19} = +1, q_{13} = +1$, the perturbation vector $Q = \{q_1 = 0, \dots, q_9 = +1, q_{10} = +1, \dots, q_{13} = +1, \dots, q_{17} = +1, \dots, q_{19} = +1\}$.

The results of pulse modeling are presented in Fig. 10,a for vertices $I-V_{13}, V_9, V_{10}, V_{19}, I-V_{15}, I-V_{16}, V_{18}, V_{17}, I-V_{11}$ and Fig. 10,b for vertices $V_{17}, V_{18}, V_{19}, I-V_{11}, V_{12}, V_{14}, I-V_{13}$.

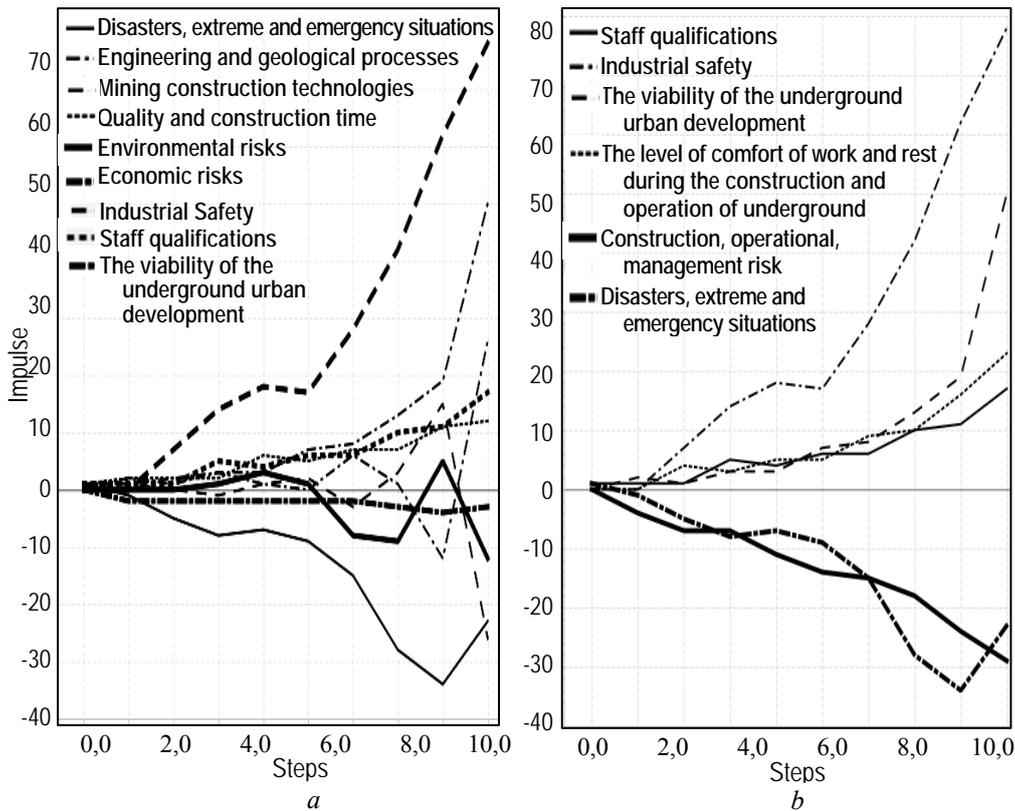


Fig. 10. Graphs of pulsed processes: 10,a and 10,b are scenarios No. 3

An analysis of the results of impulse modeling according to scenario No. 3 shows that the introduction of control actions to the vertices of Engineering and geological processes (V_9), Mining construction technologies (V_{10}), Staff qualifications (V_{17}), Quality and construction time (V_{19}), but there are Disasters, extreme and emergency situations ($I-V_{13}$) can counteract the negative impact of possible disasters and extreme situations, reducing the impact of economic, environmental and technological risks. Thus, scenario No. 3 can be considered favorable: industrial safety is increasing.

The simulation results in one more scenario No. 4 is presented. Assume that Construction, operational, management risks can be reduced. In this case, the impulse actions initiate 6 vertices of the model and the synergistic effect of their joint action is investigated. The modeling of this scenario of the situations development on the model is carried out in order to determine whether it is necessary or not to strengthen the impact on the system to achieve good indicators.

Scenario No. 4. Control actions $q_9 = +1, q_{10} = +1, q_{17} = +1, q_{19} = +1, q_{13} = +1, q_{14} = -1$, the perturbation vector $Q = \{q_1 = 0, \dots, q_9 = +1, q_{10} = +1, \dots, q_{13} = +1, q_{14} = -1, \dots, q_{17} = +1, \dots, q_{19} = +1\}$.

The results of pulse modeling are presented in Fig. 11,a for vertices $I-V_{13}, V_9, V_{10}, V_{14}, V_{17}, V_{19}$, and Fig. 11,b for vertices $V_{12}, V_{14}, V_{15}, V_{16}, V_{18}, I-V_{11}$.

Analysis of the simulation results of Scenario No. 4, which differs from scenario No. 3 by the addition of an impulse $q_{14} = -1$, simulating the possibility of reducing Construction, operational, management risks showed the following. The combined positive impact of six factors on the system leads to the possibility of the appearance of desirable trends in situations throughout the system. So, there are tendencies of improvement (growth) of the underground urban development viability, the level of comfort, work and rest during the construction and operation of underground structures, Industrial Safety while reducing all types of risk and reducing Disasters, extreme and emergency situations

Let us compare the simulation results of Scenarios No. 1, No. 2, No. 3 and No. 4, using the capabilities of the CMLS software system. The results of pulse modeling at the 10th step of modeling are selected and presented them in the form of histograms in Fig. 12.

As can be seen from Fig. 12, scenario No. 4 can be considered the best of those considered, although its results are not too different from the results of scenario No. 3. If you set the task of minimizing the cost of resources for the particular scenario implementation, then perhaps scenario No. 3 will be the best, with fewer control actions in the system.

A comparison of the results of scenarios No. 3 and No. 4 with the results of scenario No. 1, in which the control action is applied to only one vertex, shows that it is inferior to scenarios No. 3 and No. 4. So, for example, the pulse value at the vertices of Industrial safety (V10) reaches 30, and according to scenario No. 3, the pulse value at this vertex is 78, and according to scenario No. 4 pulse value is 85. If we compare the simulation results of scenario No. 2 with the results of other scenarios, it is obvious that without countering possible risks, the development scenarios of the Natural-technical geosystem system will be extremely pessimistic.

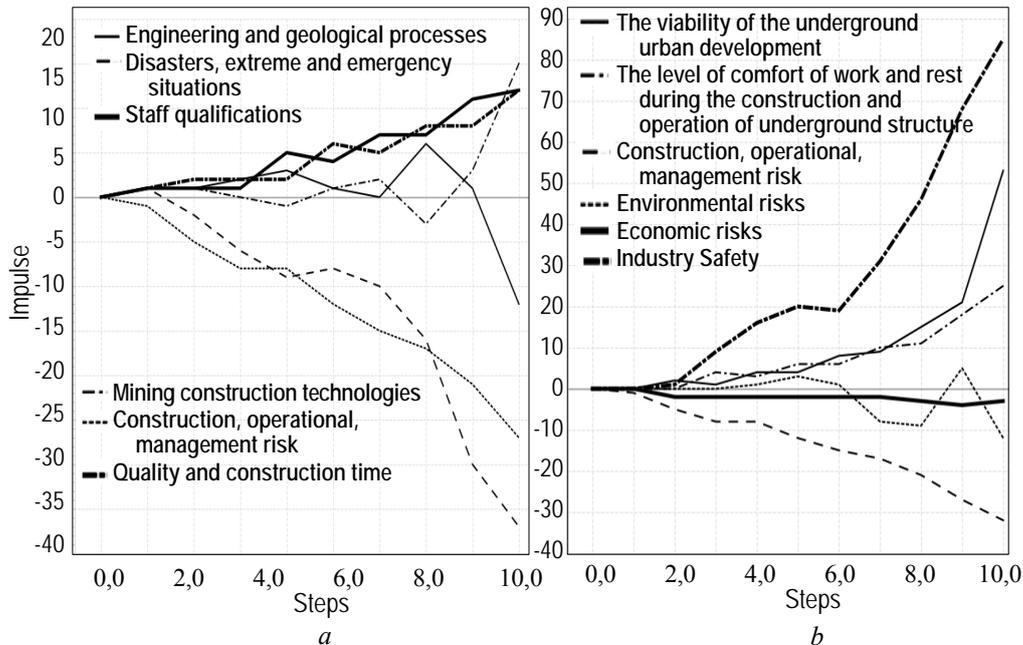


Fig. 11. Graphs of pulsed processes: 11,a and 11,b are scenarios No. 4

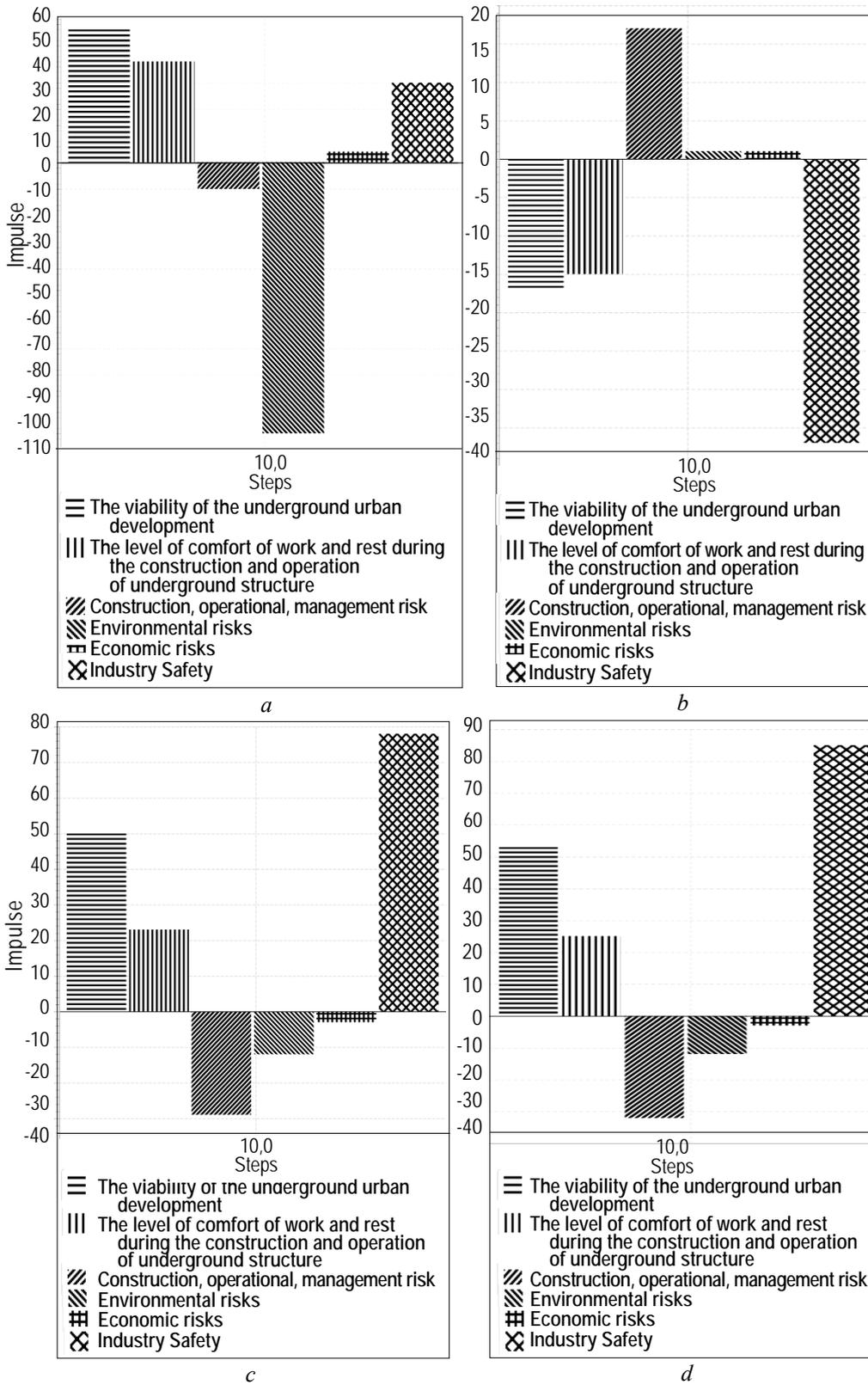


Fig. 12. Histograms of pulse values at the 10th step of modeling according to scenario No. 1(a), scenario No. 2(b), scenario No. 3(c), scenario No. 4(d)

CONCLUSION

The modelling of scenarios for possible processes of the events development in the analyzed complex system is carried out under the influence of various internal and external disturbances and control impulse effects. The results of the conducted cognitive modeling make it possible to judge that the cognitive models, which systematize and structure various information about the underground construction system, correspond to the real system and can be used to anticipate the possible processes of situations in the system under the influence of various disturbing and controlling factors. The developed author's software system CMLS allows in the process of pulse modeling and analysis of the obtained results to introduce control or exciting actions at any stage of modeling. This allows to change (correct) scenarios in the dynamics of creating a model, to determine the effects that bring the processes closer to the desired. The developed methodology and tools made it possible to combine the assessment of the impacts and relationships of geological factors, technogenic and structural-functional types for the study of the underground objects construction.

The proposed system approach to the study of the of underground objects development based on a synthesis of cognitive modeling and foresight methodologies can become the scientific and methodological basis for the development of the "Construction Geotechnology" science and its practical application to study the problems of underground construction in order to ensure the safety and quality of human life. The developed system approach is applied to the study of underground construction objects in order to select reasonable scenarios for their future development.

REFERENCES

1. *World Urbanization Prospects 2018: Highlights*. United Nations. New York, 2019. Available: <https://population.un.org/wup/Publications/Files/WUP2018-Highlights.pdf>
2. I. Vähäaho, "Underground space planning in Helsinki", *Journal of Rock Mechanics and Geotechnical Engineering*, 6, pp. 387–398, 2014. Available: <https://doi.org/10.1016/j.jrmge.2014.05.005>
3. C.L. Owen and C. Bezerra, "Evolutionary Structured Planning. A Computer-Supported Methodology for the Conceptual Planning Process", in JS Gero (ed.) *Artificial Intelligence in Design'00*. Dordrecht: Kluwer Academic Publishers, 2000, pp. 287–307.
4. A.N. Levchenko, "About a new direction of scientific research in construction geotechnology", *Mining information and analytical bulletin (scientific and technical journal)*, no. 2, pp. 15–21, 2007.
5. P. Saługa, *Ocena ekonomiczna projektów i analiza ryzyka w górnictwie* [Economic Evaluation and Risk Analysis of Mineral Projects]. Studia, Rozprawy, Monografie, nr 152, Wyd. IGSMiE PAN, Kraków, 2009.
6. B.A. Kartosia, "Mastering the underground space of large cities. New Trends", *Mining Information and Analytical Bulletin (scientific and technical magazine)*, "Construction and Architecture", pp. 615–628, 2015.
7. B. Kartosia, *Fundamentals of Underground Space Development*, (in Russian). M.: Press Department of Moscow State University for the Humanities, 2009.
8. B.A. Kartosia, "The development of the underground space is a global problem of science, production and higher mountain education", *Materials conf. prospects for the development of underground space*, pp. 12–26, 2010.

9. R. Sterling et al., “Sustainability issues for underground spaces in urban areas”, *Proceedings of ICE. Urban Design and Planning*, vol. 165, issue 4, pp. 241–254, 2012. Available: <https://doi.org/10.1680/udap.10.00020>
10. C.L. Owen and C. Bezerra, “Evolutionary Structured Planning. A Computer-Supported Methodology for the Conceptual Planning Process”, in JS Gero (ed.) *Artificial Intelligence in Design '00*. Dordrecht: Kluwer Academic Publishers, 2000, pp. 287–307.
11. P.H. Gilbert et al., *Underground Engineering for Sustainable Urban Development*. Washington: The National Academies Press, 2013. doi: <https://doi.org/10.17226/14670>
12. V.I. Resin and Yu.S. Popkov, *Large cities development in conditions of transition economy (system approach)*. Moscow: Knizhnyi dom “LIBROKOM”, 2013.
13. N. Pankratova, I. Savchenko, H. Haiko, and V. Kravets, “System approach to planning urban underground development”, *J. Information Content and Processing*, vol. 6, no. 1, pp. 3–17, 2019.
14. M.Z. Zgurovsky and V.A. Pankratov, “Strategy of innovative development of the region on the basis of the foresight and cognitive modeling methodologies synthesis”, (in Russian), *Systematic research and information technologies*, no. 2, pp. 7–17, 2014.
15. P.A. Mikhnenko, “Dynamic modification of SWOT analysis”, *Economic analysis: theory and practice*, 18 (417), pp. 60–68, 2015.
16. Alptekin Nesrin, “Integration of SWOT Analysis and TOPSIS Method In Strategic Decision Making Process”, *The Macrotheme Review*, 2 (7), 2013.
17. Garcia-Cascale M. Socorro and Lamata M. Teresa, “On rank reversal and TOPSIS method”, *Mathematical and Computer Modelling*, vol. 56, no. 5–6, pp. 123–132, 2012.
18. Mardani Abbas, Zavadskas Edmundas, Govindan Kannan, Senin Aslan, and Jusoh Ahmad, “VIKOR Technique: A Systematic Review of the State of the Art Literature on Methodologies and Applications”, *Sustainability*, 8 (37), pp. 1–38, 2016.
19. K. Gopalakrishnan and V. Vijayalakshmi, “Using Morphological Analysis for Innovation and Resource and Development: An Invaluable Tool for Entrepreneurship”, *Annual Research J. Symbiosis Centre for Management Studies, Pune*, vol. 2, issue 1, pp. 28–36, 2014.
20. T. Ritchey, “Futures Studies using Morphological Analysis”, *Adapted from an article for the UN University Millennium Project: Futures Research Methodology Series*, 2005.
21. M.Z. Zgurovsky and N.D. Pankratova, *System analysis: Theory and Applications*. New York: Springer, Berlin Heidelberg, 2007.
22. Weimer-Jehle Wolfgang, “Cross-impact balances: A system-theoretical approach to cross-impact”, *Technological Forecasting & Social Change*, 73, pp. 334–361, 2006.
23. V. Pankratov, “Development of the approach to formalization of vector’s indicators of sustainable development”, *J. Information Technologies & Knowledge. ITHEA. SOFIA*, vol. 8, no. 3, pp. 203–211, 2014.
24. *Innovative development of socio-economic systems based on foresight and cognitive modelling methodologies*, (in russian); in editors G.V. Gorelova, N.D. Pankratova. Kiev: Nauk. Dumka, 2015.
25. V. Kulba, D.A. Kononov, S.S. Kovalevsky, S.A. Kosyachenko, R.M. Nizhegorodtsev, and I.V. Chernov, *Scenario analysis of the dynamics of behavior of socio-economic systems*, (in russian). M.: IPU RAS, 2002.
26. V.I. Maksimov, “Cognitive technology – from ignorance to understanding. 1st work”, *Cognitive analysis and management of the development of situations, (CASC'2001)*, 1, pp.4–18, 2001.

27. R.H. Atkin, "Combinatorial Connectivities in Social Systems. An Application of Simplicial Complex Structures to the Study of Large Organisations", *Interdisciplinary Systems Research*, 1997.
28. J. Casti, *Connectivity, Complexity, and Catastrophe in Large-scale Systems*. Chichester – New York – Brisbane – Toronto: A Wiley – Interscience Publication International Institute for Applied Systems Analysis. JOHN WILEY and SONS, 1979.
29. F. Roberts, *Graph Theory and its Applications to Problems of Society*. Philadelphia: Society for Industrial and Applied Mathematics, 1978.
30. Program for cognitive modeling and analysis of socio-economic systems at the regional level. Certificate of state registration of computer programs N2018661, 2018.

Received 06.12.2021

INFORMATION ON THE ARTICLE

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

Vladimir A. Pankratov, ORCID: 0000-0002-8264-5835, Institute for Applied System Analysis of the National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine, e-mail: pankratov.volodya@gmail.com

СИСТЕМНИЙ ПІДХІД ДО ПЛАНУВАННЯ ОБ'ЄКТІВ ПІДЗЕМНОГО БУДІВНИЦТВА ЗА МЕТОДОЛОГІЯМИ ПЕРЕДБАЧЕННЯ ТА КОГНІТИВНОГО МОДЕЛЮВАННЯ / Н.Д. Панкратова, В.А. Панкратов

Анотація. Запропоновано системний підхід до планування об'єктів підземного будівництва на основі методологій передбачення та когнітивного моделювання. Використання методології передбачення дає змогу за допомогою процедур експертного оцінювання визначити критичні технології та побудувати альтернативи сценаріїв із кількісними характеристиками. Для обґрунтованої реалізації конкретного сценарію використовується імпульсне когнітивне моделювання, що дозволяє будувати причинно-наслідкові зв'язки на основі знань і досвіду, розуміти та аналізувати поведінку складної системи на стратегічну перспективу з великою кількістю взаємозв'язків і взаємозалежностей. Запропонований системний підхід дозволяє планувати підземні об'єкти на основі вибору обґрунтованих сценаріїв та обґрунтування пріоритетності їх створення.

Ключові слова: передбачення, імпульсне когнітивне моделювання, планування, сценарії, підземне будівництво.

СИСТЕМНЫЙ ПОДХОД К ПЛАНИРОВАНИЮ ОБЪЕКТОВ ПОДЗЕМНОГО СТРОИТЕЛЬСТВА НА ОСНОВЕ МЕТОДОЛОГИЙ ПРЕДВИДЕНИЯ И КОГНИТИВНОГО МОДЕЛИРОВАНИЯ / Н.Д. Панкратова, В.А. Панкратов

Аннотация. Предложен системный подход к планированию объектов подземного строительства на основе методологий предвидения и когнитивного моделирования. Использование методологии предвидения позволяет с помощью процедур экспертного оценивания выявить критические технологии и построить альтернативы сценариев с количественными характеристиками. Для обоснованной реализации того или иного сценария используется импульсное когнитивное моделирование, позволяющее построить причинно-следственные связи на основе знаний и опыта, понять и проанализировать поведение сложной системы на стратегическую перспективу с большим количеством взаимосвязей и взаимозависимостей. Предлагаемый системный подход позволяет планировать подземные объекты на основе выбора обоснованных сценариев и обоснования приоритетности их создания.

Ключевые слова: предвидение, импульсное когнитивное моделирование, планирование, сценарии, подземное строительство.

НЕМИНУЧІ ЗМІНИ В ІТ ІНДУСТРІЇ. ПІДГОТОВКА КАДРІВ В УМОВАХ П'ЯТОЇ ПРОМИСЛОВОЇ РЕВОЛЮЦІЇ (ІНДУСТРІЯ 5.0)

А.І. ПЕТРЕНКО

Анотація. Стрімка еволюція цифрових технологій вносить значні зміни в суспільство. Навколишнє середовище та діяльність людей стають дедалі різноманітнішими та складнішими. Суспільство вступає в новий етап свого розвитку під назвою Індустрія 5.0, орієнтований на людину, у якому реалізуються як умови економічного розвитку, так і соціальних проблем зі сфери розваг, і люди забезпечуються високою якістю життя. Така велика трансформація у новий тип суспільства істотно впливає на організацію освіти як основи Індустрії 5.0. Розглянуто неминучі зміни, що охоплюють усі суттєві елементи організації персоналізованої освіти і потребують об'єднання зусиль ключових груп, включаючи уряди, постачальників освіти та навчання, промисловість, допоміжні структури, широку спільноту та, що найважливіше, самих студентів.

Ключеві слова: Індустрія 5.0, цифрова компетентність, персоналізоване навчання, цифрові технології, цифрові трансформації, цифрове суспільство.

БАЗОВІ ОСОБЛИВОСТІ ІНДУСТРІЇ 5.0

Останні 15 років людство жило в умовах Індустрії 4.0, ері цифрової революції. Готові ми чи ні, але наступна *Індустрія 5.0* уже тут. Хоча багато виробників все ще зайняті розробленням методів і нових технологій для підвищення ефективності та продуктивності — керівного принципу, що лежить в основі Індустрії 4.0, — наступний етап індустріалізації вже на порозі. *Індустрія 5.0* доповнює існуючу парадигму Індустрії 4.0, висвітлюючи дослідження та інновації як *рушійні сили* переходу до стійкої, орієнтованої на людину, та стабільної європейської промисловості. Варто пам'ятати, що близько 90% поточних даних, які існують у світі, було створено лише за останні два роки.

Згідно з базовим документом ЄС [1] слід вирізнити такі ключові положення:

- Промисловість в Європі повинна стати *прискорювачем* і стимулятором змін та інновацій у досягненні соціальних цілей, що виходять за межі створення та зростання кількості робочих місць, щоб стати стійким джерелом процвітання, поміщаючи добробут працівника в центр виробничого процесу.
- Цифровалізація надає галузі безпрецедентні можливості. Розроблення інноваційних технологій, орієнтоване на людину, може підвищити стійкість галузей і зробити їх більш сталими.
- Промисловість повинна включити та зайняти сильну лідерську роль у переході до Green transition. Для цього необхідно розробити *циклічні процеси*, що дозволяють повторно використовувати, перенаправляти та переробляти природні ресурси, зменшувати кількість відходів та вплив на навколишнє середовище. Стійкість означає скорочення споживання енергії та

викидів парникових газів для уникнення виснаження та деградації природних ресурсів, а отже, забезпечення потреб сьогоденних поколінь, не наражаючи на небезпеку потреби майбутніх генерацій.

Перехід до Індустрії 5.0 потребує дій у цілій низці сфер. Недостатньо лише розвивати і використовувати окремі передові технології (хмарні технології, засоби збирання й аналізу Big Data, краудсорсинг, біотехнології, безпілотні автомобілі, 3D-друк, криптовалюти і технології Blockchain, інтернет речей (IoT), штучний інтелект (ШІ) та ін.), необхідне їх інтегроване комплексне **поєднання в кожній галузі та в усіх соціальних сегментах**, які радикально змінять цілі галузі економіки й суспільство в цілому, зможуть вирішити неможливі натеper проблеми, зробивши повсякденне життя більш комфортним і стійким.

Зокрема значно змінюється нова роль галузевого працівника Індустрії 5.0. Сьогодні зростаюча залежність людства від автоматизації та штучного інтелекту порушила деякі етичні питання. Вважається, наприклад, що, як і промислові революції, які відбулися до цього, Індустрія 4.0, можливо, зіграла свою роль у посиленні нерівності серед людей, оскільки найбільшими бенефіціарами від неї є невелика група інноваторів, постачальників та інвесторів, або тих, хто контролює інтелектуальний капітал чи природні ресурси. В умовах Індустрії 5.0 працівник розглядається не як «вартість», а як «інвестиційна» позиція для компанії, що дозволяє компанії та працівникові розвиватися. Це означає, що роботодавець зацікавлений в інвестуванні у компетентності, здібності та благополуччя своїх співробітників для досягнення їх цілей. Такий підхід дуже відрізняється від простого врівноваження витрат на робочу силу з фінансовими доходами: людський капітал більш шанується і цінується. Можна у зв'язку з цим процитувати генерального директора Tesla Ілона Маска, який сказав, що надмірна автоматизація в його компанії «була помилкою» і що «людей недооцінюють».

Важлива передумова Індустрії 5.0 у тому, що **технологія служить людям, а не навпаки**. У промисловому контексті це означає, що технологія, що використовується у виробництві, адаптується до потреб і різноманітності промислових робітників, замість того, щоб вони пристосовувалися до технологій, що постійно розвиваються. У працівника має бути більше можливостей, а робоче середовище має стати більш інклюзивним. Для цього робітники повинні брати активну участь у розробленні та впровадженні нових промислових технологій, включаючи робототехніку та ШІ.

Потенціал робототехніки ще далеко не вичерпано, особливо під час використання ШІ. Технології на основі ШІ, а також інструменти **віртуальної та доповненої реальності** можуть використовуватися для того, щоб спрямовувати співробітника до виконання більш спеціалізованих завдань. Тобто відкриваються можливості для залучення більшої кількості людей до робочого середовища шляхом співпраці людини з роботами на робочому місці замість просто заміни робітників роботами, як то відбувається під час сьогоденної суцільної автоматизації. Індустрія 5.0 має за мету об'єднати найкраще як із цифрового, так і з людського світу, поєднуючи чудові здатності когнітивного оброблення, швидкість та універсальність роботів разом з інтелектом, винахідливістю та креативністю людей, зосереджуючись на синергії між людьми та ШІ.

Фахівці України вже тепер мають змогу детальніше ознайомитися з новими викликами Індустрії 5.0, якщо спробують задовольнити конкурсні вимоги нової програми «Горизонт-Європа», яка почала активно приймати про-

ектні заявки за різними пріоритетами [2]. Відсоток з них, хто розуміється в запропонованих напрямках програми і здатний сформувати проєктні пропозиції, порівняльно нижчий, ніж було в часи попереднього тендеру запитів для програми «Горизонт 2020», яка базувалася на потребах і можливостях Індустрії 4.0.

ВИМОГИ ДО ЦИФРОВОЇ КОМПЕТЕНТНОСТІ ФАХІВЦІВ В УМОВАХ ІНДУСТРІЇ 5.0

Рівень компетентності є ще одним з важливих чинників для Індустрії 5.0. Потреби у компетентності розвиваються так само швидко, як і технології. Європейська промисловість бореться з нестачею компетентності, а освітні та навчальні заклади не в змозі задовольнити цей попит. Це стосується як експертного рівня, так і загальних вимог до цифрової компетентності.

У Програмі підвищення кваліфікації ЄС передбачено низку політичних ініціатив. А саме, в оновленому *Плані дій у галузі цифрової освіти (2021–2027)* викладено бачення Європейської комісії щодо високоякісних, інклюзивних та доступних систем освіти та навчання, що відповідають епосі цифрових технологій [3]. Його два пріоритети спрямовані на сприяння розвитку високоефективної екосистеми цифрової освіти та підвищення цифрової компетентності для цифрової трансформації. Можливим виходом із цієї невідповідності компетентності може бути новий підхід до розвитку технологій. Технологію можна зробити більш інтуїтивно зрозумілою та зручною для користувача, щоб працівникам не були потрібні спеціальні навички для її використання. Крім того, навчання можна розвивати одночасно з цією технологією, тим самим гарантуючи, що наявні компетентності найкраще відповідатимуть вимогам до компетентності у галузі.

У 2013 р. Європейська комісія публікує Digital Competences (DigComp 1.0) і в 2016 р. останню версію Цифрової системи компетентності для громадян (DigComp 2), розроблену відділом людського капіталу та зайнятості (Об'єднаний дослідницький центр) Генерального директорату з питань зайнятості, соціальних питань та інклюзії Європейської комісії [5]. Ключові компоненти цифрової компетентності визначено щодо знань, навичок та умінь, необхідних для володіння цифровою компетентністю. Пізніше версії 2.0 фреймворка надано 21 компетенцію, які визначають 8 рівнів кваліфікації (два базові, два середні, два передові і два вищі рівні), згрупованих у п'ять сфер компетентності (таблиця).

DigComp допомагає людям з недостатніми цифровими навичками визначати найважливіші навички для покращення свого особистого та професійного життя і відслідковувати свій прогрес. Аналогічно шукачі роботи можуть визначити та пояснити свою цифрову компетентність у своїх резюме. Вони можуть безпосередньо використовувати інструменти і методики самооцінювання, які наведено в DigComp. Більше того, роботодавці та постачальники робочих місць повинні використовувати певну «мову», щоб допомогти описати навички для своїх вакансій для створення опису роботи. Цифрова компетентність передбачає впевнене та критичне використання технологій інформаційного суспільства для роботи, відпочинку, спілкування та розвитку суспільства для наступного рівня.

Сфери цифрової компетентності у DigComp 2.0

Компетентність	Сфери компетентності
1. Інформація та цифрова грамотність	<p>1.1. Перегляд, пошук і фільтрація даних, інформації та цифровий контент <i>Формулювання інформаційних потреб, пошук даних, інформації та контенту у цифрових середовищах, здійснення доступу до даних, інформації та контенту і можливість переміщуватися між ними.</i></p> <p>1.2. Оцінювання даних, інформації та цифрового вмісту <i>Порівняльний аналіз та критичне оцінювання достовірності і надійності джерел даних, інформації та цифрового контенту. Аналіз, тлумачення, перевірка достовірності та критичне оцінювання даних, інформації та цифрового контенту.</i></p> <p>1.3. Управління даними, інформацією та цифровим вмістом <i>Організація, зберігання та вибір даних, інформації та контенту у цифрових середовищах. Організація та оброблення їх у структурованому середовищі.</i></p>
2. Спілкування і співпраця	<p>2.1. Взаємодія за допомогою цифрових технологій <i>Взаємодія та допомога широкому спектру цифрових технологій та розуміння, які засоби цифрового зв'язку доречні для даного контексту (щодо культурних, соціальних, специфічних гендерних відмінностей тощо).</i></p> <p>2.2. Обмін за допомогою цифрових технологій <i>Використання відповідних цифрових засобів та технологій для обміну даними, інформацією та цифровим контентом. Діяльність в якості посередника, знання практичних методів посилення та атрибуції.</i></p> <p>2.3. Залучення громадян до цифрових технологій <i>Участь у житті суспільства шляхом використання державних і приватних цифрових послуг. Пошук можливості самовдосконалення та залучення громадян до вживання відповідних цифрових технологій.</i></p> <p>2.4. Співпраця через цифрові технології <i>Застосування цифрових технологій та інструментів для співпраці, спільного створення та розвитку цифрових ресурсів та знань.</i></p> <p>2.5. Мережний етикет <i>Знання правил поведінки та ноу-хау щодо користування цифровими технологіями та взаємодії у цифрових середовищах. Адаптація стратегії комунікації під конкретну аудиторію та врахування різноманітності культур і поколінь у цифрових середовищах.</i></p> <p>2.6. Управління цифровою ідентифікацією <i>Створення однієї чи декількох цифрових ідентичностей та управління ними, уміння захистити власну репутацію, праця з даними, створеними за допомогою декількох цифрових засобів, середовищ і сервісів.</i></p>
3. Створення цифрового контенту	<p>3.1. Розроблення цифрового контенту <i>Створення та редагування цифрового контенту у різних форматах, самовираження цифровими засобами.</i></p> <p>3.2. Інтеграція та повторне розроблення цифрового контенту <i>Зміна, уточнення, вдосконалення та інтегрування інформації та контенту в існуючих масивах знань для створення нових, оригінальних і доречних знань та контенту.</i></p> <p>3.3. Авторські права та ліцензії <i>Розуміння того, як авторське право і ліцензії поширюються на дані, інформацію та цифровий контент. Ставлення до піратського програмного забезпечення та контенту.</i></p> <p>3.4. Програмування <i>Планування і розроблення послідовності зрозумілих інструкцій для обчислювальних систем для вирішення певної проблеми чи для виконання конкретного завдання.</i></p>

Продовження таблиці

Компетентність	Сфери компетентності
4. Безпека	<p>4.1. Захисні пристрої Захист пристроїв та цифрового контенту, розуміння ризиків та загроз у цифрових середовищах. Знання заходів безпеки та захисту і належним чином урахування вимог до надійності та приватності.</p> <p>4.2. Захист персональних даних та конфіденційності Захист персональних даних та приватності у цифрових середовищах. Розуміння того, як користуватися та обмінюватися інформацією, яка дозволяє встановити особу, зі збереженням можливості захистити себе та інших від шкоди. Розуміння того, що цифрові служби користуються «Політикою конфіденційності» для інформування про те, як використовуються персональні дані.</p> <p>4.3. Охорона здоров'я та благополуччя Уміння уникати ризиків для здоров'я і загроз для фізичного та психологічного благополуччя за користування цифровими технологіями. Уміння захистити себе та інших від можливих небезпек у цифрових середовищах (наприклад, від кіберзалякування). Знання цифрових технологій для забезпечення соціального благополуччя та соціальної інтеграції.</p> <p>4.4. Захист навколишнього середовища Усвідомлення впливу цифрових технологій та їх користування на навколишнє середовище.</p>
5. Вирішення проблем	<p>5.1. Розв'язання технічних проблем Виявлення технічних проблем під час експлуатації пристроїв і користування цифровими середовищами та розв'язання їх (від пошуку несправностей до розв'язання складніших проблем).</p> <p>5.2. Визначення потреб і технологічних заходів реагування Оцінювання потреб, визначення, відбір та використання цифрових засобів і можливих технологічних заходів реагування для задоволення цих потреб. Налаштування та пристосовування цифрових середовищ згідно з особистими потребами (наприклад, для забезпечення доступності).</p> <p>5.3. Творче використання цифрових технологій Використовування цифрових засобів та технологій для створення знань і внесення новаторських змін у процеси та продукцію. Індивідуальна і колективна участь у пізнавальному опрацюванні з метою розуміння і розв'язання концептуальних проблем і проблемних ситуацій у цифрових середовищах.</p> <p>5.4. Виявлення прогалів у цифровій компетенції Розуміння того, в яких аспектах цифрової компетентності особи необхідно підвищити або оновити. Уміння підтримувати інших у розвитку їх цифрової компетентності. Пошук можливості для саморозвитку та подальшого навчання, не відстаючи від процесу еволюції цифрових технологій.</p>

На базі компетенцій DigComp з невеликими модифікаціями у 2016 р. створено інструмент під назвою PIX10 для вимірювання та сертифікації цифрової компетентності громадян, ініційований Міністерством освіти Франції [7]. Сьогодні будь-який франкомовець може використовувати його для оцінювання рівня цифрової компетентності.

Підхід до оцінювання, застосований платформою PIX 10, базується на завданнях і реалізується онлайн-процедурами, які користувачі виконують у власному темпі та часі, практикуючись та перевіряючи свою цифрову компетентність у досить ігровій манері. Прикладом може бути завдання, коли з метою перевірити, чи користувач оволодів поняттями «Копія» та «Прихована копія» з електронної пошти, PIX 10 пропонує йому відправити безпосередньо електронний лист на певну адресу, указавши спеціальні дані в полях

«Копія» та «Прихована копія». Неважливо чи він використовує свій ноутбук або телефон, бот перевірить чи насправді він надіслав електронного листа та підтвердив свою відповідь.

Слід відзначити, що досвід Франції був використаний ЄС [8] і Україною, зокрема, у підготовці Постанови КМУ № 167-р від 03.03.21 «Про схвалення Концепції розвитку цифрових компетентностей в суспільстві України та затвердження плану заходів щодо її реалізації» [9]. При цьому в 2021 р. адаптована рамка DigComp UA для громадян включає 5 вимірів, 26 компетентностей та 6 рівнів оволодіння кожною компетенцією [10]. Додано нульовий вимір «Основи комп'ютерної грамотності». Скорочення рівнів цифрової компетентності для українців за рахунок вилучення найвищих з них видається невиправданим, бо закріплює відставання вітчизняної ІТ індустрії від зарубіжної і не сприяє підготовці до впровадження Індустрії 5.0.

НЕОБХІДНІ ЗМІНИ В ОСВІТІ

Цифрові компетентності — не єдині компетентності, які будуть актуальними для промислових робітників на підприємствах майбутнього. Всесвітній виробничий форум визначив 10 основних навичок, які будуть потрібні в майбутньому виробництві. Дивно, але лише чотири з них належать до цифрових компетентностей: «цифрова грамотність, штучний інтелект та аналітика даних», «робота з новими технологіями», «кібербезпека» та «уважність до даних». Інші компетентності — це наскрізні компетентності, пов'язані з **творчим, підприємницьким, гнучким і неупередженим мисленням**.

Компанії можуть і повинні відігравати більш важливу роль в освіті та навчанні персоналу, оскільки вони мають досвід, знання, а також прямий зв'язок з технологіями. Вони знають, яких компетентностей не вистачає, а які будуть потрібні в майбутньому.

Освіта 5.0 починається з людей, а не з технологій. Її мета безпосередньо стосується конкретних результатів, яких люди повинні досягти в результаті конкретного досвіду навчання. Мова не йде про забезпечення кожного слухача ноутбуком чи планшетом, про покращення інфраструктури та зв'язку чи про розроблення цифрових інструментів і платформ, а **йде натомість про підготовку інтелектуально, соціально та емоційно сильних особистостей**, які турбуються передусім про особистісний розвиток як загальної мети та своє здоров'я. Потім ідуть відповідні стратегічні, методичні та педагогічні підходи. Останні, серед іншого, включають способи повернути слухачам мотивацію, творчість та радість навчання. Цифрове обладнання, інфраструктура та платформи все ще можуть мати вирішальне значення тут, однак вони є *факторами, а не метою*.

Освіта 5.0 зокрема включає [4, 12]:

- визнання людських якостей центральним компонентом освіти, визначення компетентностей та ролей, які найкраще виконують люди (наприклад, пов'язані з інноваційністю, творчістю, критичним мисленням, аналітичними здібностями, дизайном, співчуттям тощо) та культивування їх;
- урахування не лише потреб ринку/компанії (можливість працевлаштування), а і потреб суспільства та слухачів;
- пропонування «загальної освіти», маючи на увазі загальну картину того, як освітня пропозиція вписується в загальну траєкторію навчання, ринок праці та розвиток у світі;

- розглядання слухачів як агентів змін і їх активне залучення до розроблення та впровадження навчальних програм;
- навчання слухачів пам'ятати про свою взаємодію з технологіями, зокрема про безпеку та ергономіку вдома, в навчальному закладі та на роботі, про необхідність підтримувати гарне фізичне та психічне здоров'я та можливі наслідки надмірного/неналежного впливу технологій, включаючи те, як можна їх уникнути;
- підготовка слухачів до навчання протягом усього життя, переконання, що освітня пропозиція розвиває здатність та готовність студентів до безперервного навчання протягом усього професійного життя;
- забезпечення свободи цілей навчальної програми та результатів навчання від звичайних рамок кваліфікацій, щоб впровадити відповідне персоналізоване та особисте навчання.

Як можна цього досягнути? Перехід до Освіти 5.0 вимагає цілісної освітньої трансформації та передбачає такі елементи [4]:

- *стратегію*: перевизначення ключових цілей та конкретних завдань освітніх пропозицій у контексті Освіти 5.0 (може, навіть переліку спеціальностей);
- *співпрацю*: підтримання практик, які виходять за межі типових інституційних моделей співпраці, зокрема створення ефективних навчальних екосистем, які залучають усі ключові групи зацікавлених сторін (окремих осіб та громад);
- *зміст*: визначення, розроблення та впровадження контенту, який відповідає елементу стратегії (включаючи адекватний баланс технічних і нетехнічних дисциплін, приділяючи особливу увагу питанням етики, соціальної інтеграції, різноманітності та сталості тощо);
- *навчальне середовище*: створення навчального середовища, яке найкращим чином відповідає конкретним цілям елемента стратегії (наприклад, підходи, що стимулюють міждисциплінарну орієнтацію, дизайнерське мислення, командний дух, колективне вирішення проблем, поведінку ризику, експериментальні підходи тощо);
- *механізми реалізації*: визначення інструментів, які найкраще підходять для досягнення цілей елемента стратегії; саме тут технологію можна вибирати або не вибирати як найбільш відповідний механізм доставки;
- *оцінювання та визнання*: дослідження та розроблення відповідних формальних та неформальних способів оцінювання та визнання для Освіти 5.0;
- *забезпечення якості*: розроблення конкретних критеріїв якості для Освіти 5.0 та безперервний моніторинг якості.

Яка роль ЄС та національних урядів? Уряди відіграють ключову роль у просуванні концепції Освіти 5.0. Можливі конкретні заходи:

- спеціально стимулюють ініціативи своєчасного переходу суспільства до цінностей Індустрії 5.0;
- сприяють розвитку високоефективної цифрової освітньої екосистеми, зокрема, замість того, щоб ставити технології в центр освітнього порядку денного, вирішують питання цифрової трансформації на досягнення бажаних результатів навчання для людей;
- заохочують ініціативи щодо визначення та поширення передового досвіду для кожного зі згаданих вище елементів Освіти 5.0;
- заохочують ініціативи щодо розроблення загальних рекомендацій для постачальників освіти та навчання щодо розроблення та впровадження Освіти 5.0 і стимулюють прийняття цих рекомендацій;

- заохочують ініціативи щодо моніторингу, аналізу та запобігання «поганим практикам» цифрової освіти, які призводять до зниження мотивації, зниження продуктивності та проблем зі здоров'ям слухачів;
- розробляють рекомендації для постачальників освіти та слухачів щодо захисту даних і конфіденційності та інших аспектів цифрової етики;
- приділяють особливу увагу створенню здорового середовища в навчальних закладах, їх технічному оснащенню у зв'язку зі стрімким зростанням використання технологій;
- вирішують виклики та можливості пандемії Covid-19, що призвела до безпрецедентного використання технологій, зокрема в цілях освіти та навчання.

ВИСНОВКИ

У зв'язку з особливостями економічного розвитку України загальне впровадження цифрових технологій відбувалось і відбувається не поступово, а стрибкоподібно. В Україні тільки почали зникати до терміна Індустрії 4.0, увівши його недавно в базові документи [9, 10, 14, 15], що визначають розвиток держави на наступні 5–15 років, включаючи заходи з адаптації до нових умов освіти [16, 17]. А тут світ вже розпочав будувати Індустрію 5.0 і знову доведеться наздоганяти і переважаюватися, хоч Індустрія 5.0 — це суспільство, яке вирішує різноманітні соціально-технічні проблеми, використовуючи різні інновації, які створено в епоху промислової Індустрії 4.0.

Основні висновки про наслідки Індустрії 5.0 для інженерної освіти, які змогли б допомогти університетам переосмислити свою освітню та дослідницьку діяльність, передбачають уведення нових, а не просто удосконалених компетентностей, які можна визначити як **об'єднані компетентності** (fusion skills) з посилення обміну знаннями, людьми, ідеями і технологіями між університетами та їхніми партнерами в спосіб, який більше узгоджується з наступним переваженням до нового типу суспільства [13]. Університети найкраще працюють як рушії економічного та соціального розвитку, коли вони систематично обмінюються знаннями зі своїми партнерами в промисловості, з академічної спільноти та в уряді, наприклад, з явним наміром покращити соціальні результати інженерії із врахуванням бачення типу суспільства та життя, яке своїм напрацюванням створюють ініціативні, талановиті та екологічно відповідальні сучасні інженери і інженери майбутнього. Тобто університети забезпечують перехрестя, на якому фундаментальні наукові дослідження та освіта перетинаються безпосередньо із суспільно важливими програмами від промисловості, урядів, неурядових організацій.

Розвиток і впровадження об'єднаних компетентностей потребують спільних досліджень у межах та між науковими дисциплінами (галузями) — або навіть відкидання ідеї дисципліни як організаційного принципу університетських досліджень — і збирання спільних команд для проведення досліджень, які є *проблемо-орієнтованими*, а не дисциплінарними, і залучають промислових та академічних партнерів і співробітників, оскільки вони є важливими для розвитку та впровадження об'єднаних компетентностей. Конче необхідні нові професійні курси, що варіюють від цифрової трансформації, стійких цифрових інновацій до наслідків штучного інтелекту для бізнесу і можливості інтелектуальної автоматизації та прийняття рішень, «блокчейніфікації» у виробничій та логістичній промисловості, квантових обчислень. Але щоб підготувати фахівців до швидких технологічних змін, ключем є також зосередження під час навчання на творчих здібностях людини, на

людських навичках, таких як комунікація, лідерство та витривалість, а також на цікавості, розумінні та навичках читання. Тобто гуманітарна складова освіти повинна бути значно посилена і, може, традиційне ділення спеціальностей на технічні і гуманітарні відійде в минуле. У світі, де технології інтегровані майже в кожен частину суспільства, цей підхід більше не буде практичним. Студенти повинні оволодіти новими об'єднаними навичками, щоб вижити в суспільстві, що змінюється, і щоб керувати цими змінами. У майбутньому, на думку фахівців Японії, яка випереджає інші країни світу в побудові суспільства Індустрії 5.0, буде переважати система освіти, в якій такі предмети, як математика, наука про дані та програмування, є базовими разом з такими предметами, як філософія та мови. «Якщо ви вивчаєте фізику за спеціальністю, ви повинні також вивчати гуманітарні науки, щоб, коли ви зіткнетеся з філософським чи етичним питанням у вашій майбутній кар'єрі, ви могли би поєднати свої наукові знання з етикою» [18, 20].

Для навчання студентів новітніх компетенцій та вдосконалювання рівня їх мислення виникає потреба у висококваліфікованих викладачах-вихователях 5.0, які здатні навчити стимулювати й застосовувати людський інтелект і мислення у процесі спільної роботи роботів і людей в комп'ютерному середовищі. Такі роботи інколи називають коботами (Cobots, Collaborative Robots) [19], а вихователі в *персоналізованому режимі* (як індійські мудрі «гуру») будуть готувати фахівців комп'ютерного середовища, у якому машини з когнітивною досконалістю можуть працювати у співпраці між собою та співпраці з людьми. А хто і як буде готувати самих вихователів?! Які зусилля треба прикласти для того, щоб викладачі в епоху промислової революції 4.0 придбали компетенції, необхідні для підготовки фахівців суспільства Індустрія 5.0? Які основні компетенції людини, необхідні для співпраці з ШІ?

Зрозуміло, що основою Індустрії 5.0 є персоналізована освіта 5.0 [21]. Перед теперішніми викладачами під час великого перезавантаження до нового типу суспільства постає завдання або отримання відповідної перепідготовки до цієї революційної зміни, або прийняти зречення. Пропонується провести відповідну дискусію з цього приводу та приділити їй серйозну увагу.

ЛІТЕРАТУРА

1. *Індустрія 5.0*. Доступно: https://ec.europa.eu/info/news/industry-50-towards-moresustainable-resilient-and-human-centric-industry-2021-jan-07_en
2. *Програма Горизонт-Європа*. Доступно: https://ec.europa.eu/deface-industry-space/horizon-europe-work-programme-2021-2022-adopted-2021-06-29_en
3. *Digital Education Action Plan (2021–2027)*. Available: https://ec.europa.eu/education/education-in-the-eu/digital-education-action-plan_en
4. *PwC Position Paper 2020*. Available: <https://www.pwc.com/gx/en/about/pdf/pwc-response-on-digital-education-action-plan.pdf>
5. *DigComp 2.1. The digital competence framework for citizens with eight proficiency levels and examples of use*. Available: https://www.researchgate.net/publication/317013679_DigComp_21_The_digital_competence_framework_for_citizens_with_eight_proficiency_levels_and_examples_of_use
6. Vuorikari Riina, Yves Punie, Stephanie Carretero Gomez, and Godelieve Van Den Brande, *DigComp 2.0: The Digital Competence Framework for Citizens. Update Phase 1: the Conceptual Reference Model. Technical Report*. Institute for Prospective Technological Studies, Joint Research Centre.
7. Jill-Jënn Vie, Benjamin Marteau, Nathalie Denos, and Françoise Tort, *PIX: A Platform for Certification of Digital Competencies*. Available: https://jill-jenn.net/_static/works/pix-a-platform-for-certification-of-digital-competencies.pdf
8. *Digital Competence Framework for Educators (DigCompEdu)*. Available: <https://ec.europa.eu/jrc/en/digcompedu>

9. Про схвалення Концепції розвитку цифрових компетентностей в суспільстві України та затвердження плану заходів щодо її реалізації, Постанова КМУ № 167-р.
10. *Опис рамки цифрової компетентності для громадян України (DigComp UA for Citizens)*. Доступно: https://thedigital.gov.ua/storage/uploads/files/news_post/2021_13/mintsifraprilyudnyue-ramku-tsifrovoi-kompetentnosti-dlya-gromadyan/%D0%9E%D0%A0%20%D0%A6%D0%9A.pdf
11. В.В. Грядуща та А.В. Денисова, “Мобільна цифрова компетентність як складова цифрової компетентності педагогів закладів професійної (професійно-технічної освіти) України”, *II Міжнародна науково-практична конференція 29.10.2021, Кембридж, GBR*. Доступно: <https://doi.org/10.36074/logos-29.10.2021.v2.17>
12. Kristina Derojeda, *Education 5.0: Rehumanizing Education in the Age of Machines*. Available: <https://www.linkedin.com/pulse/education-50-rehumanising-age-machines-kristina-derojeda/>
13. John Mitchell and David Guile, *Fusion Skills and Industry 5.0: Conceptions and Challenges*. IntechOpen, 2021. doi:10.5772/intechopen.100096. Available: <https://www.intechopen.com/online-first/78652>
14. *Стратегія розвитку вищої освіти в Україні на 2021–2031 роки*. Доступно: http://www.reform.org.ua/proj_edu_strategy_2021-2031.pdf
15. *Цифрові компетенції як умова формування якості людського капіталу: аналіт. зап.*; [В.С. Куйбіда, О.М. Петроє, Л.І. Федулова, Г.О. Андрощук]. Київ : НАДУ, 2019, 28 с.
16. Сергій Войтко, *Трансформація ЗВО в умовах Індустрії 4.0: роздуми, навіяні конференцією*. Доступно: <https://kpi.ua/2020-kr2-2>
17. *Вища освіта назустріч четвертій промисловій революції: кейси з європейського та українського досвіду*; під ред. В.Шатохи. Дніпро: «Поліграфічна акцидентна фірма», 2021, 68 с.
18. Vivek Kumar, *Industry 5.0: How It will Bring Humanity at Workplace?* Available: <https://industrywired.com/industry-5-0-how-it-will-bring-humanity-at-workplace/>
19. Kadir Alpaslan Demir, *Industry 5.0 and Human-Robot Co-working*. Available: https://www.academia.edu/42263444/Industry_5.0_and_Human-Robot_Co-working
20. *How Japan is Preparing its Students for Society 5.0*. Available: <https://foreignpolicy.com/sponsored/how-japan-is-preparing-its-students-for-society-5-0/>
21. *Educational Innovation in Society 5.0 Era: Challenges and Opportunities*. Purnomo & Herwin (Eds), 2021. Available: <https://www.routledge.com/Educational-Innovation-in-Society-50-Era-Challenges-and-Opportunities/Purnomo-Herwin/p/book/9781032053929>

Надійшла 31.01.2022

INFORMATION ON THE ARTICLE

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

INEVITABLE CHANGES IN THE IT INDUSTRY. TRAINING IN THE CONDITIONS OF THE FIFTH INDUSTRIAL REVOLUTION (INDUSTRY 5.0) / A.I. Petrenko

Abstract. The rapid evolution of digital technologies is bringing significant changes to society. The environment and human values are becoming more diverse and complex. The society is entering a new phase of its development called Industry 5.0, a human-centered one that realizes both economic development and entertainment with social issues where people can enjoy a high quality of life. Such a major transformation of the new type of a society significantly affects the organization of education as the basis of Industry 5.0. The paper examines these inevitable changes, covering all the essential elements of personalized education and requiring the combined efforts of all key stakeholder groups, including governments, education and training providers, industry, support bodies, the wider community and, most importantly, students themselves.

Keywords: Industry 5.0, digital competence, personalized learning, digital technologies, digital transformations, digital society.

НЕИЗБЕЖНЫЕ ИЗМЕНЕНИЯ В ИТ ИНДУСТРИИ. ПОДГОТОВКА КАДРОВ В УСЛОВИЯХ ПЯТОЙ ПРОМЫШЛЕННОЙ РЕВОЛЮЦИИ (ИНДУСТРИЯ 5.0) / А.И. Петренко

Аннотация. Стремительная эволюция цифровых технологий вносит значительные изменения в жизнь общества. Окружающая среда и деятельность людей становятся все более разнообразными и сложными. Общество вступает в новый этап своего развития под названием Индустрия 5.0, который ориентирован на человека и в котором реализуются как условия экономического развития, так и социальных проблем

со сферы развлечения, и люди обеспечиваются высоким качеством жизни. Такая большая трансформация в новый тип общества оказывает существенное влияние также на организацию образования как основы Индустрии 5.0. Рассмотрены неизбежные изменения, охватывающие все существенные элементы организации персонализированного образования и требующие объединения усилий ключевых групп, включая правительства, поставщиков образования и обучения, промышленность, вспомогательные структуры, общественность и, что самое важное, самих студентов.

Ключевые слова: Индустрия 5.0, цифровая компетентность, персонализированное обучение, цифровые технологии, цифровые трансформации, цифровое общество.

REFERENCES

1. *Industry 5.0*. Available: https://ec.europa.eu/info/news/industry-50-towards-more-sustainable-resilient-and-human-centric-industry-2021-jan-07_en
2. *Horizon Europe work programme 2021-2022*. Available: https://ec.europa.eu/defence-industry-space/horizon-europe-work-programme-2021-2022-adopted-2021-06-29_en
3. *Digital Education Action Plan (2021–2027)*. Available: https://ec.europa.eu/education/education-in-the-eu/digital-education-action-plan_en
4. *PwC Position Paper 2020*. Available: <https://www.pwc.com/gx/en/about/pdf/pwc-response-on-digital-education-action-plan.pdf>
5. *DigComp 2.1. The digital competence framework for citizens with eight proficiency levels and examples of use*. Available: https://www.researchgate.net/publication/317013679_DigComp_21_The_digital_competence_framework_for_citizens_with_eight_proficiency_levels_and_examples_of_use
6. Vuorikari Riina, Yves Punie, Stephanie Carretero Gomez, and Godelieve Van Den Brande, *DigComp 2.0: The Digital Competence Framework for Citizens. Update Phase 1: the Conceptual Reference Model. Technical Report*. Institute for Prospective Technological Studies, Joint Research Centre.
7. Jill-Jënn Vie, Benjamin Marteau, Nathalie Denos, and Françoise Tort, *PIX: A Platform for Certification of Digital Competencies*. Available: https://jill-jenn.net/_static/works/pix-a-platform-for-certification-of-digital-competencies.pdf
8. *Digital Competence Framework for Educators (DigCompEdu)*. Available: <https://ec.europa.eu/jrc/en/digcompedu>
9. *On approval of the Concept of development of digital competencies in the society of Ukraine and approval of the action plan for its implementation, Resolution of the Cabinet of Ministers of Ukraine, № 167-p., 03.03.2021*
10. *Description of the framework of digital competence for the citizens of Ukraine (DigComp UA for Citizens)*. Available: https://thedigital.gov.ua/storage/uploads/files/news_post/2021/3/mintsifra-oprilyudnyue-ramku-tsifrovoi-kompetentnosti-dlya-gromadyan/%D0%9E%D0%A0%20%D0%A6%D0%9A.pdf
11. V.V. Hryadushcha and A.V. Denisova, “Mobile digital competence as a component of digital competence of teachers of vocational (vocational education) institutions of Ukraine”, *Second International scientific-practical conference 29.10.2021, Cambridge, GBR*. Available: <https://doi.org/10.36074/logos-29.10.2021.v2.17>
12. Kristina Dervojeđa, *Education 5.0: Rehumanizing Education in the Age of Machines*. Available: <https://www.linkedin.com/pulse/education-50-rehumanizing-age-machines-kristina-dervojeđa/>
13. John Mitchell and David Guile, *Fusion Skills and Industry 5.0: Conceptions and Challenges*. IntechOpen, 2021. doi: 10.5772/intechopen.100096. Available: <https://www.intechopen.com/online-first/78652>
14. *Strategy for the development of higher education in Ukraine for 2021-2031*. Available: http://www.reform.org.ua/proj_edu_strategy_2021-2031.pdf
15. *Digital competences as a condition for the formation of the quality of human capital: analytic review*; [V.S. Kuybida, O.M. Petrov, L.I. Fedulova, G.O. Androschuk]. Kyiv: National Academy of Public Administration, 2019, 28 p.
16. Serhiy Voitko, *Transformation of Free Economic Zone in the Conditions of Industry 4.0: Reflections Inspired by the Conference*. Available: <https://kpi.ua/2020-kp2-2>
17. *Higher education towards the fourth industrial revolution: cases from European and Ukrainian experience (ed. V.Shatokhy)*. Dnipro: “Polihrafichna aktydentna firma”, 2021, 68 p.
18. Vivek Kumar, *Industry 5.0: How It will Bring Humanity at Workplace?* Available: <https://industrywired.com/industry-5-0-how-it-will-bring-humanity-at-workplace/>
19. Kadir Alpaslan Demir, *Industry 5.0 and Human-Robot Co-working*. Available: https://www.academia.edu/42263444/Industry_5.0_and_Human-Robot_Co-working
20. *How Japan is Preparing its Students for Society 5.0*. Available: <https://foreignpolicy.com/sponsored/how-japan-is-preparing-its-students-for-society-5-0/>
21. *Educational Innovation in Society 5.0 Era: Challenges and Opportunities*. Purnomo & Herwin (Eds), 2021. Available: <https://www.routledge.com/Educational-Innovation-in-Society-50-Era-Challenges-and-Opportunities/Purnomo-Herwin/p/book/9781032053929>

THE PROOF OF HYPOTHESIS REGARDING DISTORTION OF TIME AND SPACE USING THE NUCLEAR FUSION MODEL

Y. MATSUKI, P.I. BIDYUK

Abstract. The study is dedicated to modern topic: the analysis of conditions that lead to distortion of the time and space coordinates which results from the general theory of relativity. The main goal of this research is to prove the hypothesis regarding distortion of time and space using nuclear fusion model. For this purpose the simulation instrument is used to imitate a moving proton that hits an electron of a hydrogen atom. The methodology of simulation is based upon calculation of the probabilities of elastic scattering and charge exchange of a proton with a target electron. The distortion is modeled by the functions that relate time and space logarithmically for distorted time and exponentially for distorted space. Such geometry construction is described by the Schrödinger equation using the electron wave function. Then the probability of charge exchange is calculated as the squared coefficient of this wave function in the negative side of the geometry that is divided by the sum of the squared coefficients of all the terms of the equation. Thus, the calculation result shows that the calculated probability of the charge exchange is high when the time and space are not distorted. However, when time and space are distorted it decreases, and the probability of elastic scattering is growing. The achieved result also indicates that the discrete energy levels of electrons in hydrogen atoms shift when the distortion of time and space occurs in the nuclear fusion.

Keywords: general theory of relativity, nuclear fusion, distortion of time and space, charge exchange.

INTRODUCTION

In our previous research [2–8], we have found that a rotating ultra-heavy mass that distorts time and space produces anti-gravity. Anti-gravity is another gravity that holds the opposite sign, + or –, of the physical properties (energy intensity and angular momentum) from those of normal gravity. Then we predicted that anti-gravity should expand the size of the Universe. In addition, we developed the concept of the flying craft having a disk-shaped body for interstellar travel. The explicit images of the ultra-heavy mass that distorts time and space are a black hole and a nuclear fusion reactor as they continue producing heavier nuclei by fusing the lighter nuclei such as hydrogen, deuterium and tritium. By our previous research we have so far completed the numeric simulation of a rotating ultra-

heavy mass by applying the general theory of relativity with the mathematics of the 4-dimensional tensors. However, the distortion of time and space is still an unproven hypothesis.

A research question remains: whether the nuclear fusion relates to the distortion of time and space, or not. To answer this question, we used an approximation method. First, we have prepared an input dataset that simulates the infinite 4-dimensional time and space, by replacing them with the finite discrete values. Then we set a geometrical placement of an electron in the field of two fixed protons with a separation in two-dimensional coordinates. With this setting, we calculated the probabilities of the proton's elastic scattering and its charge exchange by the approximation method of quantum mechanics [9, pp. 95–97]. The probability of charge exchange leads to the fusion of the proton and the hydrogen-atom. Also, we simulated the distortion of time and space by the same method of our previous research [2–8], which was based on the general theory of relativity [1 pp. 32–36]. Finally, by comparing the calculated probabilities with and without the distortion of time and space, we examined the relation between nuclear fusion and the distortion of time and space.

METHOD

Probabilities of charge exchange and elastic scattering by a moving proton

We have simulated a fusion of two hydrogen-atoms. But, to simplify the simulation, we calculated the probabilities of moving proton's charge exchange and its elastic scattering with a target electron of a hydrogen atom. (Note: the moving proton is called incident proton.) The reference [9, p. 95] sets a system of an electron in the field of fixed two protons. According to the reference, we set the following linear operator that is called Hamiltonian, considering the proton and the hydrogen atom as two interacting quantum mechanical system:

$$H = -\frac{1}{2}\nabla_r^2 - \frac{1}{\left|r - \frac{1}{2}R\right|} - \frac{1}{\left|r + \frac{1}{2}R\right|} + \frac{1}{R} - \frac{1}{M}\nabla_R^2. \quad (1)$$

It is an operator that forms Schrödinger's equation shown below with wave functions, $\varphi(r, R)$, and $X(R)$. It consists of five components. First, $-\frac{1}{2}\nabla_r^2$, is for the kinetic energy of electron where electron's mass is 1 in atomic units, and where ∇_r^2 is an operator that makes $\left(\frac{\partial\varphi(r, R)}{\partial r}\right)^2$, and $\varphi(r, R)$ is a wave function of electron's coordinate r and proton's coordinate R . Second; $-\frac{1}{M}\nabla_R^2$ is for the kinetic energy of the moving proton and ∇_R^2 is an operator that makes $\left(\frac{\partial X(R)}{\partial R}\right)^2$, where $X(R)$ is a wave function of the proton coordinate R . M is the mass of the proton. (Note: $M\nabla_R X(R)$ is the momentum p of a proton, and $p^2/2M$ is the kinetic energy of a proton; then $\frac{[\nabla_R X(R)]^2}{M}$ is the total kinetic energy of two protons.) The other three terms of (1) are for the potential energies:

1) potential energy of hydrogen atom with the target electron in its initial position

$$-\frac{1}{\left|r - \frac{1}{2}R\right|};$$

2) potential energy of hydrogen atom with its electron after charge exchange

$$-\frac{1}{\left|r + \frac{1}{2}R\right|};$$

3) potential energy of hydrogen's proton

$$+\frac{1}{R}.$$

The potential energy of the target electron 1) induces the proton's elastic scattering, and the potential energy of the target electron 2) induces its charge exchange with the incident proton. The proton's potential energy 3) induces both elastic scattering and charge exchange of the slowly moving incident proton.

Because the proton is moving slower than the hydrogen atom's electron, $-\frac{1}{M}\nabla_R^2$, is called the slow subsystem H_2 of the Hamiltonian, and $-\frac{1}{2}\nabla_r^2 - \frac{1}{\left|r - \frac{1}{2}R\right|} - \frac{1}{\left|r + \frac{1}{2}R\right|} + \frac{1}{R}$, is called the fast subsystem H_1 of the Hamiltonian. Then (1) becomes as follows:

$$H = H_1 + H_2.$$

Here H_1 plays the role of potential energy while H_2 plays the role of kinetic energy in the Hamiltonian (1). Then the following Schrödinger's equation gives the solution of the problem:

$$\begin{aligned} H\varphi_n(r,R)X(R) &= \left(-\frac{1}{2}\nabla_r^2 - \frac{1}{\left|r - \frac{1}{2}R\right|} - \frac{1}{\left|r + \frac{1}{2}R\right|} + \frac{1}{R} - \frac{1}{M}\nabla_R^2 \right) \varphi_n(r,R)X(R) = \\ &= (H_1 + H_2)\varphi_n(r,R)X(R) = E_n\varphi_n(r,R)X(R). \end{aligned}$$

Henceforward we focus on the potential energy H_1 of Schrödinger's equation:

$$\begin{aligned} H_1\varphi_n(r,R)X(R) &= \left(-\frac{1}{2}\nabla_r^2 - \frac{1}{\left|r - \frac{1}{2}R\right|} - \frac{1}{\left|r + \frac{1}{2}R\right|} + \frac{1}{R} \right) \varphi_n(r,R)X(R) = \\ &= H_1\varphi_n(r,R)X(R) = \varepsilon_n\varphi_n(r,R)X(R), \end{aligned} \tag{2}$$

where $\varphi_n(r, R)$ $X(R)$ is the solution known as a wave function and ε_n is the eigenvalue of the potential energy. The wave functions are continuous, but the equation (2) leads to the discrete eigenvalues that define the energy states of the hydrogen atom as the suffix n gives discrete numbers, 0, 1, 2..., that define the discrete energy states of the hydrogen atom with an electron. Hence, the hydrogen-atom has a discrete energy spectrum in its atomic structure, while the slowly moving proton has a continuous energy spectrum.

Then the coefficients of the Schrödinger equation are calculated as solution of the problem of an electron in the field of two fixed protons with separation (distance) of R , as shown in Fig. 1 in a flat $x - y$ coordinate system. Here, r gives the distance between the origin O of the coordinates and an electron, which is indicated in the plane polar coordinates. Here, $(1/2)R$, and $-(1/2)R$ are the locations of two protons on the x-axis; and, $r - (1/2)R$ is the length of the vector between the proton located at $(1/2)R$ and the electron, and $r + (1/2)R$ is the length of the vector between the proton located at $-(1/2)R$ and the electron, indicated in plane polar coordinates.

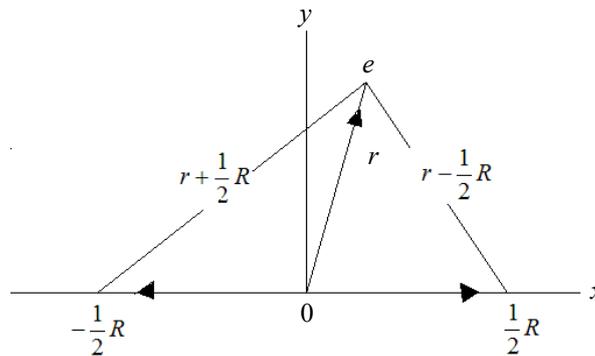


Fig. 1. Coordinates of two protons and an electron (modified from [9], p. 89, Fig. 21)

The Hamiltonian (1) is symmetric ($r \rightarrow -r$) with respect to interchange of the incident particle (a slowly moving proton) and target particle (an electron of the hydrogen-atom). Here, the interchange of the particles means that the proton's + charge changes to -. Then symmetric and anti-symmetric functions $\varphi_n^S(r, R)$ and $\varphi_n^A(r, R)$ are introduced. When the nucleons are far apart, the electron will be localized near one or the other proton, therefore we have:

$$\Phi_n^{S,A}(r, R) \xrightarrow{R \rightarrow \infty} \frac{1}{\sqrt{2}} \left[\Phi_{n_0}^Q \left(r - \frac{1}{2}R \right) \pm \Phi_{n_0}^Q \left(r + \frac{1}{2}R \right) \right], \quad (3)$$

where $\frac{1}{\sqrt{2}} \left[\Phi_{n_0}^Q \left(r - \frac{1}{2}R \right) \pm \Phi_{n_0}^Q \left(r + \frac{1}{2}R \right) \right]$ are hydrogen-atom's wave functions [8, p. 96].

The initial condition of the simulation is that the electron is attached to the proton at $+R/2$ of x -coordinates. It means that the wave function of the electron is initially:

$$\Phi = \Phi_{n_0}^Q \left(r - \frac{1}{2}R \right). \quad (4)$$

Here, $\Phi_n^Q(r-1/2R)$ means that φ_n^Q is a function of $r-1/2R$, that is the initial position of the electron, which is located at a distance from the coordinate $+1/2R$ to the electron e in the upper-right hand of Fig. 1. Here n_0 means the electron's initial state among its n energy states (4).

The uncertainty principle of quantum mechanics doesn't determine the exact position of an electron in a hydrogen-atom, but it only calculates the probability of the electron's position as the squared coefficients of Schrödinger's equation (2) that describes the system of the proton and the hydrogen-atom. Then in (2), the coefficient of $\Phi(r-1/2R)$ is for the proton's elastic scattering, and the coefficient of $\Phi(r+1/2R)$ is for the charge exchange (the capture of the proton by the hydrogen-atom). In this simulation, an electron's position is somewhere between two symmetrically placed hydrogen atoms, while their protons' positions are fixed at the $+1/2R$, and $-1/2R$ on the x -axis. Then these two protons are set in Fig. 1 to calculate the coefficients of a slowly moving proton and a hydrogen-atom of Schrödinger equation (2).

First, the position of the electron is at $+1/2R$ of x -coordinates in Fig. 1. It means the elastic scattering of the slowly moving proton because the electron of Fig. 1 stays at the electron's initial position with the hydrogen-atom's proton that is fixed at $+1/2R$. On the other hand, the move of the position of the hydrogen-atom's electron from $+1/2R$ to $-1/2R$ means the charge exchange (the hydrogen-atom's electron changes its sign from $+$ to $-$) by the interaction between the slowly moving proton and the hydrogen-atom's electron [9, p. 84].

Therefore, the coefficient of $-\frac{1}{\left| r - \frac{1}{2}R \right|} \varphi_n(r, R)$ of (2) gives the amplitude of

the elastic scattering of the proton; and, the coefficient of $-\frac{1}{\left| r + \frac{1}{2}R \right|} \varphi_n(r, R)$

gives the amplitude of the charge exchange. The coefficient of $+R/2$ gives the amplitude of both of the elastic scattering and the charge exchange to be made by the hydrogen atom's proton.

INPUT DATA FOR THE NUMERIC SIMULATION

If time and space are distorted (dependent on each other), the electron's plane polar coordinate r *must be* dependent on time. However, it contradicts the uncertainty principle of quantum mechanics because the location of the electron cannot be determined at any time by the principle. Then for the purpose of our numeric simulation, we compromise this contradiction by defining the kinetic energy as a constant term. The hydrogen-atom's electron may change its position from its initial position at $+1/2R$ of x -coordinates, to the position of the charge exchange, $-1/2R$, but the simulation doesn't calculate when it occurs, but only the probabilities of the occurrence.

Then we assigned 24 discrete values for the coordinates of r and R as shown in Fig. 2.

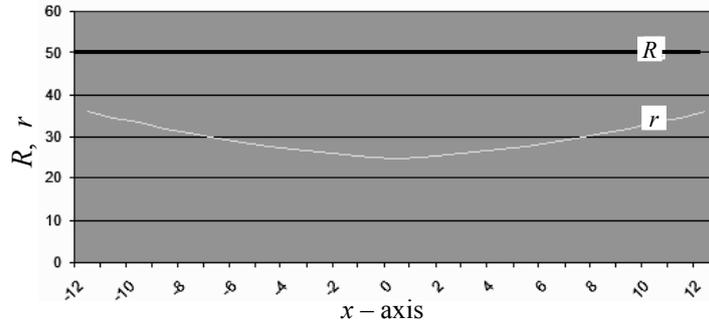


Fig. 2. Coordinates r and R without the distortion of time and space

In our simulation, symmetric and anti-symmetric functions $\varphi_n^S(r, R)$ and $\varphi_n^A(r, R)$ are substituted by the symmetric geometry of two hydrogen atoms as the mirror images on the both-sides of the origin O , as shown in Fig. 3. We assign the value of R by the empirically measured radius, 25 pico-meters, of hydrogen-atom [10]. Because two hydrogen-atoms are placed next to each other in Fig. 3, we assign 50 to the value of R . If charge exchange happens, the electron's plane polar coordinate, r , changes its position from the initial position, $+R/2$ of x -coordinates, to the position of the charge exchange, $-1/2R$. Then the relation between r and R is as follows:

$$r = \sqrt{\left(\frac{R}{2}\right)^2 + x^2}, \tag{5}$$

where x is the distance from the origin O toward $-1/2R$ and toward $+1/2R$, and the origin O is at 0 on the x -axis, $-1/2R$ is at -12 on the x -axis; and, $+1/2R$ is at +12 on the x -axis. The electron is initially attached to the proton at $+1/2R$ of x -coordinates (5); and then it will be attached to the proton at $-1/2R$ of x -coordinates after the charge exchange.

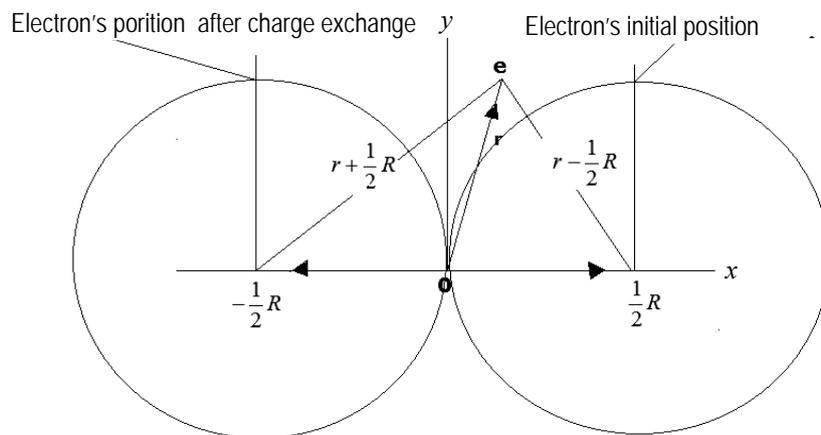


Fig. 3. Position of the electron and its coordinate r

As stated by (3), the nucleons are far apart; therefore, the electron will be localized near one or the other proton. However, it doesn't mean that R in this simulation should be far apart to infinity, but it only justifies the wave functions of hydrogen-atom that distinguish the initial state of the wave function $\Phi_{n_0}^Q\left(r - \frac{1}{2}R\right)$ and the wave function $\Phi_{n_0}^Q\left(r + \frac{1}{2}R\right)$ after the charge exchange.

Then we set sine curves as the wave functions $\varphi_n(r, R)$ of (2). And we set two frequencies, w_1 and w_2 as shown in Fig. 4, for simulating the lower energy state (the lower frequency) and the higher energy state (the higher frequency) of the electron of the hydrogen atom. These sine curves are the functions of r and R , which $\varphi_n(r, R)$ require.

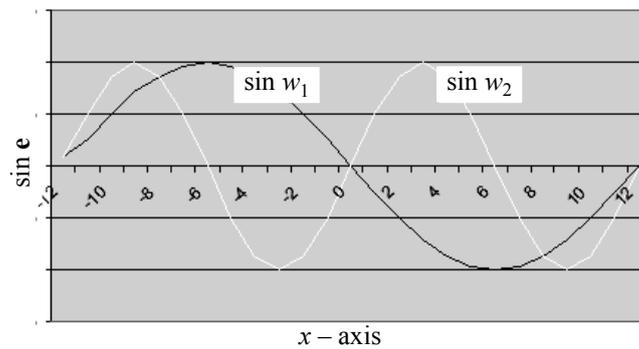


Fig. 4. Sine curves with lower frequency w_1 and the higher frequency w_2

If r and R are dependent on t , r becomes distorted distance ρ , while t becomes distorted time τ as shown in Fig. 5 and Fig. 6. In Fig. 2 shown above, r and R are distance without the distortion. However, in Fig. 5 and Fig. 6, τ and ρ are time and distance showing the interaction with each other. They are calculated by:

$$\tau = t + f(r); \tag{6}$$

$$\rho = t + g(r), \tag{7}$$

where, $f(r)$ and $g(r)$ are given functions of r [9 p. 34]. We set them as shown below, logarithmic distortion, and with exponential distortion:

$$f(r) = \log r ,$$

$$g(r) = e^r .$$

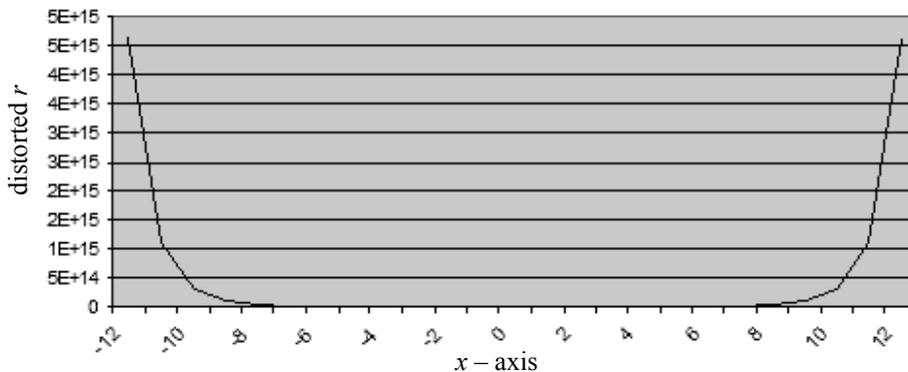


Fig. 5. Distorted coordinates of the electron's position (ρ) with interaction of time and space

We assume that logarithmic distortion for distorted time (6) and exponential distortion for distorted space (7). We think that distorted time should not vary exponentially, because distorted time coordinates should give the basis (less variable than exponential) of the 4-dimensional time and space even if they interact with space.

To calculate (6) and (7), we assigned a time coordinate t on the x -axis from +12 to -12, as the electron's initial position is at +12. According to (7), R is also affected by t , but it doesn't change as much as r does as shown in Fig. 6.

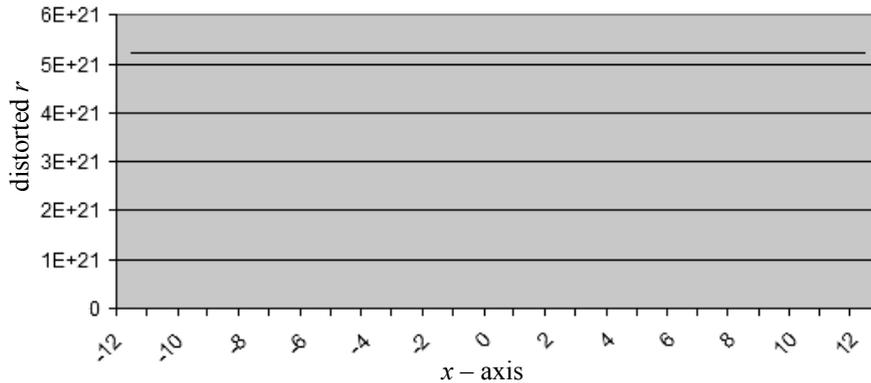


Fig. 6. Distorted coordinates of the proton's position (R) with interaction of time and space

For the distorted r and the distorted R , we used the same wave function (sine curve) as shown in Fig. 4.

ALGORITHM OF THE SIMULATION

The probability of charge exchange was calculated by the equation shown below:

$$H_1\varphi_n(r, R)X(R) = -\frac{1}{2}\nabla_r^2\varphi_n(r, R)X(R) - C_1\left(\frac{1}{r - (1/2)R}\right)\varphi_n(r, R)X(R) - C_2\left(\frac{1}{r + (1/2)R}\right)\varphi_n(r, R)X(R) + \frac{1}{R}X(R). \quad (8)$$

For this numeric simulation, we defined that $-\frac{1}{2}\nabla_r^2\varphi_n(r, R)X(R) + \frac{1}{R}X(R)$ is a constant term (let's put it as T), and it is not affected by the geometry for the potential energy of the system in Fig. 1. Then we can use the matrix algebra shown below to calculate the coefficients, C_1 and C_2 . Also, we don't need $X(R)$ henceforward, because (8) becomes independent from R . Then (8) becomes

$$H_1\varphi_n(r, R) = T - \left[-C_1\left(\frac{1}{r - (1/2)R}\right)\varphi_n(r, R) - C_2\left(\frac{1}{r + (1/2)R}\right)\varphi_n(r, R) \right] = T - \left[-C_1\left(\frac{1}{r - (1/2)R}\right)\sin w - C_2\left(\frac{1}{r + (1/2)R}\right)\sin w \right]. \quad (9)$$

When the reference [9] was published in 1969, a personal power computer of today was not available; therefore, it further described the algorithm in mathematical forms with calculus. Also, it suggested that the squared module of the coefficient C_2 of $-[1/(r+(1/2)R)]\phi_n$ gave the probability of charge exchange [9, p. 86]. In this research we used a personal computer to calculate the coefficients, C_1 and C_2 , with the matrix algebra shown below.

First, we set a two-column matrix X ,

$$X = \begin{bmatrix} \frac{-1}{r+(1/2)R} \sin w & \frac{-1}{r-(1/2)R} \sin w \end{bmatrix} = [X_1 \quad X_2],$$

and a two-row vector made of two coefficients, C_1 and C_2 :

$$c = \begin{bmatrix} C_1 \\ C_2 \end{bmatrix};$$

now, (9) becomes as follows:

$$H_1 \phi_n(r, R) = T - Xc = T - (C_1 X_1 + C_2 X_2).$$

Then we set the boundary conditions to solve the problem:

$$\frac{\partial [H_1 \cdot \phi_n(r, R)]^2}{\partial C_n} = 0.$$

For example, in case of $n = 1$, we have:

$$\frac{\partial [H_1 \phi_n(r, R)]^2}{\partial C_1} = 2H_1 \phi_n(r, R)(-X_1) = 0.$$

Then

$$X' H \phi_n(r, R) = 0,$$

where X' is a transpose matrix of X .

On the other hand,

$$H \phi_n(r, R) = T - Xc;$$

therefore,

$$X'(T - Xc) = 0.$$

Then, we can write:

$$X' Xc = X' T.$$

Therefore, the coefficients are calculated by the equation:

$$c = (X' X)^{-1} X' T,$$

where $(X' X)^{-1}$ is the inverse matrix of $X' X$. In this simulation, we set T as a 25 row-vector of unity (one), therefore the calculated coefficients are not absolute probabilities, but relative probabilities to the unity T . We think that it is sufficient to assign unity to T , because the goal of this simulation is to calculate the probabilities by means of relative squared modules of the coefficients. The calculated coefficients are proportional to the eigenvalues of the hydrogen atom.

RESULT

Fig. 7 and Fig. 8 show the calculated probabilities of the charge exchange and the elastic scattering of the slowly moving proton to the target electron. The probabilities shown in these figures are the calculated results that don't take distorted time and space. Table 1 shows the calculated values of these figures. The probability of the proton's elastic scattering is calculated by the squared coefficient C_1 divided by the sum of the squared coefficients of C_1 and C_2 , and the probability of the proton's charge exchange is calculated by squared C_2 divided by the sum of the squared coefficients of C_1 and C_2 .

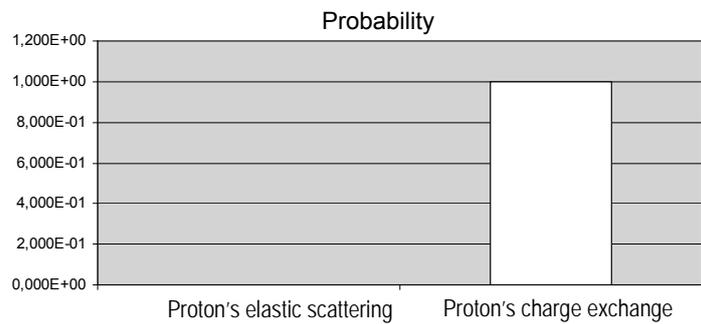


Fig. 7. Calculated probabilities of elastic scattering and change exchange with $\sin w_1$ (without the distortion of time and space)

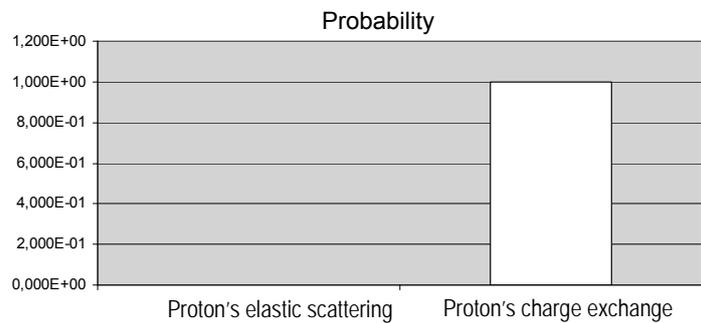


Fig. 8. Calculated probabilities of elastic scattering and change exchange with $\sin w_2$ (without the distortion of time and space)

Table 1. Calculated probability without the distortion of time and space

Parameter		C_1 for elastic scattering	C_2 for charge exchange
Coefficient	W_1	$1,252 \cdot 10^{-2}$	$-6,220 \cdot 10^{-1}$
	W_2	$7,733 \cdot 10^{-3}$	$-6,068 \cdot 10^{-1}$
Squared coefficient	W_1	$1,568 \cdot 10^{-4}$	$3,868 \cdot 10^{-1}$
	W_2	$5,979 \cdot 10^{-5}$	$3,682 \cdot 10^{-1}$
Probability	W_1	$4,051 \cdot 10^{-4}$	$9,996 \cdot 10^{-1}$
	W_2	$1,624 \cdot 10^{-4}$	$9,998 \cdot 10^{-1}$

Note: The probability is a squared coefficient divided by the sum of squared coefficients

Fig. 9, Fig. 10 and Table 2 show the result of the simulation with the distorted time τ and distorted distance ρ . The probability of the charge exchange is lower than the probability without distortion shown in Fig. 7, Fig. 8 and Table 1. Instead, the probability of elastic scattering appears and reaches the same probability of charge exchange. It is also noted that the values of the coefficients of the equation shift from those without distortion of time and space. It means that the eigenvalues of Schrödinger equation (2) shift, i.e., the discrete energy states of the hydrogen atom also shift.

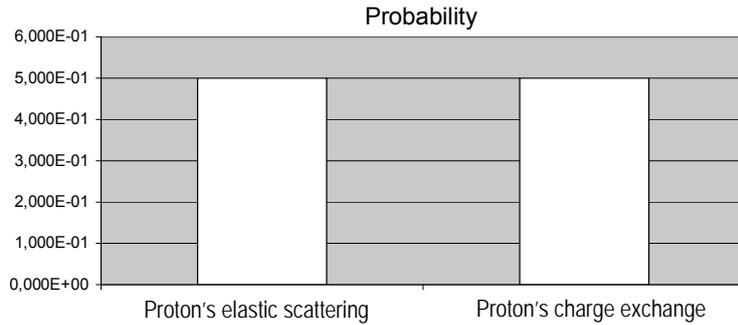


Fig. 9. Calculated probabilities of elastic scattering and change exchange with the distortion of time and space ($\sin w_1$)

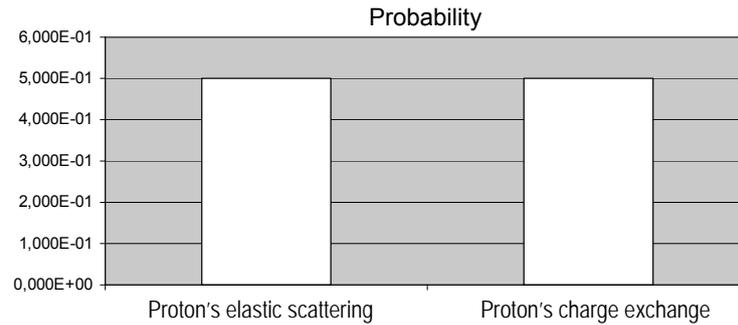


Fig. 10. Calculated probabilities of elastic scattering and change exchange with the distortion of time and space ($\sin w_2$)

Table 2. Calculated probability with the distortion of time and space

Parameter		C_1 for elastic scattering	C_2 for charge exchange
Coefficient	W_1	$1,597 \cdot 10^{27}$	$1,597 \cdot 10^{27}$
	W_2	$2,050 \cdot 10^{27}$	$2,050 \cdot 10^{27}$
Squared coefficient	W_1	$2,550 \cdot 10^{54}$	$2,550 \cdot 10^{54}$
	W_2	$4,204 \cdot 10^{54}$	$4,204 \cdot 10^{54}$
Probability	W_1	$5,000 \cdot 10^{-1}$	$5,000 \cdot 10^{-1}$
	W_2	$5,000 \cdot 10^{-1}$	$5,000 \cdot 10^{-1}$

Note: The probability is a squared coefficient divided by sum of the squared coefficients.

CONCLUSIONS AND RECOMMENDATIONS

The result of the simulation shows that the proton is captured by the hydrogen atom by the charge exchange at the beginning of nuclear fusion when the mass is

not yet heavy and when time and space are still independent. However, as the fusion process proceeds and time and space are interacted, the elastic scattering of the proton appears. It means that the nuclear fusion reaches saturation at some point of the process with the distorted time and space.

The experiment of nuclear fusion may verify the distortion of time and space. In our numeric simulation, the coefficient of the charge exchange differs between the cases with and without the distortion. It means that the discrete energy states of the electron in the hydrogen atom should also vary because the solution of discrete eigenvalues of Schrödinger equation (2) shifts from the case without the distortion to the case with the distortion. (The wave functions such as sine curves are continuous, but the equation (2) leads to the discrete eigenvalues that define the energy states of the hydrogen atom). Also, if the increased proton's elastic scattering is observed in the experiment, it verifies the distortion of time and space.

Therefore, if a laboratory experiment of nuclear fusion detects the shift of electron's energy states of the hydrogen atom and/or increase of the proton's elastic scattering, it will confirm the distortion of time and space in nuclear fusion.

REFERENCES

1. P.A.M. Dirac, *General Theory of Relativity*. New York: Florida University, A Wiley Interscience Publication, John Wiley & Sons, 1975, 69 p.
2. Y. Matsuki and P.I. Bidyuk, "Theory and Simulation of Artificial Antigravity", *IEEE 2nd International Conference on System Analysis Intelligent Computing, October 2020, Kyiv, Ukraine*. doi: 10.1109/SAIC51296.2020.9239195.
3. Y. Matsuki and P.I. Bidyuk, "Numerical Simulation of Gravitational Waves from a Black Hole, using Curvature Tensors", *System Research & Information Technology*, no. 1, pp. 54–67, 2020. doi: 10.20535/SRIT.2308.8893.2020.1.05.
4. Y. Matsuki and P.I. Bidyuk, "Simulating the Rotation of a Black Hole and Antigravity", *System Research & Information Technology*, no. 3, pp. 124–137, 2020. doi: 10.20535/SRIT.2308.8893.2020.3.09.
5. Y. Matsuki and P.I. Bidyuk, "Simulation of a Rotating Strong Gravity that Reverses Time", *System Research & Information Technology*, no. 3, pp. 7–16, 2021. doi: 10.20535/SRIT.2308.8893.2021.3.01.
6. Y. Matsuki and P.I. Bidyuk, "Numeric Simulation of Artificial Antigravity upon General Theory of Relativity", *Advances in Science, Technology and Engineering Systems Journal*, 6(3), pp. 45–53, May 2021. doi: 10.25046/aj060307.
7. Y. Matsuki and P.I. Bidyuk, "Numeric Simulation of the Waves from Artificial Antigravity upon General Theory of Relativity", *Advances in Science, Technology and Engineering Systems Journal*, 6(5), pp. 158–166, October 2021. doi: 10.25046/aj060518.
8. Y. Matsuki and P.I. Bidyuk, *A COURSE IN BLACK HOLE SIMULATION: Application of the general theory of relativity*. LAP LAMBERT Academic Publishing, October 2021, 84 p.
9. A.B. Migdal and V. Krainov, *Approximation Methods in Quantum Mechanics*. New York, Amsterdam: W.A. Benjamin Inc., 1969, 146 p.
10. J.C. Slater, "Atomic Radii in Crystals", *J. Chem. Phys.*, 41, 3199, 1964. Available: <https://aip.scitation.org/doi.org/10.1063/1.1725697>

Received 05.01.2022

INFORMATION ON THE ARTICLE

Yoshio Matsuki, ORCID: 0000-0002-5917-8263, National University of Kyiv-Mohyla Academy, Ukraine, e-mail: matsuki@wdc.org.ua

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

ДОВЕДЕННЯ ГИПОТЕЗИ СТОСОВНО ВІДХИЛЕННЯ ЧАСУ І ПРОСТОРУ НА ОСНОВІ МОДЕЛІ ЯДЕРНОГО СИНТЕЗУ / Й. Мацукі, П.І. Бідюк

Анотація. Дослідження присвячено сучасній тематиці: аналізу умов, які призводять до спотворення координат часу і простору, — явища, що є наслідком загальної теорії відносності, тобто коли час і простір стають взаємозалежними. Для аналізу використано інструментарій імітаційного моделювання з метою імітації руху протона, який вдаряє електрон атома водню. Методологія моделювання ґрунтується на обчисленні ймовірностей пружного розсіювання і обміну зарядами протона та цільового електрона. Таке спотворення моделюється функціями, які зв'язують логарифмічно координати часу і простору у випадку спотворення часу і експоненційно у випадку спотворення простору. Геометрію цієї взаємодії описано рівнянням Шредингера з використанням хвильової функції електрона. Імовірність обміну зарядом обчислено діленням квадрата коефіцієнта хвильової функції на суму квадратів коефіцієнтів усіх членів рівняння. Результати розрахунків показують: імовірність обміну зарядом висока, якщо час і простір не мають відхилень, але коли час і простір спотворюються, вона зменшується і збільшується ймовірність пружного розсіювання. Отриманий результат свідчить про те, що дискретні рівні енергії електронів атомів водню зміщуються у випадку, коли у процесі ядерного синтезу виникає спотворення часу і простору.

Ключові слова: загальна теорія відносності, ядерний синтез, спотворення часу і простору, обмін зарядами.

ДОКАЗАТЕЛЬСТВО ГИПОТЕЗЫ ОБ ИСКАЖЕНИИ ВРЕМЕНИ И ПРОСТРАНСТВА НА ОСНОВЕ МОДЕЛИ ЯДЕРНОГО СИНТЕЗА / Й. Мацуки, П.И. Бидюк

Анотация. Исследование посвящено современной тематике: анализу условий, которые приводят к искажению координат времени и пространства — явления, что является следствием общей теории относительности, т.е. когда время и пространство становятся взаимно зависимыми. Для анализа использовано инструментальный имитационный моделирование с целью имитации движения протона, который ударяет электрон атома водорода. Методология моделирования основывается на вычислении вероятностей упругого рассеяния и обмена зарядами протона и целевого электрона. Такое искажение моделируется функциями, которые связывают логарифмически координаты времени и пространства в случае искажения времени и экспоненциально в случае искажения пространства. Геометрия этого взаимодействия описана уравнением Шредингера с использованием волновой функции электрона. Вероятность обмена зарядом вычислена делением квадрата коэффициента волновой функции на сумму квадратов коэффициентов всех членов уравнения. Результаты расчетов показывают, что вероятность обмена зарядом высока, если время и пространство не имеют отклонений, а если время и пространство искажаются, то она уменьшается и увеличивается вероятность упругого рассеяния. Полученный результат свидетельствует о том, что дискретные уровни энергии электронов атомов водорода смещаются в случае, когда в процессе ядерного синтеза возникает искажение времени и пространства.

Ключевые слова: общая теория относительности, ядерный синтез, искажение времени и пространства, перезарядка.

SYSTEM ANALYSIS OF THE PROBLEM OF ESTABLISHING THE AUTHENTICITY AND AUTHORITY OF PAINTING WORKS

A.A. MARTYNNENKO, A.D. TEVYASHEV, N.E. KULISHOVA, B.I. MOROZ

Abstract. Cultural values have long been the objects of crimes, among which the export from the state stands out. Falsification hides artworks from customs control and its detection requires a long examination using a variety of methods of analysis. This article discusses the task of verifying painting's authenticity during customs inspection. A two-stage procedure is proposed, which includes a quick check based on the analysis of painting's images and a longer museum expertise. To implement the image analysis, it is proposed to use an intelligent decision-making system, which is based on a classifier that implements the k -nearest neighbors algorithm. A set of features to describe painting's properties is formed, metrics for calculating the similarity measure on objects in the course of classification is proposed. To train an algorithm, a dataset is proposed, which includes paintings by world and European artists, as well as Ukrainian painters from different centuries.

Keywords: intelligent decision-making system, automatic classification, k -nearest neighbors, customs examination, paintings.

INTRODUCTION

The rapid rates of art market growth and constantly increasing demand for works of fine art have led to the fact that the problem of authenticating works of art has become extremely urgent for all market participants: art museums and galleries, auctions, collectors and individuals, and for states customs services.

Works of fine art have long ceased to be just an expression of the artist's ideas and intentions; they often function as payment means and objects for profitable investments. For this reason, the paintings of famous masters have become associated with criminal activities — forgeries, embezzlements, illegal transportation across state borders. The canvases of famous artists are especially widely falsified. Falsifiers not only inflict enormous material damage on states, paintings owners, but also spiritually devalue the works of great masters of painting, which poses a threat to the economic security of states [1]. Falsification of works of fine art means the production of counterfeit painting objects and their sale to obtain material benefits. Depending on forger qualifications, used techniques, technical means and materials in painting, there are fakes of different complexity — from simple (copying) to super forgeries, to establish which authenticity is extremely difficult even for specialists.

Expertise procedures are used to establish the paintings authenticity, or at least to determine the degree to which they are classified as “cultural value” or “national wealth”. There are two types of expertise — customs and museum ex-

expertise [2]. The purpose of the customs examination is to ensure economic security in the country. Customs expertise is strictly structured and has a hierarchy of goals, where dating is at the top, and rest of data obtained is additional and basic for the conclusion whether this work belongs to the appropriate category.

The ultimate goal of museum expertise is to establish the authenticity and authorship of a painting. Currently, four main methods of authenticating pictures are used: forensic, attributive, technological, complex [3].

The forensic method includes: studies of author's signature on the picture; examinations of painting author fingerprints; research of handwritten notes, signatures, imprints of seals (stamps) on reverse side of the picture; analysis of provenance reliability (the work ownership history from creation moment to present). Currently, the concept of provenance has expanded: it also includes a list of checks or invoices proving the fact of purchasing an item for a certain amount, expert assessments, history of participation in auctions, reproductions in books and catalogs, participation in exhibitions, as well as any references in relevant literature.

The attributional method consists in studying the art form details to find out the specifics of master individual style.

The technological method is implemented using various technical means of analysis: microscopic, X-ray spectral, macrophotography, as well as photography in ultraviolet and reflected infrared rays, etc. In the technological method of research, all elements of picture are analyzed: base, soil, paint layer, etc. From the obtained data, it is established that at various stages of his career, what certain primers, paints, varnishes, brushes the artist used. The results show that each artist has his own manner of "painting", his own special technique, and style. To increase reliability of making a decision on the paintings authenticity, a complex forensic, technological and art history expertise is used [4]. The practical use of complex examinations for paintings authentication requires not only involvement of highly qualified experts groups equipped with the necessary technical means, but also significant financial and time costs. A systematic solution to this problem is possible based on intelligent video analytics, machine learning methods and computational intelligence.

Taking into account the effect of time factor, we can say that the customs examination should be more efficient to ensure a quick decision on the possibility of exporting art object outside the state. Museum expertise is not so strictly limited in time, since it is not performed at border crossing points. Obviously, for customs examination, the choice of methods used is rather limited in terms of speed: these are variants of technological method based on photographing works of art in different lighting conditions (studio, infrared, ultraviolet, X-ray), macro photography. The rest of technological methods, as well as forensic, art history and attributive expertise cannot be promptly performed under conditions of customs control. Therefore, a two-stage procedure for establishing the authenticity and authorship of paintings is proposed (Fig. 1).

In this work, the first stage of the examination will be considered — the customs one, which is proposed to be implemented using one of technological methods, namely, photographing works of painting with high resolution under studio lighting conditions.

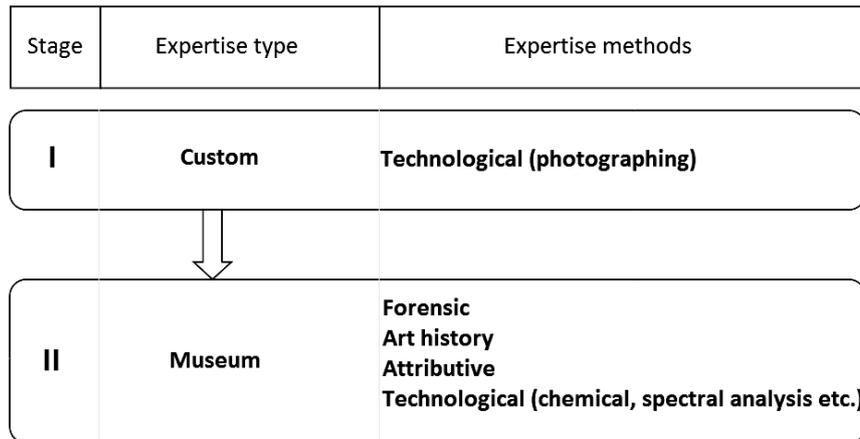


Fig. 1. Scheme of a two-stage expertise procedure for establishing the authenticity and authorship of paintings

ANALYSIS OF AREAS OF RESEARCH AND STATEMENT OF THE PROBLEM

Currently, research in this area is carried out in several directions. One direction is more focused on the creation of new and improvement of existing devices that allow the analysis of materials and substances that forms an object of art [1–6].

Another area is directly related to technologies of image digitization and their analysis using statistics, signal processing, machine learning [7–10].

Deep neural networks, which have recently become popular, have also not remained outside the attention of scientists who are developing means for comparing, identifying, and authenticating pictures. The most effective were convolutional neural networks due to their ability to distinguish a large number of heterogeneous features in images [11–14]. A number of interesting solutions were found during the use of generative networks [15–18].

This work is devoted to the problem of automatic identification of cultural values, in particular, paintings using the intelligent decision support system (IDSS) [19]. A system oriented to work in real time must have high speed and high accuracy in solving the classification problem. For this, an approach based on the weighted k -nearest neighbour’s algorithm is proposed, since for deep learning networks it is necessary to re-train if the value database is replenished by at least one object.

METHODOLOGY

Formally, the classification problem can be presented as follows. There is a set of n objects

$$I = \{i_1, i_2, \dots, i_l, \dots, i_n\}, \quad l = \overline{1, n},$$

each of them is characterized by a set of m features

$$F = \{f_1, f_2, \dots, f_q, \dots, f_m\}, \quad q = \overline{1, m}.$$

Features take values from a certain set

$$C^{[f_q]} = \{c_1^{[f_q]}, c_2^{[f_q]}, \dots, c_h^{[f_q]}, \dots, c_p^{[f_q]}\}, \quad h = 1, \dots, p,$$

where p is the number of possible discrete values of each feature.

One feature f_T is the target, its values for an objects set I make up a vector $C^{[f_T]} = C_T$. The classifier G learns by examples to establish relationships of the form

$$G(F(I)) = C_T,$$

calculating the approximated values of the target feature \hat{C}_T such that the difference between the specified and approximated values will be minimal:

$$d(C_T, \hat{C}_T) \rightarrow \min.$$

A trained classifier allows calculating target attribute values for new objects $I_{new} = \{i_{n+1}, i_{n+2}, \dots\}$ in this way:

$$G(F(I_{new})) = C_{TNew}.$$

When identifying works of painting, the classification problem can be solved for several target attributes [11]:

- determination of painting artistic style with target attribute C_{TStyle} ;
- determination of picture genre with target attribute C_{TGenre} ;
- defining the author with target attribute $C_{TArtist}$;
- determination of picture creation time with target attribute C_{TTime} .

Obviously, the dataset used to solve the problem must include the appropriate attributes. If the artist's name is an attribute required for such datasets, then defining and marking up an art style requires the participation of highly qualified art historians. The markup of the painting genre is an even more difficult problem, so this attribute may not be present in all datasets, what should be taken into account when developing an intelligent decision-making system.

In this paper, it is proposed to solve the problem of classifying paintings by the attribute of creation time C_{TTime} (Fig. 2).

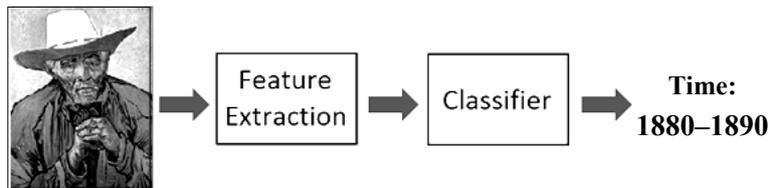


Fig. 2. Illustration of a painting classification system

As noted in [8], when marking up data, art critics often use information about the author's belonging to a particular artistic style — this increases the accuracy of identification. Therefore, it is possible to single out global and local characteristic features necessary to recognize the painting author and, accordingly, the time of the painting (Fig. 3).

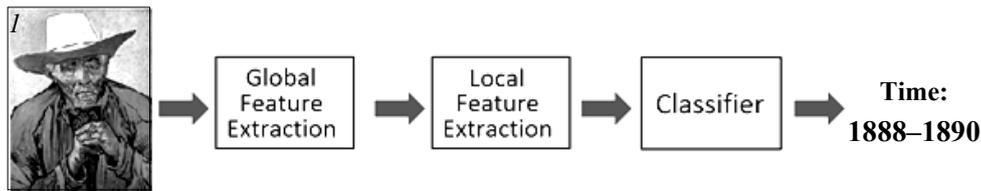


Fig. 3. Scheme of paintings classification based on global and local features

During their creative activity, many artists have changed the artistic manner of writing, moved from one style to another, so the use of signs that can characterize the artistic style will be useful. In the datasets that are currently used in the development of automatic classification systems, the following styles are most often considered: abstract expressionism, baroque, constructivism, cubism, impressionism, neo-classical, pop art, post-impressionism, realism, renaissance, romanticism, surrealism, symbolism [8]. In this work, the artistic style will be considered as one of the auxiliary attributes.

FEATURE EXTRACTION

To describe the general properties of a painting, data on color and structural properties inherent in the entire image are most often used. To form sets of such data, many different algorithms and descriptors are used: wavelet transforms, Radon, Hough, Fourier, Chebyshev transforms and their combinations [7]; Gabor filters; Local Binary Patterns (LBP); SIFT detectors [20]; textural features; first 4 moments; multidimensional histograms; edge statistics features, etc. [8]. In particular, it is by the first 4 moments and by edges, the Impressionists's works can be classified unambiguously. Contour markers convey information about brush strokes style that is specific to each artist, which makes Impressionism, stand out in comparison with other styles.

Surrealist paintings can be described more informatively with the help of contour and object statistics, which reflects the presence of significant "empty" areas in their works.

In [7] it is noted that the use of color data in descriptors compilation increases the classification accuracy by 18,1%. However, researchers most often work with color data in RGB representation, since this color space describes signals from image capture and displaying devices. In this work, it is also proposed to use the CIE Lab color space [21], since it is focused on the unambiguous description of visual stimuli in accordance with human vision.

Thus, the following set of global picture descriptors is proposed:

1. Local Binary Patterns (LBP) [22] for describing texture properties. In this work, an LBP implementation is used within a neighborhood of 20 pixels and with a radius of 2 pixels.
2. Color modification of LBP to describe the color properties of the texture. LBPs are calculated in R, G, B and CIE L, CIEa, CIEb color channels. The results are combined using concatenation to form a multivariate histogram for each image.
3. The first 4 points — mean, standard deviation, skewness, and kurtosis, calculated in the directions of 0, 45, 90, 135 degrees. Moments are calculated along the "stripes" in the image in several specified directions. A 3-bin histogram is plotted for each obtained data vector.
4. Tamura's textural features are roughness, contrast, directionality, linearity, roughness and regularity.

The texture roughness characterizes main details dimensions that form the image. Its estimate is based on average values calculation within neighborhood of pixels:

$$A_k(x, y) = \sum_i \sum_j \frac{b(i, j)}{2^{2k}},$$

where $b(i, j)$ is brightness of pixel with i, j coordinates; k is neighborhood size; the texture roughness is then

$$E_k(x, y) = A_k(x, y) - A_k(x', y), \quad x' \neq x.$$

The texture contrast is estimated based on fourth moment μ_4 relative to mathematical expectation and variance σ^2 within neighborhood:

$$C_k(x, y) = \frac{\sigma}{(\alpha_4)^{0,25}},$$

where $\alpha_4 = \frac{\mu_4}{\sigma^4}$ — kurtosis.

The texture directivity is estimated based on quantized edge directions histogram $H_{dir}(a)$:

$$D_k(x, y) = 1 - r n_{peaks} \sum_p \sum_{a \in w_p} (a - a_p)^2 H_{dir}(a),$$

where n_{peaks} is peaks number; a_p — peak angular direction; r — coefficient that depends on quantization of angles levels a_p ; $a_p = \arctan \frac{\Delta x}{\Delta y}$ calculated with

Pruitt contour detector.

Linear similarity $L_k(x, y)$ is evaluated as average coincidence of edge directions that match in pixels pairs separated by a distance along the edge direction in each pixel.

Texture regularity is a generalized feature defined as

$$R_k(x, y) = 1 - r(\sigma_{coarseness} + \sigma_{contrast} + \sigma_{directionality} + \sigma_{linelikeness}),$$

where $\sigma_{coarseness}, \sigma_{contrast}, \sigma_{directionality}, \sigma_{linelikeness}$ are standard deviations for each feature.

Roughness summarizes contrast and roughness of texture as follows:

$$Roughness_k(x, y) = E_k(x, y) + C_k(x, y).$$

5. Radon transform features, calculated for angles of 0, 45, 90, 135 degrees and then combined into 5-pocket histograms.

6. Haralik's textural features — contrast, correlation, entropy, energy and homogeneity. They are calculated based on the contingency matrix

$$P(i, j) = \frac{\neq [(p_1, p_2) \in I | (p_1 = i) \wedge (p_2 = j)]}{\neq I},$$

where p_1, p_2 are pixels belonging to image.

Then contrast will be defined as

$$C_H(x, y) = \sum_{i, j} (i - j)^2 p(i, j);$$

the correlation is as follows:

$$Corr_H(x, y) = \sum_{i, j} \frac{(i - \mu_i)(j - \mu_j)p(i, j)}{\sigma_i \sigma_j},$$

entropy:

$$Entropy_H(x, y) = \sum_i \sum_j p(i, j) \log_2 p(i, j),$$

energy:

$$Energy_H(x, y) = \sum_i \sum_j p(i, j)^2.$$

7. Palette redundancy [9]:

$$M_B = \frac{H_{\max} - H_{RGB}}{H_{\max}},$$

where H_{\max} is maximum image entropy, which for 8-bit color coding is $8 \times 3 = 24$; H_{RGB} — entropy calculated for individual R, G, B channels.

It is proposed to describe features of individual areas of image, individual details using local descriptors, which include:

- 1) euler's number, minimum, maximum, median, mathematical expectation, variance for each object;
- 2) SIFT descriptor, built for gray level images, and images in RGB and CIE Lab color spaces.

The result of local features selection are descriptors in form of multidimensional vectors.

CLASSIFICATION OF PICTURES IMAGES

A wide variety of artistic techniques, styles, colors that artists use when creating paintings, leads, when developing an intelligent decision-making system, to the need to implement such a classifier that will be highly adaptable as new samples become available, and will also allow objects to be compared using a large set very dissimilar features. In [7], a comparison of such classification methods as weighted k -nearest neighbors method and SVM was made. The comparison showed that weighted k -nearest neighbors method provides an increase in classification accuracy by about 15–20% compared to SVM.

Weights determination

To calculate weights, it is proposed to use two basic approaches:

1. Assigning weights in accordance with information gain criterion, which is calculated on class-based entropy basis for attribute values:

$$W_E(f_i) = -\sum_{c=1}^C p_c \log_2(p_c),$$

where p_c is number of objects with a feature value f_i and that belonging to class c .

2. Assignment of weights in accordance with the Fischer method:

$$W_F = \frac{\sum_{c=1}^C p_c (\mu_c - \mu)^2}{\sum_{c=1}^C p_c \sigma_c^2},$$

where μ_c , σ_c^2 is mathematical expectation and standard deviation of data points belonging to class c for a specific attribute; μ is global mathematical expectation for all data points for a particular attribute.

Metrics used

To build a classifier, can use a variety of metrics, for example, Euclidean distance, Hamming distance, Mahalanobis, Minkowski, or Chebyshev distances. However, the use of each metric is associated with peculiarities of internal data structure or algorithm properties. These factors can lead to a significant decrease in classification accuracy. The Chebyshev metric will be more useful when comparing objects by the same attribute. To determine Mahalanobis distance, it is necessary to calculate observations covariance matrix, which for considered problem becomes is a laborious task. In addition, due to significant patterns data heterogeneity, the Mahalanobis metric will reduce classification accuracy due to covariance matrix “blurring” over entire data volume.

Thus, in this problem of classifying pictures using weighted k -nearest neighbors algorithm, it is proposed to use Euclidean distance

$$d(f_i, f_j) = \sqrt{(f_{i1} - f_{j1})^2 + \dots + (f_{in} - f_{jn})^2} = \|f_i - f_j\|,$$

weighted Euclidean distance

$$d(f_i, f_j) = \sqrt{\lambda_1(f_{i1} - f_{j1})^2 + \dots + \lambda_n(f_{in} - f_{jn})^2}.$$

Hamming distance

$$d(f_i, f_j) = \sum_{l=1}^n |f_{il} - f_{jl}|$$

and the Minkowski distance

$$d(f_i, f_j) = \left(\sum_{l=1}^n |f_{il} - f_{jl}|^p \right)^{1/p},$$

where f_i, f_j are vectors of attribute values for compared objects i, j ; f_{il}, f_{jl} — values of l -th attribute for objects i, j being matched.

Obviously, all attribute values must first be normalized so that the condition

$$f_{il} \in [0, 1], \quad i = 1, \dots, n; \quad l = 1, \dots, m$$

now for each object i_p , which is characterized by a vector of features f_i , the degree of similarity with a certain class c_T is calculated in accordance with equation

$$M_{i,c_T} = \frac{1}{\min(d_{i,c_T}) \sum_{i=1}^n \frac{1}{\min(d_{i,c_T})}}.$$

Dataset for experimental research

The activities of many museums now also include the digitization of stored valuables to provide wider user access. Therefore, number of art databases now totals dozens, each with thousands of images [11, 23–28]. There are databases contain-

ing mainly classical works. Others, on the contrary, are focused on contemporary art. Still others represent different eras and artistic styles.

In this work, it is proposed to use a small-sized dataset [29] with open access. This set contains works by 50 artists who worked at different times — from the 15th to the 20th century. Their works are ranked among a variety of artistic styles: Impressionism, Post-Impressionism, Northern Renaissance, Baroque, Romanticism, Symbolism, Realism, Surrealism, Byzantine art, etc.

When studying the possibility of automatic identification of art values in our country, it is important to take into account the historical context. It so happened that for a long time access to works of world-famous artists was closed to Ukraine, here it is much more likely that you can find paintings by Russian and Ukrainian masters who worked in the 18–20 century in the appropriate artistic style. Therefore, the set [29] was supplemented with images of paintings by Russian artists of the 17–19th century in the styles of romanticism, classicism, realism [30]. These images were obtained from the official portal of the Hermitage Museum, and data on the artists — in Wikipedia. It is important to note that a considerable number of paintings are not attributed by author, but they also have art value and may be subject to identification by style.

In addition, dataset also includes paintings by Ukrainian artists of the 19th and 20th centuries, obtained from the portal of the National Art Museum of Ukraine [31].

CONCLUSIONS

The paper considers the problem of paintings automatic identification using an intelligent decision support system (IDSS). The authors proposed a solution in the form of a classifier based on a weighted k -nearest neighbor's algorithm.

The paper proposes a set of local and global features can be used to attribute objects to be classified. The set of features includes color, texture, statistical and other characteristics.

To calculate weights for k -nearest neighbors algorithm implementation, it is proposed to use the Fisher method, as well as information gain criterion. In the algorithm for similarity measure calculating, authors proposes to use several metrics: Euclidean, weighted Euclidean metrics, Minkowski and Hamming metrics.

As a dataset for experimental research, it was proposed to use a set includes works by famous world and European artists, as well as supplemented by paintings by Russian and Ukrainian masters.

REFERENCES

1. A. Mironova, F. Robache, R. Deltombe, R. Guibert, L. Nys, and M. Bigerelle, “Digital Cultural Heritage Preservation in Art Painting: A Surface Roughness Approach to the Brush Strokes”, *Sensors*, 20, 6269, 2020. doi:10.3390/s20216269.
2. K. Zreik and N. Bouhai, “Post-digital design: The Hyperheritage project”, in *Proceedings of the SHoCK—Sharing of Computable Knowledge*; A. Fioravanti, S. Corsi, S. Elahmar, S. Gargaro (Eds.); eCAADe 35: Rome, Italy, 2017, pp. 49–54.
3. W.S. Elkhuizen et al., “Comparison of three 3D scanning techniques for paintings, as applied to Vermeer’s ‘Girl with a Pearl Earring’”, *Herit. Sci.*, 7, 89, 2019.
4. M.J.W. Van Hengstum et al., “Development of a high resolution topography and color scanner to capture crack patterns of paintings”, *Proceedings of the EUROGRAPHICS Workshop on Graphics and Cultural Heritage, Vienna, Austria, 12–15 November 2018*, pp. 11–20.

5. C.-C. Carbon and V.M. Hesslinger, "On the nature of the background behind Mona Lisa", *Leonardo*, 48, pp. 183–184, 2015.
6. G.J.A.M. Eumelen, E. Bosco, A.S.J. Suiker, A. van Loon, and P.D. Iedema, "A computational model for chemo-mechanical degradation of historical oil paintings due to metal soap formation", *J. Mech. Phys. Solids*, 132, 103683, 2019.
7. L. Shamir, T. Macura, N. Orlov, M.D. Eckley, and I.G. Goldberg, "Impressionism, Expressionism, Surrealism: Automated Recognition of Painters and Schools of Art", *ACM Transactions on Applied Perception*, vol. 7 (2), pp. 1–18, 2010.
8. F.S. Khan, S. Beigpou, J. Van de Weijer, and M. Felsberg, "Painting-91: a large scale database for computational painting categorization", *Machine Vision and Applications*, 25(6), pp. 1385–1397, 2014. Available: <https://doi.org/10.1007/s00138-014-0621-6>
9. D. Wang and Y. Chen, "Calculation and Application of Xin'an Painting School Art Style Model", *ICAITA*, 1651, 012033, 2020. doi: 10.1088/1742-6596/1651/1/012033.
10. L. Liu, J. Chen, P. Fieguth, G. Zhao, R. Chellappa, and M. Pietikäinen, "From BoW to CNN: Two decades of texture representation for texture classification", *Int. J. Comput. Vis.*, 127, pp. 74–109, 2019.
11. B. Saleh and A. Elgammal, "Large-scale Classification of Fine-Art Paintings: Learning The Right Metric on The Right Feature", *Digital Art History*, no. 2, 2016. Available: <https://doi.org/10.11588/dah.2016.2.23376>
12. V. Balntas, K. Lenc, A. Vedaldi, and K. Mikolajczyk, "HPatches: A benchmark and evaluation of handcrafted and learned local descriptors", *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.
13. X. Zang, F.X. Yu, S. Karaman, and Sh.-F. Chang, "Learning discriminative and transformation covariant local feature detectors", *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.
14. K. He, Y. Lu, and S. Sclaroff, "Local descriptors optimized for average precision", *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.
15. R. Srinivasan and K. Uchino, *Quantifying Confounding Bias in Generative Art: A Case Study*. Available: arXiv: 2102.11957v1 [cs.CY] 23 Feb 2021.
16. A. Elgamma, B. Liu, M. Elhosein, and M. Mazzone, "Can: Creative adversarial networks, generating "art" by learning about styles and deviating from style norms", *International Conference on Computational Creativity (ICCC)*, 2017.
17. M. Sung, *The ai renaissance portrait generator isn't great at painting people of color*. 2019. Available: <https://mashable.com/article/ai-portrait-generator-pocs/>
18. W.R. Tan, Ch.S. Chan, and K. Tanaka, "ArtGAN: Artwork synthesis with conditional categorical GANs", *2017 IEEE International Conference on Image Processing (ICIP)*, 2017, pp. 3760–3764. doi: 10.1109/ICIP.2017.8296985.
19. A. Martynenko, V. Moroz, and I. Nulina, "An intelligent decision support system for cultural property identification", *Computer-Integrated Technologies: Education, Science, Production*, 39, pp. 78–82, 2020. Available: <http://cit-journal.com.ua/index.php/cit/article/view/126>
20. D.G. Lowe, "Object recognition from local scale-invariant features", *Proceedings of the International Conference on Computer Vision*, 2, pp. 1150–1157, 1999. doi: 10.1109/ICCV.1999.790410
21. M.D. Fairchild, *Color Appearance Models*. John Wiley and Sons, 2005, 340 p.
22. Z. Guo, L. Zhang, and D. Zhang, "A completed modeling of local binary pattern operator for texture classification", *IEEE Trans. on Image Processing*, vol. 19, issue 6, pp. 1657–1663, 2010.
23. *WikiArt. Visual art encyclopedia*. Available: <https://www.wikiart.org/en/>
24. *Arkyves*. Available: <http://arkyves.org/>
25. *Art UK*. Available: <https://artuk.org/>
26. *Artsy*. Available: <https://www.artsy.net/>
27. *Behance*. Available: <https://www.behance.net/>
28. *Artnet*. Available: <https://www.artnet.com/>
29. *Kaggle*. Available: <https://www.kaggle.com/ikarus777/best-artworks-of-all-time/tasks>

30. *Hermitage*. Available: <http://collections.hermitage.ru/entity/COLLECTION>
31. *National Art Museum of Ukraine*. Available: <http://namu.kiev.ua/ua/gallery/whole-gallery.html>

Received 15.01.2022

INFORMATION ON THE ARTICLE

Andrii A. Martynenko, ORCID: 0000-0002-5033-4696, Dnipro University of Technology, Ukraine, e-mail: martynenko.andrey.a@gmail.com

Andriy D. Tevyashev, ORCID: 0000-0002-2846-7089, Kharkiv National University of Radio Electronics, e-mail: andrew.teviashev@nure.ua

Nonna Ye. Kulishova, ORCID: 0000-0001-7921-3110, Kharkiv National University of Radio Electronics, e-mail: nonna.kulishova@nure.ua

Boris I. Moroz, ORCID: 0000-0002-5625-0864, Dnipro University of Technology, Ukraine, e-mail: moroz.boris.1948@gmail.com

СИСТЕМНИЙ АНАЛІЗ ПРОБЛЕМИ ВСТАНОВЛЕННЯ СПРАВЖНОСТІ Й АВТОРСТВА ТВОРІВ ЖИВОПИСУ / А.А. Мартиненко, А.Д. Тевяшев, Н.Є. Кулішова, Б.І. Мороз

Анотація. Культурні цінності давно є об'єктами злочинів, зокрема вивезення їх із держави. Фальсифікація приховує твори живопису від митного контролю; її виявлення потребує тривалої експертизи з використанням різноманітних методів аналізу. Розглянуто завдання встановлення справжності картин під час митної перевірки. Запропоновано двоетапну процедуру, яка передбачає швидку перевірку на основі аналізу фотографій творів живопису та більш тривалу музейну експертизу. Для реалізації аналізу фотографій запропоновано використовувати інтелектуальну систему прийняття рішень, дія якої базується на класифікаторі, що реалізовує алгоритм k -найближчих сусідів. Сформовано набір ознак опису властивостей творів живопису, запропоновано метрики для обчислення міри подібності об'єктів під час класифікації. Для навчання алгоритму пропонується набір даних, що включає картини світових, європейських художників та українських майстрів різних століть.

Ключові слова: інтелектуальна система прийняття рішень, автоматична класифікація, k -найближчих сусідів, митна експертиза, твори живопису.

СИСТЕМНЫЙ АНАЛИЗ ПРОБЛЕМЫ УСТАНОВЛЕНИЯ ПОДЛИННОСТИ И АВТОРСТВА ПРОИЗВЕДЕНИЙ ЖИВОПИСИ / А.А. Мартыненко, А.Д. Тевяшев, Н.Е. Кулишова, Б.И. Мороз

Аннотация. Культурные ценности давно являются объектами преступлений, в частности их вывоз из государства. Фальсификация укрывает произведения живописи от таможенного контроля; для ее обнаружения необходима длительная экспертиза с использованием разнообразных методов анализа. Рассмотрена задача установления подлинности картин в ходе таможенной проверки. Предложена двухэтапная процедура, которая предполагает быструю проверку на основе анализа фотографий произведений живописи и более длительную музейную экспертизу. Для реализации анализа фотографий предложено использовать интеллектуальную систему принятия решений, действие которой базируется на классификаторе, реализующем алгоритм k -ближайших соседей. Сформирован набор признаков для описания свойств произведений живописи, предложены метрики для вычисления меры сходства объектов в ходе классификации. Для обучения алгоритма предлагается набор данных, который включает картины мировых, европейских художников, а также украинских мастеров разных столетий.

Ключевые слова: интеллектуальная система принятия решений, автоматическая классификация, k -ближайших соседей, таможенная экспертиза, произведения живописи.

СИСТЕМА ПІДТРИМАННЯ ПРИЙНЯТТЯ РІШЕНЬ ДЛЯ ОЦІНЮВАННЯ ТА ПРОГНОЗУВАННЯ СТАНУ СТРАХОВОЇ КОМПАНІЇ

Р.С. ПАНІБРАТОВ

Анотація. Створено систему підтримання прийняття рішень для оцінювання та прогнозування стану страхової компанії за її фінансово-економічними показниками. Оцінювання стану даної установи розглянуто як задачу бінарної класифікації: чи є діяльність компанії ефективною чи ні. Під час дослідження реалізовано шість методів машинного навчання з учителем: метод k -найближчих сусідів, метод опорних векторів, наївний байєсівський класифікатор, випадковий ліс, XGBoost та глибоку нейронну мережу. Створена система дозволяє: виконувати кореляційний аналіз фінансово-економічних показників, перевіряти збалансованість даних, навчати обрану модель та оцінювати якість навчання, прогнозувати стан страхової компанії за обраною моделлю. За найкращими моделями спрогнозовано майбутній стан страхових компаній України.

Ключові слова: страхова компанія, система підтримання прийняття рішень, бінарна класифікація, метод k -найближчих сусідів, метод опорних векторів, наївний байєсівський класифікатор, випадковий ліс, градієнтний бустинг, нейронні мережі.

ВСТУП

Процес формування ринкової економіки будь-якої країни включає в себе надійний страховий захист суб'єктів господарювання. Саме це спонукає до пошуку ефективних засобів мобілізації фінансових ресурсів страхових компаній та їх цілеспрямованого використання. Для забезпечення надійності виконання зобов'язань даний тип установи має ефективно формувати та використовувати фінансові ресурси. Надійний захист суб'єктів господарювання потребує застосування ефективних механізмів мобілізації фінансових ресурсів страхових компаній та їх безпосереднього використання.

Натепер наявні ефективні математичні методи оцінювання фінансових ризиків у страхуванні. Наприклад, стандарт Basel II та Solvency I заохочують розвиток моделей управління та оцінювання ризиків фінансової сфери.

Розрізняють два засоби аналізу ризиків — за позицією факторів ризику та за позицією наслідків ризикової події. За таким підходом математичні методи оцінювання ризиків поділяють на дві групи: методи, що ґрунтуються на аналізі наслідків («високорівневі моделі»), та методи, що ґрунтуються на

аналізі факторів ризику («низькорівневі моделі»). До першої групи належать моделі ВІА, LDA та ІМА. Такі методи не можна використовувати для розподілу ризиків за рівнем критичності та знаходження подій, що ведуть до реалізації ризику. Друга група охоплює такі методи, як регресійний аналіз, байєсівські мережі та методи нечіткої логіки. Ці методи навпаки можуть вирішувати такі завдання, але мають певні обмеження. Так, регресійні моделі гарантують високу точність оцінювання тільки за умови достатньо великих обсягів даних. Байєсівські мережі дозволяють враховувати прямі залежності рівнів ризиків від чинників ризику і залежності між чинниками ризику, але обчислення умовних імовірностей є доволі трудомісткими.

Сьогоднішні теорії інформаційних технологій та бізнес-менеджменту і принципи побудови інтелектуальних систем вимагають того, щоб мета створення системи базувалася на цілеспрямованій та усвідомленій людській діяльності. До нинішніх напрямів розроблення людино-машинних систем належать системи автоматичного керування, експертні системи та системи підтримання прийняття рішень (СППР). Остання є найефективнішим засобом вирішення багатьох завдань [1]. Завдяки їй, особа, що приймає рішення, може, застосовуючи обчислювальні засоби, проектувати, порівнювати та вибрати альтернативні варіанти рішень найрізноманітнішими засобами.

Мета роботи — розроблення та створення СППР, що дозволяє оцінити і прогнозувати майбутній стан страхової компанії за її статистично-економічними даними. Вона розглядалася як задача бінарної класифікації, тобто прогнозування ефективності діяльності компанії.

МАТЕМАТИЧНІ МЕТОДИ ОЦІНЮВАННЯ СТАНУ СТРАХОВОЇ КОМПАНІЇ

Для оцінювання та прогнозування стану страхових компаній реалізовано шість моделей класифікації: k -найближчих сусідів, метод опорних векторів, наївний байєсівський класифікатор, випадковий ліс, XGBoost та багатошаровий перцептрон.

Метод k -найближчих сусідів

Метод k -найближчих сусідів — метричний алгоритм для задачі автоматичної класифікації об'єктів або регресії. У випадку задачі класифікації результатом є належність об'єкта до відповідного класу. Об'єкт класифікується множиною голосів його сусідів, причому об'єкт відносять до класу, який є найбільш поширеним серед k -найближчих сусідів. Параметр k — константа і задається користувачем. Вектор ознак без мітки класифікується за допомогою присвоєння мітки, яка найчастіше зустрічається серед k тренувальних зразків, що найближче розташовані до цієї точки.

Для дослідження розглядалися три найближчі сусіди та використовувалася евклідова метрика відстані, оскільки такі параметри забезпечували найвищий результат.

Метод опорних векторів

Метод опорних векторів — метод навчання з учителем, що використовується як для задачі класифікації, так і для регресії. Надається набір навчальних зразків з позначенням належності до одного з двох класів. Алгоритм будує модель, що призначає кожен новий зразок до однієї чи іншої категорії, створюючи її як найімовірніший бінарний лінійний класифікатор. Метод опорних векторів відображає навчальні зразки в точки у просторі, максимізуючи

при цьому ширину проміжку між двома категоріями. Нові зразки у цьому випадку відображаються у той же простір і прогнозується їх належність до певного класу залежно від того, на який бік проміжку вони потрапляють.

У формальному сенсі метод опорних векторів буде гіперплощину або множину гіперплощин у багатовимірному чи нескінченновимірному просторі. Придатне ділення даних досягається за рахунок гіперплощини, що має найбільшу відстань найближчих тренувальних зразків будь-якого класу (функціональна маржа), оскільки чим більша маржа, тим менша узагальнена помилка класифікатора.

Наївний байєсівський класифікатор

Наївний байєсівський класифікатор — простий імовірнісний класифікатор, який ґрунтується на застосуванні теореми Байєса з наївним припущенням щодо незалежності ознак. Для оцінювання параметрів моделей використовується метод максимальної правдоподібності, тобто працюючи з наївною байєсівською моделлю, можна і не знати теорії байєсівської ймовірності і не використовувати байєсівські методи.

Імовірнісна модель класифікатора — це модель вигляду $p(L|F_1, \dots, F_n)$ залежної змінної класу L з невеликою кількістю класів, що залежить від змінних F_1, \dots, F_n .

Використовуючи теорему Байєса, запишемо ймовірність:

$$p(L|F_1, \dots, F_n) = \frac{p(L)p(F_1, \dots, F_n|L)}{p(F_1, \dots, F_n)}.$$

Чисельник являє собою еквівалентність формули спільної ймовірності моделі вигляду $p(L, F_1, \dots, F_n)$, яку можна, використовуючи повторне застосування визначення умовної ймовірності, переписати у вигляді

$$p(L, F_1, \dots, F_n) = p(L)p(F_1|L)p(F_2|L, F_1) \dots.$$

Застосуємо наївні припущення щодо умовної незалежності: нехай кожна ознака F_i умовно незалежна від будь-якої іншої ознаки F_j , якщо $i \neq j$, тобто

$$p(F_i|L, F_j) = p(F_i|L).$$

Звідси спільну модель можна подати так:

$$p(L, F_1, \dots, F_n) = p(L)p(F_1|L)p(F_2|L) \dots p(F_n|L) = p(C) \prod_{i=1}^n p(F_i|L).$$

Відповідно до припущення про незалежності ознак імовірність за міткою класу

$$p(L|F_1, \dots, F_n) = \frac{1}{p(F_1, \dots, F_n)} p(L) \prod_{i=1}^n p(F_i|L).$$

Випадковий ліс

Випадковий ліс — алгоритм машинного навчання, який застосовується для задач регресії та класифікації. Він використовує ансамблеве навчання — метод, що комбінує велику кількість класифікаторів для вирішення складних проблем. Алгоритм випадкового лісу складається з множини дерев рішень. Ліс, згенерований даним алгоритмом, навчається за допомогою беггінгу — ансамблевого металагоритму, що покращує точність алгоритмів машинного навчання за рахунок усереднення результатів [2]. Метод установлює результати на основі прогнозів дерев рішень.

Головна відмінність алгоритму випадкового лісу від алгоритму дерева рішення полягає в тому, що встановлення кореневих вузлів і відокремлення вузлів відбуваються випадково у першому випадку. Метод використовує беггінг для генерації необхідних прогнозів. Беггінг включає використання різних наборів даних з навчальної вибірки частіше ніж один. Навчальна вибірка складається зі спостережень та ознак, що використовуються для прогнозів. Дерева рішень дають різні результати залежно від навчальних даних, що подаються до алгоритму. Результати ранжируються, а прогноз з найвищою оцінкою обирається як кінцевий результат.

Для задачі класифікації вибір остаточного результату відбувається за допомогою системи мажоритарного голосування. Вихід, обраний більшістю дерев рішень, стає кінцевим результатом випадкового лісу.

XGBoost

Бустинг — це підхід, що комбінує множину простих моделей у єдину композиційну модель. Із додаванням більш простих моделей загальна модель перетворюється у сильний предиктор. За термінологією бустингу прості моделі називаються слабкими моделями.

Градiєнтний бустинг перевизначає бустинг як проблему числової оптимізації, у якій задача полягає у мінімізації функції втрат моделі за рахунок слабких моделей з використанням алгоритму градiєнтного спуску. Оскільки градiєнтний бустинг ґрунтується на мінімізації функції втрат, можуть використовуватися різні функції втрат, що можна застосовувати для задач регресії, класифікації тощо.

Градiєнтний бустинг — покорокова адитивна модель, що генерує навчальні моделі протягом процесу навчання (додається одне дерево за один крок, дерева, що існують, у моделі не змінюються). Внесок слабкої моделі навчання до ансамблю ґрунтується на оптимізаційному процесі градiєнтного спуску. Розрахований внесок кожного дерева базується на мінімізації загальної помилки сильної моделі.

XGBoost — один з найбільш популярних варіантів градiєнтного бустингу. Це ансамблевий алгоритм машинного навчання, який базується на деревах рішень. Він використовує попередньо відсортований алгоритм та алгоритм на основі гiстограми для обчислення найкращого поділу. Алгоритм на основі гiстограм розбиває точки даних для ознаки на дискретні інтервали і використовує ці інтервали для знаходження значення поділу гiстограми.

Багат шаровий персептрон

Це глибока штучна нейронна мережа, що складається із вхідного шару для отримання сигналу, вихідного шару для прийняття рішень або передбачення щодо вхідного сигналу, а між ними — довільна кількість шарів, які являють собою обчислювальну основу багат шарового персептрона.

Архітектура нейронної мережі, що була створена для оцінювання стану страхових компаній, складалася з чотирьох прихованих шарів, де перший шар містив чотири нейрони, решта — вісім. Такої кількості достатньо для вирішення завдання оцінювання. Як функцію активації для прихованих шарів використовувалася функція гіперболічного тангенса:

$$f(x) = \tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}.$$

Оскільки розглядалася задача бінарної класифікації, то для вихідного шару нейронної мережі використовувалася функція «сигмоїд»:

$$f(x) = \frac{1}{1 + e^{-x}}.$$

Для навчання мережі як алгоритму оптимізації використовувався метод Adam, а для навчання башатошарового перцептрона — методи: batch normalization (паketне нормування) і dropout (вилучення).

Batch normalization (паketне нормування). Цей метод дає змогу підвищити продуктивність навчання і стабілізувати роботу штучних нейронних мереж. Суть методу полягає у тому, що деяким шарам нейронної мережі на вхід подаються уже оброблені дані, що мають нульове математичне сподівання та одиничну дисперсію.

За рахунок паketного нормування досягається більш швидка збіжність моделей, незважаючи на додаткові обчислення. Цей підхід дозволяє кожному шару мережі навчатися більш незалежно від інших шарів. Також виникає можливість використання більш високої швидкості навчання, оскільки паketне нормування гарантує, що виходи вузлів нейронної мережі не матимуть дуже великих чи малих значень. Крім того, метод є в деякому сенсі окремим механізмом регуляризації: такий підхід привносить у виходи прихованих шарів деякий шум. Моделі стають менш чутливими до початкової ініціалізації ваг.

Нехай навчання моделі проводиться за допомогою паketів B розмірністю n]:

$$B = \{x_1, \dots, x_n\}.$$

Алгоритм паketної нормалізації складається з таких кроків:

1. Обчислюється математичне сподівання і дисперсія пакета за формулами:

$$\mu_B = \frac{1}{n} \sum_{i=1}^n x_i; \quad \sigma_B^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu_B)^2.$$

2. Нормалізуються дані пакета:

$$\hat{x}_i = \frac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \varepsilon}},$$

де ε — константа, яка використовується для обчислювальної стійкості (налаштовується разом з іншими параметрами).

3. Застосовується операція стиску та зсуву:

$$y_i = \gamma \hat{x}_i + \beta$$

де γ — параметр стиску; β — параметр зсуву.

Ці параметри налаштовуються перед початком навчання мережі.

Dropout (вилучення). Однією з проблем, яка виникає під час будови нейронної мережі, є явище перенавчання. Метод вилучення — один з найбільш поширених засобів уникнення цієї проблеми завдяки простому алгоритму та задовільним результатам [4].

Розглянемо шар нейронної мережі, що складається з N нейронів.

Даний підхід вимикає нейрон з імовірністю p відповідно, залишає увімкненим з імовірністю $q = 1 - p$, причому ймовірність вимкнення будь-якого нейрона однакова.

Нехай $f(x)$ — функція активації, тоді застосування дропауту для i -го нейрона записується таким чином:

$$U_i = \theta_i f\left(\sum_{j=1}^N w_j x_j + b\right),$$

де w_j — ваги нейрона; b — зміщення; x_j — вхідні дані, а ймовірність $P(\theta_i = 0) = p$. Ця формула використовується під час навчання моделі. Але оскільки на даному етапі нейрон залишається в мережі з імовірністю q , то на етапі тестування потрібно емулювати поведінки нейронної мережі, що була використана під час навчання. Для цього результат вихідного значення функції активації необхідно помножити на коефіцієнт q :

$$U_i = q f\left(\sum_{j=1}^N w_j x_j + b\right).$$

На практиці також використовують техніку зворотного дропауту. На відміну від звичайного методу, множення на коефіцієнт відбувається у процесі навчання, причому цей коефіцієнт дорівнює імовірності того, що нейрон залишиться в мережі. А під час тестування значення нейрона залишається таким самим, як і в методі зворотного поширення помилки.

Тобто вихідне значення i -го нейрона на етапі навчання

$$U_i = \frac{1}{q} \theta_i f\left(\sum_{j=1}^N w_j x_j + b\right),$$

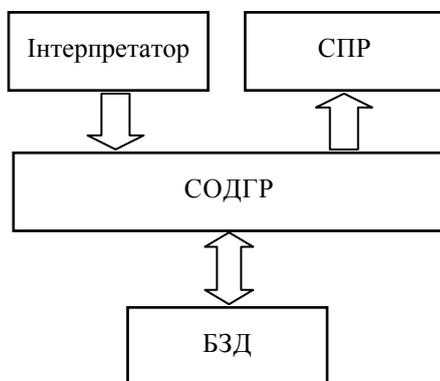
а на етапі тестування

$$U_i = f\left(\sum_{j=1}^N w_j x_j + b\right).$$

АРХІТЕКТУРА І СХЕМА ФУНКЦІОНУВАННЯ СППР

Архітектуру СППР, розроблену для оцінювання стану страхових компаній, зображено на рис. 1.

Компоненти архітектури виконують такі функції:



– інтерпретатор відповідає за безпосереднє завантаження даних джерела інформації. Як дані завантажуються файли формату *.XLSX;

– СОДГР або система оброблення даних та генерації результатів відповідає за візуалізацію даних за допомогою графіків;

– БЗД або база даних і знань відповідає за забезпечення оцінювання стану страхової компанії. Програма відображає значення метрик якості навчання моделей класифікації;

Рис. 1. Архітектура СППР

– СПР або система подання проміжних та остаточних результатів обчислень, дозволяє подавати результати. Програма забезпечує відображення прогнозу майбутнього стану страхової компанії та можливі причини в разі негативної оцінки стану.

Схему роботи СППР показано на рис. 2.

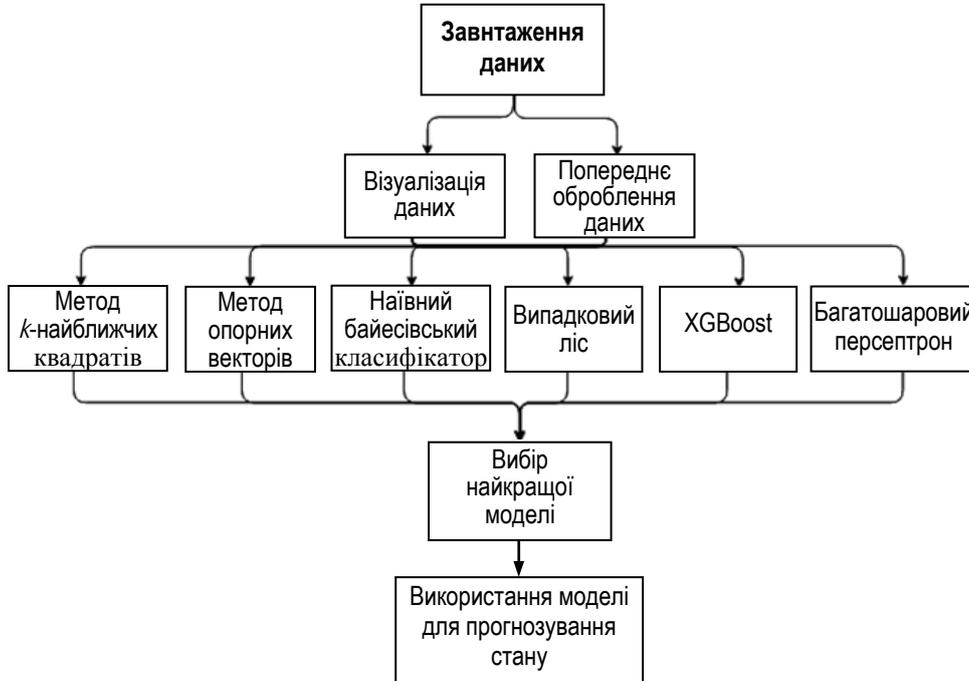


Рис. 2. Схеми функціонування СППР

ДАНІ ДЛЯ ДОСЛІДЖЕННЯ

Для дослідження роботи програми як дані взято основні показники діяльності страхових компаній (у розрізі установ), відповідно до Постановки Кабінету Міністрів України № 835 із сайту Національного банку України. Як навчально-валідаційну вибірку використано показники за 2020 р. (185 компаній), а як тестову вибірку — показники за перше півріччя 2021 року (166 компаній). Мітка оцінки діяльності для навчання моделі заповнювалася у такий спосіб: обчислювалася рентабельність капіталу страхової компанії у відсотках. Якщо цей показник перевищував 10% [5], то компанії присвоювалася оцінка «ефективна», інакше — «неефективна». Кореляційну матрицю фінансових показників та графік страхових компаній за оцінкою діяльності за навчально-валідаційною вибіркою зображено відповідно на рис. 3 і 4.

За кореляційною матрицею видно, що показник збитків до та після оподаткування є некорельованими. Під час дослідження розглядався випадок повної вибірки та скороченої (без урахування збитків). Дані є незбалансованими, оскільки кількість страхових компаній з ефективною діяльністю становить лише 24%.

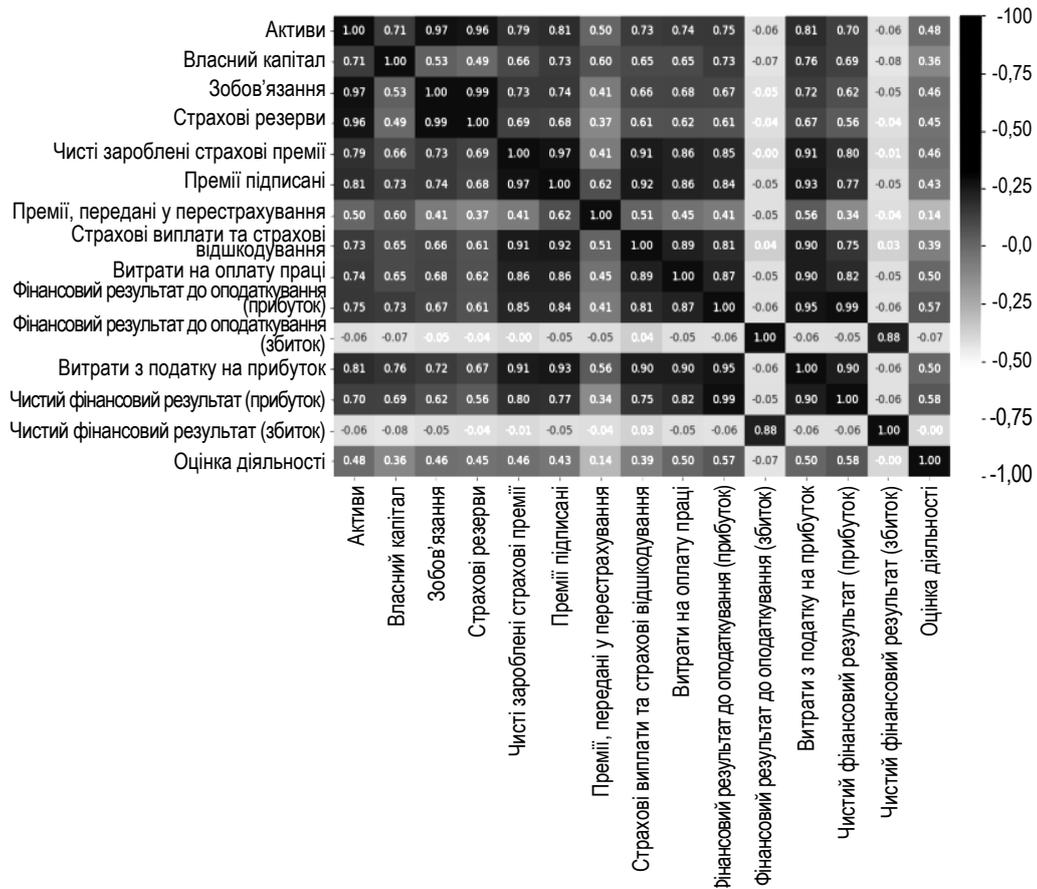


Рис. 3. Кореляційна матриця даних

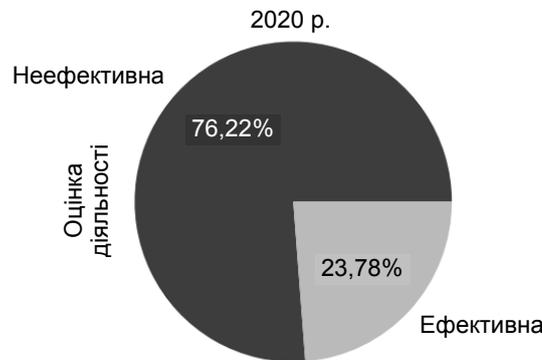


Рис. 4. Графік страхових компаній за оцінкою діяльності за 2020 р.

РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ

Для проведення дослідів навчально-валідаційна вибірка розбивалася за співвідношенням 60:40. Тренувальна вибірка мала розмірність 111, валідаційна — 74. Для оцінювання якості моделей використовувалися метрики F1-Score, Roc-Auc, Balanced Accuracy.

Результати навчання моделей наведено в табл. 1 і 2. Також будувалася матриця похибок для кожної моделі.

Таблиця 1. Результати навчання моделей класифікації для повного набору даних

Модель	F1-Score	Roc-Auc	Balanced Accuracy
<i>k</i> -найближчих сусідів	0,91	0,92	0,92
Метод опорних векторів	0,76	0,81	0,94
Наївний байєсівський класифікатор	0,69	0,78	0,83
Випадковий ліс	0,94	0,94	0,94
XGBoost	0,91	0,94	0,94
Багатошаровий перцептрон	0,91	0,92	0,92

Таблиця 2. Результати навчання моделей класифікації для скороченого набору даних.

Модель	F1-Score	Roc-Auc	Balanced Accuracy
<i>k</i> -найближчих сусідів	0,91	0,94	0,94
Метод опорних векторів	0,86	0,9	0,9
Наївний байєсівський класифікатор	0,77	0,86	0,86
Випадковий ліс	0,86	0,95	0,95
XGBoost	0,88	0,96	0,96
Багатошаровий перцептрон	0,95	0,98	0,99

Матриці похибок моделей для випадку повної та скороченої вибірок показано на рис. 5 і 6.

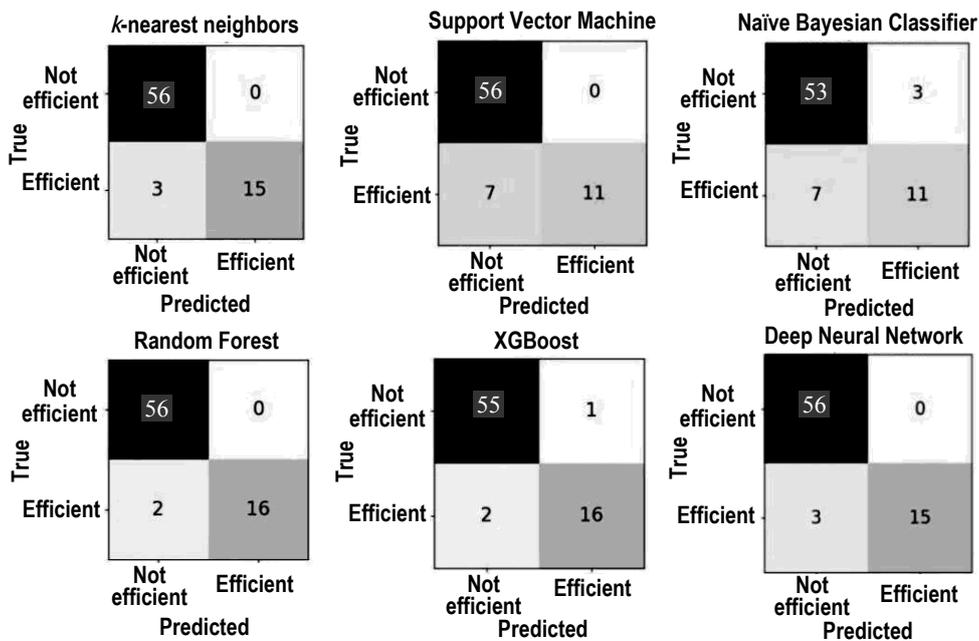


Рис. 5. Матриці похибок моделі за повного набору даних

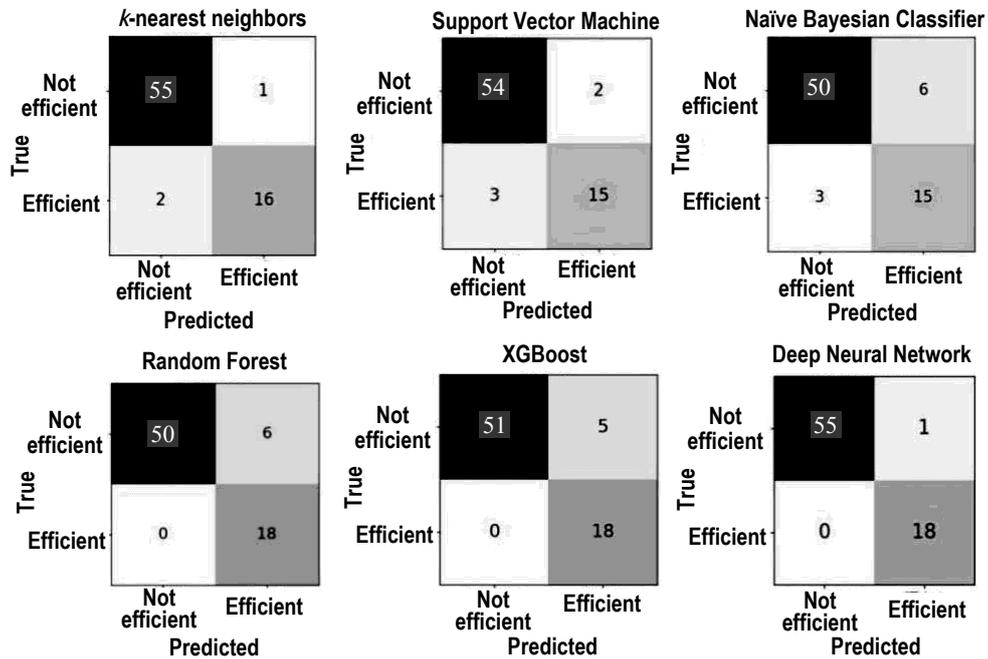


Рис. 6. Матриці похибок моделі за скороченого набору даних

Діаграму розподілу страхових компаній за прогножною оцінкою стану зображено на рис. 7.

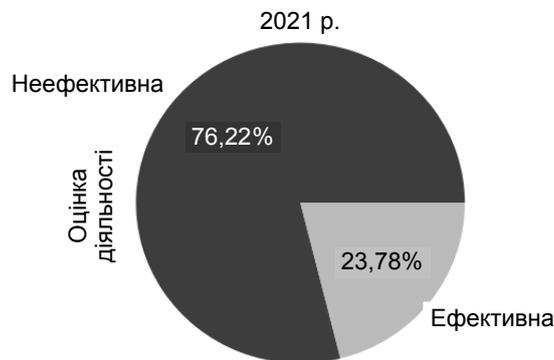


Рис. 7. Графік страхових компаній за спрогнозованою оцінкою діяльності

За результатами таблиць можна бачити, що модель випадкового лісу показала найкращі результати серед усіх моделей. За скороченого набору даних метрики якості усіх моделей покращилися. Водночас багатосаровий перцептрон продемонстрував майже ідеальні результати — його точність становила 99%. За матрицями наочно видно, що за повного набору даних найкращі результати має модель випадкового лісу. За скороченого набору даних для моделі *k*-найближчих сусідів, опорних векторів та наївного байєсівського класифікатора зменшилася кількість помилково-негативних оцінок, але збільшилася кількість помилково-позитивних оцінок. При цьому модель випадкового лісу, XGBoost та нейронної мережі взагалі не містили помилково-негативних оцінок. Для скороченого набору даних багатосаровий перцептрон показав майже безпомилкові результати. Тому, вибираючи

найкращу модель для оцінювання та прогнозування страхових компаній, користувач має вирішити чи вводити показники збитків під час навчання моделей чи не вводити.

ВИСНОВКИ

Створено систему підтримання прийняття рішення, що дозволяє оцінити та спрогнозувати стан страхової компанії за її фінансово-економічними показниками. Для оцінювання та прогнозування стану страхових компаній реалізовано шість моделей класифікації: k -найближчих сусідів, метод опорних векторів, наївний байєсівський класифікатор, випадковий ліс, XGBoost та багат шаровий перцептрон. За результатами дослідження зроблено висновок, що у випадку повної вибірки модель випадкового лісу показала найкращі результати. Водночас багат шаровий перцептрон показав майже ідеальні результати, навчаючись за скороченою вибіркою. Тому, вибираючи найкращу модель для оцінювання та прогнозування страхових компаній, потрібно прийняти рішення щодо того, чи використовувати показники збитків під час навчання моделей чи не використовувати, оскільки вони некорельовані.

У подальших дослідженнях доцільно розглянути задачу побудови комерційної системи підтримання прийняття рішень з функціями ретроспективного аналізу попередніх сесій використання системи, а також адаптації моделей, що використовуються для класифікації досліджуваних об'єктів.

ЛІТЕРАТУРА

1. О.С. Олексюк, *Системи підтримки прийняття фінансових рішень*. К.: Наук. думка, 1998, 508 с.
2. J. Ali, R. Khan, N. Ahmad, and I. Maqsood, "Random Forests and Decision Trees", *International Journal of Computer Science Issues (IJCSI)*, vol. 9, no. 3, pp. 272–278, 2012.
3. S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift", *ICML*, vol. 37, pp. 448–456, 2015.
4. N. Srivatsava et al., "Dropout: A Simple Way to Prevent Neural Networks from Overfitting", *Journal of Machine Learning Research*, vol. 15, pp. 1929–1958, 2014.
5. А.А. Шірінян та Л.В. Шірінян, "Показники ефективності страхового ринку України з позицій власників страхових компаній", *Ефективна економіка*, № 10, 2019. doi: 10.32702/2307-2105-2019.10.172

Надійшла 12.01.2022

INFORMATION ON THE ARTICLE

Roman S. Panibratov, ORCID: 0000-0002-8604-4420, Institute for Applied System Analysis of the National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine, e-mail: roman.panibratov@gmail.com

СИСТЕМА ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ ДЛЯ ОЦЕНИВАНИЯ И ПРОГНОЗИРОВАНИЯ СОСТОЯНИЯ СТРАХОВОЙ КОМПАНИИ / Р.С. Панибратов

Аннотация. Создано систему поддержки принятия решений для оценивания и прогнозирования состояния страховой компании за её финансово-экономическими показателями. Оценивание состояния данного учреждения рассмотрено как задачу бинарной классификации: является ли деятельность страховой компании эффективной или нет. Во время исследования реализованы шесть методов машинного обучения с учителем: метод k -ближайших соседей, метод опорных векторов, наивный байесовский классификатор, случайный лес, XGBoost и глубокая нейронная сеть. Созданная система позволяет: выполнить корреляционный анализ финансово-экономических показателей, проверять сбалансированность данных, обучать выбранную модель и оценивать качество обучения, прогнозировать состояние страховой компании по выбранной модели. По лучшим моделям спрогнозировано будущее состояние страховых компаний Украины.

Ключевые слова: страховая компания, система поддержки принятия решений, бинарная классификация, метод k -ближайших соседей, метод опорных векторов, наивный байесовский классификатор, случайный лес, градиентный бустинг, нейронные сети.

DECISION SUPPORT SYSTEM FOR ESTIMATING AND FORECASTING STATE OF INSURANCE COMPANY / R.S. Panibratov

Abstract. The decision support system was created for estimating and forecasting the state of an insurance company according to its financial and economic indicators. The task of estimating the state of this type of an institution was considered as a problem of a binary classification: whether the company's activity is efficient or not. During the research, six supervised machine learning methods were implemented: k -nearest neighbors, support vector machine, naive Bayes classifier, random forest, XGBoost and deep neural network. The created system allows the following: to perform correlation analysis of financial and economic indicators, to check the balance of data, to perform training of the selected model and to estimate quality of training, to predict the state of the insurance company according to the selected model. According to the best model, the future state of insurance companies in Ukraine was predicted.

Keywords: insurance company, decision support system, binary classification, k -nearest neighbors, support vector machine, naive Bayes classifier, random forest, gradient boosting, neural networks.

REFERENCES

1. O.S. Oleksiuk, *Financial Decision Support Systems*, (in ukr.). Kyiv: Naukova Dumka, 1998, 508 p.
2. J. Ali, R. Khan, N. Ahmad, and I. Maqsood, "Random Forests and Decision Trees", *International Journal of Computer Science Issues (IJCSI)*, vol. 9, no. 3, pp. 272–278, 2012.
3. S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift", *ICML*, vol. 37, pp. 448–456, 2015.
4. N. Srivatsava et al., "Dropout: A Simple Way to Prevent Neural Networks from Overfitting", *Journal of Machine Learning Research*, vol. 15, pp. 1929–1958, 2014.
5. A.A. Shirinyan and L.V. Shirinyan, "Efficiency indicators of ukrainian insurance market from the position of owners of insurance companies", *Efektivna ekonomika*, vol. 10, 2019. doi: 10.32702/2307-2105-2019.10.172

**HYBRID GMDH DEEP LEARNING NETWORKS – ANALYSIS,
OPTIMIZATION AND APPLICATIONS IN FORECASTING
AT FINANCIAL SPHERE**

Yu. ZAYCHENKO, He. ZAICHENKO, G. HAMIDOV

Abstract. In this paper, the new class of deep learning (DL) neural networks is considered and investigated — so-called hybrid DL networks based on self-organization method Group Method of Data Handling (GMDH). The application of GMDH enables not only to train neural weights, but also to construct the network structure as well. Different elementary neurons with two inputs may be used as nodes of this structure. So the advantage of such a structure is the small number of tuning parameters. In this paper, the optimization of parameters and the structure of hybrid neo-fuzzy networks was performed. The application of hybrid DL networks for forecasting market indices was considered with various forecasting intervals: one day, one week, and one month. The experimental investigations of hybrid GMDH neo-fuzzy networks were carried out and comparison of its efficiency with FNN ANFIS in the forecasting problem was performed which enabled to estimate their efficiency and advantages.

Keywords: hybrid deep learning networks, self-organization, parameters and structure optimization, forecasting.

INTRODUCTION

Nowadays deep learning (DL) networks are widely used in different problems of artificial intelligence: forecasting, pattern recognition, medical diagnostics, etc.[1–4]. For its training various algorithms were developed usually based on Back propagation method. Presence of many layers when using gradient algorithm usually leads to occurrence drawbacks as vanishing or explosion of gradient. Therefore, the approach was suggested how to exclude this drawback to perform layer after layer training using stacked encoder-decoder or stacked restricted Boltzmann machines [1, 2]. However, the problem is left how to choose the number of layers in DL network. The existing DL methods don't enable to generate structure of DL networks. But the training process will be more efficient if to adapt not only neuron weights but the structure of network as well. For this goal the application of GMDH method seems very promising. GMDH is based on principle of self- organization and enables to construct network structure automatically in the process of algorithm run [5–7]. In the previous years GMDH-neural networks having active neurons [5–7], R-neurons [19], Q-neurons [3] as

nodes were developed; in the area integrating fuzzy GMDH and neural networks the GMDH neuro-fuzzy and GMDH neo-fuzzy systems [13] were developed.

The very important property of GMDH is that as building blocks for construction of a structure of DL networks elementary models with only two inputs, so-called partial descriptions, are used. This allows to cut substantially training time for hybrid DL network as compared with conventional DL networks.

Therefore, new generation of deep learning — GMDH-hybrid neuro-fuzzy networks were developed in [16] that combine advantages of the traditional GMDH and DL fuzzy networks and may be trained with simple learning procedures. The nodes of this network are Wang-Mendel elementary neural networks with only two inputs. The experimental investigations of this class of hybrid DL networks have shown their efficiency and preference over conventional DL networks. But the drawbacks of application of Wang-Mendel networks as nodes of hybrid DL networks lies herein that it's necessary to train not only neural weights but membership functions as well.

Later another class of hybrid DL networks — GMDH neo-fuzzy networks were developed wherein as nodes of network neo-fuzzy neurons with two inputs are used [17]. For their training its necessary to adapt only neuron weights that demands less computational resources and cuts training time. That's very important for DL networks with a large number of hidden layers. The experimental investigations of hybrid neo-fuzzy networks and comparison with conventional DL network have shown their efficiency and less computational calculations for training. But the problem is left to find the optimal parameters and structure of hybrid neo-fuzzy networks and investigate them in practical applications.

The goal of this paper is to find optimal parameters and structure of hybrid deep learning networks and investigate their efficiency in forecasting problem at financial markets.

HYBRID NETWORK STRUCTURE OPTIMIZATION BASED ON GMDH METHOD

The GMDH method was used to synthesize the structure of the hybrid network based on the principle of self-organization. The principal idea of generation optimal structure is the successive increase in the number of layers until the value of the external criterion of optimality MSE begins to increase for the best model of the current layer. In this case it is necessary to return to the previous layer, to find there the best model with the minimum value of criterion. Then moving backward, go through its connections, find the corresponding neurons of the previous layer. This process continues until we reach the first layer and the corresponding structure is automatically determined.

The process of synthesis of the network structure in the forward direction is shown in Fig. 1 where in grey color the outputs which passed through selection block (SB) are shown while in black color -outputs which were dropped (excluded) by SB.

The process of restoring the desired structure in the backward direction is shown in Fig. 2. In the grey color nodes and their connections selected by this process are indicated.

The corresponding optimal constructed structure of the hybrid network for this forecasting problem is shown in Fig. 3.

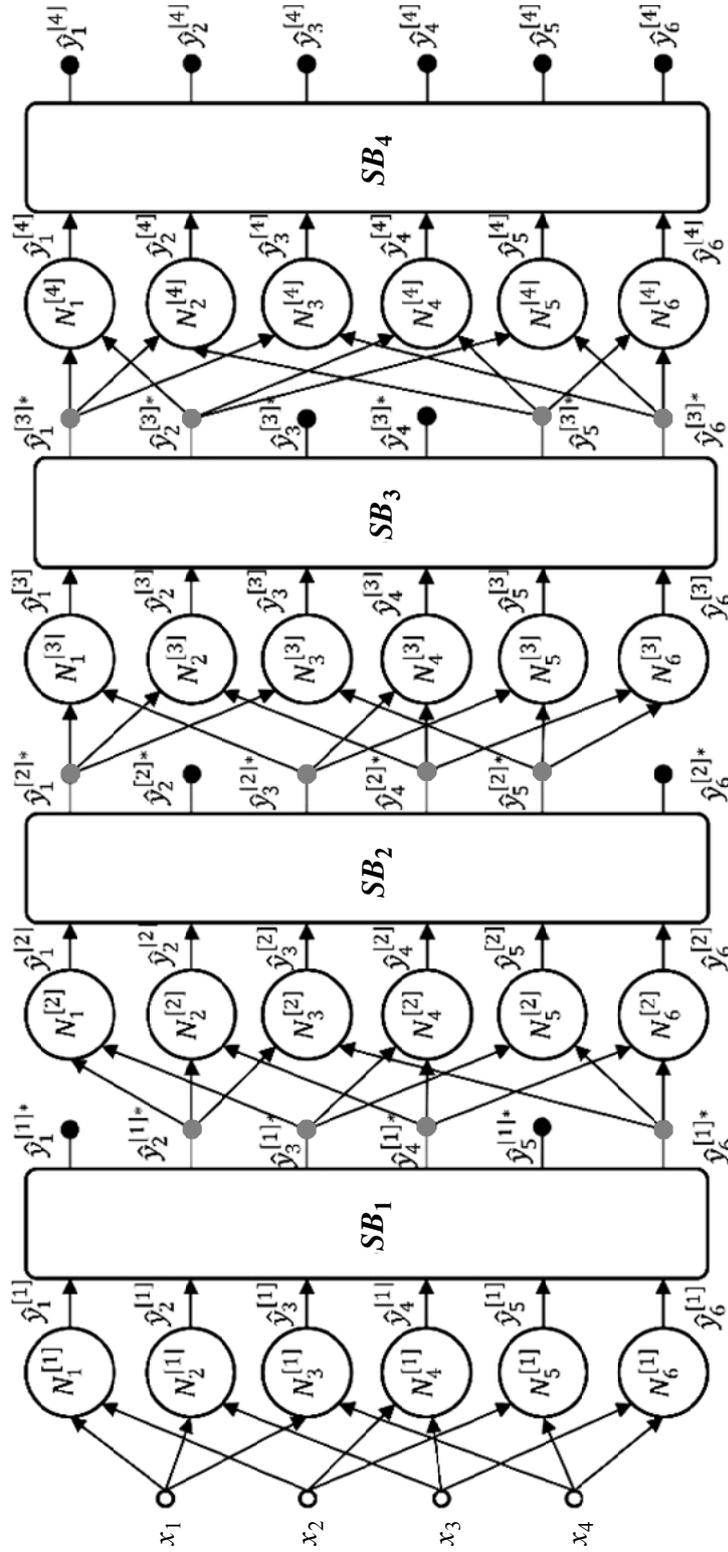


Fig. 1. Hybrid network structure construction using GMDH method

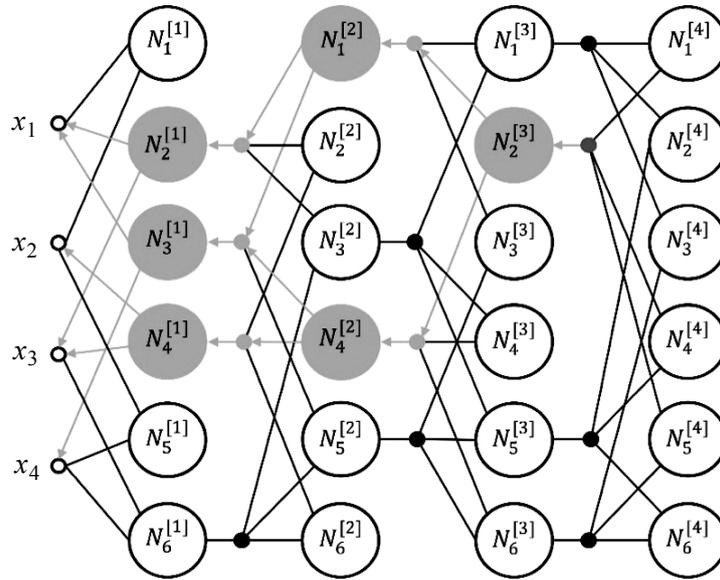


Fig. 2. Process of restoring found optimal structure in backward direction

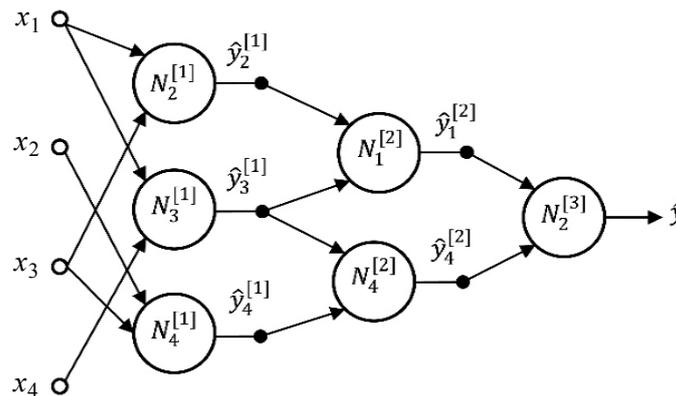


Fig. 3. Optimal Structure of hybrid network for covid forecast constructed by GMDH

It consists of 3 layers: first layer has 3 neo-fuzzy neurons, second layer- two neurons and the last- one neuron.

EXPERIMENTAL INVESTIGATIONS FOR SEARCH OPTIMAL PARAMETERS OF HYBRID GMDH NEO-FUZZY NETWORK

The experimental investigations of hybrid GMDH neo-fuzzy network were performed in the problem of Dow Jones and Nasdaq Index forecasting and compared with FNN ANFIS. In the process of experiments optimal parameters and structure of hybrid GMDH networks were found. The experiments were performed with different forecasting intervals: one day, one week and one month. For each forecasting interval optimal parameters of hybrid neo-fuzzy networks were found and investigated.

The Dow Jones is the stock index of the 30 largest American companies, which was founded in 1896. The initial data was taken from Yahoo, a leading financial information provider owned by Yahoo!

To prepare the initial data, data were uploaded at various intervals, namely the value of the stock index by days, weeks and months. Each of the sets contains the following data:

- Date — data period;
- Open — opening price;
- High — the highest price for the period;
- Low — the lowest price for the period;
- Close — the price at the end of the period;
- Adj Close — average closing price;
- Volume — sales for the period.

The data set for the interval of one day contains 4867 records, of which non-zero records are 4788 ones. The data set for the interval one month contains 1001 records, of which 1000 records are non-zero. The data set for the interval of one month contains 195 records, of which 195 are non-zero.

Data normalizing. Reduction to a single scale is provided by normalization of each variable to the range of its values. In the simplest case, it is a linear transformation

$$mx_i = \frac{x - x_{i \min}}{x_{i \max} - x_{i \min}}, \text{ in the interval } x_i \in [0,1].$$

To find the most informative features as an input vector the network was alternately trained on data sets that transmit only the following features subsets:

('Open', 'High', 'Low', 'Volume', 'Close'); ('Open', 'High', 'Low', 'Volume'); ('Open', 'High', 'Low', 'Close'); ('Open', 'High', 'Low'); ('Open', 'High', 'Close'); ('Open', 'High', 'Volume'); ('Open', 'Close', 'Low'); ('Open', 'Volume', 'Low'); ('High', 'Low', 'Close'); ('Open', 'High'); ('High', 'Close'); ('Low', 'Close'); ('Open', 'Volume').

The main network parameters that can be configured and are to be optimized include the size of the input vector, the number of rules, and the function that sets them, the number of parameters that are transferred to the next layer.

The size of the input vector is determined by the number of informative features that are transmitted for training, and the number of days on the basis of which the network gives the predicted value. Also, the number of network functions that can be set includes the number of membership functions and their appearance, as well as the degree of freedom of choice of the system.

To select these parameters, it is necessary to conduct an experiment, training the system, setting these parameters in the interval, and keeping those that give the best results in the test sample.

The following parameters were investigated:

- n — number of preceding days, based on which the forecasting is performed (sliding window size) $N \in [1;6]$;

- h — number of membership functions in each node, $N \in [2;9]$;
- s — membership function parameter, where $\sigma = \frac{(b-a)}{h}(sh-1)$;
- b — an interval end;
- a — an interval beginning;
- h — membership functions number, which cover the interval;
- $s \in [0,01; 1,5]$;
- f — number of parameters which are transferred to the network next layer (freedom of choice).

The set of initial data was divided into a training sample and test sample in the ratio of 70% and 30%, respectively.

Having launched GMDH neo-fuzzy system for training, values of MAE and MAPE criteria were obtained with different combinations of these parameters.

For the Dow Jones stock index with different forecast intervals, the best parameters for the different set of informative features were obtained as a result of training and testing, which are shown in Table 1.

Table 1. The results of the selection of the optimal parameters of GMDH-neo-fuzzy system for Dow Jones index with different prediction intervals

Sets of informative features	1 month						1 week					
	n	h	f	s	MAE	MAPE	n	h	f	s	MAE	MAPE
'Open', 'High', 'Low', 'Volume', 'Close'	1	2	2	1,0	0,0147	0,0452	2	4	2	0,7	0,0077	0,0295
'Open', 'High', 'Low', 'Volume'	1	2	3	1,3	0,0156	0,0476	2	4	3	0,9	0,0086	0,0332
'Open', 'High', 'Low', 'Close'	1	2	2	1,0	0,0147	0,0453	2	4	2	0,7	0,0077	0,0295
'Open', 'High', 'Low'	1	2	3	1,3	0,0156	0,0476	2	4	3	0,9	0,0086	0,0332
'Open', 'High', 'Close'	1	2	3	1,2	0,0153	0,0467	2	4	3	0,9	0,0079	0,0309
'Open', 'High', 'Volume'	5	2	5	0,1	0,0177	0,0654	2	4	3	1,0	0,0098	0,0380
'Open', 'Low', 'Close'	1	2	3	1,2	0,0147	0,0456	2	4	3	0,7	0,0081	0,0308
'Open', 'Volume', 'Low'	5	3	7	0,1	0,0171	0,0644	4	2	6	0,1	0,0095	0,0348
'High', 'Low', 'Close'	1	2	2	1,0	0,0147	0,0453	2	4	2	0,7	0,0077	0,0295
'Open', 'High'	5	2	5	0,1	0,0177	0,0654	2	4	3	1,0	0,0098	0,0380
'Open', 'Close'	1	2	2	1,3	0,0165	0,0498	2	4	3	0,6	0,0085	0,0331
'High', 'Close'	1	2	2	1,2	0,0154	0,0467	2	4	3	0,9	0,0079	0,0309
'Low', 'Close'	1	2	2	1,2	0,0147	0,0456	2	4	2	0,7	0,0081	0,0306
'Open', 'Volume'	5	2	2	0,8	0,0189	0,0689	3	4	2	0,1	0,0112	0,0445

Thus, analyzing presented results one may conclude that the most informative for GMDH neo-fuzzy system are the following sets of features: ['Open', 'High', 'Close'], ['Open', 'Low', 'Close'], ['High', 'Low', 'Close'], ['High', 'Close'], ['Low', 'Close'].

For the Dow Jones stock index for one month forecast period, the following optimal configurations of GMDH neo-fuzzy network were obtained:

- the number of informative features — 3;
- the number of periods on the basis of which the forecast is made — 1;
- the number of membership functions in each of the nodes — 2;
- the number of layers — 2;
- the number of nodes in the first layer — 3;
- number of nodes on the second layer — 1.

For the Dow Jones stock index for the one week forecast period, the following optimal configurations of the GMDH neo-fuzzy system were obtained:

- the number of informative features — 3;
- the number of periods on the basis of which the forecast is made — 2;
- the number of membership functions in each of the nodes — 4;
- the number of layers — 2;
- the number of nodes on the first layer — 24;
- the number of nodes on the second layer — 1.

The form of the membership function for forecasting interval of one week is shown in the Fig. 4.

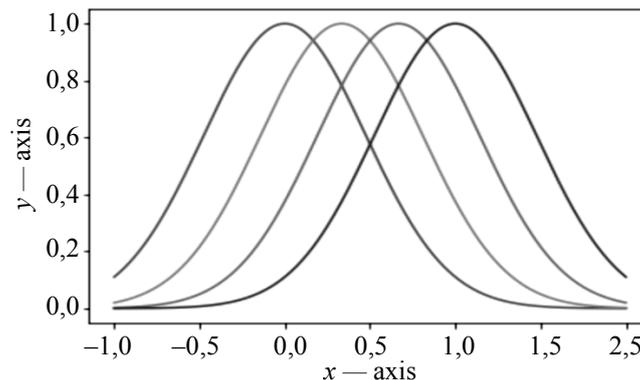


Fig. 4. Forms of the membership function of Dow Jones index for the forecast period of 1 week

For the Dow Jones stock index for one day forecast period, the following optimal configurations of GMDH neo-fuzzy network were obtained:

- number of informative features — 3;
- the number of periods on the basis of which the forecast is made — 5;
- the number of membership functions in each of the nodes — 2;
- the number of layers — 2;
- the number of nodes in the first layer — 30;
- the number of nodes in the second layer — 1.

In the next series of experiments the optimal parameters of hybrid GMDH neo-fuzzy network were searched for the problem of Nasdaq index forecast with different forecasting intervals. The optimal parameters and sets of informative features for interval one month and one week are presented in the Table 2, while for the interval one day — in the Table 3.

Table 2. The results of the selection of the optimal parameters of GMDH neo-fuzzy system for Nasdaq index with different prediction intervals

Sets of informative features	1 month						1 week					
	<i>n</i>	<i>h</i>	<i>f</i>	<i>s</i>	MAE	MAPE	<i>n</i>	<i>h</i>	<i>f</i>	<i>s</i>	MAE	MAPE
'Open', 'High', 'Low', 'Volume', 'Close'	1	3	2	0,58	0,0090	0,0796	3	3	2	0,1	0,0043	0,0400
'Open', 'High', 'Low', 'Volume'	1	3	3	0,78	0,0088	0,0812	5	2	3	0,7	0,0048	0,0445
'Open', 'High', 'Low', 'Close'	1	3	3	0,58	0,090	0,0796	3	3	2	0,1	0,0044	0,0400
'Open', 'High', 'Low'	1	3	3	0,78	0,0088	0,0812	5	2	3	0,7	0,0048	0,0445
'Open', 'High', 'Close'	1	3	3	0,68	0,0095	0,0824	3	3	2	0,1	0,0045	0,0427
'Open', 'High', 'Volume'	2	3	3	0,18	0,0109	0,1124	4	8	3	0,9	0,0054	0,0520
'Open', 'Low', 'Close'	1	3	3	0,88	0,0085	0,0850	3	5	2	0,5	0,0044	0,0414
'Open', 'Volume', 'Low'	2	3	3	0,48	0,0095	0,0941	5	2	4	0,1	0,0051	0,0472
'High', 'Low', 'Close'	2	5	3	0,08	0,0089	0,0796	3	3	2	0,1	0,0043	0,0400
'Open', 'High'	2	3	3	0,18	0,0109	0,1128	4	8	3	0,9	0,0054	0,0520
'Open', 'Close'	2	2	3	0,18	0,0093	0,1011	4	6	2	0,5	0,0046	0,0421
'High', 'Close'	1	3	2	0,68	0,0095	0,1066	3	3	2	0,1	0,0045	0,0427
'Low', 'Close'	1	3	2	0,88	0,0085	0,085	3	5	2	0,5	0,0044	0,0414
'Open', 'Volume'	2	5	4	1,38	0,0121	0,1503	4	7	3	0,9	0,0064	0,0597

Table 3. The results of the selection of the optimal parameters of GMDH neo-fuzzy system for Nasdaq index with one day prediction interval

Sets of informative features	1 day					
	<i>n</i>	<i>h</i>	<i>f</i>	<i>s</i>	MAE	MAPE
'Open', 'High', 'Low', 'Volume', 'Close'	6	8	2	0,1	0,0023	0,0193
'Open', 'High', 'Low', 'Volume'	6	7	3	0,7	0,0026	0,0232
'Open', 'High', 'Low', 'Close'	6	8	2	0,1	0,0023	0,0193
'Open', 'High', 'Low'	6	7	3	0,7	0,0026	0,0232
'Open', 'High', 'Close'	6	7	2	0,1	0,0024	0,0204
'Open', 'High', 'Volume'	6	10	3	0,1	0,0030	0,0262
'Open', 'Low', 'Close'	6	7	5	0,1	0,0024	0,0200
'Open', 'Volume', 'Low'	6	9	5	0,1	0,0028	0,0242
'High', 'Low', 'Close'	6	8	2	0,1	0,0023	0,0193
'Open', 'High'	1	7	2	0,1	0,0029	0,0241
'Open', 'Close'	6	8	6	0,1	0,0025	0,0213
'High', 'Close'	6	7	2	0,1	0,0024	0,0205
'Low', 'Close'	6	9	6	0,1	0,0024	0,0202
'Open', 'Volume'	6	7	2	0,1	0,0034	0,0288

For Nasdaq stock index for one month forecast period, the following optimal configurations of GMDH-neo-fuzzy network were obtained:

- the number of informative features — 4;
- the number of periods on the basis of which the forecast is made — 1;

- the number of membership functions in each of the nodes — 3;
- the number of layers — 2;
- the number of nodes in the first layer — 12;
- number of nodes on the second layer — 1.

For Nasdaq stock index for the one week forecast period, the following optimal configurations of the GMDH-neo-fuzzy system were obtained:

- the number of informative features — 4;
- the number of periods on the basis of which the forecast is made — 3;
- the number of membership functions in each of the nodes — 3;
- the number of layers — 2;
- the number of nodes on the first layer — 36;
- the number of nodes on the second layer — 1.

In the Fig. 5 forms of membership functions of Nasdaq index for one month forecast are presented.

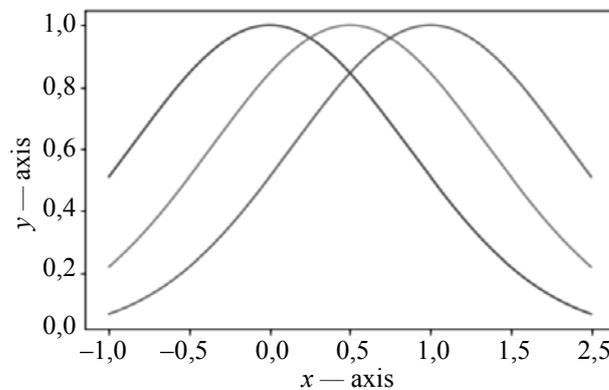


Fig. 5. Forms of the membership function of Nasdaq index for the forecast period of 1 month

For Nasdaq index with forecasting interval 1 month the following results were obtained:

- MAE — 0,02812;
- MAPE — 0,03165;
- Forecasting time — 0,0005815 s.

For Nasdaq index with forecasting interval 1 week the following results were obtained:

- MAE — 0,0099397;
- MAPE — 0,0109336;
- Forecasting time — 0,0003004 s.

For Nasdaq index with forecasting interval 1 day the following results were obtained

- MAE — 0,005740;
- MAPE — 0,0063267;
- Forecasting time — 0,000287 s.

Next, experiments were performed to find the optimal values of the parameters of FNN ANFIS. The size of the input vector is determined by the number of informative features that are transmitted for training, and the number of days of prehistory, on the basis of which the forecasting is performed.

To select these parameters, an experiment was performed, including training of the network, setting these parameters in the interval, and choosing those that give the best results at the test sample.

The set of initial data was divided into a training sample and test data in the proportion of 70% and 30%, respectively. By launching the ANFIS network with different combinations of these parameters, data on MAE and MAPE criteria were obtained.

For the Dow Jones stock index one month forecast period, the following optimal ANFIS network configurations were obtained:

- number of informative features — 3;
- number of nodes — 6;
- the number of periods on the basis of which the forecast is made — 2;
- the number of membership functions in each of the nodes — 6.

The optimal parameters of FNN ANFIS for Dow Jones index forecast are shown in Table 4.

Table 4. The results of the selection of the optimal characteristics of ANFIS network for Dow Jones index with different forecast intervals

Sets of informative features	1 month				1 week				1 day			
	<i>n</i>	<i>h</i>	MAE	MAPE	<i>n</i>	<i>h</i>	MAE	MAPE	<i>n</i>	<i>h</i>	MAE	MAPE
'Open', 'High', 'Low'	2	6	0,222	0,0710	1	9	0,0091	0,0334	1	10	0,0037	0,0142
'Open', 'High', 'Close'	2	3	0,0223	0,0727	2	8	0,0080	0,0303	1	11	0,0034	0,0129
'Open', 'Low', 'Close'	2	6	0,0192	0,0680	2	10	0,0804	0,0307	1	5	0,0045	0,0154
'High', 'Low', 'Close'	2	8	0,0209	0,0720	2	9	0,0903	0,0325	2	10	0,0036	0,0134
'High', 'Close'	2	9	0,0223	0,0750	1	3	0,0077	0,0282	1	7	0,0035	0,0135
'Low', 'Close'	2	7	0,0201	0,0691	1	5	0,0094	0,0338	1	5	0,0035	0,0136

After finding all the optimal parameters of GMDH neo-fuzzy system and training parameters, the system was trained, and then the data for prediction was provided. Training and testing of the system took place on data for the period up to 01.01.2021 for monthly periods, and until 01.06.2021 for weekly and day periods.

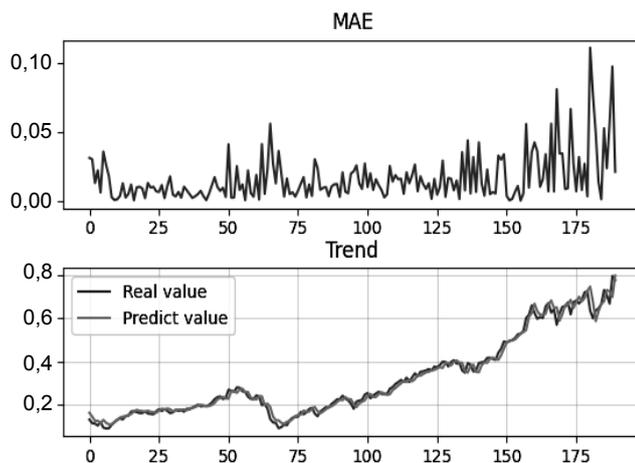


Fig. 6. Results of training and forecasting Dow Jones Index with interval one month by hybrid GMDH neo-fuzzy system

Forecasting was based on data for the period after 01.01.2021 for monthly periods and after 01.06.2021 for day and week periods. For Dow Jones index with a forecast period of one month, the following forecasting data were obtained: MAE — 0,02952; MAPE — 0,0335; forecasting time — 0,00025

Learning and forecasting results are shown in Fig. 6.

COMPARISON OF FORECASTING RESULTS OF GMDH NEO-FUZZY SYSTEM AND ANFIS NETWORK

Experimental investigations of the accuracy of market indexes Dow Jones and Nasdaq forecasting with forecasting intervals of one month, one week and one day were performed, using a hybrid GMDH neo-fuzzy network. For each prediction interval the optimal parameters found in previous experiments were selected. A comparative analysis with the forecasting results obtained by FNN ANFIS was performed.

According to the results of forecasting, values of MAE, MAPE and training time for each type of neural network were obtained. All comparison results are summarized in Tables 5–7 for Dow Jones index and in Tables 8–10 for Nasdaq index.

Table 5. Comparison of the forecasting results of GMDH neo-fuzzy neural network and FNN ANFIS for Dow Jones Index with forecasting interval 1 month

Criterion	GMDH neo-fuzzy neural network	FNN ANFIS	Difference
MAE at training sample	0,016938	0,016135	4,70%
MAPE at training sample	0,061866	0,052607	14,97%
MAE at test sample	0,02952	0,096734	-227,68%
MAPE at test sample	0,03350	0,107397	-220,59%
Training time (sec)	0,0023246	75,258	32375x
Forecasting time (sec)	0,0003123	0,02652	84,92x

Table 6. Comparison of the forecasting results of GMDH neo-fuzzy neural network and FNN ANFIS for Dow Jones Index with forecasting interval 1 week

Criterion	GMDH neo-fuzzy neural network	FNN ANFIS	Difference
MAE at training sample	0,007949	0,008564	-7,74%
MAPE at training sample	0,029890	0,029291	2,00%
MAE at test sample	0,011476	0,019279	-67,99%
MAPE at test sample	0,012468	0,020923	-67,82%
Training time (sec)	0,012840	194,3520	14980x
Forecasting time (sec)	0,00027132	0,028604	105,42x

Table 7. Comparison of the forecasting results of GMDH neo-fuzzy neural network and FNN ANFIS for Dow Jones Index with forecasting interval 1 day

Criterion	GMDH neo-fuzzy neural network	FNN ANFIS	Difference
MAE at training sample	0,003618	0,004234	-17,03%
MAPE at training sample	0,013981	0,014067	-0,615%
MAE at test sample	0,005348	0,005822	-8,86%
MAPE at test sample	0,005812	0,005822	-0,172%
Training time (sec)	0,19944	876,3658	4394,13x
Forecasting time (sec)	0,00040317	0,038055	94,39x

Table 8. Comparison of the forecasting results of GMDH neo-fuzzy neural network and FNN ANFIS for Nasdaq Index with forecasting interval 1 month

Criterion	GMDH neo-fuzzy neural network	FNN ANFIS	Difference
MAE at training sample	0,011264	0,011140	1,10%
MAPE at training sample	0,098307	0,088272	10,21%
MAE at test sample	0,006635	0,008617	-59,87%
MAPE at test sample	0,060995	0,097332	-59,57%
Training time (sec)	0,0065255	34,5328	5291,9x
Forecasting time (sec)	0,0005815	0,024286	41,76x

Table 9. Comparison of the forecasting results of GMDH neo-fuzzy neural network and FNN ANFIS for Nasdaq Index with forecasting interval 1 week

Criterion	GMDH neo-fuzzy neural network	FNN ANFIS	Difference
MAE at training sample	0,0052929	0,0055274	-4,43%
MAPE at training sample	0,041831	0,052723	-26,04%
MAE at test sample	0,009940	0,012973	-30,51%
MAPE at test sample	0,010933	0,014203	-29,91%
Training time (sec)	0,0411811	175,5418	4262,7x
Forecasting time (sec)	0,00030041	0,02489	82,85x

Table 10. Comparison of the forecasting results of GMDH neo-fuzzy neural network and FNN ANFIS for Nasdaq Index with forecasting interval 1 day

Criterion	GMDH neo-fuzzy neural network	FNN ANFIS	Difference
MAE at training sample	0,002349	0,002798	-19,11%
MAPE at training sample	0,019121	0,025317	-32,40%
MAE at test sample	0,005740	0,007161	-24,76%
MAPE at test sample	0,0063267	0,0079001	-24,87%
Training time (sec)	3,8612	823,90	213,39x
Forecasting time (sec)	0,0004616	0,085263	184,72x

Analyzing the presented results one may conclude, the best forecasting results for all forecasting intervals were obtained for hybrid GMDH neo-fuzzy system for both indexes Dow Jones and Nasdaq. The worst forecasting result for ANFIS network was obtained for one month forecasting period. The largest difference in the accuracy of forecasting by both criteria was obtained for the forecasting period of one month (over 200%). As the forecasting period decreases, the gap between the networks accuracy also decreases.

In addition, training and direct prediction times were also significantly less for hybrid GMDH neo-fuzzy system as compared with ANFIS.

CONCLUSION

In the paper new generation of Deep learning networks-hybrid GMDH neo-fuzzy networks are considered, optimized and investigated.

The algorithm of hybrid network structure synthesis is presented and demonstrated at the problem of forecasting.

The experimental investigations of the hybrid networks were carried out and compared with conventional DL networks. The problem of forecasting Dow Jones

and Nasdaq Index with application of hybrid neo-fuzzy networks was considered, investigated and compared with FNN ANFIS at the different forecasting intervals: one month, one week and day.

The optimal parameters of hybrid neo-fuzzy networks and sets of informative features for forecasting problems were found. The experimental results have shown the forecasting accuracy of hybrid neo-fuzzy networks is much better than for FNN ANFIS.

The training time is the least for hybrid neo-fuzzy network as compared with alternative ANFIS network.

In a whole the hybrid DL networks based on GMDH are free from drawbacks of conventional DL networks- decay or explosion of gradient. Besides, they enable to construct optimal network structure automatically in the process of algorithm GMDH run and additionally they demand less computational costs for training due to small number of tunable parameters (only two) in every hidden node as compared with DL networks of general structure. That's is especially significant for DL networks with large number of layers.

REFERENCES

1. I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.
2. G. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets", *Neural Computation*, vol. 18, no. 7, pp. 1527–1554, May 2006.
3. Y. Bengio, Y. LeCun, and G. Hinton, "Deep learning", *Nature*, no. 521, pp. 436–444, May 2015.
4. J. Schmidhuber, "Deep learning in neural networks: an overview", *Neural Networks*, no. 61, pp. 85–117, 2015.
5. A.G. Ivakhnenko, G.A. Ivakhnenko, and J.A. Mueller, "Self-organization of the neural networks with active neurons", *Pattern Recognition and Image Analysis*, 4, 2, pp. 177–188, 1994.
6. A.G. Ivakhnenko, D. Wuensch, and G.A. Ivakhnenko, "Inductive sorting-out GMDH algorithms with polynomial complexity for active neurons of neural networks", *Neural Networks*, 2, pp. 1169–1173, 1999.
7. G.A. Ivakhnenko, "Self-organization of neuronet with active neurons for effects of nuclear test explosions forecasting", *System Analysis Modeling Simulation*, 20, pp. 107–116, 1995.
8. M. Zgurovsky and Yu. Zaychenko, *Fundamentals of computational intelligence: System approach*. Springer, 2016.
9. L.-X. Wang and J.M. Mendel, "Fuzzy basis functions, universal approximation, and orthogonal least-squares learning", *IEEE Trans. on Neural Networks*, vol. 3, no. 5, pp. 807–814, 1992.
10. J.-S. Jang, "ANFIS: Adaptive-network-based fuzzy inference systems", *IEEE Trans. on Systems, Man, and Cybernetics*, 23, pp. 665–685, 1993.
11. T. Yamakawa, E. Uchino, T. Miki, and H. Kusanagi, "A neo-fuzzy neuron and its applications to system identification and prediction of the system behavior", in *Proc. 2nd Intern. Conf. Fuzzy Logic and Neural Networks «LIZUKA-92»*, Lizuka, 1992, pp. 477–483.
12. Ye. Bodyanskiy, N. Teslenko, and P. Grimm, "Hybrid evolving neural network using kernel activation functions", in *Proc. 17th Zittau East-West Fuzzy Colloquium, Zittau/Goerlitz, HS, 2010*, pp. 39–46.
13. Ye. Bodyanskiy, Yu. Zaychenko, E. Pavlikovskaya, M. Samarina, and Ye. Viktorov, "The neo-fuzzy neural network structure optimization using the GMDH for the solving forecasting and classification problems", *Proc. Int. Workshop on Inductive Modeling, Krynica, Poland, 2009*, pp. 77–89.
14. Ye. Bodyanskiy, O. Vynokurova, A. Dolotov, and O. Kharchenko, "Wavelet-neuro-fuzzy network structure optimization using GMDH for the solving forecasting tasks", in *Proc. 4th Int. Conf. on Inductive Modelling ICIM 2013, Kyiv, 2013*, pp. 61–67.
15. Ye. Bodyanskiy, O. Vynokurova, and N. Teslenko, "Cascade GMDH-wavelet-neuro-fuzzy network", in *Proc. 4th Int. Workshop on Inductive Modeling "IWIM 2011", Kyiv, Ukraine, 2011*, pp. 22–30.

16. Ye. Bodyanskiy, O. Boiko Yu. Zaychenko, and G. Hamidov, "Evolving Hybrid GMDH-Neuro-Fuzzy Network and Its Applications", in *Proceedings of the International conference SAIC 2018, Kiev, Ukraine, 2018*.
17. Evgeniy Bodyanskiy, Yuriy Zaychenko, Olena Boiko, Galib Hamidov, and Anna Zelikman, "The hybrid GMDH-neo-fuzzy neural network in forecasting problems in financial sphere", in *Proceedings of the International conference IEEE SAIC 2020, Kiev, Ukraine, 2020*.
18. T. Ohtani, "Automatic variable selection in RBF network and its application to neuro-fuzzy GMDH", *Proc. Fourth Int. Conf. on Knowledge-Based Intelligent Engineering Systems and Allied Technologies, 2000*, vol. 2, pp. 840–843.

Received 17.01.2022

INFORMATION ON THE ARTICLE

Yuriy P. Zaychenko, ORCID: 0000-0001-9662-3269, Institute for Applied System Analysis of the National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine, e-mail: zaychenkoyuri@ukr.net

Helen Yu. Zaychenko, ORCID: 0000-0002-4546-0428, Institute for Applied System Analysis of the National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine, e-mail: syncmaster@bigmir.net

Galib Hamidov, "Azerishiq", Azerbaijan, e-mail: galib.hamidov@gmail.com

ГІБРИДНІ МГУА-МЕРЕЖІ ГЛИБОКОГО НАВЧАННЯ — АНАЛІЗ, ОПТИМІЗАЦІЯ ТА ЗАСТОСУВАННЯ ДЛЯ ПРОГНОЗУВАННЯ У ФІНАНСОВІЙ СФЕРІ / Ю.П. Зайченко, О.Ю. Зайченко, Г. Гамідов

Анотація. Розглянуто та досліджено новий клас мереж глибокого навчання — гібридні мережі глибокого навчання на основі методу самоорганізації МГУА. Застосування МГУА дозволяє навчати не тільки ваги зв'язків, але і конструювати структуру мережі. Як вузли мережі можуть бути використані елементарні нейрони з двома входами. Перевага такої структури — мала кількість налаштовуваних параметрів. Виконано оптимізацію параметрів та структури гібридних неофаззі мереж. Розглянуто застосування гібридних мереж глибокого навчання з оптимізованими параметрами для прогнозування біржових індексів з різними інтервалами упередження — один день, тиждень та місяць. Проведено експериментальні дослідження гібридних МГУА неофаззі мереж та порівняння їх з нечіткою нейронною мережею ANFIS, що дозволило оцінити ефективність та переваги гібридних мереж порівняно звичайними мережами глибокого навчання.

Ключові слова: гібридні мережі глибокого навчання, самоорганізація, оптимізація параметрів і структури, прогнозування.

ГИБРИДНЫЕ МГУА-СЕТИ ГЛУБОКОГО ОБУЧЕНИЯ — АНАЛИЗ, ОПТИМИЗАЦИЯ И ПРИМЕНЕНИЯ ДЛЯ ПРОГНОЗИРОВАНИЯ В ФИНАНСОВОЙ СФЕРЕ / Ю.П. Зайченко, Е.Ю. Зайченко, Г. Гамидов

Аннотация. Рассмотрен и исследован новый класс сетей глубокого обучения — гибридные сети глубокого обучения на основе метода самоорганизации МГУА. Применение МГУА позволяет обучать не только веса связей, но и конструировать структуру сети. В качестве узлов сети могут быть использованы элементарные нейроны с двумя входами. Преимущество такой структуры — малое количество настраиваемых параметров. Выполнена оптимизация параметров и структуры гибридных неофаззи сетей. Рассмотрено применение гибридных сетей глубокого обучения с оптимизированными параметрами для прогнозирования биржевых индексов с различными интервалами упреждения — один день, неделя и месяц. Проведены экспериментальные исследования гибридных МГУА неофаззи сетей и сравнение их с нечеткой нейронной сетью ANFIS, что позволило оценить эффективность и преимущества гибридных сетей по сравнению обычными сетями глубокого обучения.

Ключевые слова: гибридные сети глубокого обучения, самоорганизация, оптимизация параметров и структуры, прогнозирование.

ADAPTIVE HYBRID ACTIVATION FUNCTION FOR DEEP NEURAL NETWORKS

Ye. BODYANSKIY, S. KOSTIUK

Abstract. The adaptive hybrid activation function (AHAF) is proposed that combines the properties of the rectifier units and the squashing functions. The proposed function can be used as a drop-in replacement for ReLU, SiL and Swish activations for deep neural networks and can evolve to one of such functions during the training. The effectiveness of the function was evaluated on the image classification task using the Fashion-MNIST and CIFAR-10 datasets. The evaluation shows that the neural networks with AHAF activations achieve better classification accuracy comparing to their base implementations that use ReLU and SiL. A double-stage parameter tuning process for training the neural networks with AHAF is proposed. The proposed approach is sufficiently simple from the implementation standpoint and provides high performance for the neural network training process.

Keywords: adaptive hybrid activation function, double-stage parameter turning process, deep neural networks.

INTRODUCTION

In the recent years deep neural networks (DNNs) have got a wide proliferation for solving ranges of problems in virtually all areas of human activity, including the fields of Data Mining, Big Data, Data Science, digital video and audio signal processing, natural language processing, forecasting and control of complex systems [1–6].

The common property of all neural networks is their learning ability which consists of tuning the parameters (and, possibly, architectures) during the processing of available information and their universal approximation capabilities [7, 8] that allows to analyze and recover arbitrary complex nonlinear dependencies in the source data.

The most popular neural node of the DNN is the elementary perceptron of F. Rosenblatt which uses so-called squashing functions as their activation functions [7], such as sigmoid σ -functions, which are the most common squashing functions, tanh, Softsign, Satlin, aretan and others. At the same time the application of squashing functions runs against computational difficulties (so-called effect of vanishing gradient) when their derivatives approach zero while the input signal moves further from the origin.

Thereby instead of the squashing functions various DNN implementations commonly use piece-wise activation functions that belong to the so-called “rectified unit” family [9] which includes ReLU, ELU, PReLU, LReLU, NReLU and other similar functions [10–12]. It shall be noted that piece-wise activation functions allow only piece-wise approximation, i.e., the number of nodes and layers in the neural network shall be significantly increased to provide the required approximation capacity for non-trivial dependencies.

At the same time, there is a relatively wide group of recurrent neural networks [13] such as long-short-term memories, transformers and similar networks

that use squashing functions in their gated recurrent units [3], so the hybrid activation functions were introduced that combine the properties of both the rectifiers and sigmoid functions. The list of hybrid functions includes [14], Swish [15], S-shaped [16], WiG [17] and other similar functions [18, 19].

All such hybrid activation functions have some free parameters that define their exact shape, amplitude and singular points which shall be in some way selected and adjusted for solving specific tasks. In this regard, it is advisable to introduce some additional procedures for automatic adjustment of the activation functions parameters. [20–25] address the off-line procedures that allow to find the required function parameters after the synaptic weights of the network are already set up. It is clear this approach significantly increases the training time.

In [26] the adaptive parametric rectified linear activation function (AdPReLU) was introduced where the parameters were adjusted simultaneously with the synaptic weights during the error backpropagation procedure. This approach allowed to reduce the training time and improve the quality of the obtained solution compared to Adaline, ReLU and tanh on the prediction task.

It is advisable to implement a similar approach for hybrid activation functions [14–19] and synthesize on their basis an adaptive activation function that is a generalization of the ones that are already used in the DNN applications.

ARCHITECTURE OF A NEURON WITH ADAPTIVE HYBRID ACTIVATION FUNCTION

Elementary perceptron of F. Rosenblatt as node of a neural network performs a non-linear transformation of the following form:

$$\hat{y}_j(k) = \psi_j \left(\theta_{j0} + \sum_{i=1}^n w_{ji} x_i(k) \right) = \psi_j \left(\sum_{i=0}^n w_{ji} x_i(k) \right) = \psi_j(w_j^T x(k)) = \psi_j(u_j(k)),$$

where $\hat{y}_j(k)$ — output signal of the j -th neuron of the network on the k -th data processing step, $k = 1, 2, 3, \dots, N, \dots$, $\psi_j(u_j(k))$ — non-linear transformation that is performed by the activation function on the signal of internal activation $u_j(k)$, θ_{j0} — threshold signal, w_{ji} — synaptic weight on the i -th input of the j -th neuron, $i = 0, 1, 2, \dots, n$, $w_{j0} \equiv \theta_{j0}$, $w_j = (w_{j0}, w_{j1}, \dots, w_{jn})^T \in R^{n+1}$, $x(k) = (1, x_1(k), \dots, x_n(k))^T$ — $(n+1) \times 1$ — dimensional vector of the input signals.

One of the most popular activation functions in the neural networks is a so-called sigmoid one that is studied by G. Cybenko [7] and has the following form:

$$\psi_j(u_j) = \sigma(\gamma_j u_j) = \frac{1}{1 + e^{-\gamma_j u_j}}, \quad (1)$$

where γ_j — so-called gain parameter [20] that defines the shape of this function. The gain parameter value is often assumed to be equal to 1.

While the usage of sigmoid activation functions allows to provide universal approximation capabilities for the neural network, its application in DNNs runs up against computational complexities when the signal of internal activation starts

to rise in its amplitude. In those cases, the derivative of the σ -function approaches zero, i.e., the effect of “vanishing gradient” increases.

To overcome this problem, we propose using a hybrid activation function of the following form:

$$\psi_j(u_j) = \beta_j u_j \sigma(\gamma_j u_j) = \frac{\beta_j u_j}{1 + e^{-\gamma_j u_j}}, \quad (2)$$

where β_j and γ_j — parameters that shall be determined together with the synaptic weights during the training process. Being a modification of (1), activation function (2) does not suffer from the vanishing gradient effect. Note that the derivative of (2) by the signal of internal activation:

$$\frac{\partial \psi_j(u_j)}{\partial u_j} = \beta_j \sigma(\gamma_j u_j) (1 + u_j \gamma_j (1 - \sigma(\gamma_j u_j)))$$

produces small by amplitude values only when $u_j \ll 0$ that can be compensated by dialing the gain parameter γ_j .

Fig. 1 shows the architecture of an artificial neuron with adaptive hybrid activation function (2) (AHAF) in which function parameters β_j and γ_j are trained together with the vector of synaptic weights.

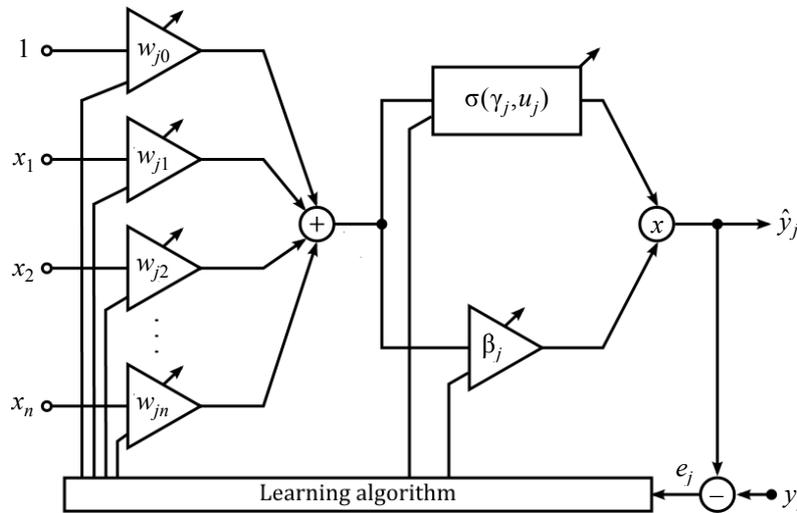


Fig. 1. Neuron with adaptive hybrid activation function (AHAF)

Here γ_j — external reference signal, $e_j = y_j - \hat{y}_j = y_0 - \psi_j(u_j) = y_j - \beta_j u_j (1 + e^{-\gamma_j u_j})^{-1}$ — learning error.

TRAINING ALGORITHM FOR A NEURON WITH AHAF

For training artificial neurons with AHAF we use the standard δ -rule [9] that for a regular perceptron of F. Rosenblatt and the error squared loss criteria:

$$E_j(k) = \frac{1}{2} e_j^2(k) = \frac{1}{2} (y_j(k) - \psi_j(u_j(k)))^2 = \frac{1}{2} \left(y_j(k) - \psi_j \left(\sum_{i=0}^n w_{ji} x_i(k) \right) \right)^2$$

allows to refine the synaptic weights with a recurrent procedure:

$$\begin{aligned} w_{ji}(k) &= w_{ji}(k-1) - \eta_w(k) \frac{\partial E_j(k)}{\partial e_j(k)} \frac{\partial e_j(k)}{\partial w_{ji}} = \\ &= w_{ji}(k-1) - \eta_w(k) e_j(k) \frac{\partial e_j(k)}{\partial w_{ji}} = w_{ji}(k-1) - \eta_w(k) e_j(k) \frac{\partial e_j(k)}{\partial u_j(k)} \frac{\partial u_j(k)}{\partial w_{ji}} = \\ &= w_{ji}(k-1) + \eta_w(k) e_j(k) \psi'_j(u_j(k)) x_i(k) = w_{ji}(k-1) + \eta_w(k) \delta_j(k) x_i(k), \end{aligned}$$

where $\eta_w(k)$ — learning rate parameter the choice of which determines the convergence rate and the filtering (smoothing) abilities of the algorithm, $\delta_j(k) = e_j(k) \psi'_j(u_j(k))$ — so-called δ -error, based on which the error back-propagation procedure is implemented for training of multilayer neural networks.

For a neuron with AHAF that has a two-layer architecture (i.e., the first layer — synaptic weights $w_{ji} = 0, 1, \dots, n$, the second — tunable parameters β_j and γ_j), backpropagation is implemented on a per-neuron level: parameters of the activation function are tuned first, then — the synaptic weights. This training procedure is referenced in this paper as the double-stage parameter tuning procedure (the DSPT procedure).

Considering that the δ -rule for tuning the activation function parameters $\psi_j(u_j, \beta_j, \gamma_j)$:

$$\begin{aligned} \frac{\partial \psi_j}{\partial \beta_j} &= u_j \sigma(\gamma_j u_j) = \frac{u_j}{1 + e^{-\gamma_j u_j}}, \\ \frac{\partial \psi_j}{\partial \gamma_j} &= \beta_j u_j^2 \sigma(\gamma_j u_j) (1 - \sigma(\gamma_j u_j)) = \beta_j \frac{u_j^2}{1 + e^{-\gamma_j u_j}} \frac{e^{-\gamma_j u_j}}{1 + e^{-\gamma_j u_j}} \end{aligned}$$

can be written in the form of:

$$\begin{aligned} \beta_j(k) &= \beta_j(k-1) - \eta_\beta(k) \frac{\partial E_j(k)}{\partial \beta_j} = \beta_j(k-1) + \eta_\beta(k) e_j(k) \frac{\partial \psi_j(k)}{\partial \beta_j} = \\ &= \beta_j(k-1) + \eta_\beta(k) (y_j(k) - \psi_j(u_j(k), \beta_j(k-1), \gamma_j(k-1))) \times \\ &\quad \times u_j(k) \sigma(\gamma_j(k-1) u_j(k)), \end{aligned}$$

where $u_j(k) = w_j^T(k-1)x(k)$, and:

$$\begin{aligned} \gamma_j(k) &= \gamma_j(k-1) - \eta_\gamma(k) \frac{\partial E_j(k)}{\partial \gamma_j} = \gamma_j(k-1) + \eta_\gamma(k) e_j(k) \frac{\partial \psi_j(k)}{\partial \gamma_j} = \\ &= \gamma_j(k-1) + \eta_\gamma(k) (y_j(k) - \psi_j(u_j(k), \beta_j(k-1), \gamma_j(k-1))) \times \\ &\quad \times \beta_j(k-1) u_j^2(k) \sigma(\gamma_j(k-1) u_j(k)) (1 - \sigma(\gamma_j(k-1) u_j(k))) = \\ &= \gamma_j(k-1) + \eta_\gamma(k) e_j(k) u_j^2(k) \sigma(\gamma_j(k-1) u_j(k)) (1 - \sigma(\gamma_j(k-1) u_j(k))), \end{aligned}$$

the training error can be recalculated after the tuning is performed for β_j and γ_j :

$$\tilde{e}_j(k) = y_j(k) - \psi_j(u_j(k), \beta_j(k), \gamma_j(k)) =$$

$$= y_j(k) - \frac{\beta_j(k)u_j(k)}{1 + e^{-\gamma_j(k)u_j(k)}} = y_j(k) - \frac{\beta_j(k)w_j^T(k-1)x(k)}{1 + e^{-\gamma_j(k)w_j^T(k-1)x(k)}},$$

and the synaptic weights are turned:

$$\begin{aligned} w_{ji}(k) &= w_{ji}(k-1) + \eta_w(k)\tilde{e}_j(k) \frac{\partial \psi(u_j(k), \beta_j(k), \gamma_j(k))}{\partial u_j(k)} x_i(k) = \\ &= w_{ji}(k-1) + \eta_w(k)\tilde{e}_j(k)\beta_j(k)\sigma(\gamma_j(k)u_j(k)) \times \\ &\times (1 + u_j(k)\gamma_j(k))(1 - \sigma(\gamma_j(k)u_j(k)))x_i(k) = w_{ji}(k-1) + \eta_w(k)\tilde{\delta}_j(k)x_i(k), \end{aligned}$$

where

$$\begin{aligned} \tilde{\delta}_j(k) &= \tilde{e}_j(k)\psi'_j(u_j(k), \beta_j(k), \gamma_j(k)) = \\ &= \tilde{e}_j(k)\beta_j(k)\sigma(\gamma_j(k)u_j(k))(1 + u_j(k)\gamma_j(k))(1 - \sigma(\gamma_j(k)u_j(k))). \end{aligned}$$

With regards to selection of the learning rate parameters η_β , η_γ , η_w , the adaptive training algorithms like Adam [27], that are popular in DNNs, can be successfully replaced by the ones with the filtering and tracking properties [28] that have a sufficiently high speed of convergence.

For training of multi-layer networks, the hybrid error back propagation procedure can be used that, comparing to the standard one, calculates the training error and the δ -error twice per each hybrid layer of the network: $e_j(k)$, $\tilde{e}_j(k)$, $\delta_j(k)$, $\tilde{\delta}_j(k)$.

EVALUATION

Performance of the adaptive hybrid activation function was evaluated on the image classification task on two different datasets with two base neural network architectures in a similar way to [29]. The base architectures were modified to use AHAF activations instead of “classic” activations like ReLU and SiL. The performance of the modified networks was compared to the reference implementations. The neural network implementations together with the valuation and training environment were coded in Python 3.8 using PyTorch 1.9.0 [30]. The implementation is publicly available on GitHub: <https://git.io/JDBIZ>.

A. Dataset

The models with adaptive hybrid activation function were evaluated on two datasets: Fashion-MNIST [31] and CIFAR-10 [32].

Fashion-MNIST is a dataset that contains 60000 monochrome images, each 28×28 pixels in size, with associated class labels. Out of all images, 50000 images are used for training and 10000 are used for validation. The classes are exclusive, the one-hot encoding was used for the class labels. The pixel values were divided by 255 to rescale them to the $[0, 0 \dots 1, 0]$ range. The images were augmented using the random horizontal flip with the flip probability of 0,5 and the random shift by both width and height with the maximum shift factor of 0,1.

CIFAR-10 is a dataset of 60000 RGB images, each 32×32 pixels in size and each having a one of 10 class labels associated with it. The train to test distribution is 5:1, where all images are randomly selected from the whole dataset. The

classes are exclusive, the one-hot encoding was used for the class labels. Pixel values on all color channels were rescaled to the $[0,0..1,0]$ range using division by 255. The training set was augmented using the random horizontal flip with probability of 0,5 and the random horizontal and vertical shift by the maximum factor of 0,1.

B. Neural Networks and Activations

Two base neural networks architectures were used in the experiment: LeNet-5 [33] and KerasNet from Keras version 1.2.2 [34].

LeNet-5 is a simple convolutional neural network consisting of 4 layers: 2 convolutional layers with pooling and activation functions, 1 linear layer with an activation function and 1 output linear layer with Softmax. The convolutional layers use 5×5 filters with 20 output channels for the first layer and 50 output channels for the second layer. Max pooling with the kernel of 2×2 is used as the pooling implementation. The hidden linear layer has 500 output features, the output layer has 10, one per each class. Several variants of LeNet-5 were used for evaluation: one with ReLU activations for the hidden layers, one with SiL, one AHAF activation initialized as ReLU and one with AHAF activation initialized as SiL. The total number of parameters depends on the size of the input images: 431000 and 657000 for Fashion-MNIST and CIFAR-10 correspondingly. The total number of parameters does not count the parameters of AHAF activations.

KerasNet is a neural network that is partially similar to VGG. The network has 6 layers: 4 convolutional layers with activation functions with each second layer followed by max pooling with dropout, 1 hidden linear layer with an activation function and dropout, 1 output linear layer with Softmax activation. The first and the second convolutional layers have 32 output channels with 3×3 filters, the first layer applies 1×1 padding to its input, while the second one does not apply any padding. Max pooling with 2×2 kernels and the dropout with the probability of 0,25 follow the first two convolutional layers. The third and the fourth convolutional layers use 3×3 filters and have 64 output channels, the third layer applies 1×1 padding while the fourth does not apply any padding. Max pooling with the kernel size of 2×2 and the dropout with the probability of 0,25 are used after the third and fourth convolutional layers. The hidden linear layer has 512 output features, dropout with the probability of 0,5 is applied after the hidden linear layer. The output layer has 10 output features, one per each class. Several variants of KerasNet were used for evaluation: one with ReLU activations for the hidden layers, one with SiL, one AHAF activation initialized as ReLU and one with AHAF activation initialized as SiL. The total number of parameters depends on the size of the input images: 889834 and 1250858 for Fashion-MNIST and CIFAR-10 correspondingly. The total number of parameters does not count the parameters of AHAF activations.

C. Training Procedures

The neural networks were trained on the Fashion-MNIST and CIFAR-10 datasets with the batch size of 64 for 100 epochs on a laptop with NVIDIA GeForce GTX 1650 Max-Q. The RMSprop optimizer was used for training with the initial learning rate of 10^{-4} and the learning rate decay of 10^{-6} applied per one minibatch.

The neural network variants with AHAF activations were trained using the “classic” training procedure (when all trainable parameters are updated in one go) and the DSPT procedure. Implementation of the DSPT procedure uses separate instances of the optimizer class per each set of parameters one per all AHAF parameters, one per the trainable parameters outside of AHAF activations.

The training set loss and the test set accuracy were recorded per for each of the training runs. The results of the training are analyzed and presented in the following section.

D. Analysis of Results

The network variants with AHAF activations outperform the base implementations with ReLU and SiL activations on both CIFAR-10 and Fashion-MNIST. LeNet-5 achieves the best results on the Fashion-MNIST dataset with AHAF activations initialized as ReLU and the DSPT procedure. KerasNet achieves the best results on the CIFAR-10 dataset with AHAF activations initialized as SiL and the DSPT procedure. Table presents the best achieved test set accuracy and the epoch number when this result was achieved for each of the network variants, datasets and parameter tuning procedures used for evaluation.

Best test set accuracy, up to 100 epochs

Network	Activ.	Init.	Proc.	Fashion-MNIST		CIFAR-10	
				Acc.,%	Epoch	Acc.,%	Epoch
LeNet-5	ReLU	N/A	Classic	91,43	98	75,89	96
LeNet-5	SiL	N/A	Classic	90,60	95	73,76	95
LeNet-5	AHAF	ReLU	Classic	91,55	99	76,69	95
LeNet-5	AHAF	SiL	Classic	91,16	99	74,47	99
LeNet-5	AHAF	ReLU	DSPT	91,73	93	74,44	95
LeNet-5	AHAF	SiL	DSPT	90,95	100	74,05	95
KerasNet	ReLU	N/A	Classic	91,29	100	79,36	97
KerasNet	SiL	N/A	Classic	91,76	93	79,83	99
KerasNet	AHAF	ReLU	Classic	91,30	84	79,71	100
KerasNet	AHAF	SiL	Classic	92,02	97	80,31	98
KerasNet	AHAF	ReLU	DSPT	91,35	55	79,30	96
KerasNet	AHAF	SiL	DSPT	91,96	98	80,37	98

Analysis of the dependency between the training loss, test set accuracy and the training epoch shows the potential for performance improvements using longer training runs (running the training for more epochs), different optimizers and learning rates. For KerasNet on the CIFAR-10 dataset the SiL-initialized AHAF activation function consistently shows lower training loss and higher test set accuracy comparing to the base implementation with SiL. Fig. 2 illustrates the

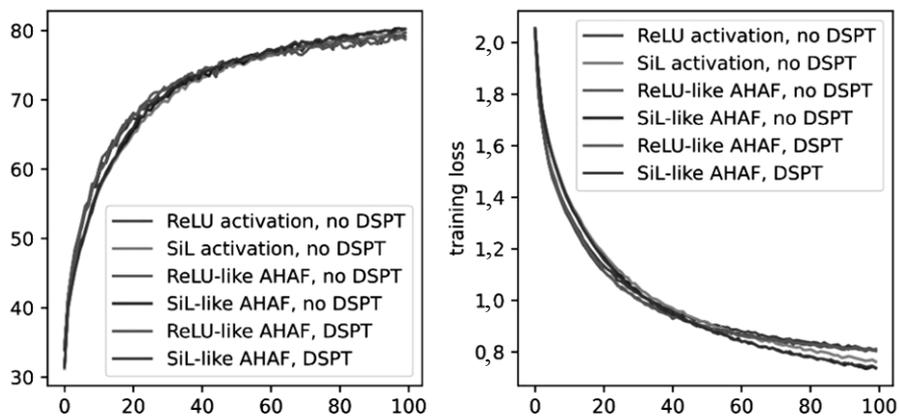


Fig. 2. Dependency between the loss, accuracy and the training epoch for KerasNet network on CIFAR-10

dependency between the training loss, the test set error, and the training epoch for the KerasNet network trained on the CIFAR-10 dataset.

For neural networks with AHAF initialized as ReLU, AHAF keeps its ReLU-like form, but changes the amplitude during the training process. This observation can be explained by the values of the gradient with respect to the γ parameter — the gradient decreases with the increase of the γ parameter. For neural networks with AHAF initialized as SiL, AHAF changes its form and amplitude during the training process. Fig. 3 and Fig. 4 show the form of the activation functions for the two final neurons of the KerasNet network trained on the CIFAR-10 dataset with ReLU-like and SiL-like AHAF activations correspondingly.

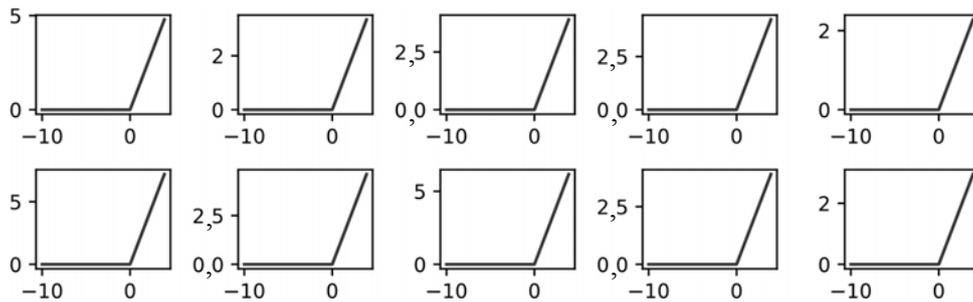


Fig. 3. The activation function form for AHAF initialized as ReLU

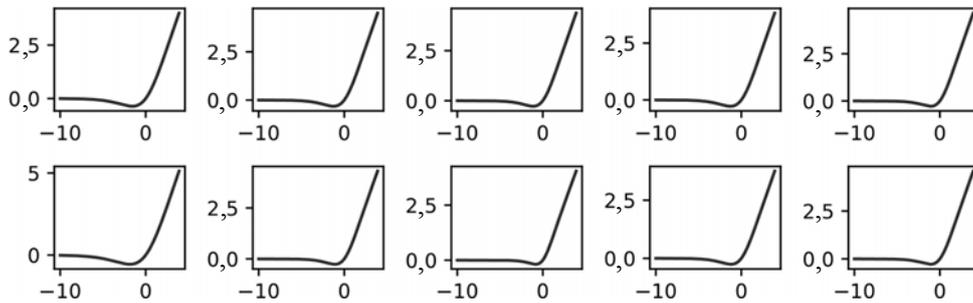


Fig. 4. The activation function form for AHAF initialized as SiL

CONCLUSIONS

Proposed an adaptive hybrid activation function (AHAF) that is applicable for usage in feed-forward and recurrent deep neural networks and combines the properties of both squashing functions and the ones from the rectified unit family. This function does not suffer from the effect of “vanishing gradient” and its parameters are trained together with the synaptic weights. Introduced a training algorithm for a neuron based on AHAF.

The proposed approach is sufficiently simple from the implementation standpoint and provides high performance for the neural network training process.

REFERENCES

1. Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning”, *Nature*, vol. 521, no. 7553, pp. 436–444, 2015. doi: 10.1038/nature14539.
2. J. Schmidhuber, “Deep learning in neural networks: An overview”, *Neural Networks*, vol. 61, pp. 85–117, 2015. doi: 10.1016/j.neunet.2014.09.003.

3. I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.
4. D. Graupe, *Deep Learning Neural Networks: Design and Case Studies*. USA: World Scientific Publishing Co., Inc., 2016.
5. A.L. Caterini and D.E. Chang, *Deep Neural Networks in a Mathematical Framework*, 1st ed. Springer Publishing Company, Incorporated, 2018.
6. C.C. Aggarwal, *Neural Networks and Deep Learning: A Textbook*, 1st ed. Springer Publishing Company, Incorporated, 2018.
7. G. Cybenko, "Approximation by superpositions of a sigmoidal function", *Mathematics of Control, Signals and Systems*, vol. 2, no. 4, pp. 303–314, 1989. doi: 10.1007/BF02551274.
8. K. Hornik, "Approximation capabilities of multilayer feedforward networks", *Neural Networks*, vol. 4, no. 2, pp. 251–257, 1991. doi: 10.1016/0893-6080(91)90009-T.
9. A. Cichocki and R. Unbehauen, *Neural Networks for Optimization and Signal Processing*, 1st ed. USA: John Wiley & Sons, Inc., 1993.
10. K. He, X. Zhang, S. Ren, and J. Sun, "Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification", in *2015 IEEE International Conference on Computer Vision (ICCV)*, 2015, pp. 1026–1034. doi: 10.1109/ICCV.2015.123.
11. D.-A. Clevert, T. Unterthiner, and S. Hochreiter, "Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs)", *arXiv [cs.LG]*, 2016. doi: 10.1162/neco.1997.9.8.1735.
12. K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition", in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778. doi: 10.1109/CVPR.2016.90.
13. S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory", *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997. doi: 10.1162/neco.1997.9.8.1735.
14. S. Elfwing, E. Uchibe, and K. Doya, "Sigmoid-Weighted Linear Units for Neural Network Function Approximation in Reinforcement Learning", *arXiv [cs.LG]*, 2017.
15. P. Ramachandran, B. Zoph, and Q.V. Le, "Searching for Activation Functions", *arXiv [cs.NE]*, 2017.
16. X. Jin, C. Xu, J. Feng, Y. Wei, J. Xiong, and S. Yan, "Deep Learning with S-shaped Rectified Linear Activation Units", *arXiv [cs.CV]*, 2015.
17. M. Tanaka, "Weighted Sigmoid Gate Unit for an Activation Function of Deep Neural Network", *arXiv [cs.CV]*, 2018.
18. B. Yuen, M.T. Hoang, X. Dong, and T. Lu, "Universal Activation Function For Machine Learning", *arXiv [cs.LG]*, 2020.
19. D. Misra, "Mish: A Self Regularized Non-Monotonic Activation Function", *arXiv [cs.LG]*, 2020.
20. J.K. Kruschke and J.R. Movellan, "Benefits of gain: speeded learning and minimal hidden layers in back-propagation networks", *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 21, no. 1, pp. 273–280, 1991. doi: 10.1109/21.101159.
21. Z. Hu and H. Shao, "The study of neural network adaptive control systems", *Control and Decision*, no. 7, pp. 361–366, 1992.
22. C.-T. Chen and W.-D. Chang, "A Feedforward Neural Network with Function Shape Autotuning", *Neural Netw.*, vol. 9, no. 4, pp. 627–641, 1996. doi: 10.1016/0893-6080(96)00006-8.
23. E. Trentin, "Networks with Trainable Amplitude of Activation Functions", *Neural Netw.*, vol. 14, no. 4–5, pp. 471–493, 2001. doi: 10.1016/S0893-6080(01)00028-4.
24. F. Agostinelli, M. Hoffman, P. Sadowski, and P. Baldi, "Learning Activation Functions to Improve Deep Neural Networks", *arXiv [cs.NE]*, 2015.
25. L.R. Sütfeld, F. Brieger, H. Finger, S. Füllhase, and G. Pipa, "Adaptive Blending Units: Trainable Activation Functions for Deep Neural Networks", *arXiv [cs.LG]*, 2018.
26. Y.V. Bodyanskiy, A. Deineko, I. Pliss, and V. Slepanska, "Formal Neuron Based on Adaptive Parametric Rectified Linear Activation Function and its Learning", in *Proc. 1st Int. Workshop on Digital Content & Smart Multimedia "DCSMART 2019"*, vol. 2533, pp. 14–22.
27. D.P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization", *arXiv [cs.LG]*, 2017.

28. P. Otto, Y. Bodyanskiy, and V. Kolodyazhnyi, "A new learning algorithm for a forecasting neuro-fuzzy network", *Integrated Computer-Aided Engineering*, vol. 10, pp. 399–409, 2003. doi: 10.3233/ICA-2003-10409.
29. F. Manessi and A. Rozza, "Learning Combinations of Activation Functions", *CoRR*, vol. abs/1801.09403, 2018.
30. A. Paszke et al., "PyTorch: An Imperative Style, High-Performance Deep Learning Library", in *Advances in Neural Information Processing Systems 32*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, Red's Curran Associates, Inc., 2019, pp. 8024–8035.
31. H. Xiao, K. Rasul, and R. Vollgraf, "Fashion-MNIST: a Novel Image Dataset for Benchmarking Machine Learning Algorithms", *arXiv [cs.LG]*, 2017.
32. A. Krizhevsky, *Learning multiple layers of features from tiny images*, 2009.
33. Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition", *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998. doi: 10.1109/5.726791.
34. F. Chollet et al., "Keras", 2015. [Online]. Available: <https://github.com/fchollet/keras>.

Received 17.12.2021

INFORMATION ON THE ARTICLE

Yevgeniy V. Bodyanskiy, ORCID: 0000-0001-5418-2143, Kharkiv National University of Radio Electronics, Ukraine, e-mail: yevgeniy.bodyanskiy@nure.ua

Serhii O. Kostiuk, ORCID: 0000-0003-4196-2524, Kharkiv National University of Radio Electronics, Ukraine, e-mail: serhii.kostiuk@nure.ua

АДАПТИВНА ГІБРИДНА ФУНКЦІЯ АКТИВАЦІЇ ДЛЯ ГЛИБОКИХ НЕЙРОННИХ МЕРЕЖ / Є.В. Бодянський, С.О. Костюк

Анотація. Запропоновано адаптивну гібридну функцію активації (АНАФ), що поєднує особливості випрямних блоків (rectifier units) та стискальних (squashing) функцій. Запропонована функція може бути використана як пряма заміна активаційних функцій ReLU, SiL і Swish для глибоких нейронних мереж, а також набути форми однієї з цих функцій в процесі навчання. Ефективність функції досліджено на задачі класифікації зображень на наборах даних Fashion-MNIST і CIFAR-10. Результати дослідження показують, що нейронні мережі з активаційними функціями АНАФ показують точність класифікації кращу, ніж їх базові реалізації на основі ReLU та SiL. Запропоновано двоетапний процес налаштування параметрів для навчання нейронних мереж з АНАФ. Запропонований підхід достатньо простий в реалізації та забезпечує високу продуктивність у навчанні нейронної мережі.

Ключові слова: адаптивна гібридна функція активації, двоетапний процес налаштування параметрів, глибокі нейронні мережі.

АДАПТИВНАЯ ГИБРИДНАЯ ФУНКЦИЯ АКТИВАЦИИ ДЛЯ ГЛУБОКИХ НЕЙРОННЫХ СЕТЕЙ / Е.В. Бодянский, С.А. Костюк

Аннотация. Предложена адаптивная гибридная функция активации (АНАФ), которая объединяет свойства выпрямительных блоков (rectifier units) и сжимающих (squashing) функций. Предложенная функция может быть использована как прямая замена активационных функций ReLU, SiL и Swish для глубоких нейронных сетей, а также принимать форму одной из этих функций в процессе обучения. Эффективность функции исследована на задаче классификации изображений на наборах данных Fashion-MNIST и CIFAR-10. Результаты исследования показывают, что нейронные сети с активационными функциями АНАФ показывают точность классификации лучшую, чем их базовые реализации на основе ReLU и SiL. Предложено двухэтапный процесс настройки параметров для обучения нейронных сетей с АНАФ. Предложенный подход достаточно простой в реализации и обеспечивает высокую продуктивность в обучении нейронной сети.

Ключевые слова: адаптивная гибридная функция активации, двухэтапный процесс настройки параметров, глубокие нейронные сети.

GENERATIVE TIME SERIES MODEL BASED ON ENCODER-DECODER ARCHITECTURE

N.I. NEDASHKOVSKAYA, D.V. ANDROSOV

Abstract. Encoder-decoder neural network models have found widespread use in recent years for solving various machine learning problems. In this paper, we investigate the variety of such models, including the sparse, denoising and variational autoencoders. To predict non-stationary time series, a generative model is presented and tested, which is based on a variational autoencoder, GRU recurrent networks, and uses elements of neural ordinary differential equations. Based on the constructed model, the system is implemented in the Python3 environment, the TensorFlow2 framework and the Keras library. The developed system can be used for modeling continuous time-dependent processes. The system minimizes a human factor in the process of time series analysis, and presents a high-level modern interface for fast and convenient construction and training of deep models.

Keywords: prediction, variational autoencoder, GRU recurrent neural network, neural ordinary differential equation, latent space, nonstationary time series.

INTRODUCTION

Classical methods of autoregression with moving average (ARMA) [1, 2] are used to analyze and predict stationary time series. Autoregressive models with integrated moving average (ARIMA) [1, 3], heteroskedastic (ARCH/GARCH) [1, 4, 5] and other [6] are designed to analyze a wider class of nonstationary processes. GARCH models, in particular, help to provide the volatility analysis of financial time series [7]. ARIMA models are based on numerical differentiation technique and an operator of finite differences to make time series stationary. Moving variance is applied in GARCH models to model heteroskedasticity. The choice of degree of autoregression and moving average in ARMA, ARIMA, ARCH and GARCH models, when analyzing the autocorrelation, is often carried out manually.

Recurrent neural networks (RNNs) of the long short-term memory (LSTM) type have also been used in recent years to predict time series [8–11]. Gated recurrent unit (GRU), proposed in 2014 [11], is a simplified version of the LSTM network, probably shows as good results as LSTM [12], and therefore is widely used in recent years. Machine learning and deep learning techniques [8–15] mainly require scaling of the input data and presentation of the series in the form of “values for previous periods – values for the current period” or “features – the resulting value”. The main problem of such a representation is the invariance of the fixed values of the series with respect to time. This representation of the series assumes that each value of the series is fixed at the same interval, although in practice this is not always the case.

The paper aims to develop a generative model on basis of the autoencoder for time series prediction, which will be sensitive to different intervals of fixing

values of the series and will be able to find hidden patterns in the data. The goal is also to minimize human interference in the data processing, leaving only the requirement to scale the input data.

AUTOENCODER MODELS

An autoencoder is an artificial neural network that, without a teacher, based on an unmarked data is able to recognize encodings – effective representations of input data [8, 13]. Such encodings often have a much smaller dimension compared to the input data, so autoencoders are also a means to reduce dimensionality.

An important feature of the autoencoder is that it can be a *generative model*, capable of randomly generating new data that is very similar to the input. Goals of the autoencoder are as follows: to reconstruct the input data, as well as to identify features hidden in the input data. The typical autoencoder model consists of two parts (Fig. 1): the *coder* and the *decoder* networks. The coder has to recognize and convert the input data into a latent space, that is the internal representation of the input data. The decoder, in turn, is seen as a generating network that converts the internal representation into outputs. Typically, the decoder has the same architecture as the coder, but symmetrically mapped relative to the layer responsible for creating the latent space (Fig. 1).

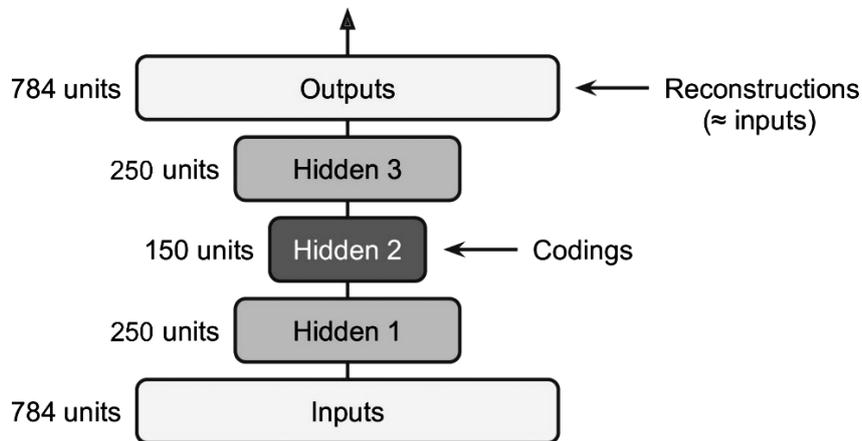


Fig. 1. An example of a deep autoencoder model for the mnist dataset reconstruction, adapted from [8]

To achieve the first goal, namely, to provide the reconstruction of input data, training of the autoencoder is performed by minimizing the loss function, which is called the *reconstruction error*:

$$E(x, g(f(x))), \tag{1}$$

where x is an input vector, E is a function that penalizes $g(f(x))$ for dissimilarity to x , $g(h)$ is the decoder output, $h = f(x)$ is the coder output.

In order to better identify the features hidden in the input data, a regularization is added to the autoencoder model. This allows the model to obtain more properties in addition to the ability to copy input data. The desirable properties of the model are as follows:

- presentation of sparsity of data;
- resistance to noise in the input data and the absence of part of the inputs;
- small values of the derivatives of codings relative to the input data.

The *regularized* autoencoder is a model with the loss function presented in the form [13]:

$$E(x, g(h)) + \Omega(h, x), \quad (2)$$

where E is the reconstruction loss (1), h is a coder layer, $\Omega(h, x)$ is a coder layer penalty.

The peculiarity of regularized autoencoders is the absence of an obvious Bayesian interpretation. Thus, other known regularized models, for example the ridge regression and other, are an approximation of the Bayesian maximum of the a posteriori probability with the addition of a regularizing penalty, which corresponds to the a priori probability distribution of the model parameters. Regularized autoencoders have a different interpretation, because the $\Omega(h, x)$ penalty depends on the x – an input data and therefore cannot be formally considered as a priori distribution. However, it is still believed that the introduction of the $\Omega(h, x)$ regularizer helps to implicitly prefer certain functions.

Let us consider several models of regularized autoencoders depending on how the penalty $\Omega(h, x)$ in (2) is defined.

Model of sparse autoencoder. One of the key reasons for the high energy efficiency of the human brain is the sparse activation of its neurons: only a small part of the neurons is active in the brain at any given time.

To model the sparseness in an artificial neural network, we consider the probabilistic interpretation of neuronal activation. Let the artificial neuron of the hidden layer be a Bernoulli random variable, and the average value of activation of this neuron corresponds to the probability of obtaining a unit in the Bernoulli test. The probability of activation of each such individual neuron should be low to increase the sparseness of the neurons of the hidden (latent) layer. Let the desired probability of neuron activation be equal to ρ , and let the empirical average value of neuron activation on the basis of train data be equal to $\hat{\rho}$. Sparsity loss [16, 17] is considered as the $\Omega(h, x)$ penalty in expression (2), a measure of dissimilarity between distributions and is based on the Kullback–Leibler divergence between the model distribution and the data distribution:

$$KL(\rho \parallel \hat{\rho}) = \rho \ln \frac{\rho}{\hat{\rho}}.$$

Learning criterion for the sparse autoencoder can be considered as follows:

$$(1 - \alpha)E(x, g(f(x))) + \alpha\Omega(h, x),$$

$$\Omega(h, x) = \sum_{i=1}^N KL(\rho_i \parallel \hat{\rho}_i) = \sum_{i=1}^N \rho_i \ln \frac{\rho_i}{\hat{\rho}_i},$$

where $\Omega(h, x)$ is the sparsity loss, N is a number of neurons in a coder layer, $\alpha \in [0,1]$ is a sparsity weight, which is a hyperparameter of the model.

If α is large, the model will pay more attention to the target sparsity, but will not be able to reconstruct the inputs properly. If, on the contrary, the weight is too low, the model will mostly ignore the sparsity and will not find interesting features in the data. Methods of decision support [18, 19] can be used to determine the most acceptable α value based on quantitative and qualitative decision criteria in a particular practical problem.

An important property of the sparse autoencoder is that it can be considered as a generative model with latent variables, which approximates the maximum likelihood. Let us consider a model [13] with an input vector of visible variables x , latent variables h and a common probability distribution

$$p_{\text{model}}(x, h) = p_{\text{model}}(h)p_{\text{model}}(x | h).$$

$p_{\text{model}}(h)$ is called the a priori distribution of latent variables and represents the a priori belief of the model that it will “see” the input vector x , where h is still the output of the coder. This interpretation differs from the traditional use of the term “a priori”, which denotes the distribution $p(w)$, which describes the hypotheses about the parameters of the model before reading the training data.

The logarithm of the plausibility of the model can be represented as:

$$\ln p_{\text{model}}(x) = \ln \sum_h p_{\text{model}}(x, h).$$

The autoencoder is considered as an approximation of this sum by a point estimate for only one value of h , which has a high probability. With this choice of h , the following function is maximized:

$$\ln p_{\text{model}}(x, h) = \ln p_{\text{model}}(h) + \ln p_{\text{model}}(x | h).$$

The term $\ln p_{\text{model}}(h)$ can cause sparsity. For example, the a priori Laplace distribution

$$p_{\text{model}}(h_i) = \frac{\lambda}{2} \exp(-\lambda |h_i|)$$

corresponds to the sparsity penalty in terms of the L_1 norm.

Denoising autoencoder. Another way to make the autoencoder show interesting features is to add noise to the inputs and teach it to restore the initial not noisy input [8, 13, 20, 21]. There are two ways:

- a random variable is added to the input vector, normally distributed with a small variance, which determines the noise level;
- part of the input neurons is set to zero. The level of noise is determined by what part it is. This method is more used in image processing problems.

In the denoising autoencoder models, a conditional distribution $C(\hat{x} | x)$ of noisy examples under the condition of true examples is introduced. Next, the autoencoder learns the distribution of the reconstruction $p_{\text{reconstr}}(x | \hat{x})$, which is estimated on the basis of training pairs (x, \hat{x}) as follows [13]:

- select example x from the training set;
- select the noisy version \hat{x} with $C(\hat{x} | x)$;

- use (x, \hat{x}) as a training example to estimate the distribution of reconstruction $p_{\text{reconstr}}(x|\hat{x}) = p_{\text{decoder}}(x|h)$, where h is the output of the coder, and p_{decoder} is determined by the decoder $g(h)$;
- minimize the following loss function using the mini-batch gradient descent:

$$-\ln p_{\text{reconstr}}(x|\hat{x}) = -\ln p_{\text{decoder}}(x|h).$$

If the encoder is deterministic, then the autoencoder is often a feedforward neural network, and the same methods can be used to train it as for any feedforward neural network, for example, the mini-batch stochastic gradient descent.

The *variational autoencoder (VAE) model* was proposed in 2014 and is designed to reconstruct the law of distribution of training data for artificial generation of samples from the general distribution [22]. This is a probabilistic model, because its output after training is determined randomly. VAE has a basic architecture common to all autoencoders (Fig. 2): the first part corresponds to the encoder network (it consists of the hidden layers 1 and 2 in the example in Fig. 2), followed by the decoder network (the hidden layers 3 and 4 in Fig. 2). The difference from deterministic encoders is that the VAE encoder for a given input results in the average encoding μ and the standard deviation σ . The coding is then chosen randomly from the Gaussian distribution with mean μ and standard deviation σ . A standard distribution other than Gaussian can also be used. Next, the decoder decodes the received encoding in the usual way.

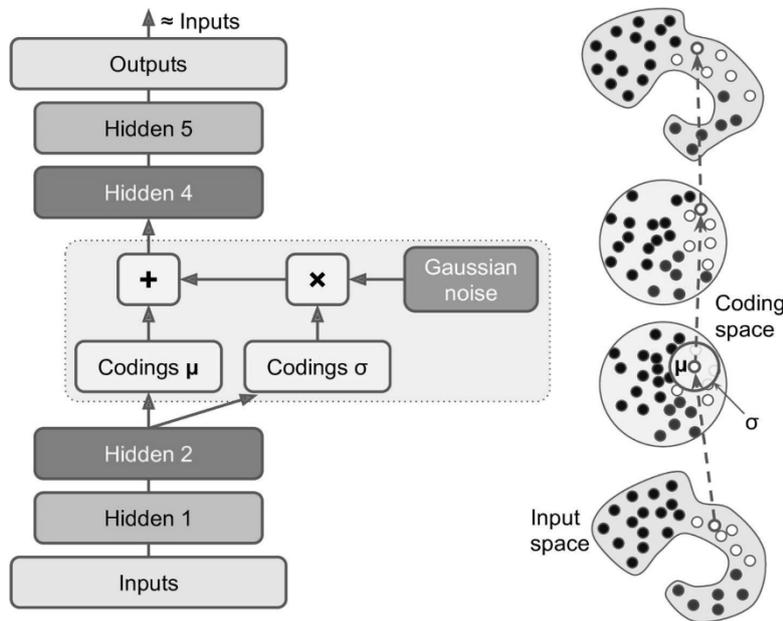


Fig. 2. An example of the variational autoencoder model, VAE [8]

Inputs (on the right in Fig. 2) can have a complex distribution. During training, the coding moves inside the latent space, the coding space, and occupies an approximately spherical region similar to a cloud of Gaussian points. The loss function of the variational autoencoder is the sum of two terms [8]:

$$E(x, g(h)) + L(h, x),$$

where E is the reconstruction error (1) of the input vector x , as before; $L(h, x)$ is the latent loss, which is often the Kullbak–Leibler divergence between the target distribution, such as Gaussian, and the actual coding distribution.

It is easy to generate a new example based on a trained variational autoencoder: you need to choose a random encoding from the Gaussian distribution and decode it.

LODE-GRU-VAE VARIATIONAL AUTOENCODER MODEL FOR TIME SERIES PREDICTION

Fig. 3 shows a simplified architecture of the proposed model. It consists of an encoder – a LODE-GRU network, and a decoder network. Two modifications are proposed: with simulation of timestamp distribution and without it. In both cases, the pair $\langle x_i, t_i \rangle$ – the data tensor together with the corresponding timestamps is the input for the model (Fig. 3). In the second case we suggest the uniform distribution of these time slices on the observation interval.

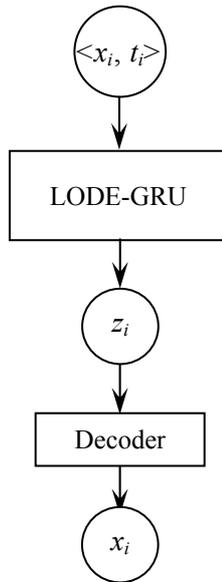


Fig. 3. A simplified LODE-GRU-VAE model

The proposed model is based on the ideas of variational encoder and GRU recurrent neural networks. Features of the GRU unit in comparison with the known LSTM are as follows. Firstly, a single state vector $h(t)$ is used. Secondly, the forgetting gateway and the input gateway are controlled by a single controller – a fully connected layer with a logistical activation function. If the result of this controller is equal to 1, then the forget gateway opens and the input gateway closes. If the result of the controller is equal to 0, then the opposite action occurs. This means that the place is first cleared before saving a certain memory. Thirdly, there is no output gateway in the GRU, so a complete state vector is the result at each time step. The main fully connected layer analyzes the current inputs and some parts of the previous state, which are determined by the additional gateway controller.

The LODE-GRU encoder in Fig. 3 is a modification of recurrent networks of the GRU type. The modification is to use the ordinary differential equations (ODEs) to predict the values of hidden trajectories. As a result, the LODE-GRU network layers are defined as follows:

$$h''_{\tau} = Sol(f_{\theta}, h_{\tau-1}, t_{\tau-1}, t_{\tau}), \tag{3}$$

$$u_{\tau} = \sigma(W_{xu}x_{\tau} + W_{hu}h''_{\tau} + b_u), \tag{4}$$

$$r_{\tau} = \sigma(W_{xr}x_{\tau} + W_{hr}h''_{\tau} + b_r), \tag{5}$$

$$h'_{\tau} = \tanh(W_{xh'}x_{\tau} + W_{hh'}(r_{\tau} \odot h''_{\tau}) + b_{h'}), \tag{6}$$

$$h_{\tau} = u_{\tau} \odot h''_{\tau} + (1 - u_{\tau}) \odot h'_{\tau}. \tag{7}$$

The block described by equation (3) generates a set of points according to the trajectory of the studied process under the conditions (4)–(7) of process transformations, which project the trajectory to another space called latent or hidden. The combination of neural ordinary differential equations and neural RNN networks began to be studied in 2018 to model irregularly observed time series [23], and was further developed in [24–26]. Equation (3) is called the latent ordinary differential equation (Latent ODE). Model (3)–(7) helps to generate new series values at intermediate points between observations, in contrast to standard neural RNN LSTM networks. This is especially useful when the time interval between adjacent observations is large.

Let us modify the described model (3)–(7) to obtain the probability distributions of the hidden trajectories. Let us consider the case when the source vector for the encoder is a multidimensional Gaussian vector with a mathematical expectation equal to μ_{z_0} and the variance equal to σ_{z_0} during the last observation τ . Therefore, the additional layer described by the next formula is added:

$$z_0 = g(h_\tau, \theta) \sim \mathcal{N}(\mu_{z_0}, \sigma_{z_0}). \quad (8)$$

Thus, the LODE-GRU-VAE model is described by formulas (3)–(8) and represents a variation encoder. The coding of the input data x in this model is characterized by a conditional probability distribution $q(z/x)$, where z is a random vector in the latent space. The loss function in this case is the mathematical expectation of losses relative to z :

$$\mathbb{E}_{z \sim q(z/x)} L(z, q) = \mathbb{E}_{z \sim q(z/x)} \ln p(x, z) - D_{KL}(q(z/x) \| p(z)), \quad (9)$$

where $p(x, z)$ is the joint distribution of x and z , $D_{KL}(A \| B)$ is the Kullback–Leibler divergence, which determines the degree of “dissimilarity” of distributions A and B .

The rationale for the variation encoder is based on the Expectation Maximization (EM) method. For the autoencoder, the input and target are the same vector x . Therefore, the decoder returns the conditional distribution $p(x/z)$ when the code z is an input, and the error function determines the plausibility of the “binding” of the error function to the output of the last layer of the network.

The EM algorithm assumes that we can calculate the distribution $p(x, z/\theta) = p(z/x, \theta)p(x/\theta)$. The problem is to maximize $p(x/\theta)$ with respect to the parameters θ . Let us take the logarithm of both parts of the equality above and express $\ln p(x/\theta)$:

$$\log p(x/\theta) = \ln p(x, z/\theta) - \ln p(z/x, \theta).$$

Next let us take the mathematical expectation with respect to z :

$$\int_Z q(z) \ln p(x/\theta) dz = \int_Z q(z) \ln p(x, z/\theta) dz - \int_Z q(z) \ln p(z/x, \theta) dz.$$

After the transformations we get:

$$\ln p(x/\theta) = \int_Z q(z) \ln \left(\frac{p(x/z, \theta)p(z/\theta)}{q(z)} \right) dz - \int_Z q(z) \ln \left(\frac{p(z/x, \theta)}{q(z)} \right) dz.$$

The last term on the right is the Kullbak–Leibler divergence, which is always non-negative quantity. Therefore, the expression $\int_Z q(z) \ln \left(\frac{p(x/z, \theta) p(z/\theta)}{q(z)} \right) dz$ can be considered as the lower estimate of the value of $\ln p(x/\theta)$.

Let us denote $\int_Z q(z) \ln \left(\frac{p(x/z, \theta) p(z/\theta)}{q(z)} \right) dz = ELBO(q, \theta)$. Then

$$ELBO(q, \theta) = \ln p(x/\theta) - D_{KL}(q(z) \| p(z/x)). \tag{10}$$

In general, maximizing (10) by the parameter θ , the approximation $q(z)$ to $p(z/x)$ is performed. Let us show that the loss function (9) is equal to $ELBO(q, \theta)$ (10). Using the formulas of conditional probability, let us rewrite (9) in the form:

$$\mathcal{L}(q) = \mathbb{E}_{z \sim q(z/x)} \ln p(x/z) - D_{KL}(q(z/x) \| p(z)). \tag{11}$$

Since $D_{KL}(q(z/x) \| p(z))$ is close to zero for the same distributions $q(z/x)$ and $p(z)$, and $\ln p(x/z) \leq \ln p(x)$, then the upper limit of the loss function will also be $\ln p(x)$. That is, we came to the same result.

The described LODE-GRU-VAE model can be trained by Adam, Rprop or similar methods with respect to the sample $z \sim q(z/x) = q(z; f(x; \theta))$ in order to obtain a gradient with respect to θ and subsequent generation of distribution parameters. With this approximation, the Monte Carlo method can be used to calculate the mathematical expectation for $\ln p(x/z)$ and the Kullbak–Leibler divergence at a fixed θ .

Let us consider how the described LODE-GRU-VAE model can be used to generate a new sequence of series. For example, the initial sequence of n values is first fed to the input of the model, and the prediction of one next value is performed on the basis of the model. Next, the predicted value is joined to the sequence. The prediction for the next value is calculated on basis of the last n values, which are given to the model. This process can generate a new sequence that is similar to the original time series.

STATEMENT AND RESULTS OF THE EXPERIMENT. SELECTION OF MODEL HYPERPARAMETERS

For the time series prediction experiment, let us choose a synthetic nonstationary time series, which is a function of $x = \sin(2\pi\omega t)$, $\omega \in \mathbb{R}$, $t \in [0, T]$. To bring the data closer to the real ones, the Gaussian noise is added to each value of the series:

$$x = \sin(\pi\omega t) + \epsilon, \epsilon \sim \mathcal{N}(0, \sigma), \sigma \in \mathbb{R}^+.$$

Not all historical data is always available in real forecasting problems. Therefore, additional values are introduced: d is the sampling step and i is the indicator set of points to be left in the sample. The simulated series can be formally written as follows:

$$\hat{X} = X^T i = \left\| (t_k, \sin(\pi \omega t_k)) \right\|_{k=1, d}^T i.$$

The following parameters were selected for the experiment:

- $\omega = 0,25$.
 - $d = 1000$.
 - Number of points to stay is 150.
 - The number of layers in the network that simulates the behavior of ODE and their dimension are 1 and 6, respectively.
 - Learning rate is adaptive, initially equal to 0,01.
 - The method for solving the differential equation is Dormand–Prince.
 - The optimization method for learning the neural network is Adamax.
- Gradients within the ODE-layers are calculated by the method of conjugate equations.
- The dimension of the hidden space is 6.
 - The initial value of the variance is 0,1.
 - Number of epochs is 200.

The next step is the analysis and processing of the experimental results. Metrics for assessing the quality of the model can be components of the error function. For example, the Kullbak–Leibler divergence can be chosen as a metric. But the limitation of the latter is that it is difficult to interpret. In this work, the mean square error MSE and the coefficient of determination R^2 are used.

Several experiments have been carried out to prove the efficiency of the model (3) – (8) (Fig. 4, 5). In the first experiment, acceptable values of model

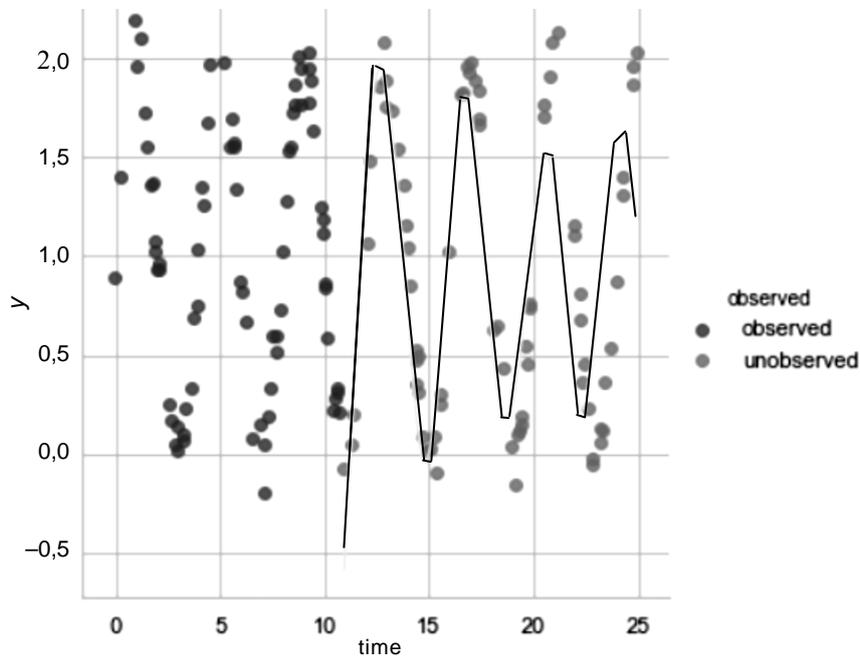


Fig. 4. Real and predicted values of the target variable (experiment 1)

quality metrics were obtained, namely $MSE = 0,11$ and $R^2 = 0,78$. In the second experiment, the values of the model quality metrics were almost the same and

also acceptable: $MSE = 0,10$, $R^2 = 0,80$. The observed points on which the model was built are marked in Fig. 4, 5 in blue color. The unobserved values of the test set, which were used to assess the quality of the obtained model are marked in orange color. The graph of the predicted values is marked in yellow. Model's predicted values are of admissible MAPE and RMSE rates.

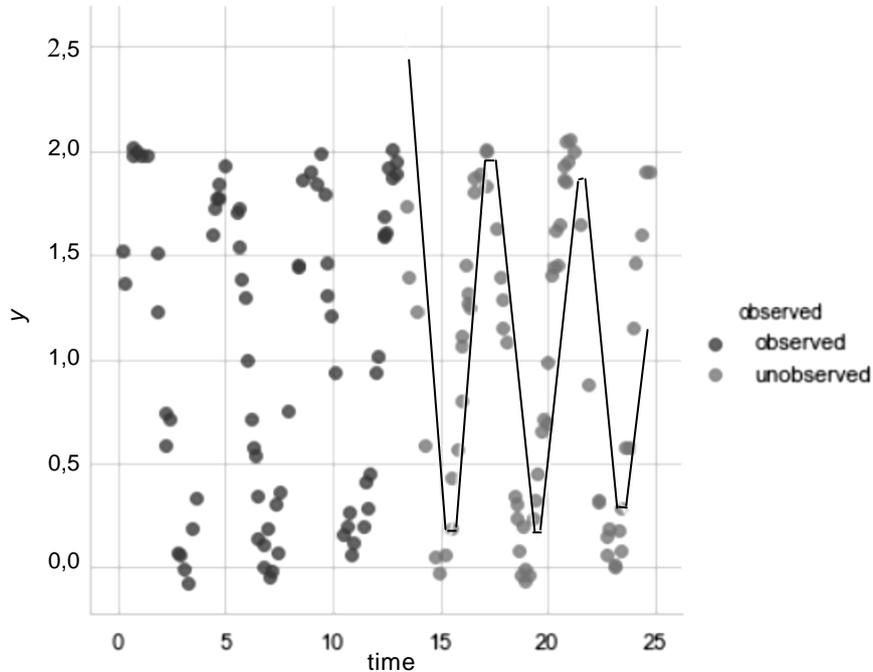


Fig. 5. Real and predicted values of the target variable (experiment 2)

The software in the Python3 environment using the TensorFlow2 framework and the Keras library was developed, which implements the proposed encoder-decoder system. The main arguments in favor of choosing the Python 3 programming language were the speed of writing code and the popularity of this language. Tensorflow2 uses the capabilities of the Nvidia CUDA and has an integrated Keras library, and thus was chosen as the best framework in terms of hardware and performance resources. It is a high-level modern API for fast and easy design and learning of deep models.

CONCLUSIONS

The work is devoted to the research and development of a neural network model based on the encoder-decoder architecture and recurrent blocks for predicting the values of nonstationary time series. Models and methods of machine and deep learning used for sequence processing are studied: LSTM and GRU models of recurrent neural networks, generative models, such as VAE, encoder-decoder models for detecting hidden patterns in data, Adam and Adamax stochastic optimization learning methods.

The LODE-GRU-VAE model was built and tested to reconstruct the dynamics of nonstationary time series. The model allows to generate values at intermediate points between observations, and therefore it is possible to generate new

values of a series where the time interval between two adjacent observations is large. In standard GRU-type RNN networks, the latent state is updated with each observation and remains constant between them. Conversely, within the framework of the LODE-GRU structure, a neural ordinary differential equation learns to model a continuous change in the latent state of a network between two observations.

The encoder-decoder system is implemented based on the proposed model in the Python3 environment using the TensorFlow2 framework and the Keras library. Experiments have been carried out to prove the efficiency of this system in the problems of modeling processes that depend on continuous time.

REFERENCES

1. P.I. Bidyuk, V.D. Romanenko, and O.L. Timoshchuk, *Time series analysis*. Kyiv: Polytechnika, NTUU “KPI”, 2013.
2. Terence C. Mills, “Chapter 3 - ARMA Models for Stationary Time Series”, *Applied Time Series Analysis. A Practical Guide to Modeling and Forecasting*, pp. 31–56, 2019. Available: <https://doi.org/10.1016/B978-0-12-813117-6.00003-X>.
3. Terence C. Mills, “Chapter 4 - ARIMA Models for Nonstationary Time Series”, *Applied Time Series Analysis. A Practical Guide to Modeling and Forecasting*, pp. 57–69, 2019. Available: <https://doi.org/10.1016/B978-0-12-813117-6.00004-1>.
4. Amélie Charles and Olivier Darné, “The accuracy of asymmetric GARCH model estimation”, *International Economics*, vol. 157, pp. 179–202, May 2019. Available: <https://doi.org/10.1016/j.inteco.2018.11.001>
5. O.L. Tymoshchuk, V.H. Huskova, and P.I. Bidyuk, “A combined approach to modeling nonstationary heteroscedastic processes”, *Radio Electronics, Computer Science, Control*, (2), pp. 80–89, 2019. Available: <https://doi.org/10.15588/10.15588/1607-3274-2019-2-9>.
6. P. Bidyuk, T. Prosyankina-Zharova, and O. Terentiev, “Modelling nonlinear nonstationary processes in macroeconomy and finances”, *Advances in Computer Science for Engineering and Education*, vol. 754, pp. 735–745, 2019. Available: https://doi.org/10.1007/978-3-319-91008-6_72.
7. Anoop S. Kumar and S. Anandarao, “Volatility spillover in crypto-currency markets: Some evidences from GARCH and wavelet analysis”, *Physica A: Statistical Mechanics and its Applications*, vol. 524, pp. 448–458, 15 June 2019. Available: <https://doi.org/10.1016/j.physa.2019.04.154>.
8. Aurelien Geron, *Hands-On Machine Learning with Scikit-Learn and TensorFlow*. Sebastopol, CA: O’Reilly Media Inc., 2017, 760 p.
9. Alex Sherstinsky, “Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network”, *Physica D: Nonlinear Phenomena*, vol. 404, 132306, March 2020. Available: <https://doi.org/10.1016/j.physd.2019.132306>.
10. Mikel Canizo, Isaac Triguero, Angel Conde, and Enrique Onieva, “Multi-head CNN–RNN for multi-time series anomaly detection: an industrial case study”, *Neurocomputing*, vol. 363, pp. 246–260, 21 October 2019. Available: <https://doi.org/10.1016/j.neucom.2019.07.034>.
11. Kyunghyun Cho et al., “Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation”, *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), October 25–29, 2014*, pp. 1724–1734, Available: <https://aclanthology.org/D14-1179.pdf>.
12. Klaus Greff et al., *LSTM: A Search Space Odyssey*. 2015. Available: arXiv:1503.04069.

13. Ian Goodfellow, Yoshua Bengio, and Aaron Courville, *Deep Learning*. Massachusetts London, England: The MIT Press Cambridge, 2016, 802 p.
14. Henrik Brink, Joseph Richards, and Mark Fetherolf, *Machine Learning*. StPb.: Piter, 2017, 336 p.
15. Jake VanderPlas, *Python Data Science Handbook*. O'Reilly Media, Inc, 2016, 576 p.
16. Z. Deng et al., "Sparse stacked autoencoder network for complex system monitoring with industrial applications", *Chaos, Solitons & Fractals*, vol. 137, August 2020. Available: <https://doi.org/10.1016/j.chaos.2020.109838>.
17. H. Zhu et al., "Stacked pruning sparse denoising autoencoder based intelligent fault diagnosis of rolling bearings", *Applied Soft Computing*, vol. 88, March 2020. Available: <https://doi.org/10.1016/j.asoc.2019.106060>.
18. N.I. Nedashkovskaya, "Method for Evaluation of the Uncertainty of the Paired Comparisons Expert Judgements when Calculating the Decision Alternatives Weights", *Journal of Automation and Information Sciences*, vol. 47, no. 10, pp. 69–82, 2015. Available: <https://doi.org/10.1615/JAutomatInfScien.v47.i10.70>.
19. N.I. Nedashkovskaya, "A system approach to decision support on basis of hierarchical and network models", *System research and information technologies*, no. 1, pp. 7–18, 2018. Available: <https://doi.org/10.20535/SRIT.2308-8893.2018.1.01>.
20. J.Yu, "Manifold regularized stacked denoising autoencoders with feature selection", *Neurocomputing*, vol. 358, pp. 235–245, 17 September 2019. Available: <https://doi.org/10.1016/j.neucom.2019.05.050>.
21. N. Abiri et al., "Establishing strong imputation performance of a denoising autoencoder in a wide range of missing data problems", *Neurocomputing*, vol. 365, pp. 137–146, 6 November 2019. Available: <https://doi.org/10.1016/j.neucom.2019.07.065>.
22. Diederik P. Kingma and Max Welling, *Auto-Encoding Variational Bayes*, 2014. Available: arXiv:1312.6114v10.
23. Ricky T.Q. Chen, Yulia Rubanova, Jesse Bettencourt, and David Duvenaud, *Neural ordinary differential equations*. NeurIPS, 2018.
24. Yulia Rubanova, Ricky T.Q. Chen, and David K. Duvenaud, *Latent ordinary differential equations for irregularly-sampled time series*. NeurIPS, 2019.
25. Calypso Herrera, Florian Krach, and Josef Teichmann, *Neural Jump Ordinary Differential Equations: Consistent Continuous-Time Prediction and Filtering*, 2020. Available: arXiv:2006.04727.
26. J. Lu et al., "Neural-ODE for pharmacokinetics modeling and its advantage to alternative machine learning models in predicting new dosing regimens", *iScience*, vol. 24, issue 7, 23 July 2021. Available: <https://doi.org/10.1016/j.isci.2021.102804>.

Received 09.08.2021

INFORMATION ON THE ARTICLE

Nadezhda I. Nedashkovskaya, ORCID: 0000-0002-8277-3095, Institute for Applied System Analysis of the National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine, e-mail: n.nedashkivska@gmail.com

Dmytro V. Androsov, Institute for Applied System Analysis of the National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine, e-mail: androsovdmitry80@gmail.com

ГЕНЕРАТИВНА МОДЕЛЬ ДЛЯ ПРОГНОЗУВАННЯ ЧАСОВИХ РЯДІВ НА ОСНОВІ АРХІТЕКТУРИ КОДУВАЛЬНИК-ДЕКОДУВАЛЬНИК / Н.І. Недашківська, Д.В. Андросов

Анотація. Моделі нейронних мереж на основі архітектури кодувальник- декодувальник знайшли широке застосування в останні роки для розв'язання різноманітних задач машинного навчання. Досліджено різновиди таких моделей,

серед яких розріджений, шумопригнічувальний та варіаційний автокодувальники. Для прогнозування нестационарного часового ряду подано і протестовано модель, що базується на варіаційному автокодувальнику, блоках рекурентних мереж типу GRU і використовує елементи нейронних звичайних диференціальних рівнянь. На основі побудованої моделі реалізовано систему у середовищі Python3 з використанням фреймворку TensorFlow2 та бібліотеки Keras. Розроблена система може використовуватися для моделювання процесів, що залежать від неперервного часу. Система мінімізує втручання людини у процес аналізу часових рядів, представляє високорівневий сучасний інтерфейс для швидкого і зручного конструювання та навчання глибоких моделей.

Ключові слова: прогнозування, варіаційний автокодувальник, рекурентна нейронна мережа типу GRU, нейронне звичайне диференціальне рівняння, латентний простір, нестационарний часовий ряд.

ГЕНЕРАТИВНАЯ МОДЕЛЬ ДЛЯ ПРОГНОЗИРОВАНИЯ ВРЕМЕННЫХ РЯДОВ НА ОСНОВЕ АРХИТЕКТУРЫ КОДИРОВЩИК-ДЕКОДИРОВЩИК /

Н.И. Недашкова, Д.В. Андросов

Аннотация. Модели нейронных сетей на основе архитектуры кодировщик-декодировщик нашли широкое распространение в последние годы при решении различных задач машинного обучения. Исследованы разновидности таких моделей, среди которых разреженный, шумоподавляющий и вариационный автокодировщик. Для прогнозирования нестационарного временного ряда представлена и протестирована порождающая модель, которая основана на вариационном автокодировщике, блоках рекуррентных сетей типа GRU и использует элементы нейронных обыкновенных дифференциальных уравнений. На основе построенной модели реализована система в среде Python3 с использованием фреймворка TensorFlow2 и библиотеки Keras. Разработанная система может использоваться для моделирования процессов, зависящих от непрерывного времени. Система минимизирует вмешательство человека в процесс анализа временных рядов, представляет высокоуровневый современный интерфейс для быстрого и удобного конструирования и обучения глубоких моделей.

Ключевые слова: прогнозирование, вариационный автокодировщик, рекуррентная нейронная сеть типа GRU, нейронное обыкновенное дифференциальное уравнение, латентное пространство, нестационарный временной ряд.

**МАТЕМАТИЧНЕ МОДЕЛЮВАННЯ КОНТАКТНОЇ
ВЗАЄМОДІЇ ДВОХ ПРУЖНИХ ТРАНСВЕРСАЛЬНО-
ІЗОТРОПНИХ ПІВПРОСТОРІВ, ОДИН З ЯКИХ МІСТИТЬ
ПРИПОВЕРХНЕВУ ВІЙМКУ ЕЛІПТИЧНОГО ПЕРЕРІЗУ****В.С. КИРИЛЮК, О.І. ЛЕВЧУК, В.В. ГАВРИЛЕНКО**

Анотація. На основі математичної моделі розглянуто задачу про стискання двох пружних трансверсально-ізотропних півпросторів, один з яких містить похилу приповерхневу виїмку еліптичного перерізу. Розв'язок задачі отримано за допомогою подання Еліота для трансверсально-ізотропного тіла через гармонічні функції, класичних гармонічних потенціалів та зведення граничної задачі до розгляду інтегро-диференціального рівняння з невідомою областю інтегрування. Як частинний випадок зі знайдених аналітичних виразів впливають основні параметри контакту трансверсально-ізотропних півпросторів за наявності в одному з них виїмки вісесиметричної форми, а також параметри контактної взаємодії двох пружних ізотропних півпросторів, один з яких містить виїмку еліптичного перерізу. Отримано числові результати, вивчено вплив пружних властивостей півпросторів, геометричних параметрів виїмки та навантаження на контактну взаємодію та закриття зазору між тілами.

Ключові слова: математичне моделювання, трансверсально-ізотропний півпростір, приповерхнева виїмка, еліптичний переріз, параметри контактної взаємодії.

ВСТУП

Під час проектування елементів конструкцій широко застосовуються біматеріали, складові яких мають різні пружні властивості, у тому числі анізотропні. Це стимулює дослідження і аналіз розподілу напружень у пружних тілах з біматеріалів поблизу концентраторів напружень, а також у разі їх контактної взаємодії. Водночас розв'язання просторових задач теорії пружності для анізотропних тіл ускладнюється, оскільки у цьому випадку необхідно розв'язувати граничну задачу для системи рівнянь рівноваги анізотропного тіла, яка має суттєво більш складну структуру (порівняно з відповідною системою рівнянь для ізотропного пружного тіла).

Для трансверсально-ізотропних пружних та електропружних тіл з концентраторами напружень з більшою глибиною досліджено двовимірні задачі [9, 16, 17, 27]. Подання загальних розв'язків систем рівнянь рівноваги у тривимірній постановці для трансверсально-ізотропних пружних та електро-

пружних тіл через гармонічні функції запропоновано у працях [11] і [24, 28] відповідно. Цікаві результати досліджень напруженого стану у просторових задачах для трансверсально-ізотропних пружних та електропружних тіл отримано у працях [1, 3, 4–6, 8, 10–12, 15, 18, 22, 23, 26] та [2, 7, 19, 20, 24, 29].

Відзначимо, що задачі контактної взаємодії для двох пружних ізотропних півпросторів за стискання через наявність між півпросторами включень чи похилої приповерхневої виїмки еліптичного перерізу досліджувались у працях [14, 21, 25]. Але задача контактної взаємодії для двох пружних трансверсально-ізотропних півпросторів за наявності в одному з них приповерхневої виїмки еліптичного перерізу не вивчалась.

У цій роботі за допомогою математичного моделювання на основі строгої моделі досліджено контактну взаємодію двох пружних трансверсально-ізотропних півпросторів, один з яких містить приповерхневу виїмку еліптичного перерізу, під час стискання. У постановці задачі припускається, що поверхня поділу двох пружних тіл розташована у площині ізотропії трансверсально-ізотропних матеріалів. Також вважається, що між тілами відбувається гладкий (без тертя) контакт. За допомогою подання розв'язку рівнянь рівноваги для трансверсально-ізотропного тіла через гармонічні функції (подання Еліота) і подальшого зведення задачі до розгляду інтегродиференціального рівняння отримано аналітичний розв'язок задачі. Знайдено основні параметри контактної взаємодії трансверсально-ізотропних пружних півпросторів (за наявності в одному з них приповерхневої виїмки еліптичного перерізу) під час стискання. Як частинний випадок з отриманих виразів впливають параметри контактної взаємодії двох трансверсально-ізотропних пружних півпросторів за наявності в одному з них виїмки кругового перерізу, а також параметри контакту двох ізотропних пружних півпросторів, один з яких містить виїмку еліптичного перерізу.

Постановка задачі. Розглянемо математичну модель, на основі якої вивчимо контактну взаємодію трансверсально-ізотропних пружних півпросторів, один з яких (тіло 2) містить приповерхневу виїмку еліптичного перерізу (рис. 1). Припускаємо, що форма виїмки описується таким виразом:

$$f_0(x, y) = \begin{cases} -h_0(1 - x^2/a_0^2 - y^2/b_0^2)^{3/2}, & \sqrt{x^2/a_0^2 + y^2/b_0^2} \leq 1, \quad (h_0 \ll \max(a_0, b_0)), \\ 0, & \sqrt{x^2/a_0^2 + y^2/b_0^2} > 1. \end{cases} \quad (1)$$

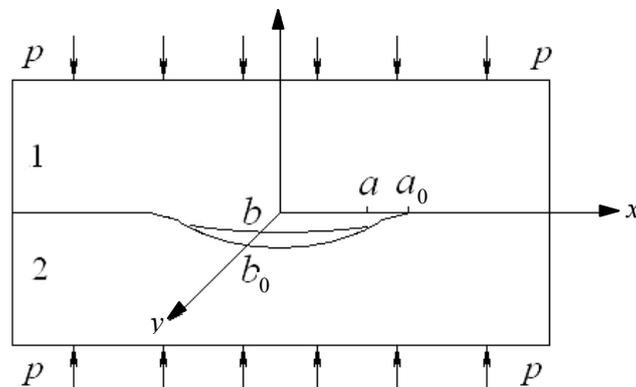


Рис. 1. Контактна взаємодія трансверсально-ізотропних пружних півпросторів, один з яких містить приповерхневу виїмку еліптичного перерізу

z

Розглядаючи задачу, будемо вважати, що площина контакту пружних трансверсально-ізотропних півпросторів розташована у площині ізотропії обох трансверсально-ізотропних матеріалів. Припускаємо, що до контактуючих тіл прикладено стискальні зусилля p і між тілами відбувається гладкий (без тертя) контакт. За наявності виїмки контакт між тілами здійснюється не по всій площині $z = 0$, а по деякій її частині $\sqrt{x^2/a^2 + y^2/b^2} > 1$, де a, b — невідомі параметри ділянки еліптичної форми (рис. 1), які залежать від значення стискальних зусиль p , геометрії початкової виїмки (описується параметрами a_0, b_0 і h_0), пружних властивостей обох трансверсально-ізотропних матеріалів півпросторів. Припускаємо, що у процесі деформування виїмка набуває вигляду, який можна описати у вигляді

$$f(x, y) = \begin{cases} -h(1 - x^2/a^2 - y^2/b^2)^{3/2}, & x^2/a_0^2 + y^2/b^2 \leq 1, (h_0 \ll \max(a, b)), \\ 0, & x^2/a_0^2 + y^2/b^2 > 1. \end{cases}$$

На поверхні поділу (у площині $z = 0$) отримуємо такі граничні умови:

$$\sigma_{zz}^{(i)} = 0, x^2/a^2 + y^2/b^2 \leq 1;$$

$$u_z^{(1)} = u_z^{(2)} + f_0(x, y), x^2/a^2 + y^2/b^2 > 1 \quad (i=1, 2);$$

$$\sigma_{zx}^{(i)} = \sigma_{zy}^{(i)} = 0, z = 0; \sigma_{zz}^{(1)} = \sigma_{zz}^{(2)}, x^2/a^2 + y^2/b^2 > 1 \quad (i=1, 2). \quad (2)$$

У граничних умовах (2) індекси 1 і 2 відповідають першому та другому пружним півпросторам, а також використовується функція, що описується виразом (1).

Припускаємо, що площина $z = 0$ (рис. 1), яка обмежує два півпростори, є площиною ізотропії обох трансверсально-ізотропних матеріалів, тобто вісь Oz є віссю симетрії для цих матеріалів. Також вважаємо, що у площині контактної взаємодії $z = 0$ виконуються умови гладкого (без тертя) контакту пружних тіл. Розміри ділянки розшарування двох пружних півпросторів (рис. 1) невідомі і визначаються з розв'язку задачі. Параметри контактної взаємодії пружних тіл залежать від значення стискальних зусиль p , пружних властивостей обох трансверсально-ізотропних матеріалів півпросторів і геометричних параметрів виїмки.

Виразимо напружений стан у трансверсально-ізотропних півпросторах суперпозицією двох станів, перший з яких — стискання вздовж осі симетрії матеріалів Oz , тобто $\sigma_{zz} = -p$, а для визначення другого стану суперпозиції (збуреного стану) на поверхні півпростору $z = 0$ отримуємо відповідні граничні умови.

Розглядаючи задачу, припускаємо, що в разі деформування півпросторів форма виїмки описується неперервно-диференційованою функцією $f(x, y)$, яка задовольняє умови:

$$\frac{\partial f(x, y)}{\partial x_1} \ll 1, \quad \frac{\partial f(x, y)}{\partial x_2} \ll 1 \quad (x, y) \in S; \quad (3)$$

$$f(x, y) = 0, \frac{\partial f(x, y)}{\partial x_1} = 0, \frac{\partial f(x, y)}{\partial x_2} = 0 \quad (x, y) \in \partial S.$$

У формулах (3) S — поверхня виїмки. Із віддаленням від ділянки контакту виконуються умови на нескінченності для трансверсально-ізотропних півпросторів:

$$\sigma_{zz} \rightarrow -p, \quad \sigma_{xx}, \sigma_{xy}, \sigma_{xz}, \sigma_{yy}, \sigma_{yz} \rightarrow 0, \quad \text{якщо } z \rightarrow \pm\infty.$$

Основні рівняння і співвідношення. Статичні рівняння теорії пружності для трансверсально-ізотропного тіла у переміщеннях набувають вигляду [4, 11]:

$$\begin{aligned} & c_{11} \frac{\partial^2 u_x}{\partial x^2} + \frac{1}{2}(c_{11} - c_{12}) \frac{\partial^2 u_x}{\partial y^2} + c_{44} \frac{\partial^2 u_x}{\partial z^2} + \\ & + \frac{\partial}{\partial x} \left[\frac{1}{2}(c_{11} + c_{12}) \frac{\partial u_y}{\partial y} + (c_{13} + c_{44}) \frac{\partial u_z}{\partial z} \right] = 0; \\ & \frac{1}{2}(c_{11} - c_{12}) \frac{\partial^2 u_y}{\partial x^2} + c_{11} \frac{\partial^2 u_y}{\partial y^2} + c_{44} \frac{\partial^2 u_y}{\partial z^2} + \\ & + \frac{\partial}{\partial y} \left[\frac{1}{2}(c_{11} + c_{12}) \frac{\partial u_x}{\partial x} + (c_{13} + c_{44}) \frac{\partial u_z}{\partial z} \right] = 0; \\ & c_{44} \left(\frac{\partial^2 u_z}{\partial y^2} + \frac{\partial^2 u_z}{\partial x^2} \right) + c_{33} \frac{\partial^2 u_z}{\partial z^2} + (c_{13} + c_{44}) \frac{\partial}{\partial z} \left(\frac{\partial u_x}{\partial x} + \frac{\partial u_y}{\partial y} \right) = 0, \end{aligned} \quad (4)$$

де $c_{11}, c_{12}, c_{13}, c_{33}, c_{44}$ — незалежні пружні сталі трансверсально-ізотропного матеріалу.

Розв'язок системи рівнянь (4) [11] можна виразити через три потенціальні функції Φ_i ($i=1,2,3$):

$$\begin{aligned} u_x &= \partial\Phi_1/\partial x + \partial\Phi_2/\partial x + \partial\Phi_3/\partial y; \\ u_y &= \partial\Phi_1/\partial y + \partial\Phi_2/\partial y - \partial\Phi_3/\partial x; \\ u_z &= m_1\partial\Phi_1/\partial z + m_2\partial\Phi_2/\partial z, \end{aligned} \quad (5)$$

де Φ_1, Φ_2, Φ_3 — функції, що входять до формул (5), задовольняють рівняння

$$(\partial^2/\partial x^2 + \partial^2/\partial y^2 + v_j \partial^2/\partial z^2)\Phi_j = 0, \quad (6)$$

а також $v_3 = 2c_{44}/(c_{11} - c_{12})$; v_1, v_2 — корені квадратного рівняння:

$$c_{11}c_{44}v^2 - [c_{44}^2 + c_{33}c_{11} - (c_{13} + c_{44})^2]v + c_{33}c_{44} = 0; \quad (7)$$

$$m_j = \frac{c_{11}v_j - c_{44}}{c_{13} + c_{44}} = \frac{v_j(c_{13} + c_{44})}{c_{33} - v_j c_{44}} \quad (j=1,2). \quad (8)$$

Увівши позначення $z_j = zv_j^{-1/2}$ ($j=1,2,3$) з використанням виразів (6)–(8), легко встановити, що функції $\Phi_1(x,y,z_1), \Phi_2(x,y,z_2), \Phi_3(x,y,z_3)$ будуть гармонічними функціями у відповідній системі координат (x,y,z_i) .

Надалі скористаємось виразами

$$k_1 = m_1 = \frac{c_{11}v_1 + c_{44}}{c_{13} + c_{44}} - 1; \quad k_2 = m_2 = \frac{c_{11}v_2 + c_{44}}{c_{13} + c_{44}} - 1;$$

$$a_j = c_{44}(1 + m_j) \quad (j=1,2). \quad (9)$$

Метод розв’язання. Потенціальні функції $\Phi_i(x,y,z_i)$ у поданні (5) для збуреного стану візьмемо, використовуючи класичні потенціали, у вигляді

$$\Phi_i^{(k)}(x,y,z_i) = -\frac{\alpha_i^{*(k)}}{2\pi} \left(\iint_{S_1} \frac{h(\vec{\xi})d_\xi S}{\sqrt{(x-\xi_1)^2 + (y-\xi_2)^2 + z_i^2}} + \iint_{S_0} \frac{r(\vec{\xi})d_\xi S}{\sqrt{(x-\xi_1)^2 + (y-\xi_2)^2 + z_i^2}} \right), \quad (10)$$

де

$$h(\xi_1, \xi_2) = h_1(1 - \xi_1^2/a^2 - \xi_2^2/b^2)^{3/2}, \quad r(\xi_1, \xi_2) = -h_0(1 - \xi_1^2/a_0^2 - \xi_2^2/b_0^2)^{3/2};$$

S_1 і S_0 — еліптичні ділянки, значення індексу $k=1$ відповідає першому півпростору, значення $k=2$ — другому півпростору. Також покладемо функцію $\Phi_3^{(k)} \equiv 0$. Сталі $\alpha_i^{*(k)}$ у формулах (10) визначимо так: спочатку знайдемо значення $\alpha_j^{*(k)}$ ($j,k=1,2$) із системи рівнянь

$$\sum_{j=1}^2 \alpha_j^{*(k)} (c_{44}^{(k)} (1 + k_j^{(k)})) = 1; \quad \sum_{j=1}^2 \alpha_j^{*(k)} (c_{44}^{(k)} (1 + k_j^{(k)})) / \sqrt{v_j^{(k)}} = 0, \quad k=1, 2, \quad (11)$$

а потім

$$\alpha_j^{*(k)} = \alpha_j^{*(k)} / M^*, \quad \text{де } M^* = \sum_{j=1}^2 \alpha_j^{*(1)} k_j^{(1)} / \sqrt{v_j^{(1)}} + \sum_{j=1}^2 \alpha_j^{*(2)} k_j^{(2)} / \sqrt{v_j^{(2)}}.$$

У наведених виразах використано позначення (9). Відзначимо, що потенціали вигляду (10) використовувались у дослідженні контактної взаємодії двох пружних ізотропних півпросторів, один з яких містив виїмку еліптичного перерізу [30]. За такого вибору $\alpha_j^{*(k)}$ задовольняються граничні умови по дотичних напруженнях уздовж всієї поверхні поділу, по переміщеннях, а також справджується рівність нормальних напружень поза виїмкою. Невідомими залишаються a, b — значення півосей ділянки контакту еліптичної форми; h_1 — максимальна висота проміжку (зазору) в результаті контакту двох трансверсально-ізотропних пружних півпросторів, один з яких містить виїмку. Значення цих параметрів визначимо з розв’язку контактної задачі.

За допомогою потенціальних функцій $\Phi_i^{(k)}(x,y,z_i)$ згідно з виразами (10), (11), задовольняючи граничні умови, що залишилися, отримуємо інтегро-диференціальне рівняння

$$\begin{aligned} & \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right) \iint_{S_1} \frac{h(\bar{\xi}) d_{\xi} S}{\sqrt{(x - \xi_1)^2 + (y - \xi_2)^2 + z_i^2}} = \\ & = - \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right) \iint_{S_0} \frac{r(\bar{\xi}) d_{\xi} S}{\sqrt{(x - \xi_1)^2 + (y - \xi_2)^2 + z_i^2}} + 2\pi M^* p, \quad (x, y) \in S_1; \quad (12) \\ & M^* = \sum_{j=1}^2 \alpha_j^{*(1)} k_j^{(1)} / \sqrt{v_j^{(1)}} + \sum_{j=1}^2 \alpha_j^{*(2)} k_j^{(2)} / \sqrt{v_j^{(2)}}, \end{aligned}$$

де

$$\begin{aligned} h(\xi_1, \xi_2) &= h_1 (1 - \xi_1^2 / a^2 - \xi_2^2 / b^2)^{3/2}, \\ r(\xi_1, \xi_2) &= -h_0 (1 - \xi_1^2 / a_0^2 - \xi_2^2 / b_0^2)^{3/2}. \end{aligned}$$

Після проведення диференціювання у (12) скористаємося значеннями двовимірних інтегралів по еліптичній ділянці [30]. У результаті маємо

$$\frac{1}{4\pi} \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right) \iint_{S_1} \frac{h(\xi) d_{\xi} S}{|x - \xi|} = \alpha_0 + A_0 x_1^2 + B_0 x_2^2, \quad x \in S, \quad (13)$$

де

$$d_{\xi} S = d\xi_1 d\xi_2; \quad k_0 = 3a_0 r_0 / 4e_0^2 b_0^4 f_0; \quad r_0 = h_0;$$

$$A_0 = k_0 [E(e_0)(1 + e_0^2) / f_0 - K(e_0)]; \quad B_0 = k_0 [(2e_0^2 - 1)E(e_0) + f_0 K(e_0)];$$

$$\alpha_0 = M^* p - k_0 e_0^2 b_0^2 E(e_0); \quad f_0 = 1 - e_0^2; \quad e_0 = \sqrt{1 - a_0^2 / b_0^2}.$$

У наведених виразах $K(e_0)$, $E(e_0)$ є повними еліптичними інтегралами першого та другого роду.

Припускаючи, що ділянка контакту S_1 у формулах (13) має форму еліпса з невідомими поки півосями a, b , розшукуємо висоту зазору між контактуючими тілами у вигляді $h(x) = \beta(1 - x_1^2 / a^2 - x_2^2 / b^2)^{3/2}$. Скориставшись значеннями інтегралів по еліптичній ділянці [31], отримуємо

$$\alpha + Ax_1^2 + Bx_2^2 = a_0 + A_0 x_1^2 + B_0 x_2^2,$$

де

$$\alpha = -ke^2 b^2 E(e), \quad A = k[E(e)(1 + e^2) / f - K(e)], \quad B = k[(2e^2 - 1)E(e) + fK(e)],$$

$$f = 1 - e^2, \quad k = 3\alpha h / (4e^2 b^4 f), \quad e = \sqrt{1 - a^2 / b^2}.$$

Прирівнюючи коефіцієнти за однакових структурних складових, знаходимо

$$a = a_0 \sqrt{1 - N_0}; \quad b = b_0 \sqrt{1 - N_0}; \quad h = r_0 [1 - N_0]^{3/2};$$

$$N_0 = 2M^* p b_0 \sqrt{1 - e_0^2} / (r_0 3E(e_0)). \quad (14)$$

Відзначимо, що із формул (14) можна визначити значення стискальних зусиль p , за якого виїмка між пружними тілами повністю заповнюється матеріалом. Видно, що при $N_0 = 1$ значення півосей a, b та максимальної висоти виїмки h_1 дорівнюють нулю. Отже, якщо $p \geq p^* = \frac{r_0 3E(e_0)}{2M^* b_0 \sqrt{1-e_0^2}}$, виїмка еліптичного перерізу у пружному півпросторі повністю зникає (заповнюється матеріалом).

Як частинний випадок зі знайдених аналітичних виразів для виїмки еліптичного перерізу випливають результати для виїмки вісесиметричної форми.

Таким чином, з виразів (14) за відомими розмірами початкової виїмки (параметри a_0, b_0 і h_0) та значення стискальних зусиль p , десяти незалежних пружних сталих двох трансверсально-ізотропних матеріалів півпросторів (входять через величину M^*) знаходимо значення півосей ділянки контакту a, b і максимальну висоту виїмки.

Перетворимо у виразі M^* кожний доданок, що залежить від властивостей трансверсально-ізотропного матеріалу відповідного півпростору. Після елементарних операцій дістаємо

$$\sum_{j=1}^2 \alpha_j^* k_j / \sqrt{v_j} = \frac{1}{(\sqrt{v_1} - \sqrt{v_2})} \left[\frac{-m_2}{c_{44}(1+m_2)} + \frac{m_1}{c_{44}(1+m_1)} \right] = \frac{c_{11} (v_1^{1/2} + v_2^{1/2})(c_{13} + c_{44})}{c_{44} (c_{11} v_1 + c_{13})(c_{11} v_2 + c_{13})}. \quad (15)$$

У результаті подальших перетворень виразу (15) (з використанням теореми Вієта для коренів квадратного рівняння (7)) маємо

$$\sum_{j=1}^2 \alpha_j^* k_j / \sqrt{v_j} = \frac{\sqrt{c_{11}}}{(c_{11}c_{33} - c_{13}^2)\sqrt{c_{44}}} \left[\sqrt{c_{11}c_{33} - c_{13}^2 - 2c_{44}c_{13} + 2c_{44}\sqrt{c_{11}c_{33}}} \right]. \quad (16)$$

Отриманий вираз (16) дозволяє знаходити шукане значення безпосереднім підставленням у нього пружних сталих трансверсально-ізотропного матеріалу, не знайшовши заздалегідь коренів квадратного рівняння (7).

У випадку абсолютно жорсткої основи з еліптичною виїмкою маємо

$$M^* \rightarrow \sum_{j=1}^2 \alpha_j^{*(1)} k_j^{(1)} / \sqrt{v_j^{(1)}}. \quad (17)$$

У виразі (17) залишився тільки перший доданок з двох складових. Під час наступного граничного переходу від трансверсально-ізотропного до пружного ізотропного матеріалу півпростору отримуємо

$$c_{11} = \lambda + 2\mu; c_{13} = \lambda; c_{44} = \mu; v_1 = v_2 = 1;$$

$$A^{Iso} \rightarrow \frac{\lambda + 2\mu}{2\mu(\lambda + \mu)} = \frac{1 - \nu}{\mu}.$$

За допомогою перетворень можна показати, що з виразів (12), (14) випливають результати роботи [30] для контактної взаємодії двох пружних

ізотропних півпросторів (один — з виїмкою еліптичного перерізу) як частинний випадок.

Розглянемо частинний випадок попередньої задачі — виїмку кругового перерізу. Спрямувавши ексцентриситет еліпса до нуля, отримуємо, що $a_0 = b_0$, $a = b$, а зі знайдених виразів (15) маємо

$$N_0 = 4M^* p b_0 / (r_0 3\pi); \quad b = b_0 \sqrt{1 - \frac{4p b_0 M^*}{3\pi r_0}}; \quad h_1 = h_0 \left(1 - \frac{4p b_0 M^*}{3\pi r_0}\right)^{3/2}, \quad (18)$$

$$\text{де } M^* = \sum_{j=1}^2 \alpha_j^{*(1)} k_j^{(1)} / \sqrt{v_j^{(1)}} + \sum_{j=1}^2 \alpha_j^{*(2)} k_j^{(2)} / \sqrt{v_j^{(2)}}.$$

Із виразів (18) за відомими розмірами початкової вісесиметричної виїмки (параметри b і h_0), значеннями стискальних зусиль p , десяти незалежних пружних сталей для двох трансверсально-ізотропних пружних матеріалів (входять через величину M^*) визначаємо радіус контакту a і максимальну висоту зазору h_1 (після контактної взаємодії). Також з рівності нулю виразу a знаходимо значення сили стиснення $p^* = \frac{3\pi h_0}{4bM^*}$, за якого виїмка у півпросторі повністю заповнюється матеріалом. Отже, під час стиснення $p \geq p^*$ приповерхнева виїмка в результаті контактної взаємодії повністю зникає.

Відзначимо, що під час розгляду частинного випадку задачі — випадку приповерхневої виїмки кругового перерізу, з'являється можливість використання альтернативного алгоритму розв'язання задачі, пов'язаного із застосуванням парних інтегральних рівнянь (алгоритм 2 у праці [2]).

Аналіз результатів числових досліджень. На рис. 2, 3 зображено залежності радіуса і висоти вісесиметричного зазору (частинного випадку зазору еліптичного перерізу) між пружними трансверсально-ізотропними півпросторами від діючих навантажень. У розрахунках використано значення пружних сталей із праці [13], наведені у таблиці, де номер матеріалу відповідає номеру кривої на рис. 2, 3.

Властивості матеріалів

Номер матеріалу	Матеріал	c_{11} , ГПа	c_{33} , ГПа	c_{44} , ГПа	c_{12} , ГПа	c_{13} , ГПа
1	Be	292,3	336,4	162,5	26,7	14,0
2	Co	307,0	358,1	78,3	165,0	103,0
3	Hf	181,1	196,9	55,7	77,2	66,1
4	ZnO	209,7	210,9	42,5	121,1	105,1
5	Zr	143,4	164,8	32,0	72,8	65,3

На рисунках використано позначення p_1^* , яке обчислювалось для матеріалу 1 з наведеної таблиці за допомогою виразу $p^* = 4a_0 N^* p / (3\pi r_0)$. Показано зміну радіуса і висоти зазору за навантаження, які залежать також від десяти пружних сталей двох трансверсально-ізотропних матеріалів (реалізується за допомогою величини N^*).

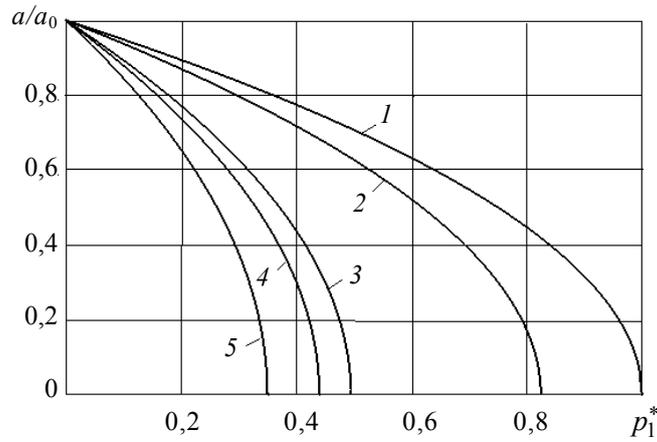


Рис. 2. Зміна радіуса виїмки за навантаження

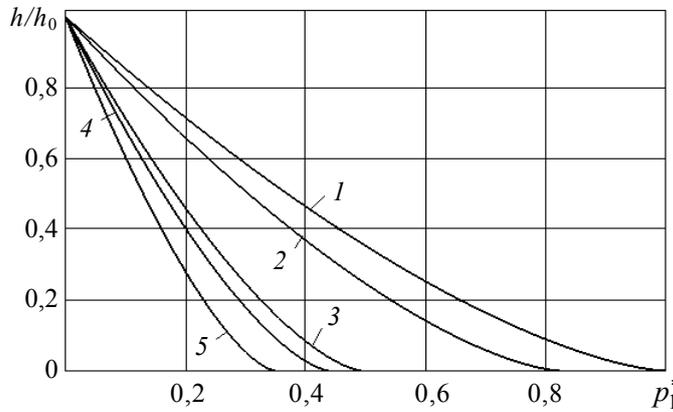


Рис. 3. Зміна висоти виїмки за навантаження

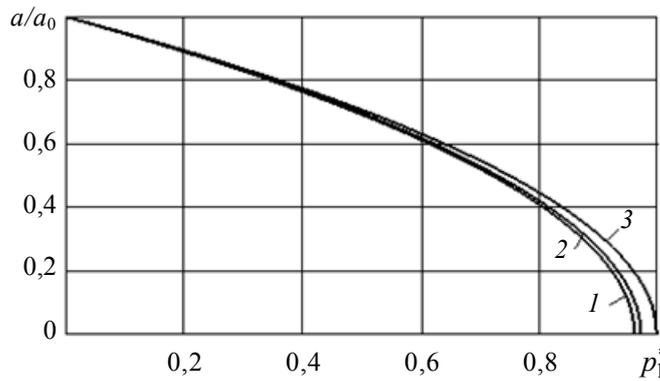


Рис. 4. Зміна розмірів перерізу еліптичної виїмки за навантаження

Виїмка повністю заповнюється матеріалом, якщо $p = p^* = \frac{3\pi h_0}{4a_0 N^*}$. Для

значень навантажень $p < p^*$ утворюється новий зазор. Для розрахунків пружні властивості обох півпросторів вибирались однаковими (відповідно до наведеного номера у таблиці).

Зміну геометричних параметрів зазору для еліптичної у перерізі виїмки за навантаження зображено на рис. 4, 5, на яких лінії 1, 2, 3 відповідають

значенням ексцентриситету еліпса $e = 0,1; 0,2; 0,4$. Використано позначення $p^{*(0,4)}$, яке відповідало значенню навантажень, що закривають виїмку, якщо $e = 0,4$.

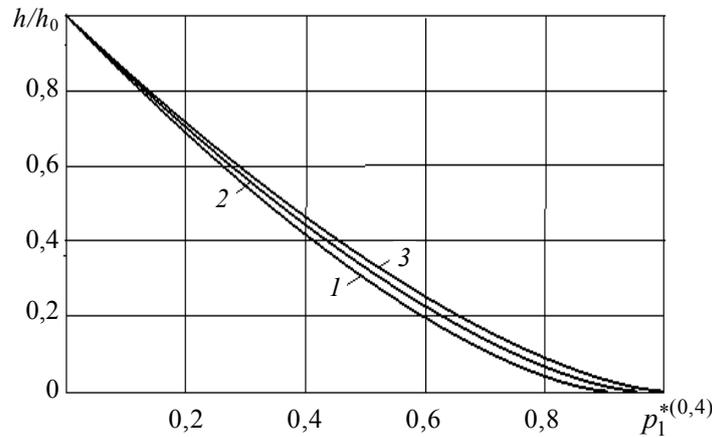


Рис. 5. Зміна висоти еліптичної виїмки за навантаження

ВИСНОВОК

У роботі за допомогою математичної моделі отримано аналітичний розв'язок задачі про контактну взаємодію складної системи, що складається з двох трансверсально-ізотропних пружних півпросторів (за наявності похилої приповерхневої виїмки еліптичного перерізу в одному з них), під час стискання. Виконано числові розрахунки, досліджено вплив геометричних розмірів виїмки та властивостей пружних трансверсально-ізотропних матеріалів півпросторів на параметри контактної взаємодії.

ЛІТЕРАТУРА

1. В.С. Кирилюк, О.І. Левчук, О.В. Гавриленко, та М.Б. Вітер, “Моделювання контактної взаємодії нагрітого плоского жорсткого еліптичного штампу з трансверсально-ізотропним півпростором”, *Системні дослідження та інформаційні технології*, № 3, с. 138–148, 2020. doi: 10.20535/SRIT.2308-8893.2020.3.10.
2. В.С. Кирилюк и О.И. Левчук, “Моделирование контактного взаимодействия пьезоэлектрического полупространства и упругой изотропной основы с приповерхностной выемкой кругового сечения”, *Системні дослідження та інформаційні технології*, № 4, с.120–132, 2016.
3. В.С. Кирилюк и О.И. Левчук, “Моделювання контактної взаємодії двох трансверсально-ізотропних пружних півпросторів за наявності жорсткого дископодібного включення між ними і тиску на ділянці розшарування”, *Системні дослідження та інформаційні технології*, № 1, с.107–119, 2020. doi: 10.20535/SRIT.2308-8893.2020.1.10.
4. Ю.Н. Подильчук, *Граничные задачи статики упругих тел*. Киев: Наук. думка, 1984, 304 с.
5. Ю.Н. Подильчук, “Точные аналитические решения пространственных граничных задач статики трансверсально-изотропного тела канонической формы (обзор)”, *Прикл. механика*, **33**, № 10, с. 3–30, 1997.

6. F.M. Borodich, B.A. Galanov, L.M. Keer, and M.M. Suarez-Alvarez, “The JKR-type adhesive contact problems for transversely isotropic elastic solids”, *Mechanics of Materials*, **75**, pp. 34–44, 2014.
7. Y.S. Chai and I.I. Argatov, “Local tangential contact of elastically similar, transversely isotropic elastic bodies”, *Meccanica*, **53**, no. 11-12, pp. 3137–3143, 2018.
8. W.Q. Chen, J. Zhu, and X.Y. Li, “General solutions for elasticity of transversely isotropic materials with thermal and other effects: A review”, *J. Thermal Stresses*, **42**, no. 1, pp. 90–106, 2019.
9. L. Dai, W. Guo, and X. Wang, “Stress concentration at an elliptic hole in transversely isotropic piezoelectric solids”, *Int. J. Solids and Struct.*, **43**, no. 6, pp. 1818–1831, 2006.
10. D.B. Davtyan and D.A. Pozharskii, “Action of an elliptic punch on a transversally isotropic half-space”, *Mechanics of Solids*, **49**, no. 5, pp. 578–586, 2014.
11. H.A. Elliott and N.F. Mott, “Three-dimensional stress distributions in hexagonal aeolotropic crystals”, *Mathematical Proceedings of the Cambridge Philosophical Society*, **44**, no. 4, pp. 522–533, 1948.
12. V.I. Fabrikant, “Contact problem for an arbitrarily oriented transversely isotropic half-space”, *Acta Mechanica*, **228**, no. 4, pp. 1541–1560, 2017.
13. L.B. Freund and S. Suresh, *Thin Film Materials*. Cambridge: Cambridge University Press, 2003, 802 p.
14. G.M.L. Gladwell, “On Inclusions at a Bi-Material Elastic Interface”, *Journal of Elasticity*, **54**, no. 1, pp. 27–41, 1999.
15. P.F. Hou, W.H. Zhang, and J.-Y. Chen, “Three-dimensional exact solutions of homogeneous transversely isotropic coated structures under spherical contact”, *Int. J. Solids Structures*, **161**, pp. 136–173, 2019.
16. S.A. Kaloerov and A.A. Samodurov, “Problem of Electromagnetoviscoelasticity for Multiply Connected Plates”, *International Applied Mechanics*, **51**, no. 6, pp. 623–639, 2015.
17. S.A. Kaloerov, “Determining the intensity factors for stresses, electric-flux density, and electric-field strength in multiply connected electroelastic anisotropic media”, *Int. Appl. Mech.*, **43**, no. 6, pp. 631–637, 2007.
18. V.S. Kirilyuk and O.I. Levchuk, “On stressed state of transversely isotropic medium with an arbitrary orientated spheroidal void or penny-shaped crack under internal pressure”, *Strength of Materials*, **37**, no. 5, pp. 480–488, 2005.
19. V.S. Kirilyuk, “Elastic state of a transversely isotropic piezoelectric body with an arbitrarily oriented elliptic crack”, *Int. Appl. Mech.*, **44**, no. 2, pp. 150–157, 2008.
20. V.S. Kirilyuk, “Stress state of a piezoceramic body with a plane crack opened by a rigid inclusion”, *Int. Appl. Mech.*, **44**, no. 7, pp. 757–768, 2008.
21. A. Kotousov, L.B. Neto, and A. Khanna, “On a rigid inclusion pressed between two elastic half spaces”, *Mechanics of Materials*, **68**, no. 1, pp. 38–44, 2014.
22. R. Kumar and V. Gupta, “Green’s function for transversely isotropic thermoelastic diffusion bimetals”, *Journal of Thermal Stresses*, **37**, no. 10, pp. 1201–1229, 2014.
23. F. Marmo, F. Toraldo, and L. Rosati, “Analytical formulas and design charts for transversely isotropic half-spaces subject to linearly distributed pressures”, *Meccanica*, **51**, no. 11, pp. 2909–2928, 2016.
24. Yu.N. Podil’chuk, “Representation of the general solution of statics equations of the electroelasticity of a transversally isotropic piezoceramic body in terms of harmonic functions”, *International Applied Mechanics*, **34**, no. 7, pp. 623–628, 1998.
25. A.P.S. Selvadurai, “A unilateral contact problem for a rigid disc inclusion embedded between two dissimilar elastic half-spaces”, *Q. J. Mech. Appl. Math.*, no. 3, pp. 493–509, 1994.

26. Yu.V. Tokovyy and C.C. Ma, “Three-Dimensional Elastic Analysis of Transversely-Isotropic Composites”, *Journal of Mechanics*, **33**, no. 6, pp. 821–830, 2018.
27. Y.J. Wang, C.F. Gao, and H.P. Song, “The anti-plane solution for the edge cracks originating from an arbitrary hole in a piezoelectric material”, *Mechanics Research Communications*, vol. 65, pp. 17–23, 2015.
28. Z.K. Wang and B.L. Zheng, “The general solution of three-dimension problems in piezoelectric media”, *Int. J. Solids Structures*, **32**, no. 1, pp. 105–115, 1995.
29. M.H. Zhao, Y.B. Pan, C.Y. Fan, and G.T. Xu, “Extended displacement discontinuity method for analysis of cracks in 2D piezoelectric semiconductors”, *International Journal of Solids and Structures*, vol. 94–95, pp. 50–59, 2016.
30. Г.С. Кіт та Р.М. Мартиняк, “Просторові контактні задачі для пружного півпростору і жорсткої основи з поверхневими виїмками”, *Мат. методи та фіз.-мех. поля*, **42**, № 6, с. 7–11, 1999.
31. М.В. Хай, *Двумерные интегральные уравнения ньютоновского потенциала и их приложения*. Киев: Наук. думка, 1993, 256 с.

Надійшла 31.05.2021

INFORMATION ON THE ARTICLE

Vitaly S. Kirilyuk, ORCID: 0000-0002-8513-0378, S.P. Timoshenko Institute of mechanics of NAS of Ukraine, e-mail: kirilyuk_v@ukr.net.

Olga I. Levchuk, ORCID: 0000-0002-6514-6225, S.P. Timoshenko Institute of mechanics of NAS of Ukraine, e-mail: 2013levchuk@gmail.com

Valeriy V. Gavrilenko, ORCID: 0000-0001-9682-4204, National Transport University, Ukraine, e-mail: v_gavr@ukr.net

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ КОНТАКТНОГО ВЗАИМОДЕЙСТВИЯ ДВУХ УПРУГИХ ТРАНСВЕРСАЛЬНО-ИЗОТРОПНЫХ ПОЛУПРОСТРАНСТВ, ОДНО ИЗ КОТОРЫХ СОДЕРЖИТ ПРИПОВЕРХНОСТНУЮ ВЫЕМКУ ЭЛЛИПТИЧЕСКОГО СЕЧЕНИЯ / В.С. Кирилюк, О.И. Левчук, В.В. Гавриленко

Аннотация. На основе математической модели рассмотрена задача о сжатия двух упругих трансверсально-изотропных полупространств, одно из которых содержит пологую приповерхностную выемку эллиптического сечения. Решение задачи получено с помощью представления Эллиота для трансверсально-изотропного тела через гармонические функции, классических гармонических потенциалов и сведения краевой задачи к рассмотрению интегродифференциального уравнения с неизвестной областью интегрирования. Как частный случай из найденных аналитических выражений вытекают основные параметры контакта трансверсально-изотропных полупространств при наличии в одном из них выемки осесимметричной формы, а также параметры контактного взаимодействия двух упругих изотропных полупространств, одно из которых содержит выемку эллиптического сечения. Получены численные результаты, изучено влияние упругих свойств полупространств, геометрических параметров выемки и погрузки на контактное взаимодействие и закрытия зазора между телами.

Ключевые слова: математическое моделирование, трансверсально-изотропные полупространства, приповерхностная выемка, эллиптического сечения, параметры контактного взаимодействия.

MATHEMATICAL MODELING OF THE CONTACT INTERACTION OF TWO ELASTIC TRANSVERSELY ISOTROPIC HALF-SPACES, ONE OF WHICH CONTAINS A NEAR-SURFACE GROOVE OF AN ELLIPTICAL SECTION / V.S. Kirilyuk, O.I. Levchuk, V.V. Gavrilenko

Abstract. On the basis of a mathematical model, the problem of compression of two elastic transversely isotropic half-spaces, one of which contains a shallow near-surface groove of an elliptical section, is considered. The solution to the problem is obtained using the Elliott representation for a transversely isotropic body in terms of harmonic functions, classical harmonic potentials and reducing the boundary value problem to considering an integro-differential equation with an unknown domain of integration. As a special case, the obtained analytical expressions yield the basic parameters of the contact of transversely isotropic half-spaces in the presence of an axisymmetric groove in one of them, as well as the parameters of the contact interaction of two elastic isotropic half-spaces, one of which contains an elliptical cross-section groove. Numerical results are obtained, the influence of elastic properties of half-spaces, geometrical parameters of groove and loading on contact interaction and closing of the gap between bodies is studied.

Keywords: mathematical modeling, transversely isotropic half-space, near-surface groove, elliptical section, parameters of contact interaction.

REFERENCES

1. V.S. Kirilyuk, O.I. Levchuk, O.V. Gavrilenko, and M.B. Viter, "Simulation of contact interaction of a heated flat rigid elliptical stamp with a transversely isotropic half-space", *System research and information technologies*, no. 3, pp.138–148, 2020. doi: 10.20535/SRIT.2308-8893.2020.3.10.
2. V.S. Kirilyuk and O.I. Levchuk, "Modeling of contact interaction of piezoelectric half-space and elastic isotropic base with surface groove of circle section", *System research and information technologies*, no. 4, pp. 120–132, 2016.
3. V.S. Kirilyuk and O.I. Levchuk, "Simulation of the contact interaction of two transversely isotropic spring half-spaces for the presence of a hard disk-like inclusion between them and pressure on the stratification area", *System research and information technologies*, no. 1, pp. 107–119, 2020. doi: 10.20535/SRIT.2308-8893.2020.1.10.
4. Yu.N. Podil'chuk, *Boundary value problems of statics of elastic bodies*. Kyiv: Nauk. dumka, 1984, 304 p.
5. Yu.N. Podil'chuk, "Exact analytical solutions of spatial boundary value problems of statics of a transversely isotropic body of canonical form (Review)", *Int. Appl. Mech.*, **33**, no. 10, pp. 3–30, 1997.
6. F.M. Borodich, B.A. Galanov, L.M. Keer, and M.M. Suarez-Alvarez, "The JKR-type adhesive contact problems for transversely isotropic elastic solids", *Mechanics of Materials*, **75**, pp. 34–44, 2014.
7. Y.S. Chai and I.I. Argatov, "Local tangential contact of elastically similar, transversely isotropic elastic bodies", *Meccanica*, **53**, no. 11-12, pp. 3137–3143, 2018.
8. W.Q. Chen, J. Zhu, and X.Y. Li, "General solutions for elasticity of transversely isotropic materials with thermal and other effects: A review", *J. Thermal Stresses*, **42**, no. 1, pp. 90–106, 2019.
9. L. Dai, W. Guo, and X. Wang, "Stress concentration at an elliptic hole in transversely isotropic piezoelectric solids", *Int. J. Solids and Struct.*, **43**, no. 6, pp. 1818–1831, 2006.
10. D.B. Davtyan and D.A. Pozharskii, "Action of an elliptic punch on a transversally isotropic half-space", *Mechanics of Solids*, **49**, no. 5, pp. 578–586, 2014.
11. H.A. Elliott and N.F. Mott, "Three-dimensional stress distributions in hexagonal aeolotropic crystals", *Mathematical Proceedings of the Cambridge Philosophical Society*, **44**, no. 4, pp. 522–533, 1948.
12. V.I. Fabrikant, "Contact problem for an arbitrarily oriented transversely isotropic half-space", *Acta Mechanica*, **228**, no. 4, pp. 1541–1560, 2017.
13. L.B. Freund and S. Suresh, *Thin Film Materials*. Cambridge: Cambridge University Press, 2003, 802 p.

14. G.M.L. Gladwell, "On Inclusions at a Bi-Material Elastic Interface", *Journal of Elasticity*, **54**, no. 1, pp. 27–41, 1999.
15. P.F. Hou, W.H. Zhang, and J.-Y. Chen, "Three-dimensional exact solutions of homogeneous transversely isotropic coated structures under spherical contact", *Int. J. Solids Structures*, **161**, pp. 136–173, 2019.
16. S.A. Kaloerov and A.A. Samodurov, "Problem of Electromagnetoviscoelasticity for Multiply Connected Plates", *International Applied Mechanics*, **51**, no. 6, pp. 623–639, 2015.
17. S.A. Kaloerov, "Determining the intensity factors for stresses, electric-flux density, and electric-field strength in multiply connected electroelastic anisotropic media", *Int. Appl. Mech.*, **43**, no. 6, pp. 631–637, 2007.
18. V.S. Kirilyuk and O.I. Levchuk, "On stressed state of transversely isotropic medium with an arbitrarily orientated spheroidal void or penny-shaped crack under internal pressure", *Strength of Materials*, **37**, no. 5, pp. 480–488, 2005.
19. V.S. Kirilyuk, "Elastic state of a transversely isotropic piezoelectric body with an arbitrarily oriented elliptic crack", *Int. Appl. Mech.*, **44**, no. 2, pp. 150–157, 2008.
20. V.S. Kirilyuk, "Stress state of a piezoceramic body with a plane crack opened by a rigid inclusion", *Int. Appl. Mech.*, **44**, no. 7, pp. 757–768, 2008.
21. A. Kotousov, L.B. Neto, and A. Khanna, "On a rigid inclusion pressed between two elastic half spaces", *Mechanics of Materials*, **68**, no. 1, pp. 38–44, 2014.
22. R. Kumar and V. Gupta, "Green's function for transversely isotropic thermoelastic diffusion bimaterials", *Journal of Thermal Stresses*, **37**, no. 10, pp. 1201–1229, 2014.
23. F. Marmo, F. Toraldo, and L. Rosati, "Analytical formulas and design charts for transversely isotropic half-spaces subject to linearly distributed pressures", *Meccanica*, **51**, no. 11, pp. 2909–2928, 2016.
24. Yu.N. Podil'chuk, "Representation of the general solution of statics equations of the electroelasticity of a transversally isotropic piezoceramic body in terms of harmonic functions", *International Applied Mechanics*, **34**, no. 7, pp. 623–628, 1998.
25. A.P.S. Selvadurai, "A unilateral contact problem for a rigid disc inclusion embedded between two dissimilar elastic half-spaces", *Q. J. Mech. Appl. Math.*, no. 3, pp. 493–509, 1994.
26. Yu.V. Tokovyy and C.C. Ma, "Three-Dimensional Elastic Analysis of Transversely-Isotropic Composites", *Journal of Mechanics*, **33**, no. 6, pp. 821–830, 2018.
27. Y.J. Wang, C.F. Gao, and H.P. Song, "The anti-plane solution for the edge cracks originating from an arbitrary hole in a piezoelectric material", *Mechanics Research Communications*, **65**, pp. 17–23, 2015.
28. Z.K. Wang and B.L. Zheng, "The general solution of three-dimension problems in piezoelectric media", *Int. J. Solids Structures*, **32**, no. 1, pp. 105–115, 1995.
29. M.H. Zhao, Y.B. Pan, C.Y. Fan, and G.T. Xu, "Extended displacement discontinuity method for analysis of cracks in 2D piezoelectric semiconductors", *International Journal of Solids and Structures*, vol. 94–95, pp. 50–59, 2016.
30. G.S. Keith and R.M. Martyniak, "Spatial contact problems for an elastic half-space and a rigid base with surface recesses", *Math. methods and physical and mechanical fields*, **42**, no. 6, pp. 7–11, 1999.
31. M.V. Hai, Two-dimensional integral equations of Newtonian potential and their applications. Kiev: Naukova dumka, 1993, 256 p.

OVERVIEW OF THE DETECTION AND TRACKING METHODS OF THE LAB ANIMALS

M.A. SHVANDT, V.V. MOROZ

Abstract. This article presents an overview of several most common techniques and approaches for object detection and tracking. Today, the tracking task is a very common problem and it can appear in many aspects of our life. One particular case of using object tracking techniques can appear during a lab animal behavior study. Different experimental conditions and the need of certain data collection can require some special tracking techniques. Thus, a set of general approaches to object tracking techniques were considered, and their functionality and possibilities were tested in a real life experiment. In this paper, their basis and main aspects are presented. The experiment has demonstrated the advantages and disadvantages of the studied methods. Considering this, conclusions and recommendations to their usage cases were made.

Keywords: object tracking, object detection, algorithm, video, frame, image, background, foreground, experiment, color space, thresholding, background estimation, segmentation.

INTRODUCTION

Over the last two decades, strong development of tech and technologies had to their widespread implementation in all areas of human life. One of such technologies is image processing and visual analysis [1]. Lots of processes in the world that surrounds us, including street traffic control, the tasks of terrorism prevention and war operations, need to be monitored, analyzed and, very often, controlled. In most such cases photo and video analysis comes in handy. Object detection and tracking usually play important roles in it. A certain object has to be detected on life video stream or recorded video, and then it is necessary to observe and trace the object's movements and position and, presumably, perform some analysis of these movements.

Object detection and tracking are in fact the key tasks of computer vision, as they allow one to gather consecutive information about the object which later can be analyzed [2]. As one knows most of information a person receives through its eyes, that is why computer vision also play an important role in data analysis. The tasks of computer vision include information acquisition, processing of the acquired information, processed data analysis and useful data acquisition. Computer vision is focused on the processing of two- and three-dimensional images. One of the tasks of 2D-processing is optical flow processing (video processing). It includes three key steps:

- 1) detection of moving objects;
- 2) object tracking from frame to frame;
- 3) analysis of an object to determine its characteristics.

In a simple way object tracking can be determined as the task of object trajectory estimation in the image plane.

The detection and tracking methods rely on many features, but the key ones are object shape and video background state. Thus the problem of tracking/detection can be quite challenging due to several factors. For example, there can be a lack of visual information due to the projection of a three-dimensional object on a two-dimensional plane. At the detection phase, the object can be partially occluded or it gets occluded later during tracking process. Also, the tracked object can change its shape or scale, which can also lead to tracking errors. And sometimes its movements can be quite complicated and hard-to-predict. The change of lighting can cause tracking errors, as well as the image noise. Image background can be a major source of difficulties. The easy situation is when there is a static background or it changes very slightly – then it is simple enough to pick out the tracked object. But if the background on each frame changes quite severely, it can also lead to situation when the tracking algorithm fails to pick out the necessary object correctly and loses it. In addition to that the tracking algorithm must be applicable for real-time video processing. In order to solve all these problems, different approaches have been suggested.

In biology there often is a necessity to study the life processes and behavior of lab animals, for example mice or fish. Such studies in this field have been carried out for a long time and they are still of sufficient scientific interest [3–5]. Different lab conditions may require specific approaches for automatic animal behavior examination. Thus the problem of object detection and tracking can be studied well on the particular example (Fig. 1) of such activity study. In the first case (Fig. 1, *a* and 1, *b*) the test environment is represented by a box with circle holes in the bottom (the holes denote the center of the test stand). In this case the task is to track the lab mice, note their movements between holes, time spent in the test stand center and moments when several mice contact with each other. In the second case there is an aquarium (Fig. 1, *c*). The task is quite similar: to track fish movements and notice their contacts. In both cases the camera is placed above the test environment. In order to solve these particular tasks a wide research was carried out to find the most suitable object detection and tracking approaches that could be used separately or combined. Thus various methods were examined, their advantages, disadvantages and algorithmic aspects have been considered. The complete analysis is presented further in this paper.

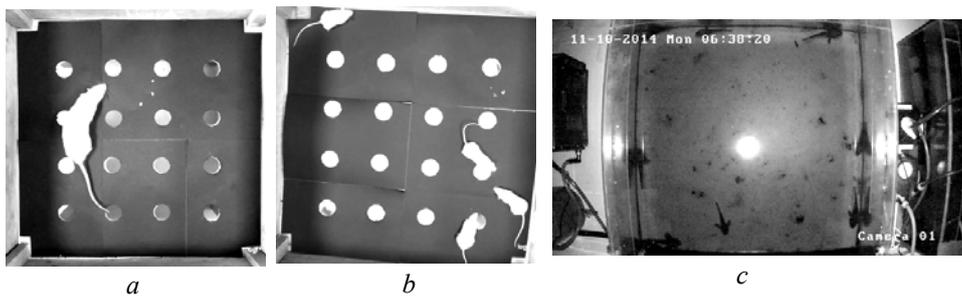


Fig. 1. Lab animals behavior study: *a* — lab rats; *b* — lab mice; *c* — fish

THE PROBLEM OF OBJECT DETECTION

The key task that appears during the object tracking process on video stream is their detection. Some methods require full object detection only on the first frame, some use continuous full detection on each frame.

Segmentation-based detection. *Image segmentation* is a process of digital image division on multiple sets of pixels. This process can also assign special markers to each pixel, so that the pixels with similar markers could have common visual characteristics [6]. For example, such approach can be based on the Watershed Transform [7] and results in the Watershed Algorithm combined with the Distance Transform. Image segmentation allows one to simplify the image analysis. It results in the highlighting of the borders and object itself. Thus, pixels belonging to the same segments are similar by some calculated feature (color, brightness value, etc.), with the rest of elements being significantly different by that feature. The result of such segmentation can be seen on Fig. 2. This approach is easy to use when objects of interest significantly differ from the background by some parameter. But the main problem is that it has low versatility and requires too accurate algorithm parameters setup in each particular case. It also is very sensitive to lighting conditions.

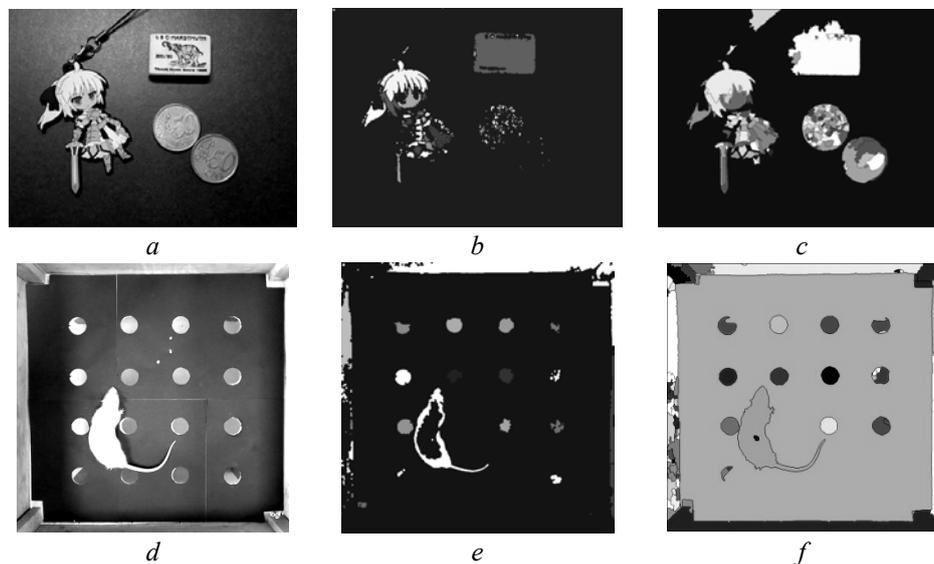


Fig. 2. Image segmentation: *a* — test 1: original image; *b* — test 1: markers; *c* — test 1: segmentation result; *d* — test 2: original image; *e* — test 2: markers; *f* — test 2: segmentation result

Another approach that works with image segments is *Template Matching* [8]. The algorithm compares the given object template with the sub-regions of processed image. To do this it simply slides the template along the image and checks if it matches to some region. The template (or patch) is sliding one pixel at a time (left to right, up to down) [9, 10]. At each location the algorithm calculates a metric that allows one to understand how similar the patch is to that particular area of the source image. For each location T over input image I it stores the metric in the result matrix R . Each cell (x, y) from R contains the match metric. Thus it is possible to find the best match by searching for the highest value (or lower, depending on the type of matching method) in the R matrix.

It is worth noticing, that while the patch must be a rectangle it may be that not the whole area of the rectangle is relevant. In this case the algorithm uses mask to isolate the portion of the patch that should be used to find the match. The mask is a grayscale image that masks the template image and must have the same

dimensions and number of channels. The match $R(x, y)$ can be calculated in several ways:

1. As a Square Difference:

$$R_{sq}(x, y) = \sum_{x', y'} (T(x', y') - I(x + x', y + y'))^2. \quad (1)$$

2. As Normed Square Difference:

$$R_{nsq}(x, y) = R_{sq}(x, y) / \sqrt{\sum_{x', y'} T(x', y')^2 \sum_{x', y'} I(x + x', y + y')^2}. \quad (2)$$

3. As Cross Correlation:

$$R_{ccorr}(x, y) = \sum_{x', y'} (T(x', y') I(x + x', y + y')). \quad (3)$$

4. As Normed Cross Correlation:

$$R_{nccorr}(x, y) = R_{ccorr}(x, y) / \sqrt{\sum_{x', y'} T(x', y')^2 \sum_{x', y'} I(x + x', y + y')^2}. \quad (4)$$

5. As Correlation Coefficient:

$$R_{ccoeff}(x, y) = \sum_{x', y'} (T'(x', y') I'(x + x', y + y')), \quad (5)$$

where

$$T'(x', y') = T(x', y') - \frac{\sum T(x'', y'')}{wh},$$

$$I'(x + x', y + y') = I(x + x', y + y') - \frac{\sum I(x + x'', y + y'')}{wh}.$$

6. As Normed Correlation Coefficient:

$$R_{nccoeff}(x, y) = R_{ccoeff}(x, y) / \sqrt{\sum_{x', y'} T(x', y')^2 \sum_{x', y'} I(x + x', y + y')^2}. \quad (6)$$

The result of Template matching algorithm is presented on Fig. 3. This approach can be used in case of some scene analysis when camera is static and objects of interest look almost identical, for example, detection of some products on a factory assembly line. But on the other hand, such method does not work stable in case of rotation or scaling and when an object is partially occluded. If the searched objects are scaled the most simple way could be to enlarge the template image as much as possible and that consequently scale it down at each search stage, hoping that at some point the template image will be scaled to the correct size. If the objects are rotated, the easiest way is to create a set of rotated by 1 degree template images and then iteratively check each sample. But both such approaches will deliver poor performance, especially in case of high resolution images. In case when the objects are both scaled and rotated, the performance can get even worse.

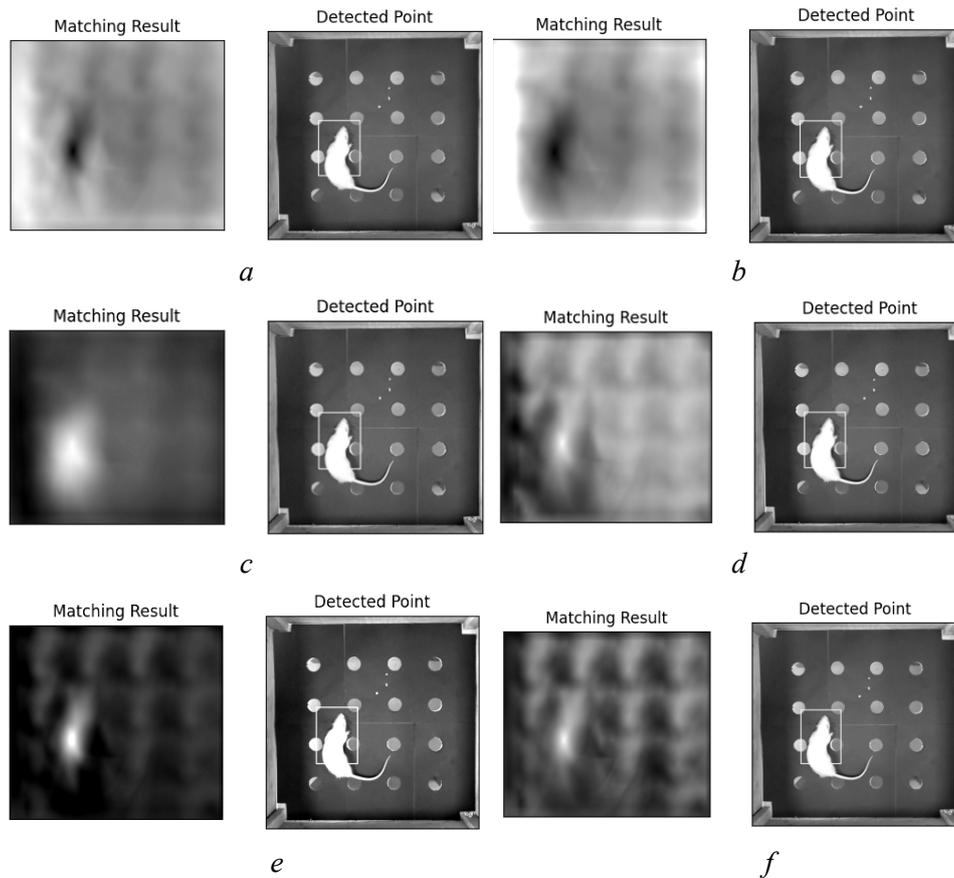


Fig. 3. Template matching: *a* — Square Difference (1); *b* — Normed Square Difference (2); *c* — Cross Correlation (3); *d* — Normed Cross Correlation (4); *e* — Correlation Coefficient (5); *f* — Normed Correlation Coefficient (6)

Feature-based detection. One more way to detect an object on image is to find it by some features. A feature is some element or part that is more distinguished than the other parts/elements, some local image particle. As simple example of such features are corners and borders. The search of an object in this case is based on the comparison of the characteristic features of the processed frame and a template showing the object one is looking for [11]. Local features should be repetitive (stable to change the angle or lighting during the video series), compact (their number should be much less than the total number pixels of the image), unique (each feature must have their own description).

To identify the characteristic features special detectors are used. One of the most common is the Harris (corner) detector (Fig. 4,c), which recognizes the features of the type “corner” in the image. As corner detectors are not very sensitive to image scaling, the concept of so-called drops (Blob) was introduced - teardrop-shaped neighborhoods with a special point located in the center. One of the most common blob methods is *LoG* (The Laplacian of Gaussian) [12]. *LoG* is

a filter $LoG(x, y) = \frac{x^2 + y^2 - 2\sigma^2}{\pi\sigma^4} e^{-\frac{x^2 + y^2}{2\sigma^2}}$ that applies the Gaussian operator

$L(x, y; \sigma) = G(x, y; \sigma) * I(x, y)$, $G(x, y; \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2+y^2}{2\sigma^2}}$, and the Laplace operator $\nabla^2 = \frac{\partial^2 I}{\partial x^2} + \frac{\partial^2 I}{\partial y^2}$ to the image, respectively (Fig. 4,a):

Here σ the standard deviation, $L(x, y; \sigma)$ is the Gaussian scale-space representation of an image $I(x, y)$, and $*$ is the convolution operator. DoG detector

$$DoG \widehat{=} (G_{\sigma_1} - G_{\sigma_2}) = \frac{1}{\sqrt{2\pi}} \left(\frac{1}{\sigma_1} e^{-\frac{x^2+y^2}{2\sigma_1^2}} - \frac{1}{\sigma_2} e^{-\frac{x^2+y^2}{2\sigma_2^2}} \right)$$

based on the Gaussian

difference $L_1(x, y; \sigma_1) = G_{\sigma_1}(x, y; \sigma_1)I(x, y)$, $L_2(x, y; \sigma_2) = G_{\sigma_2}(x, y; \sigma_2)I(x, y)$;
 $L_1(x, y; \sigma_1) - L_2(x, y; \sigma_2) = G_{\sigma_1}(x, y; \sigma_1)I(x, y) - G_{\sigma_2}(x, y; \sigma_2)I(x, y) =$
 $= (G_{\sigma_1}(x, y; \sigma_1) - G_{\sigma_2}(x, y; \sigma_2)) * I(x, y) = DoG * I(x, y)$. [13] is also common (Fig. 4,b). The difference between the two smoothing is as follows.

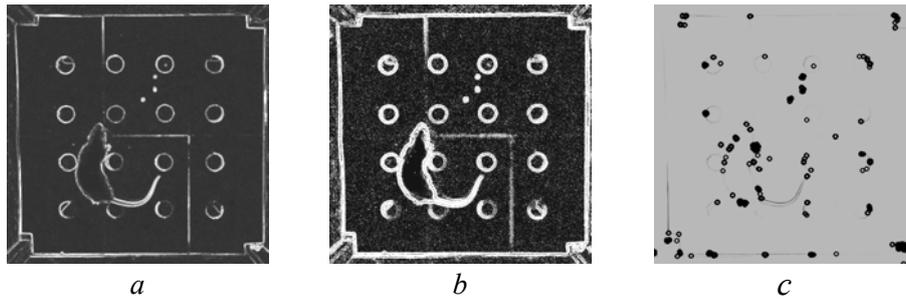


Fig. 4. Blob detection using Gaussian filter: a — Laplacian of Gaussian (LoG); b — Gaussian difference (DoG); c — Harris detector

After finding special points, it is necessary to compare them. This task requires a way of compact characteristic features representation. In practical tasks, the *SIFT* (Scale-Invariant feature transform) descriptor [14] and its derivatives, such as *SURF* [15], are considered to be the best methods. Despite being invariant to small turns, scaling of objects and changes in stage lighting, the feature-based approach actually makes it impossible to define an object as instance of some class and it also provide false results in case of object dynamic shape change (Fig. 5).

Categorical recognition. Methods for detecting characteristic features are well suited to solve the problem of searching across the database of images [16]. However, in our particular case it is necessary not simply to reveal some object on the frames of a video corresponding to some template, but also to recognize all objects of certain class. The considered problem could be solved by methods of feature detection but at the same time it would be necessary to create a large number of templates and it would take a long time to compare the frames with each of them. The approach that allows us to avoid this is based on the classification of objects, i.e. categorical recognition. It consists of two main elements: the definition of a set of features or descriptors and machine learning of

the classifier. As a set of features the Histogram of Oriented Gradients (*HOG*) or Haar features can be used. The HOG features [17] are based on the calculation of the number of gradient directions in the local areas of the image (Fig. 6).

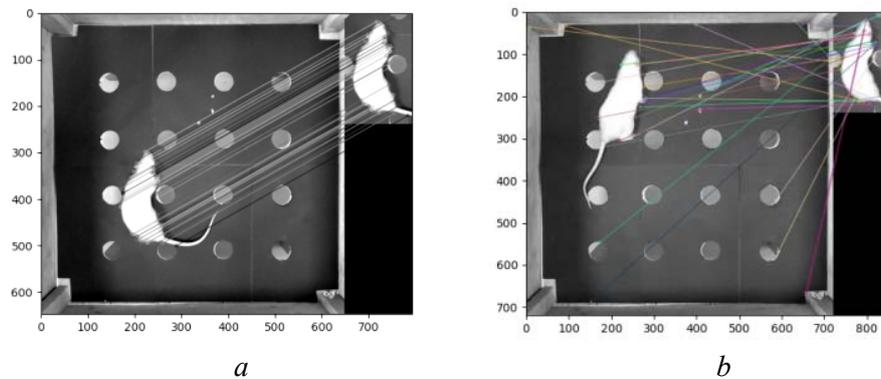


Fig. 5. SIFT feature object detection: *a* — perfect match; *b* — mismatch case

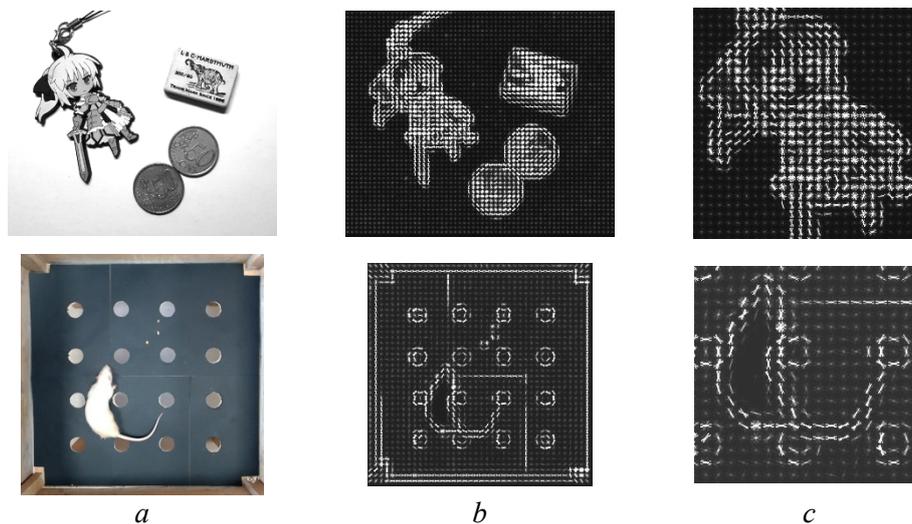


Fig. 6. HOG features detection: *a* — input images; *b* — extracted HOG features; *c* — HOG features (magn.)

Haar signs [18] or primitives are rectangles consisting of adjacent areas (see Fig. 7, *a*). These areas get positioned on the image, then the intensity of pixels in the areas is summed, and then the difference between the obtained sums is calculated, which is the value of a certain feature of a certain size, located on the image in a certain way. An example of the use of Haar features is shown on Fig. 7, *b*. The advantage of Haar features is a relatively high computational speed. Machine learning is used to create a class clarifier. The classifier is used to indicate which features belong to the object. Thus for training purpose some base of these features is used.

HOG is calculated on a dense grid of evenly distributed cells (Fig. 6). This method highlights well the objects with multiple details, but in case when the object is mostly a single piece without any significant details, in most cases it will

only highlight the borders, which can be not enough for complete detection. The Haar approach on the other hand is suitable for face detection and recognition. But this approach still requires a pre-trained classifier which sometimes can be problematic and will not work well with objects that tend to change shapes.

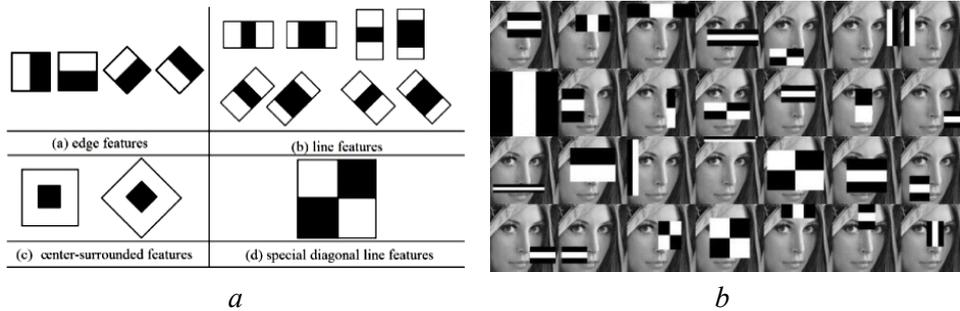


Fig. 7. Haar features: *a* — Haar features types (source: www.spiedigitallibrary.org); *b* — general representation of training the Haar classifier (source: medium.com)

THE PROBLEM OF OBJECT TRACKING

As mentioned above the process of tracking of moving objects is one of the components of many real-time systems such as observation systems, video analysis and others. The input data of any tracking algorithm is a sequence of images (video frames) I_1, I_2, \dots, I_n with an increasing amount of information that needs to be processed and analyzed. The task of tracking is to construct the trajectories of the target objects on the input sequence of frames. If we assume that the position of the object on the image numbered k is denoted by P_k . Then the trajectory of the object is sequence of its positions $P_s, P_{s+1}, \dots, P_{s+l-1}$, where s is the number of the first frame in which the object was detected, l is the number of frames in the sequence where the object is observed.

Some methods of object detection allow us to detect the entire object, but usually they are not suitable for continuous work, especially for real-time video processing. In most cases performing the detection “from scratch” for each frame can be very costly in terms of performance and speed, thus the detection process should be optimized in some way, especially if some frames have already been processed and we received some additional information from them. That is way several different object tracking approaches have been introduced. Note that depending on the method of tracking the position of the object can be determined differently (coordinates and size of the sides of the surrounding rectangle, coordinates of the center of mass of the contour, etc.).

Color-based tracking approaches. The idea of simple color-based tracking consists of the following steps [17]: first, the algorithm takes each frame and converts it from RGB to HSV color model. It is necessary because the RGB representation is not very suitable for selection of some specific color range. It can be performed in the following way [1, 19]: the given R, G, B values are scaled to change the range from $0 \dots 255$ to $0 \dots 1$. Then one calculates

$$C_{\max} = \max(R', G', B'), \quad C_{\min} = \min(R', G', B'), \quad \Delta = C_{\max} - C_{\min},$$

where R', G', B' are scaled R, G, B values. The Hue (H) can be calculated using the following formula:

$$H = \begin{cases} 0^\circ & \text{if } \Delta = 0; \\ 60^\circ \cdot \left(\frac{G' - B'}{\Delta} \right), & \text{if } C_{\max} = R'; \\ 60^\circ \cdot \left(\frac{B' - R'}{\Delta} + 2 \right), & \text{if } C_{\max} = G'; \\ 60^\circ \cdot \left(\frac{R' - G'}{\Delta} + 4 \right), & \text{if } C_{\max} = B'; \end{cases}$$

$$H = H + 360, \text{ if } H < 0.$$

The Saturation (S) calculation:

$$S = \begin{cases} 0, & C_{\max} = 0; \\ \frac{\Delta}{C_{\max}}, & C_{\max} \neq 0. \end{cases}$$

The Value V calculation:

$$V = C_{\max}.$$

It is worth noticing that before the RGB to HSV conversion a Gaussian blur is applied, as described in [20] to remove noise in order to receive better output. The result is seen on Fig. 8, *a, b*. The second step after a successful conversion is the color thresholding. The lower and upper boundaries of the desired color are set in the HSV color space. This allows to filter out the rest of the colors from the image. The thresholding process for an input image I can be described as follows:

$$\begin{aligned} dst(I) = & (lowerB(I)_1 \leq \\ & \leq src(I)_1 \leq upperB(I)_1) \wedge \dots \wedge (lowerB(I)_n \leq src(I)_n \leq upperB(I)_n), \end{aligned}$$

where $lowerB(I)_i \leq src(I)_i \leq upperB(I)_i$ stands for the i_{th} input array channel, $i = 1, \dots, n$. Thus:

- For every element of a single-channel input array:

$$dst(I) = lowerB(I)_1 \leq src(I)_1 \leq upperB(I)_1;$$

- For two-channel arrays:

$$dst(I) = lowerB(I)_1 \leq src(I)_1 \leq upperB(I)_1 \wedge lowerB(I)_2 \leq src(I)_2 \leq upperB(I)_2.$$

The resulting image is a binary image, i.e. all its pixel values are 1 or 0. For the resulting image after thresholding the operations of *erosion* and *dilatation* are applied, as described in [21, 22]. This allows us to get rid of most separated areas that managed to pass the threshold (Fig. 8, *c, d, f*). The final step is a centroid calculation for each blob using the binary image moments [23]:

$$M_{i,j} = \sum_x \sum_y x^i y^j I(x,y). \quad (7)$$

If one denotes a blob area as M_{00} , then the centroid can be calculated as follows:

$$\{\bar{x}; \bar{y}\} = \left\{ \frac{M_{10}}{M_{00}}; \frac{M_{01}}{M_{00}} \right\}. \quad (8)$$

For each blob its point $\{\bar{x}; \bar{y}\}$ can be used as object position on current frame (Fig. 8, *e*).

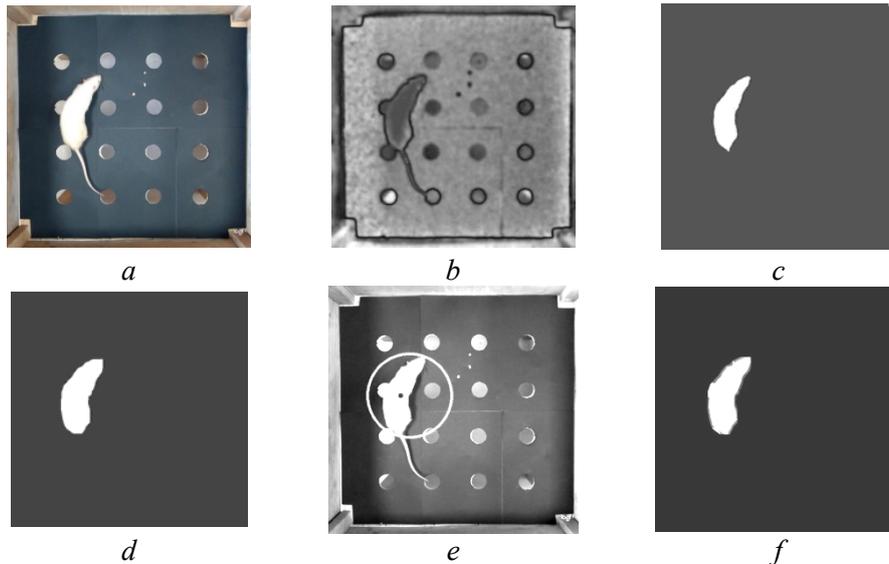


Fig. 8. Color-based tracking: *a* — RGB frame; *b* — HSV frame; *c* — raw binary image; *d* — after erosion & dilatation operations; *e* — detected object; *f* — HSV thresholding result

The advantages of such tracking approach are that in fact the target object gets detected automatically it works well with objects, which change their shape. In addition the overall realization is very simple and it has good performance speed, which makes it suitable for real-time video capturing. But this method has some serious disadvantages. Firstly, it is more of a detection than tracking technics, so if there are several object of interest and they occasionally get occluded, after repeated detection there is no guarantee, that these objects' positions were not messed up (Fig. 9, *a, b*). This point requires additional control in addition to tracking technics. Also it requires from the user a manual selection of lower and upper HSV color threshold boundaries, which is not a very easy task by itself, and the tracked objects have to be distinguished from the background by color (Fig. 9, *c*). In addition, this technic will work well mostly only with the colorful images, because the grayscale color space is much more poor for color differentiation, thus it will not be suitable for usage on videos like one on Fig. 1, *c*. An finally, if several tracked objects of one occasionally come very close to each other, they merge into just one object and the algorithm begins treating them as a single object. This fact devalues the accuracy of the method.

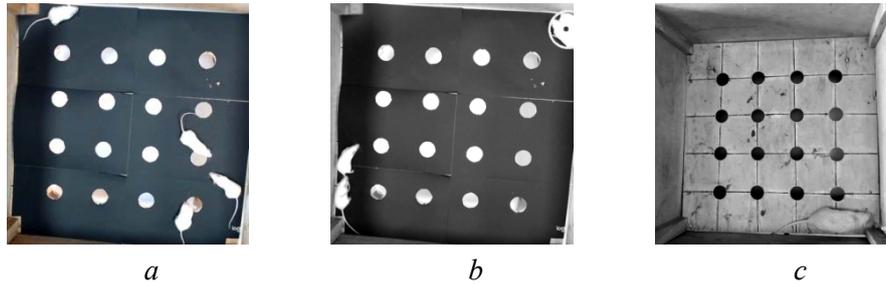


Fig. 9. Color-based tracking errors: *a* — 4 mice; *b* — 2 mice are merged into a single blob; *c* — identical color range case

Another approach is the background subtraction. The idea of this method is similar to the color tracking algorithm as it also directly separates the tracking object from the background [21, 24]. The main difference is that one requires an image of observed location without any moving objects on it. In case of rat/mice tracking task this location is the test box. This method was implemented and tested in [20]. In brief, it consists of the following steps:

1. The algorithm receives an image of empty observed location, it is converted from RGB to grayscale and cleared from noise with Gaussian or Median filters (Fig. 10, *a*).
2. Each video frame is also converted from RGB to grayscale and cleared from noise (Fig. 10, *b*).
3. For each frame the background subtraction operation is performed on both grayscale frame and grayscale image of an empty box. The main formula is:

$$d_{i,j} = |b_{i,j} - f_{i,j}|, \quad i = \overline{1,n}, \quad j = \overline{1,m},$$

where $d_{i,j}$ is pixel value of the resulting image (i.e. background subtraction output / image difference), $b_{i,j}$ and $f_{i,j}$ are the pixel values of empty grayscale background image and each grayscale video frame respectively, $i = \overline{1,n}$ and $j = \overline{1,m}$ are the dimensions image/frames. Notice, that these dimensions must be equal for both empty background image and frame for obvious reasons. The result is presented on Fig. 10, *c*.

4. Next step is thresholding [25]: all pixel values, that are higher than some *threshold* are put to 0, the rest is set to 255. The result is a binary image (Fig. 10, *d*), i.e. it is only black and white. If necessary, operations of erosion and dilatation are applied (Fig. 10, *e*).

5. Finally, similarly to color-based tracking, for binary image blobs centroid calculation is performed [23]. It is done using image moments calculation (formulas 7 and 8). The result can be seen on Fig. 10, *f* (circles were detected using Hough transform as the box central area [25]).

This methods has similar to color-based tracking advantages, as it also requires a threshold value, but it is more convenient as it requires only one such value instead of a range. Thus it is more stable. But the main disadvantage is the mandatory existence of the background image. In case of difficulties with

providing such image this algorithm should not be used. The rest of possible problems are also similar to color-based tracking.

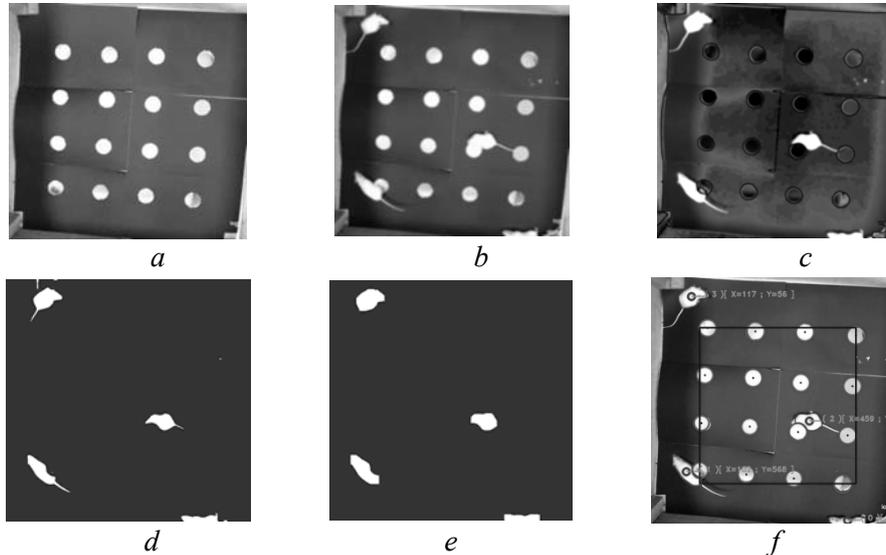


Fig. 10. Background subtraction tracking: *a* — grayscale box image smoothed; *b* — grayscale frame smoothed; *c* — image difference; *d* — binary image; *e* — after erosion/dilatation; *f* — background subtraction result

It can happen that there is no background image provided. Then *background estimation* method comes in handy. It can be used as an addition to background subtraction algorithm. The main idea is that the background image gets calculated from input video. It can be performed using the Approximation Median Algorithm [26, 27]. As it is described in [26], it finds the difference of values of the current pixel's intensity and the median of some recent pixel's intensity. For this task an n -size buffer is used, it contains n last frames whose pixel values are used for calculating the median value for background image. The main formula for this method can be written as follows:

$$|F_{i,j,k} - Med_{i,j,k}| > thresh, \quad i = \overline{1, side_1}, \quad j = \overline{1, side_2}, \quad k = \overline{1, frNum},$$

with F representing the current frame and Med being the median of last n frames. For each new frame k for each pixel (i, j) the difference of current frame pixel with pixel of median of last n frames decides whether this value is foreground or background. The median value gets updated for last n recent pixel values.

The described method can be also performed in the following way [28]. Assuming, that the camera is static and most of the time every pixel shows the same piece of the background, every moving object will occlude the background. In this case for the video on can randomly sample n frames. Thus for every pixel, now there are n estimates of the background. As long as a pixel is not occluded by the moving object, more than 50% of the time, the median of the pixel over these n can be a good estimate of the background at that pixel. This process can be repeated this for every pixel and thus it recovers the entire background.

As an alternative, Mixture of Gaussian can be used for background estimation [26, 27]. It uses a Gaussian probability density function to evaluate the pixel

intensity value. This method calculates the difference of values of the current pixel's intensity and cumulative average of the previous values. It means that the algorithm keeps a cumulative average μ_t of the recent pixel values, and if the difference of the current image's pixel and the cumulative pixel values is greater than the product of a constant value c and standard deviation σ , then this difference it is classified as foreground. Thus for each frame t the F_t pixel value can be denoted as foreground pixel, if the following inequality holds:

$$|F_t - \mu_t| > c\sigma_t, \quad t = \overline{1, fr Num},$$

In other case, this value can be classified as background. Also, this algorithm updates background image as the running average using formulas:

$$\mu_{t+1} = \mu_t F_t + (1 - \alpha)\mu_t,$$

$$\sigma_{t+1}^2 = \alpha(F_t - \mu_t)^2 + (1 - \alpha)\sigma_t^2,$$

where α is the learning rate (typically $\alpha = 0,05$); F_t is the pixel current value; μ_t is the previous average.

The result of such technics is shown on Fig. 11, 12. In case of mice/rat tracking, there is an empty box image, so it can be compared with the resulting background estimation (Fig. 11). Notice, that three white dots were static on each video frame, thus they managed to pass to estimated background (Fig. 11, *b*). Fig. 12, *a* shows estimated background image, acquired from the corresponding aquarium video. In this particular case no empty aquarium image had been provided, thus this is exactly the case when background estimation can be applied. Also notice, that one fish in the top corners remained static during the whole video, so they were also classified as background and was later missed by the algorithm. This case shows the main drawback of such approach (Fig. 12, *b*).

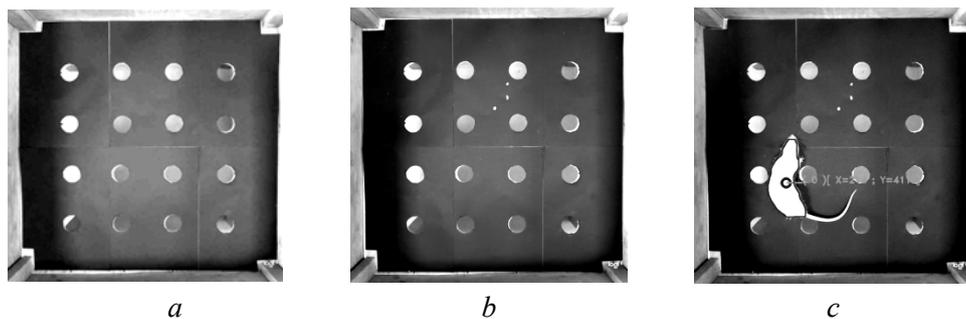


Fig. 11. Background estimation: *a* — original empty box image; *b* — estimated box image; *c* — tracking result

The positive side of this approach is that when one uses the background subtraction and there is no empty background image, in most cases this technic can compensate this need. But as it is shown on Fig. 12, *a*, if one of the tracked objects remains static during the whole video, it will be classified as a part of the background and thus the tracking method will not be able to detect and track it. It is also effective only if the entire background is static.

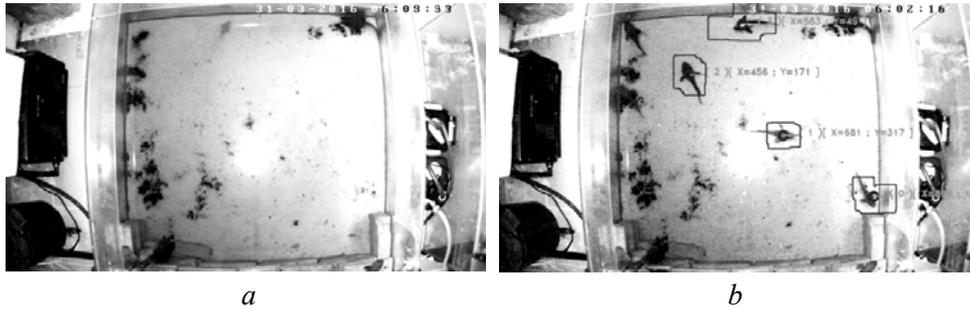


Fig. 12. Background subtraction operation test with fish: *a* — empty aquarium estimation; *b* — tracking result

Kernel tracking & optical flow. A kernel component is the shape of an object. In the simplest case, the component can be represented by a rectangular or oval shape, in more complex ones by three-dimensional model of the object projected on the plane of the image. The methods of this group are usually used if the motion is determined by a normal displacement, rotation, or affine transformation. Component tracking is an iterative localization procedure based on maximizing some similarity criterion. In practice, it is realized using mean shift and its continuous modification (Continuous Adaptive Mean Shift, CAM Shift).

The idea of the *Mean Shift* [29, 30] is that for each special point (in the general case, for each object) the search window is selected, the center of masses of the intensity distribution (i.e. of the histogram) is calculated. Accordingly, the center of the window is shifted to the center of mass, which is the position of the point on the current frame. Determining the position of the point in the following frames is reduced to the application of the next step of the method of “average shift”. The method stops when the center of mass stops shifting (Fig. 13).



Fig. 13. Mean shift tracking

The problem with Mean Shift is that the window (ROI) always has the same size whether the object is very far or very close to the camera, it needs to be adapted during the tracking process. The solution to this is CAM Shift (Continuously Adaptive Mean Shift) [31]. This approach applies the Mean Shift first, then once Mean Shift converges, it updates the size of the window as

$$s = 2 \sqrt{\frac{M_{00}}{256}}.$$

CAM Shift also calculates the orientation of the best fitting ellipse to it. It applies the Mean Shift with new scaled search window again and previous window location. This process continues until the required accuracy is met (Fig. 14). This approach shows fine work speed and is more stable than Mean Shift, but the

main problem with CAM Shift is that it is connected to color range, thus it is sensitive to lighting conditions and can fail with objects that change their shape.



Fig. 14. CAM Shift tracking

Optical flow estimation can be used as the alternative to all previous methods [32, 33]. Optical flow itself can be described as a trace of visible object movement between two consecutive frames [34]. It can be caused by moving object itself or by camera movement and it is represented by 2D vector field where each vector is a displacement vector showing the movement of points from first frame to second. There are several applications, where optical flow can be used, especially motion detection, or video stabilization.

There are several assumptions that optical flow works with [33, 34]: firstly, the pixel intensities of an object do not change between consecutive frames, and secondly, pixels in neighborhood must have similar motion. Let $I(x, y, t)$ be a pixel from the first frame (t is time), and it gets moved by distance (dx, dy) in the next frame taken after dt time. Assuming, that the pixel intensity does not change, the following holds:

$$I(x, y, t) = I(x + dx, y + dy, t + dt).$$

By using the Taylor series approximation of right-hand side, removing common terms and dividing by dt one gets the following equation:

$$f_x u + f_y v + f_t = 0$$

$$\text{with } f_x = \frac{\partial f}{\partial x}; \quad f_y = \frac{\partial f}{\partial y}; \quad u = \frac{dx}{dt}; \quad v = \frac{dy}{dt}. \quad (9)$$

The equation (9) is called *Optical Flow* equation, where f_x, f_y are image gradients, which can be found, and f_t is the gradient along time. The u and v components are unknown and thus equation (9) cannot be solved with two unknown variables. There are several solutions to this problem. One of them is Lucas-Kanade method [32, 34]. The Lucas-Kanade approach uses the 3×3 patch around the point, so that 9 points have the same motion. It is possible to calculate f_x, f_y, f_t for these 9 points, thus there appears a task to solve 9 equations with two unknown variables which is over-determined. It can be solved with least square fit method:

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} \sum_i f_{x_i}^2 & \sum_i f_{x_i} f_{y_i} \\ \sum_i f_{x_i} f_{y_i} & \sum_i f_{y_i}^2 \end{bmatrix}^{-1} \begin{bmatrix} -\sum_i f_{x_i} f_{t_i} \\ -\sum_i f_{y_i} f_{t_i} \end{bmatrix}.$$

It is worth noticing, that the inverse matrix is similar to Harris corner detector, as corners are better points to be tracked. Also, as it can be seen, this approach allows to detect only small motions, but not the big ones. In order to solve this problem the pyramids are used: when going up in the pyramid, small motions

are removed and large motions become small motions. Thus when one applies Lucas–Kanade there, one gets optical flow along with the scale. The result is presented on Fig. 15, *a*. As Lucas–Kanade method computes optical flow for a sparse feature set (*sparse optical flow*), using, for example, Shi-Tomasi corner detection technic, another approach, based on the Gunner Farneback’s algorithm (*dense optical flow*) [34, 35] computes the optical flow for all the points in the frame (Fig. 15, *b*). For vectors (u, v) it is possible to find their magnitude and direction. Thus it allows us to trace the moving object and its movement directions (color shows the direction).

Both Lucas–Kanade and Farneback’s algorithms perform well in case of a static background. They also do not require any manual object selection, the object of interest can be found by its motion. But in case of object occlusion redetection is required, this fact makes these technics suitable mostly only for laboratory conditions, like in this particular case (Fig. 15). They also perform fine in case of object’s shape change.

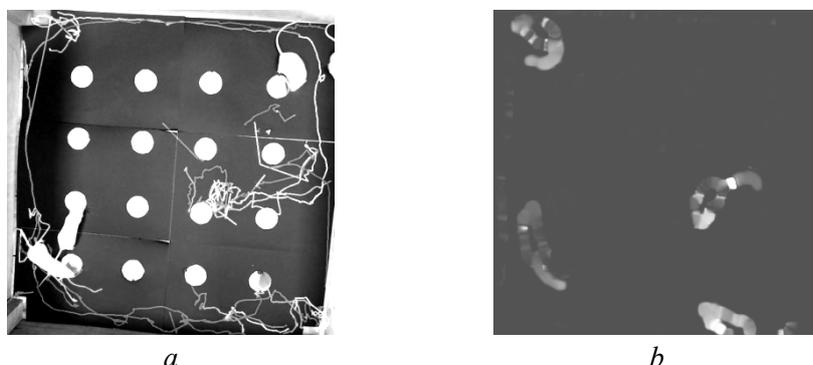


Fig. 15. Optical flow: *a* — Lucas–Kanade method; *b* — Farneback’s method

Point tracking methods. In such approaches, it is assumed that the position of the object is determined by the location of a set of characteristic points. The same object in consecutive frames is represented by sets of corresponding pairs of points. This group of methods is divided into two subgroups:

- Deterministic methods [36] use qualitative heuristics of motion (a small change in velocity, the invariance of the distance in three-dimensional space between a pair of points belonging to object), in essence, the task is reduced to minimizing the function correspondence of sets of points. Methods based on the calculation of dense and sparse optical flux, as well as methods of matching key point descriptors are typical representatives of deterministic methods.
- Probabilistic methods use an approach based on the concept of state space. It is believed that a moving object has a certain internal state, which is measured on to each frame. To estimate the next state of the object, it is necessary to generalize as much as possible the received measurements, that is, to determine the new state provided that the set is obtained measurements for states on previous frames. Typical examples of such methods are methods based on the Kalman filter [37, 38] or Particle filter [39].

The Kalman filter is used to track single objects in noisy images. Each state of the system can be described by a vector of its parameters. By some influence the system passes from one state to another. The set of all states of the system and transitions form a model. There is a concept of observation data vector. This is

a set of system parameters that we can extract from the observation of behavior of the system. In most cases, the dimension of the vector states of the system exceeds the dimension of the observation data vector. In this case, the Kalman filter is able to estimate with a certain probability the complete internal state of the system.

The Kalman filter works with time-discrete linear dynamical systems. Such systems are modeled by Markov chains with the help of linear operators and terms with normal distribution. At each discrete moment of time, the linear operator acts on the state and translates it into another state, adding some random variable in the form of normal noise and, in the general case, a control vector that simulates the influence of the control signal. Mathematical model of this process in matrix form:

$$x_k = F_k x_{k-1} + B_k u_k + w_k,$$

$$z_k = H_k x_k + v_k,$$

where F_k is a $n \times n$ matrix that describes how the state changes systems in transition from $k-1$ to k without control; B_k is a $n \times l$ matrix that describes how the control effect u_k changes state from $k-1$ to k , l is the dimension of the control effect; H_k is a $c \times n$ matrix that describes how the state x_k is transformed into an observation z_k , c is the dimension of the observation vector; w_k , v_k are arbitrary values representing the normally distributed noise when measuring the state c by the corresponding covariance matrices $Q_k, R_k, w_k \sim N(0, Q_k)$, $v_k \sim N(0, R_k)$.

The algorithm consists of two repeating phases: extrapolation phase and correction phase. During the operation of the first phase, a prediction of values of the state variables takes place (extrapolation) based on state estimation on the previous step, as well as their uncertainty. This assessment often also called *a priori* because it is given to perform any measurements and is based on mathematical model only. The second phase is responsible for refining the result of extrapolation using the appropriate measurements, possibly obtained with some error. This assessment is called *a posteriori*.

In the classical operation of the algorithm, these phases alternate, i.e. the prediction happens in relation to the results of adjustment with past iteration, and the adjustment specifies the result of the extrapolation phase. However, in some cases, the correction phase may be missed and the prediction will be based on an unspecified estimate. This situation can occur if for some reason we do not have information from the measuring sensors at this stage. To understand further processes, it is necessary to enter the following notation:

- x_k — the actual state of the system at the time k ;
- \hat{x}_k — estimated state at time k ;
- \hat{x}_k^- — predicted system state at time k ;
- P_k — estimated matrix of error covariance of condition measurement;
- P_k^- — predicted matrix of error covariance of condition measurement.

Table 1. Kalman Filter algorithm

Extrapolation		Correction
1) State extrapolation: $\hat{x}_k^- = F_k \hat{x}_{k-1}^- + B_k u_k$	→	1) Kalman amplification: $K_k = P_k^- H_k^T (H_k P_k^- H_k^T + R_k)^{-1}$
2) Covariance matrix extrapolation: $P_k^- = F_k P_{k-1}^- F_k^T + Q_k$	←	2) State vector correction: $\hat{x}_k = \hat{x}_k^- + K_k (z_k - H_k \hat{x}_k^-)$
		3) Covariance matrix calculation: $P_k = (I - K_k H_k) P_k^-$

Interactive tracking. The general idea of the set of suggested approaches is the motion and appearance models [40]. As one remembers, the task of tracking is to detect an object in the current frame given this object was successfully detected and tracked in all (nearly all) previous frames. If the object was tracked up until current frame, it means that it has been moving, i.e. the parameters of the motion model are known. This term means that object’s location and the velocity (speed and motion direction) in previous frames are also known. If there is no other information on the object, it can be possible to estimate its new location based on the currently existing motion model and thus one can get close to the object real position.

Thus if the object is simple and its appearance did not change too much, it is possible to use some simple template as an appearance model and look for it. But as the object appearance can change pretty much, the model can be represented as a classifier that is trained during the whole tracking process. The main task for the classifier is to classify a rectangular region of interest (ROI) of an image as either an object or background. In order to do this, it takes as the input an image patch and returns an estimation value in range [0, 1]. This value is the probability that the image patch contains the object. As one can see here the binary classification is used, thus if the estimation score is 0, it means that the classifier thinks that the image patch is the background, and if the score is 1, it says that the patch is the object. The training (learning) is performed during the tracking process, as the classifier “learns” to detect the object. This approach is similar to the work of the neural networks, but in this particular case the training set is quite small, as it is just the set of video frames.

There are several interactive training methods [40–44], that uses this methodology. First group includes BOOSTING, MIL, KCF trackers. The BOOSTING tracker is based on the AdaBoost algorithm and uses HAAR cascade based face detector. The user should provide the initial bounding box, that is used as a positive example for the object and many other image patches outside this box are treated as the background. Also, this algorithm cannot detect the tracking failure. MIL (Multiple Instance Learning) is based on the same idea, but instead of considering only the current location of the object as a positive example, it looks in a small neighborhood around the current location to estimate several potential positive examples. The KCF (Kernelized Correlation Filters) tracker also supports the ideas from BOOSTING and MIL. The difference is, that this tracker uses the fact that the multiple positive samples used in the MIL tracker have large overlapping regions. The fact of overlapping is used for performance enhancement. This method reports a tracking failure and can recover from partial occlusion.

The experiment had shown that both BOOSTING (Fig. 16) and MIL (Fig. 17, *a*) tracker had shown similar performance and tracking quality. The main problem was that in current condition they began failing and losing the objects (Fig. 17, *b*). The KCF tracker indeed had shown much faster frame processing due to its technic of usage of the overlapping regions. But it also resulted in much worse tracker quality – the tracker tends to loose objects very quickly.



Fig. 16. BOOSTING tracker

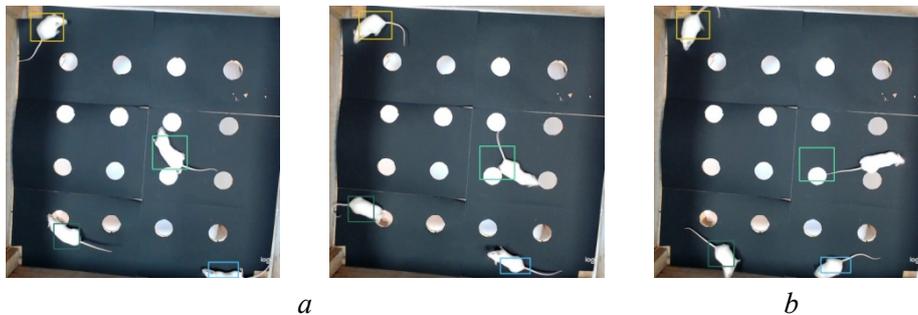


Fig. 17. MIL tracker: *a* — successful tracking; *b* — tracker failure

The other set of trackers include TLD, MEDIANFLOW, MOSSE and CSRT. The TLD (Tracking, learning, and detection) as its name suggests separates the tracking process into three subtasks, i.e. tracking, learning, detection [40]. According to its creators, the algorithm tracks the object from frame to frame, while its detector localizes all object's appearances that have been found so far and performs tracker's self-correction if required. During the learning process the algorithm estimates errors of the tracker's object detector and then updates it in order to avoid them further on. This results in tracker jumping around, which one hand, in case of sudden occlusions allows the tracker to return back to initial object. But on the other hand as result of such jumps quite often TLD tends to lose its target and focus on another object. Thus despite this tracker performs fine under occlusion over multiple frames or scale changes, it provides lots of false positives results, which making it almost unusable. The testing of TLD is shown on Fig. 18.

The MEDIANFLOW tracker [40] follows the object in both forward and backward directions and estimates the divergence between object's two trajectories. Thus it calculates forward-backward error and tries to minimize it. This technic allows to detect tracking failures and keep a more or less stable trajectory. The test has shown (and it matches the earlier results [40]), that this

tracker works well only with predictable and small movements with no occlusions. But in case of lab animals which tend to move unpredictably it fails almost immediately (Fig. 19).



Fig. 18. TLD tracker: *a* — successful tracking; *b* — tracker jumps to other object



Fig. 19. MEDIANFLOW tracker: *a* — successful tracking; *b* — tracker failure due to chaotic mice movements

The final two trackers are MOSSE (Minimum Output Sum of Squared Error) and CSRT (Channel and Spatial Reliability Tracker). The MOSSE tracker is based on the calculation of adaptive correlation, as it produces stable correlation filters when initialized using a single frame. This tracker can operate fast at very high framerates, and it couples fine with lighting, scale, pose changes and non-rigid deformations. But its overall performance is lower than learning-based trackers, for example, like MIL or KCF. The CSRT tracker uses the spatial reliability map for adjusting the filter support to the part of the selected region from the frame for tracking [1]. This allows to resolve situations with enlarging and localization of the selected, thus it can track fine the non-rectangular regions or objects. But in current static background conditions and unpredictable movements it also tends to loose objects (Fig. 20).

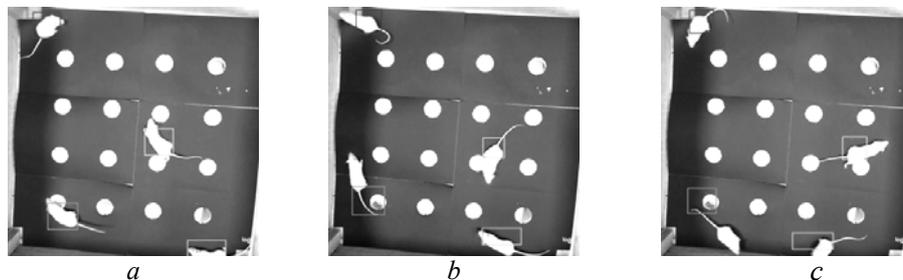


Fig. 20. CSRT tracker: *a* — successful tracking; *b* — tracker begins to lose objects; *c* — tracker failure: object lost

CONCLUSIONS

As the experiment results show, the interactive trackers may perform well under certain conditions, but in this particular case of lab animal tracking none of them can track all objects up to the very end of the video. The main advantage of them is that they usually do not require any additional data, like empty background image and do not need to estimate it (as minor preprocessing, some noise reduction or contrast change can be applied). Also, some of them can resolve minor occlusion situations. But all of these methods require manual object selection and none of them managed to demonstrate any stable work, which can say that they are not completely suitable for the current task in its original form.

The more simple approaches, like background subtraction (with or without background estimation) in case of static location can perform quite well, as they detect the objects automatically and their computational complexity is not high. Kernel tracking approaches, as well as optical flow and point tracking can provide some additional information about object motion, which can be used in combination with the simple/interactive tracking methods for performance improvement and additional motion data acquisition.

Some of object detection methods can also serve as some addition to the tracking methods during the tracking process itself. Different use cases can require different detection technics. For example, in case of lab mice behavior observations with specific environment conditions (a box with the dark floor as a test stand) an automatic detection by color can be applied, but in case of some more complex environment a combination of several detection approaches may be required. The detailed comparison of tested object detection and tracking approaches is presented on Table 2, 3.

The testing results also allow one to assume that for lab mice activity study most likely background subtraction in combination with image segmentation and interactive tracker can be used. Fish tracking may require some interactive tracker combined with optical flow and image segmentation. Our further research will include the usage of the composition of interactive trackers and the simple approaches. The idea is to use the positive sides of both sets of methods to compensate each other's disadvantages. Also, the neural network based tracking is planned to be applied in attempt to create a completely stable tracker.

Table 2. Object detection approaches performance comparison

Method	Advantages	Disadvantages / Features
Image segmentation	Easy to use when the objects of interest significantly differ from the background by some parameter Can be used as an addition to other methods to highlight the main image parts	Low versatility Requires too accurate algorithm parameters setup in each particular case Sensitive to lighting conditions
Template Matching	Useful for scene analysis when camera is static and objects of interest look almost identical Can be used for detection of some products on a factory assembly line	Can fail in case of rotation, scaling or partial occlusions If rotation/scaling take place, additional search steps will be required
Feature-based detection	Invariant to minor turns, scaling of objects and changes in stage lighting Suitable for rough object search	Requires a way of compact characteristic features representation Impossible to define an object as instance of some class Provides false results in case of object dynamic shape change

Continued Table 2

Method	Advantages	Disadvantages / Features
Categorical recognition	HOG highlights well the objects with multiple details Haar approach is more suitable for face detection and recognition	When the object is mostly a single piece without any significant details, HOG will only highlight the borders Both approaches still require a pretrained classifier Fails when objects tend to change shape

Table 3. Object tracking approaches performance comparison

Method	Advantages	Disadvantages / Features
Color-based tracking	Object gets detected automatically Simple realization Good performance speed Works well with objects, which change their shape	Cannot handle occlusions Cannot handle object 'merging' Requires additional algorithms for centroid tracking Requires manual color threshold setup Unstable if colors are too similar
Background subtraction	Same benefits as color tracking Requires only one threshold value instead of a range	Mandatory existence of the background image Other possible problems are similar to those ones from color-based tracking
Background estimation	Useful if no empty background image provided Can be used as addition to Background subtraction	Objects of interest that do not move actively can be classified as background Effective only if the entire background is static
Kernel tracking (Mean Shift/CAM Shift)	Shows fine work speed CAM Shift ROI can adjust its size during the process	The Mean Shift ROI has fixed size CAM Shift is connected to color range CAM Shift is sensitive to lighting conditions and object shape changes
Optical flow	Good work in case of static background The object of interest can be found by its motion automatically Performs fine in case of object's shape change	In case of object occlusion redetection is required Object gets lost when its movements are getting slower Cannot handle 'object merging' problem
Point tracking (Kalman filter)	Good performance when tracking single objects on noisy images	Has complicated computations and implementation Not good at handling object merging/occlusions
Interactive tracking (BOOSTING / MIL / KCF)	Fine work speed (BOOSTING/ MIL) Best work speed (KCF) Suitable for non-static background (moving camera)	KCF tends to loose objects more often than MIL/BOOSTING Not good at handling object merging/occlusions
Interactive tracking (TLD)	Can handle object occlusions/ merging Works with non-static background (moving camera)	Provide too many false positives, tends to loose object of interest
Interactive tracking (MEDIANFLOW / MOSSE / CSRT)	CSRT has fine work speed at high framerates CSRT handles lighting, scale, pose changes Suitable for non-static background (moving camera)	MEDIANFLOW works well only with predictable and small movements with no occlusions Methods works well mostly with predictable object movements Methods cannot handle occlusions

ACKNOWLEDGEMENTS

Special thanks for the research assistance and provided test videos and images of lab animals to Faculty of Biology of Odesa I.I. Mechnikov National University.

REFERENCES

1. A.R. Smith, "Color Gamut Transform Pairs", in *SIGGRAPH '78: Proceedings of the 5th annual conference on Computer graphics and interactive techniques*, pp. 12–19, 1978. doi:10.1145/800248.807361.
2. W.K. Pratt, *Digital Image Processing*; 4th edition. Wiley-Interscience, A John Wiley & Sons, Inc., Publication, 2007, 807 p.
3. X. Zhao, S. Yan, and Q. Gao, "An Algorithm for Tracking Multiple Fish Based on Biological Water Quality Monitoring", *IEEE Access*, vol. 7, pp. 15018–15026, January 2019. doi:10.1109/ACCESS.2019.2895072.
4. J. Delcourt, M. Denoel, M. Ylieff, and P. Poncin, "Video multitracking of fish behaviour: A synthesis and future perspectives", *Fish and Fisheries*, vol. 14, no. 2, pp. 186–204, June 2013. doi:10.1111/j.1467-2979.2012.00462.x.
5. H.E.-D. Mohamed et al. "MSR-YOLO: Method to Enhance Fish Detection and Tracking", *The 11th International Conference on Ambient Systems, Networks and Technologies (ANT), April 6–9, 2020, Warsaw, Poland*.
6. P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik, "Contour Detection and Hierarchical Image Segmentation", *IEEE TPAMI*, vol. 33, no. 5, pp. 898–916, 2011. doi: 10.1109/TPAMI.2010.161.
7. S. Beucher, "The Watershed Transformation Applied to Image Segmentation", *Scanning microscopy*, vol. 6, July 2000, 26 p.
8. R. Brunelli, *Template Matching Techniques in Computer Vision. Theory and Practice*. Wiley, 2009, 339 p.
9. Dr. A.S. Khedher, Dr. A.M. Alkababji, and O. Hadi, "Improving the Reliability of Object Recognition Based On Template Matching", *AL Rafdain Engineering Journal, Computer Science*, pp. 81–88, 2015.
10. *Template Matching*. Available: https://docs.opencv.org/3.4/de/da9tutorial_template_matching.html
11. D.G. Lowe, "Object recognition from local scale-invariant features", *International Conference on Computer Vision – ICCV, 1999*, vol. 2, pp. 1150–1157.
12. H. Kong, H.C. Akakin, and S.E. Sarma, "A Generalized Laplacian of Gaussian Filter for Blob Detection and Its Applications", *IEEE Transactions on Cybernetics*, vol. 43, no. 6, pp. 1719–1733, January 2013. doi: 10.1109/TSMCB.2012.2228639.
13. L. Assirati, N.R. Silva, L. Berton, A.A. Lopes, and O.M. Bruno, "Performing edge detection by Difference of Gaussians using q-Gaussian kernels", *2nd International Conference on Mathematical Modeling in Physical Sciences 2013, Journal of Physics, Conference Series 490(2014) 012020*, IOP Publishing, 2014, 4 p. doi: 10.1088/1742-6596/490/1/012020.
14. T. Lindeberg, "Scale Invariant Feature Transform", *Scholarpedia*, vol. 7, no. 5: 10491, 2012, 17 p. doi: 10.4249/scholarpedia.10491.
15. H. Bay, T. Tuytelaars, and L.V. Gool, "SURF: Speeded up robust features", in *Proceedings of the 9th European conference on Computer Vision*, vol. 1, 2006, 14 p. doi: 10.1007/11744023_32.
16. N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection", *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '05), San Diego, United States*, vol. 1, pp. 886–893, June 2005. doi: 0.1109/CVPR.2005.177.
17. T. Kobayashi, A. Hidaka, and T. Kurita, "Selection of Histograms of Oriented Gradients Features for Pedestrian Detection", *Neural Information Processing, 14th International Conference, ICONIP 2007, Kitakyushu, Japan, November 13–16, 2007, Revised Selected Papers, Part II*, pp. 598–607. doi: 10.1007/978-3-540-69162-4_62.
18. P. Viola and M. Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features", *IEEE Conference on Computer Vision Pattern Recognition*, vol. 1, 2001. doi: I-511.10.1109/CVPR.2001.990517.
19. H. Hunud, A. Kadouf, and Y.M. Mustafah, "Colour-based Object Detection and Tracking for Autonomous Quadrotor UAV", *2013 IOP Conf. Ser.: Mater. Sci. Eng.*, vol. 53, 2013, 9 p. doi:10.1088/1757-899X/53/1/012086.

20. V.V. Moroz and M.A. Shvandt, "Study of movement and behavior of laboratory animals by methods of object detection and tracking", *Herald of the National Technical University "KhPI", Series of "Informatics and Modeling"*, Kharkiv: NTU "KhPI", Kharkiv, vol. 13, no. 1338, pp. 93–103, 2019. doi: 10.20998/2411-0558.2019.13.09.
21. A.M. Raid, W.M. Khedr, M.A. El-dosuky, and Mona Aoud, "Image restoration based on morphological operations", *International Journal of Computer Science, Engineering and Information Technology (IJCEIT)*, vol. 4, no. 3, 2014, pp. 9–21. doi: 10.5121/ijceit.2014.4302.
22. S. Ravi and A.M. Khan, "Morphological Operations for Image Processing: Understanding and its Applications", in *NCVSComs-13 Conference Proceedings*, 2015, pp. 17–19.
23. Y. Zhang, "Pathological Brain Detection based on wavelet entropy and Hu moment invariants", *Bio-Medical Materials and Engineering*, no. 26, pp. 1283–1290, 2015.
24. P. Joshi, D.M. Escrivá, and V. Godoy, *OpenCV By Example*. Birmingham: Packt Publishing Ltd, 2016, 297 p.
25. M. Nixon and A. Aguado, *Feature Extraction & Image Processing for Computer Vision*; 3d ed. London: Elsevier Ltd, 2012, 623 p.
26. S.-C.S. Cheung and C. Kamath, "Robust Background Subtraction with Foreground Validation for Urban Traffic Video", *EURASIP Journal on Applied Signal Processing*, vol. 14, pp. 2330–2340, Hindawi Publishing Corporation, 2005.
27. M.A. Alawi, O.O. Khalifa, and M.D.R. Islam, "Performance Comparison of Background Estimation Algorithms for Detecting Moving Vehicle", *World Applied Sciences Journal 21 (Mathematical Applications in Engineering)*, IDOSI Publications, 2013, pp. 109–114. doi: 10.5829/idosi.wasj.2013.21.mae.99934.
28. S. Mallick, *Simple Background Estimation in Videos using OpenCV (C++/Python)*. 2019. Available: <https://learnopencv.com/simple-background-estimation-in-videos-using-opencv-c-python/>.
29. Y. Cheng, "Mean Shift, Mode Seeking, and Clustering", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, no. 8, pp. 790–799, August 1995.
30. D. Comaniciu and P. Meer, "Mean Shift: A Robust Approach Toward Feature Space Analysis", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, pp. 603–619, May 2002. doi: 10.1109/34.1000236.
31. G. Bradski, "Computer Vision Face Tracking For Use in a Perceptual User Interface", *Archived 2012-04-17 at the Wayback Machine, Intel Technology Journal*, no. Q2, 1998.
32. B. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision", in *Proceedings of the 7th International Joint Conference on Artificial Intelligence (IJCAI '81)*, pp. 674–679.
33. A. Radgui, C. Demonceaux, E. Mouaddib, D. Aboutajdine, and M. Rziza, "An adapted Lucas–Kanade's method for optical flow estimation in catadioptric images", *The 8th Workshop on Omnidirectional Vision, Camera Networks and Non-classical Cameras - OMNIVIS, Marseille, France, Marseille, 2008*, 12 p.
34. *Optical Flow*. Available: https://docs.opencv.org/3.4/d4/dee/tutorial_optical_flow.html.
35. G. Farneback, "Two-Frame Motion Estimation Based on Polynomial Expansion", *Lecture Notes in Computer Science*, 8 p., 2003.
36. C. Veenman, M. Reinders, and E. Backer, "Resolving motion correspondence for densely moving points", *IEEE Transactions on Pattern Analysis Machine Intelligence*, vol. 23, no. 1, pp. 54–72, 2001. doi: 10.1109/34.899946.
37. A. Salarpour, A. Salarpour, M. Fathi, and M.H. Dezfoulian, "Vehicle tracking using Kalman filter and features", *Signal and Image Processing: An International Journal (SIPIJ)*, vol. 2, no. 2, pp. 45–67, 2011. doi: 10.5121/sipij.2011.2201.
38. S. Dan, Zh. Baojun, and T. Linbo, "A Tracking Algorithm Based on SIFT and Kalman Filter", in *Proceedings The 2nd International Conference on Computer Application and System Modeling*, 2012, pp. 1563–1566.
39. F. Gunnarsson, N. Bergman, U. Forssell, J. Jansson, R. Karlsson, and P.J. Nordlund, "Particle Filters for Positioning, Navigation and Tracking", *IEEE Transactions on Signal Processing*, vol. 2, no. 2, pp. 425–437, 2002.

40. S. Mallick, *Object Tracking using OpenCV (C++/Python)*. 2017. Available: <https://learnopencv.com/object-tracking-using-opencv-cpp-python/>.
41. Z. Kalal, K. Mikolajczyk, and J. Matas, "Forward-Backward Error: Automatic Detection of Tracking Failures", in *Proceedings of International Conference on Pattern Recognition, 23-26 August, 2010, Istanbul, Turkey*, 4 p. doi: 10.1109/ICPR.2010.675.
42. Z. Kalal, K. Mikolajczyk, and J. Matas, "Tracking-Learning-Detection", in *Proceedings of IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, no. 1, 2010, 14 p.
43. S. Zhou, Y. Peng, K. Gong, and L. Shu, "An Improved TLD Tracking Algorithm for Fast-moving Object", *International Conference on Computer Science, Electronics and Communication Engineering (CSECE 2018), Advances in Computer Science Research*, vol. 80, pp. 69–73.
44. D.S. Bolme, J.R. Beveridge, B.A. Draper, and Y.M. Lui, "Visual Object Tracking using Adaptive Correlation Filters", *CVPR*, 2010, 10 p. doi: 10.1109/CVPR.2010.5539960.

Received 07.07.2021

INFORMATION ON THE ARTICLE

Maksym A. Shvandt, ORCID: 0000-0002-4580-3961, Odesa I.I. Mechnikov National University, Ukraine, e-mail: maxim.shvandt@gmail.com

Volodymyr V. Moroz, ORCID: 0000-0002-3240-4590, Odesa I.I. Mechnikov National University, Ukraine, e-mail: v.moroz@onu.edu.ua

ОГЛЯД МЕТОДІВ ВИЯВЛЕННЯ ТА ВІДСТЕЖЕННЯ ЛАБОРАТОРНИХ ТВАРИН / М.А. Швандт, В.В. Мороз

Анотація. Подано огляд та аналіз кількох найпоширеніших методів та алгоритмів виявлення і відстеження об'єктів. Окремий випадок використання техніки відстеження об'єктів може виникнути під час лабораторного дослідження поведінки тварин. Різні експериментальні умови та необхідність збирання певних корисних даних можуть потребувати спеціальних методів відстеження. Тому розглянуто набір загальних підходів до відстеження об'єктів, а їх функціональність та можливості перевірено в реальному експерименті. Наведено їх основу та базові аспекти. Експеримент продемонстрував переваги та недоліки досліджуваних методів. Зроблено висновки та рекомендації щодо випадків їх використання.

Ключові слова: відстеження (трекінг) об'єктів, детектування об'єктів, алгоритм, відео, кадр, зображення, задній план, передній план, експеримент, кольоровий простір, порогове значення, обчислення заднього плану, сегментація.

ОБЗОР МЕТОДОВ ОБНАРУЖЕНИЯ И ОТСЛЕЖИВАНИЯ ЛАБОРАТОРНЫХ ЖИВОТНЫХ / М.А. Швандт, В.В. Мороз

Аннотация. Представлены обзор и анализ нескольких распространенных методов и алгоритмов обнаружения и отслеживания объектов. Частный случай использования методики отслеживания объектов может возникнуть во время лабораторного исследования поведения животных. Различные экспериментальные условия и необходимость сбора определенных полезных данных могут потребовать специальных методов отслеживания. Поэтому рассмотрен набор общих подходов к отслеживанию объектов, а их функциональность и возможности проверены в ходе реального эксперимента. Представлены их основа и базовые аспекты. Эксперимент продемонстрировал преимущества и недостатки исследуемых методов. Сделаны выводы и рекомендации по поводу случаев их использования.

Ключевые слова: отслеживание (трекинг) объектов, обнаружение объектов, алгоритм, видео, кадр, изображение, задний план, передний план, эксперимент, цветовое пространство, пороговое значение, вычисление заднего плана, сегментация.

ВІДОМОСТІ ПРО АВТОРІВ

Андросов Дмитро Васильович,

магістр ПСА КПІ ім. Ігоря Сікорського, Україна, Київ

Бідюк Петро Іванович,

професор, доктор технічних наук, професор кафедри математичних методів системного аналізу ПСА КПІ ім. Ігоря Сікорського, Україна, Київ

Бодянський Євгеній Володимирович,

професор, доктор технічних наук, професор кафедри штучного інтелекту Харківського національного університету радіоелектроніки, Україна, Харків

Гавриленко Валерій Володимирович,

професор, доктор фізико-математичних наук, завідувач кафедри інформаційних систем і технологій факультету транспортних та інформаційних технологій Національного транспортного університету, Україна, Київ

Гамідов Галіб,

доктор філософії, керівник департаменту інформаційних технологій «Azerishiq», Баку, Азербайджан

Зайченко Юрій Петрович,

професор, доктор технічних наук, професор кафедри математичних методів системного аналізу ПСА КПІ ім. Ігоря Сікорського, Україна, Київ

Зайченко Олена Юрївна,

доцент, доктор технічних наук, професор кафедри математичних методів системного аналізу ПСА КПІ ім. Ігоря Сікорського, Україна, Київ

Кирилюк Віталій Семенович,

старший науковий співробітник, доктор фізико-математичних наук, провідний науковий співробітник відділу теорії коливань Інституту механіки імені С.П.Тимошенка НАН України, Київ

Костюк Сергій Олександрович,

аспірант кафедри штучного інтелекту Харківського національного університету радіоелектроніки, Україна, Харків

Кулішова Нонна Євгенівна,

доцент, кандидат технічних наук, професор кафедри медіасистем та технологій Харківського національного університету радіоелектроніки, Україна, Харків

Левчук Ольга Іванівна,

старший науковий співробітник, кандидат фізико-математичних наук, старший науковий співробітник відділу теорії коливань Інституту механіки імені С.П.Тимошенка НАН України, Київ

Мартиненко Андрій Анатолійович,

старший викладач кафедри програмного забезпечення комп'ютерних систем Національного технічного університету «Дніпровська політехніка», Україна, Дніпро

Мацукі Йошіо,

професор, доктор технічних наук, професор Національного університету «Києво-Могилянська академія», Україна, Київ

Мороз Борис Іванович,

професор, доктор технічних наук, член-кореспондент Академії прикладної електроніки, професор кафедри програмного забезпечення комп'ютерних систем Національного технічного університету «Дніпровська політехніка», Україна, Дніпро

Мороз Володимир Володимирович,

доцент, кандидат технічних наук, професор факультету математики, фізики та інформаційних технологій Одеського національного університету імені І.І. Мечникова, Україна, Одеса

Недашківська Надія Іванівна,

доцент, доктор технічних наук, доцент кафедри математичних методів системного аналізу ІПСА КПІ ім. Ігоря Сікорського, Україна, Київ

Панібратов Роман Сергійович,

магістр ІПСА КПІ ім. Ігоря Сікорського, Україна, Київ

Панкратов Володимир Андрійович,

кандидат технічних наук, молодший наукович співробітник 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

Тевяшев Андрій Дмитрович,

професор, доктор технічних наук, завідувач кафедри прикладної математики Харківського національного університету радіоелектроніки, Україна, Харків

Швандт Максим Альбертович,

аспірант факультету математики, фізики та інформаційних технологій Одеського національного університету імені І.І. Мечникова, Україна, Одеса