



ISSN 2079-0023 (Print)
ISSN 2410-2857 (Online)

ВІСНИК

Національного технічного університету «ХПІ».
Серія: Системний аналіз, управління
та інформаційні технології

№ 1 (9) 2023

BULLETIN

of the National Technical University "KhPI".
Series: System analysis, control
and information technology

No. 1 (9) 2023

Харків Kharkiv

МІНІСТЕРСТВО ОСВІТИ
І НАУКИ УКРАЇНИ

Національний технічний університет
«Харківський політехнічний інститут»

MINISTRY OF EDUCATION
AND SCIENCE OF UKRAINE

National Technical University
"Kharkiv Polytechnic Institute"

**Вісник Національного
технічного університету
«ХПІ». Серія: Системний
аналіз, управління та
інформаційні технології**

№ 1 (9) 2023

Збірник наукових праць

Видання засноване у 1961 р.

**Bulletin of the National
Technical University
"KhPI". Series: System
analysis, control and
information technology**

No. 1 (9) 2023

Collection of Scientific papers

The edition was founded in 1961

Харків
НТУ «ХПІ», 2023

Kharkiv
NTU "KhPI", 2023

Вісник Національного технічного університету «ХПІ». Серія: Системний аналіз, управління та інформаційні технології = Bulletin of the National Technical University "KhPI". Series: System analysis, control and information technology : зб. наук. пр. / Нац. техн. ун-т «Харків. політехн. ін-т». — Харків : НТУ «ХПІ», 2023. — № 1 (9) 2023. — 118 с. — ISSN 2079-0023.

Видання публікує нові наукові результати в області системного аналізу та управління складними системами, отримані на основі сучасних прикладних математичних методів і прогресивних інформаційних технологій. Публікуються роботи, пов'язані зі штучним інтелектом, аналізом великих даних, сучасними методами високопродуктивних обчислень у системах підтримки прийняття рішень.

Издание публикует новые научные результаты в области системного анализа и управления сложными системами, полученные на основе современных прикладных математических методов и прогрессивных информационных технологий. Публикуются работы, связанные с искусственным интеллектом, анализом больших данных, современными методами высокопродуктивных вычислений в системах поддержки принятия решений.

Для научных работников, преподавателей высшей школы, аспирантов, студентов и специалистов в области системного анализа, управления и компьютерных технологий.

Edition publishes new scientific results in the field of system analysis and control of complex systems, based on the application of modern mathematical methods and advanced information technology. Works related to artificial intelligence, big data analysis and modern methods of high-performance computing in decision support systems are publishing.

For scientists, teachers of higher education, post-graduate students, students and specialists in the field of systems analysis, management and computer technology.

Свідоцтво Міністерства юстиції України
КВ № 23778-13618Р від 14 лютого 2019 р.

Мова статей – українська, російська, англійська.

Наказом МОН України № 1643 від 28 грудня 2019 року «Про затвердження рішень Атестаційної колегії Міністерства щодо діяльності спеціалізованих вчених рад від 18 грудня 2019 року» «Вісник Національного технічного університету "ХПІ". Серія: Системний аналіз, управління та інформаційні технології» внесено до категорії Б «Переліку наукових фахових видань України, в яких можуть публікуватися результати дисертаційних робіт на здобуття наукових ступенів доктора і кандидата наук».

Вісник Національного технічного університету «ХПІ». Серія: Системний аналіз, управління та інформаційні технології включений до зовнішніх інформаційних систем, у тому числі в наукометричну базу даних Index Copernicus (Польща), бібліографічну базу даних OCLC WorldCat (США), індексується пошуковими системами Google Scholar і Crossref; зареєстрований у світовому каталозі періодичних видань бази даних Ulrich's Periodicals Directory (New Jersey, USA).

Офіційний сайт видання: <http://samit.khpi.edu.ua/>

Засновник
Національний технічний університет
«Харківський політехнічний інститут»

Founder
National Technical University
"Kharkiv Polytechnic Institute"

Редакційна колегія

Головний редактор:

Годлевський М. Д., проф., НТУ «ХПІ», Україна

Заступник головного редактора

Кущенко О. С., проф., НТУ «ХПІ», Україна

Члени редколегії:

Ахієзер О. Б., проф., НТУ «ХПІ», Україна

Безменов М. І., проф., НТУ «ХПІ», Україна

Бентаб Ф., доц., Ліонський університет-2, Франція

Богомолів С., доц., Австралійський національний

університет, Австралія

Галуза О. А., проф., НТУ «ХПІ», Україна

Дорофєєв Ю. І., проф., НТУ «ХПІ», Україна

Керстен В., проф., Гамбурзький технологічний університет,

Німеччина

Любчик Л. М., проф., НТУ «ХПІ», Україна

Москаленко В. В., проф., НТУ «ХПІ», Україна

Павлов О. А., проф., НТУУ «КПІ», Україна

Раскін Л. Г., проф., НТУ «ХПІ», Україна

Ткачук М. В., проф., ХНУ ім. В. Н. Каразіна, Україна

Хайрова Н. Ф., проф., НТУ «ХПІ», Україна

Чередніченко О. Ю., проф., НТУ «ХПІ», Україна

Шаронова Н. В., проф., НТУ «ХПІ», Україна

Відповідальний секретар

Безменов М. І., проф., НТУ «ХПІ», Україна

Editorial

Editor-in-chief

Godlevskyi M. D., prof., NTU "KhPI", Ukraine

Deputy editor-in-chief

Kutsenko O. S., prof., NTU "KhPI", Ukraine

Editorial staff members:

Akhiezer O. B., prof., NTU "KhPI", Ukraine

Bezmenov M. I., prof., NTU "KhPI", Ukraine

Bentayeb F., Associate Professor, University of Lyon-2, France

Bogomolov S., Assistant Professor, Australian National

University, Australia

Galuzha O. A., prof., NTU "KhPI", Ukraine

Dorofiev Yu. I., prof., NTU "KhPI", Ukraine

Kersten Wolfgang, Prof., Hamburg University of Technology,

Germany

Lyubchik L. M., prof., NTU "KhPI", Ukraine

Moskalenko V. V., prof., NTU "KhPI", Ukraine

Pavlov O. A., prof., NTUU "KPI", Ukraine

Raskin L. G., prof., NTU "KhPI", Ukraine

Tkachuk M. V., prof., V. N. Karazin KhNU, Ukraine

Khairova N. F., prof., NTU "KhPI", Ukraine

Cherednichenko O. O., prof., NTU "KhPI", Ukraine

Sharonova N. V., prof., NTU "KhPI", Ukraine

Executive secretary:

Bezmenov M. I., prof., NTU "KhPI", Ukraine

Рекомендовано до друку Вченою радою НТУ «ХПІ».
Протокол № 6 від 7 липня 2023 р.

СИСТЕМНИЙ АНАЛІЗ І ТЕОРІЯ ПРИЙНЯТТЯ РІШЕНЬ

SYSTEM ANALYSIS AND DECISION-MAKING THEORY

DOI: 10.20998/2079-0023.2023.01.01

UDC 004:519.24:681.3.06

A. A. PAVLOV, Doctor of Technical Sciences, Full Professor, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine, Professor of Informatics and Software Engineering Department; e-mail: pavlov.fiot@gmail.com; ORCID: <https://orcid.org/0000-0002-6524-6410>

M. N. HOLOVCHENKO, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine, Senior Lecturer of Informatics and Software Engineering Department; e-mail: ma4ete25@ukr.net; ORCID: <https://orcid.org/0000-0002-9575-8046>

V. V. DROZD, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Bachelor of Informatics and Software Engineering Department, Kyiv, Ukraine, e-mail: drozdllera@gmail.com, ORCID: <https://orcid.org/0000-0003-0418-1139>

EFFICIENCY SUBSTANTIATION FOR A SYNTHETICAL METHOD OF CONSTRUCTING A MULTIVARIATE POLYNOMIAL REGRESSION GIVEN BY A REDUNDANT REPRESENTATION

In recent years, the authors in their publications have developed two different approaches to the construction of a multivariate polynomial (in particular, linear) regressions given by a redundant representation. The first approach allowed us to reduce estimation of coefficients for nonlinear terms of a multivariate polynomial regression to construction of a sequence of univariate polynomial regressions and solution of corresponding nondegenerate systems of linear equations. The second approach was implemented using an example of a multivariate linear regression given by a redundant representation and led to the creation of a method the authors called a modified group method of data handling (GMDH), as it is a modification of the well-known heuristic self-organization method of GMDH (the author of GMDH is an Academician of the National Academy of Sciences of Ukraine O. G. Ivakhnenko). The modification takes into account that giving a multivariate linear regression by redundant representation allows for construction of a set of partial representations, one of which has the structure of the desired regression, to use not a multilevel selection algorithm, but an efficient algorithm for splitting the coefficients of the multivariate linear regression into two classes. As in the classic GMDH, the solution is found using a test sequence of data. This method is easily extended to the case of a multivariate polynomial regression since the unknown coefficients appear in the multivariate polynomial regression in a linear way. Each of the two approaches has its advantages and disadvantages. The obvious next step is to combine both approaches into one. This has led to the creation of a synthetic method that implements the advantages of both approaches, partially compensating for their disadvantages. This paper presents the aggregated algorithmic structure of the synthetic method, the theoretical properties of partial cases and, as a result, the justification of its overall efficiency.

Keywords: univariate polynomial regression, multivariate polynomial regression, redundant representation, least squares method, test sequence, repeated experiment.

Introduction. Multivariate linear and non-linear regressions, constructed based on the results of active or passive experiments, are widely used in modern diagnostic information systems, in particular, medical ones, and in information management systems with a wide range of applications [1–10]. Universal methods for multivariate regressions construction vary from classical statistical methods to heuristics, such as the group method of data handling (GMDH) or genetic algorithms. But none of them, due to the complexity of the problem, dominates the others. Existing methods complement each other. Therefore, scientific research in this field is still relevant.

We give in the abstract, at a qualitative level, the characteristics of a synthetic method of constructing a multivariate polynomial regression (MPR) given by a redundant

representation. In this paper, we explain the aggregated algorithmic structure of the synthetic method, substantiate the logic of its construction, its theoretical properties, the properties of partial cases of redundant representations of MPRs that lead to the finding of coefficient estimates for nonlinear members of the MPR with acceptable accuracy.

1. General theoretical provisions that we use.

1.1. Univariate polynomial regression (UPR) [11]. Let a UPR be given in the form:

$$Y(x) = \theta_0 + \theta_1 x + \dots + \theta_r x^r + E, \quad (1)$$

where E is a random variable with an arbitrary distribution, its mathematical expectation $ME = 0$, its variance $DE = \sigma^2 < \infty$, the variance or its upper bound is known.

© Pavlov A. A., Holovchenko M. N., Drozd V. V., 2023



Research Article: This article was published by the publishing house of NTU "KhPI" in the collection "Bulletin of the National Technical University "KhPI" Series: System analysis, management and information technologies." This article is distributed under a Creative Commons [Creative Commons Attribution \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/). **Conflict of Interest:** The author/s declared no conflict of interest.



According to the results of the experiment $(x_i \rightarrow y_i, i = \overline{1, n})$, $r < n$, $x_i \in [c, d]$, where

$$y_i = \theta_0 + \theta_1 x_i + \dots + \theta_r x_i^r + \delta_i, \quad (2)$$

where δ_i is an implementation of a random variable E in the i -th test, estimates of unknown coefficients can be found using normalized orthogonal polynomials of Forsythe (NOPFs) constructed from the values of the input variable $x_i, i = \overline{1, n}$.

$$Q_j(x) = q_{j0} + q_{j1}x + \dots + q_{jj}x^j, \quad j = \overline{0, r}.$$

In this case, the UPR (1) has the form

$$Y(x) = \sum_{j=0}^r \omega_j Q_j(x) + E, \quad (3)$$

$$\hat{\omega}_j = \sum_{i=1}^n y_i Q_j(x_i), \quad j = \overline{0, r}, \quad (4)$$

$$\hat{\theta}_j = \hat{\omega}_j q_{rj} + \dots + \hat{\omega}_j q_{jj}, \quad j = \overline{0, r}, \quad (5)$$

$$M \hat{\omega}_j = \omega_j, D \hat{\omega}_j = \sigma^2, M \hat{\theta}_j = \theta_j, D \hat{\theta}_j = \sigma^2 \sum_{j=1}^r q_{jj}^2. \quad (6)$$

1.2. *A repeated experiment.* A. Pavlov and D. Kovalenko proved that the results of a repeated experiment with k repetitions of the main experiment $(x_i \rightarrow y_i, i = \overline{1, n})$ are equivalent to the experiment

$$\left(x_i \rightarrow \frac{\sum_{p=0}^{k-1} y_{p,n+i}}{k}, i = \overline{1, n} \right),$$

and the variances of estimates $\hat{\theta}_j, j = \overline{0, r}$, decrease by the factor of k .

1.3. *Estimation of coefficients for nonlinear members of a UPR using a single set of NOPFs.* It was shown in [12] that the estimates of the coefficients for nonlinear terms of a UPR (1) at arbitrary values of $c < d$ can be found using only a single set of NOPFs based on the results of the following virtual experiment. We introduce a virtual deterministic scalar variable z and for its values

$$z_1 < z_2 < \dots < z_n \quad (7)$$

we find with the specified accuracy a set of NOPFs $Q_j(z), j = \overline{0, r}$. Having put

$$a = \frac{d-c}{z_n - z_1} > 0, b = c - \frac{d-c}{z_n - z_1} z_1, \quad (8)$$

$$x_i = az_i + b, i = \overline{1, n} (x_1 = c, x_n = d),$$

substituting $x = az + b$, we reduce the UPR (1) to a virtual UPR

$$Y(z) = \theta_0 + \theta_1 (az + b) + \dots + \theta_r (az + b)^r + E = \gamma_0 + \gamma_1 z + \dots + \gamma_r z^r + E. \quad (9)$$

The coefficients $\theta_j, \gamma_j, j = \overline{0, r}$, are in a mutually unambiguous correspondence

$$\begin{pmatrix} 1 \\ a & a_{ij} \\ a^2 & & \ddots \\ 0 & & & a^r \end{pmatrix} \begin{pmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_r \end{pmatrix} = \begin{pmatrix} \gamma_0 \\ \gamma_1 \\ \vdots \\ \gamma_r \end{pmatrix}. \quad (10)$$

According to results of the main experiment $(x_i \rightarrow y_i, i = \overline{1, n})$, where $x_i, i = \overline{1, n}$, satisfy (8), we design a virtual experiment $(z_i \rightarrow y_i, i = \overline{1, n})$. Based on the set of NOPFs constructed for $z_i, i = \overline{1, n}$, we find estimates $\hat{\gamma}_j, j = \overline{0, r}$, and based on the system of equations (10) we find estimates $\hat{\theta}_j, j = \overline{0, r}$. It is shown in [12] that

$$\sum_{i=1}^n \left(y_i - \sum_{j=0}^r \hat{\theta}_j x_i^j \right)^2 = \min_{\theta_j, j=0, r} \sum_{i=1}^n \left(y_i - \sum_{j=0}^r \theta_j x_i^j \right)^2.$$

2. Estimation with acceptable accuracy of the coefficients for nonlinear members of a UPR (1) using a single set of NOPFs.

2.1. *Justification of the number of tests n of the main experiment and the choice of values of $z_i, i = \overline{1, n}$, of the virtual scalar variable z .* With the appropriate selection of the values of the virtual scalar variable z , expression (10) significantly simplifies the preliminary analysis and the input data formation for the main active experiment to obtain, with acceptable accuracy, estimates of the coefficients for nonlinear terms of the UPR and, as will be shown later, an MPR. Such a requirement is for $r_{\max} < n$ a compromise between the number of tests n of the main experiment and the value of variances of the estimates of the coefficients for nonlinear members of the virtual UPR (9). As a result of the analysis of the conducted experiments, the following compromise solution is proposed for $r_{\max} \leq 5$: $n = 10, z_1 = -50, z_{10} = 50, \Delta z = (z_i - z_{i-1}) = \text{const}$. In this case,

$$D \hat{\gamma}_2 = 4.26 \cdot 10^{-6} \sigma^2, D \hat{\gamma}_3 = 7.55 \cdot 10^{-9} \sigma^2, \quad (11)$$

$$D \hat{\gamma}_4 = 1.4 \cdot 10^{-12} \sigma^2, D \hat{\gamma}_5 = 1.28 \cdot 10^{-15} \sigma^2,$$

that is, with an increase of j by one, $D \hat{\gamma}_j$ is decreased by three orders of magnitude, starting from $D \hat{\gamma}_2$.

2.2. *Justification of conditions for obtaining estimates for nonlinear members of a UPR (1) with acceptable accuracy.* In [12], analytical expressions are given for variances of estimates $\hat{\theta}_j, j \geq 2$, for a UPR (1). The expressions were obtained from the results of estimation by a virtual UPR (9). This makes it possible to determine, depending on the values of c, d, a, b (8), the number k of repetitions of the main experiment (variances of the coefficient estimates $\hat{\theta}_j$ are reduced by the factor of k) that, using the three-

sigma rule, leads to obtaining the estimates $\hat{\theta}_j$, $j \geq 2$, with acceptable accuracy. Thus, when the repeated experiment turned out to be practically feasible, the problem has a solution.

Remark 1. The three-sigma rule for an arbitrary random variable X is formulated as follows. With a probability of 0,89, any realization of X belongs to the segment $[MX - 3\sigma, MX + 3\sigma]$, where $\sigma = \sqrt{DX}$.

Remark 2. Due to the roughness of the three-sigma rule, the actual number of repetitions k is significantly less.

3. A decomposition method of estimating the coefficients of an MPR given by a redundant representation.

3.1. The problem formulation. Let an MPR be given by the following redundant representation [13]:

$$Y(\bar{x}) = \sum_{\forall (i_1, \dots, i_l) \in K, \forall (j_1, \dots, j_l) \in K(i_1, \dots, i_l)} b_{i_1, \dots, i_l}^{j_1, \dots, j_l} (x_{i_1})^{j_1} \dots (x_{i_l})^{j_l} + E, \quad (12)$$

where $\bar{x} = (x_1, \dots, x_m)^T$ is a deterministic vector of input variables,

$$x_i \in [c_i, d_i], \quad c_i > 0, \quad i = \overline{1, m}, \quad (13)$$

where E is a random variable, $ME = 0$, $DE = \sigma^2 < \infty$. The values of the coefficients $b_{i_1, \dots, i_l}^{j_1, \dots, j_l}$ are unknown (b_0^0 is a constant).

Remark 3. The condition $c_i > 0$, $i = \overline{1, m}$, corresponds to the most common practical case. All the results obtained below are trivially extended to arbitrary real values of the numbers $c_i < d_i$, $i = \overline{1, m}$.

3.2. The decomposition method [13] implements the methodology of reducing the estimation of nonlinear members of an MPR (12) to the sequential construction of UPRs and the solution of the corresponding systems of linear equations, the variables of which are the estimates for the nonlinear members of the MPR (12).

The general algorithmic scheme for obtaining estimates for nonlinear members of the MPR (12) consists of two sub-algorithms [13].

3.2.1. Aggregated modified algorithmic scheme of the first sub-algorithm [13]. The l -th step ($l \leq L_1$, where L_1 is the total number of nonlinear components in (12), each of which contains at least one scalar variable raised to a power greater than or equal to two) is implemented for the next nonlinear term of the MPR (12), whose coefficient was not estimated at the previous steps of the first sub-algorithm and that contains a scalar variable to the maximum power. Let's denote it as x_{i_p} . In the MPR (12), the scalar variable x_{i_p} is replaced with a virtual scalar variable z : $x_{i_p} = a_{i_p} z + b_{i_p}$, where a_{i_p} , b_{i_p} are found according to (8) for $d = d_{i_p}$, $c = c_{i_p}$. In the real main experiment, the scalar variable x_{i_p} takes the value according to (7), (8), and the other scalar variables in all tests take fixed values. In this case, the MPR (12) is transformed into a UPR, and the data of the main virtual experiment are found based on the main real experiment

$$\left(x_{i_p, i} \quad \forall_{j \neq i_p} x_{j, i} = x_j^F \rightarrow y_i, \quad i = \overline{1, n} \right),$$

namely:

$$\left(z_i \rightarrow y_i - \sum_{\forall b_{i_1, \dots, i_l}^{j_1, \dots, j_l} \in \bigcup_{m=1}^{l-1} \{J_m\}} b_{i_1, \dots, i_l}^{j_1, \dots, j_l} (x_{i_1, i})^{j_1} \dots (x_{i_l, i})^{j_l}, \quad i = \overline{1, n} \right), \quad (14)$$

where $\bigcup_{m=1}^{l-1} \{J_m\}$ is the set of coefficients estimated with acceptable accuracy at the previous steps of the first sub-algorithm.

Remark 4. In a similar way, we found the data of a repeated virtual experiment, in which the number of output data is $y_i, i = \overline{1, kn}$, and the input data of the main experiment is repeated k times.

Remark 5. The maximum degree of a UPR is j_p .

The number of UPRs constructed at the l -th step can be more than one if the corresponding coefficient(s) of the first UPR is (are) expressed linearly by several coefficients for the nonlinear terms of the MPR (12) not evaluated at the previous steps of the first sub-algorithm.

The right-hand parts of the obtained nondegenerate systems of linear equations are the estimates for nonlinear terms of the UPRs. Their solutions are the estimates of the corresponding coefficients for nonlinear terms of the MPR (12).

3.2.2. Aggregated modified scheme of the second sub-algorithm [13]. The l -th step ($l \leq L_2$, where L_2 is the number of nonlinear components in (12) of the form $b_{i_1, \dots, i_l}^{1, \dots, 1} \times x_{i_1} \dots x_{i_l}$) is implemented for a nonlinear coefficient $b_{i_1, \dots, i_l}^{1, \dots, 1}$ that was not evaluated at the previous steps and has a maximum value of t_l . Each input variable x_{i_j} , $j = \overline{1, t_l}$, is expressed linearly by a virtual variable z : $x_{i_j} = a_{i_j} z + b_{i_j}$ according to (8) for $c = c_{i_j}$, $d = d_{i_j}$, $j = \overline{1, t_l}$. In the main experiment, the variables x_{i_j} , $j = \overline{1, t_l}$, vary according to (7), (8). Other scalar input variables take fixed values in each test. The data of the virtual main experiment are determined based on the data of the main experiment

$$\left(x_{i_1, i}, \dots, x_{i_{t_l}, i} \quad \forall_j x_{j, i} = x_j^F, \quad j \notin \{i_1, \dots, i_{t_l}\} \rightarrow y_i, \quad i = \overline{1, n} \right),$$

namely:

$$\left(z_i \rightarrow y_i - \sum_{\forall b_{i_1, \dots, i_l}^{j_1, \dots, j_l} \in \bigcup_{m=1}^{K_1} \{J_m\}} \hat{b}_{i_1, \dots, i_l}^{j_1, \dots, j_l} (x_{i_1, i})^{j_1} \dots (x_{i_{t_l}, i})^{j_{t_l}} - \sum_{\forall b_{i_1, \dots, i_l}^{j_1, \dots, j_l} \in \bigcup_{m=1}^{l-1} \{G_m\}} \hat{b}_{i_1, \dots, i_l}^{1, \dots, 1} \prod_{l=1}^{t_l} x_{i_l, i}, \quad i = \overline{1, n} \right), \quad (15)$$

where K_1 is the number of steps of the first sub-algorithm; G_m is a set of coefficients estimated with sufficient accuracy at the previous steps of the second sub-algorithm.

Remark 6. The data for the repeated virtual experiment is found in a similar way.

Remark 7. The maximum degree of a UPR is t_l .

3.3. *Classes of redundant representations for which the coefficients for nonlinear terms of an MPR (12) are estimated completely or partially with acceptable accuracy.*

3.3.1. The redundant representation of (12) satisfies the following three conditions.

1) $\forall (x_{i_l})^{j_l}, j_l \geq 2$ can be included in only one component of (12);

2) for any two nonlinear components of (12) with coefficients $b_{i_1 \dots i_{j_1}}^{j_1 \dots j_1}, b_{i_1 \dots i_{j_2}}^{j_2 \dots j_2}$ the following is true:

$$\{i_1, \dots, i_{j_1}\} \neq \{i_1, \dots, i_{j_2}\}; \quad (16)$$

3) $0 \in [c_i, d_i], i = \overline{1, m},$ (17)

where m is the number of input variables.

Then the l -th step of the first sub-algorithm of the decomposition method implemented for the input scalar variable $(x_{i_p})^{j_p}, j_p \geq 2$, of the nonlinear component of (12)

$$b_{i_1 \dots i_{j_l}}^{j_1 \dots j_l} (x_{i_1})^{j_1} \dots (x_{i_l})^{j_l} \quad (18)$$

must satisfy the following additional conditions. In all tests of the active experiment, in each component of (12) $b_{i_1 \dots i_{j_l}}^{j_1 \dots j_l} (x_{i_1})^{j_1} \dots (x_{i_l})^{j_l}, b_{i_1 \dots i_{j_l}}^{j_1 \dots j_l} \in \bigcup_{m=1}^{l-1} \{J_m\}$ (see (14)), one input scalar variable whose index is not included in the set $\{i_1, \dots, i_{j_l}\}$ (18) is equal to zero. In this case, the linear equation for estimating $b_{i_1 \dots i_{j_l}}^{j_1 \dots j_l}$ (18) has the form

$$b_{i_1 \dots i_{j_l}}^{j_1 \dots j_l} (a_{i_p})^{j_p} \prod_{\substack{m=1 \\ m \neq p}}^l x_{i_m}^F = \hat{\gamma}_{j_p} = \gamma_{j_p} \pm |\varepsilon_{\hat{\gamma}_{j_p}}|, \quad (19)$$

where $\hat{\gamma}_{j_p}$ is the estimate of γ_{j_p} — the coefficient of the corresponding virtual UPR (depending on its meaning, it is a realization or a random variable), and $\varepsilon_{\hat{\gamma}_{j_p}}$ is the realization of a random variable with zero mathematical expectation and the variance equal to $D\hat{\gamma}_{j_p}$ (if $\hat{\gamma}_{j_p}$ is considered a random variable). Let $\varepsilon(\sum_{l=1}^t j_l) > 0$ (this is an expertly set upper bound on the value of $|b_{i_1 \dots i_{j_l}}^{j_1 \dots j_l}|$ below which the corresponding component (18) is excluded from the redundant representation (12) as not essential). Then the values of $x_{i_m}^F \forall i_m, i_m \neq p$, are set equal to $x_{i_m}^F = d_{i_m} > 0, m \neq p$, and thereby the number of repetitions of the main experiment that guarantees the fulfillment of the inequality

$$|\varepsilon_{\hat{\gamma}_{j_p}}| \leq 10^{-1} (a_{i_p})^{j_p} \varepsilon(\sum_{l=1}^t j_l) \prod_{\substack{m=1 \\ m \neq p}}^l d_{i_m} \quad (20)$$

(with the corresponding probability, using the three-sigma rule), is the minimum possible.

Remark 8. To obtain the estimate (20), the three-sigma rule is applied to the random variable $\hat{\gamma}_{j_p}$, which realization is used to find the estimate of the coefficient $\hat{b}_{i_1 \dots i_{j_l}}^{j_1 \dots j_l}$ (18).

If

$$|\hat{\gamma}_{j_p}| < (a_{i_p})^{j_p} \varepsilon(\sum_{l=1}^t j_l) \prod_{\substack{m=1 \\ m \neq p}}^l d_{i_m}, \quad (21)$$

then the component (18) is excluded from the redundant representation (12). In the opposite case, we have found an estimate of the coefficient for the component (18) with an acceptable accuracy (20).

The l -th step of the second sub-algorithm of the decomposition method implemented for the component

$$b_{i_1 \dots i_t}^{1 \dots 1} \prod_{l=1}^t x_{i_l}, t = \max, \quad (22)$$

must satisfy the following additional conditions. In all tests of the active experiment in each component of (12) of the form $b_{i_1 \dots i_t}^{j_1 \dots j_t} (x_{i_1})^{j_1} \dots (x_{i_t})^{j_t}, b_{i_1 \dots i_t}^{j_1 \dots j_t} \in \{\bigcup_{m=1}^{K_l} \{J_m\}\} \cup \{\bigcup_{m=1}^{l-1} \{G_m\}\}$ (see (15)), one input scalar variable whose index does not belong to the set $\{i_1, \dots, i_t\}$ (22) is equal to zero. The linear equation for estimating $b_{i_1 \dots i_t}^{1 \dots 1}$ (22) has the form

$$b_{i_1 \dots i_t}^{1 \dots 1} \prod_{l=1}^t a_{i_l} = \hat{\gamma}_t = \gamma_t \pm |\varepsilon_{\hat{\gamma}_t}|, \quad (23)$$

where $\hat{\gamma}_t$ is the estimate of γ_t — the coefficient of the corresponding virtual UPR. Let the number of repetitions of the main experiment be feasible to fulfill, based on the three-sigma rule, the limitation

$$|\varepsilon_{\hat{\gamma}_t}| \leq 10^{-1} \varepsilon(t) \prod_{l=1}^t a_{i_l}. \quad (24)$$

Then, if

$$|\hat{\gamma}_t| < \varepsilon(t) \prod_{l=1}^t a_{i_l}, \quad (25)$$

then the component (22) is excluded from the redundant representation (12). In the opposite case, we have found an estimate of the coefficient $b_{i_1 \dots i_t}^{1 \dots 1}$ (22) with an acceptable accuracy (24).

Thus, if all the repeated experiments are feasible, the decomposition method allows you to find all estimates for nonlinear terms of the MPR (12) with acceptable accuracy.

Corollary. If for some steps of the first or second sub-algorithms the required number of repetitions of the main experiment is unfeasible, then the corresponding coefficients are not estimated, and conditions (15), (16) allow such components of the redundant representation (12) to be excluded from other steps of the first and second sub-algorithms.

Remark 9. Each step of the first and second sub-algorithms estimates only one coefficient of the MPR (12).

3.3.2. The redundant representation (12) is given by the following four conditions.

The first condition: all input variables included in the redundant representation (12) of the form $b_{i_1 \dots i_t}^{j_1 \dots j_t} \times (x_{i_1})^{j_1} \dots (x_{i_t})^{j_t}$, $\sum_{l=1}^t j_l \geq 2$, have a range of acceptable values $[1, d_{i_l}]$, $l = \overline{1, t}$.

The second condition: if an arbitrary input variable is included in an arbitrary component of (12) to a power greater than or equal to two, it is no longer included in any other component of (12) of the form $b_{i_1 \dots i_t}^{j_1 \dots j_t} (x_{i_1})^{j_1} \dots (x_{i_t})^{j_t}$, $\sum_{l=1}^t j_l \geq t + 1$.

The third condition: all nonlinear components of the form $b_{i_1 \dots i_t}^{1 \dots 1} \prod_{l=1}^t x_{i_l}$ do not have common input variables.

The fourth condition: no input variable included in any component of (12) that has $\exists j_p \geq 2$, is included in any component of (12) whose coefficient has the form $b_{i_1 \dots i_t}^{1 \dots 1}$, $t \geq 2$.

Let the number of repetitions of the main experiment for an arbitrary step of the first and second sub-algorithms be feasible for obtaining, using the three-sigma rule, the estimates:

a) for the first sub-algorithm:

$$|\varepsilon_{\hat{y}_p}| \leq 10^{-1} (a_{i_p})^{j_p} \varepsilon(\sum_{l=1}^t j_l); \quad (26)$$

b) for the second sub-algorithm: estimate (24).

Then, if during the implementation of an arbitrary step of the first and second sub-algorithms, the fixed input variables included in the nonlinear terms of the MPR in expressions (14), (15) are set equal to one, then in expressions (14), (15) from the values of y_i will be subtracted the values of $\sum_{\forall(\cdot)} \hat{b}_{i_1 \dots i_t}^{j_1 \dots j_t}$, which will differ in modulus from the values of $\sum_{\forall(\cdot)} b_{i_1 \dots i_t}^{j_1 \dots j_t}$, due to estimates (24), (26) and the roughness of the three-sigma rule, by values that can be practically neglected. Then all estimates for non-linear terms of the MPR (12) are found with acceptable accuracy.

Remark 10. The number of repetitions of the main experiment can be real if $a_i > \frac{1}{2}$, $i = \overline{1, m}$. This is also true for the second sub-algorithm for the class 3.3.1.

Remark 11. The number of repetitions of the main experiment at some steps of the first sub-algorithm can be significantly smaller, if for this step the estimate for $|\varepsilon_{\hat{y}_p}|$ is of the form (20). That is, some input variables that are not included in (14), (15) can take maximum values.

Remark 12. If we exclude the third and fourth conditions imposed on the redundant representation (12), then the decomposition method implements only the first sub-algorithm.

Remark 13. Each step of the decomposition method for the redundant representation (12) that satisfies the four or the first two conditions of the class 3.3.2, estimates only one coefficient of the redundant representation (12).

3.3.3 (generalization of the class 3.3.2). A redundant representation (12) satisfies the first condition of the class

3.3.1. $c_i < d_i$, $c_i > 0$, $i = \overline{1, m}$, are arbitrary numbers. The second condition is set based on the results of the implementation of the following version of the decomposition method. When performing the l -th ($l \geq 2$) step of the first sub-algorithm for the term $b_{i_1 \dots i_l}^{j_1 \dots j_l} (x_{i_1})^{j_1} \dots (x_{i_l})^{j_l}$, the fixed values of the input variables in (14), which are not part of the set $\{x_{i_1}, \dots, x_{i_l}\}$, are set modulo the minimum possible.

When implementing an arbitrary step of the second sub-algorithm, all fixed variables included in (15) take modulo minimum values.

Let the number of repetitions of the main experiment for each step of the first and second sub-algorithms be feasible to fulfill constraints (20), (24). Then the second condition imposed on the redundant representation (12) is

$$\forall \left| \hat{b}_{i_1 \dots i_t}^{j_1 \dots j_t} \right| \geq \varepsilon \left(\sum_{l=1}^t j_l \right) \cdot 10^p, \quad p \geq 2, \quad \text{or}$$

$$\left| \hat{b}_{i_1 \dots i_t}^{j_1 \dots j_t} \right| < \varepsilon \left(\sum_{l=1}^t j_l \right).$$

It is obvious that for any real values of $c_i < d_i$, $i = \overline{1, m}$, r_{\max} (the maximum degree of the virtual UPR) there is such a small enough value of the natural number p that the replacement of exact values of $b_{i_1 \dots i_t}^{j_1 \dots j_t}$ in expressions (14), (15) with their estimates is statistically guaranteed to have practically no effect on the values of coefficient estimates for nonlinear terms of the MPR (12).

Remark 14. Only one coefficient of the MPR (12) is estimated at each step of the decomposition algorithm.

3.3.4. The redundant representation of the MPR (12) is arbitrary, the random variable $E = 0$. Such a case occurs when E is a measurement error and is neglected having sufficient measurement accuracy. In this case, the general algorithmic procedure of the decomposition method accurately finds the values of all coefficients for the nonlinear terms of (12). The exact values of the coefficients for linear terms, including the constant, are found by the usual interpolation procedure, which consists in solving a system of linear equations with the number of variables $m + 1$, where m is the number of input variables.

The advantage of the decomposition method over an arbitrary interpolation method lies in two-fold: the coefficients for nonlinear terms of the redundant representation are estimated using only a single set of NOPFs found with a given accuracy; the needed-to-solve nondegenerate systems of linear equations have a significantly smaller dimension than the number of nonlinear components of (12), and in most cases the dimension is equal to one.

3.3.5. The redundant representation (12) is arbitrary. In general case, the detailed formalization of the first and second sub-algorithms for finding estimates with sufficient accuracy is inefficient. A visual analysis of the specific redundant representation (12) using the above theoretical results allows for each specific case to create an efficient sequence of steps of an individual algorithm of the decomposition method (which can be a step of the first or second

sub-algorithm in any order), which leads to estimation with acceptable accuracy of the maximum number of coefficients for nonlinear terms of the MPR (12).

4. A modified group method of data handling (MGMDH). As a result of the implementation of the decomposition method, we obtained a set $\{J\}$ of coefficients for nonlinear terms of the MPR (12) estimated with acceptable accuracy.

Remark 15. Insignificant coefficients are excluded from the redundant representation (12) and the set $\{J\}$.

To obtain the final result, we propose to use the MGMDH described in [14] to construct a multivariate linear regression (MLR) given by a redundant representation. Indeed, the problem of constructing an MPR given by a redundant representation (12) was reduced to the following:

$$Y(\bar{x}) = \sum_{\forall \{b_{i_1 \dots i_k}^{j_1 \dots j_k}\} \in \{J\}} b_{i_1 \dots i_k}^{j_1 \dots j_k} (x_{i_1})^{j_1} \dots (x_{i_k})^{j_k} + f(\bar{x}) + E, \quad (27)$$

$$\text{where } f(\bar{x}) = \sum_{\forall \{b_{i_1 \dots i_k}^{j_1 \dots j_k}\} \in \{J\}} \hat{b}_{i_1 \dots i_k}^{j_1 \dots j_k} (x_{i_1})^{j_1} \dots (x_{i_k})^{j_k}.$$

The regression problem without any changes (to an accuracy of the content of the columns of the matrix A in the expression $(A^T A)^{-1} A^T$ of the general formula of the least squares method) can be solved by the method [14], since all unknown coefficients are included linearly in expression (27).

Remark 16. As shown in [14], the use of k repetitions of the main experiment significantly increases the computational efficiency of the MGMDH. Namely, the variance of estimates is reduced by a factor of k , and only the matrices of the main experiment are used in the general formula of the least squares method.

Remark 17. The MGMDH can solve the problem of constructing an MPR given by a redundant representation based also on the results of a passive experiment. In this case, disappear only the advantages of an active experiment: the possibility of using repetitions of the main experiment.

Conclusions. 1. We considered the aggregated algorithmic scheme of the universal synthetic method of constructing an MPR given by a redundant representation. The synthetic method organically combines the decomposition method of estimating coefficients for nonlinear terms of the MPR with acceptable accuracy and the modified group method of data handling.

2. We presented theoretically justified classes of redundant representation of MPR, which allow to estimate fully or partially coefficients for nonlinear terms of an MPR with acceptable accuracy.

3. We substantiated the possibility of extending the modified group method of data handling, created for estimating the coefficients of an MLR given by a redundant representation, to the case of estimating the coefficients of an MPR given by a redundant representation.

References

1. Yu L. Using negative binomial regression analysis to predict software faults: a study of Apache Ant. *International Journal of Information*

- Technology and Computer Science (IJITCS). 2012. Vol. 4, No. 8. P. 63–70. DOI: 10.5815/ijitcs.2012.08.08.
2. Shahrel M.Z., Mutalib S., Abdul-Rahman S. PriceCop – price monitor and prediction using linear regression and LSVM-ABC methods for e-commerce platform. *International Journal of Information Engineering and Electronic Business (IJIEEB)*. 2021. Vol. 13, no. 1. P. 1–14. DOI: 10.5815/ijieeb.2021.01.01.
3. Satter A., Ibtehaz N. A regression based sensor data prediction technique to analyze data trustworthiness in cyber-physical system. *International Journal of Information Engineering and Electronic Business (IJIEEB)*. 2018. Vol. 10, no. 3. P. 15–22. DOI: 10.5815/ijieeb.2018.03.03.
4. Isabona J., Ojuh D. O. Machine learning based on kernel function controlled gaussian process regression method for in-depth extrapolative analysis of Covid-19 daily cases drift rates. *International Journal of Mathematical Sciences and Computing (IJMSC)*. 2021. Vol. 7, no. 2. P. 14–23. DOI: 10.5815/ijmsc.2021.02.02.
5. Sinha P. Multivariate polynomial regression in data mining: methodology, problems and solutions. *International Journal of Scientific & Engineering Research*. 2013. Vol. 4, iss. 12. P. 962–965
6. Kalivas J. H. Interrelationships of multivariate regression methods using eigenvector basis sets. *Journal of Chemometrics*. 1999. Vol. 13 (2). P. 111–132. DOI: 10.1002/(SICI)1099-128X(199903/04)13:2<111::AID-CEM532>3.0.CO;2-N.
7. Ortiz-Herrero L., Maguregui M. I., Bartolomé L. Multivariate (O)PLS regression methods in forensic dating. *TrAC Trends in Analytical Chemistry*. 2021. Vol. 141. 116278. DOI: 10.1016/j.trac.2021.116278.
8. Guo G., Niu G., Shi Q., Lin Q., Tian D., Duan Y. Multi-element quantitative analysis of soils by laser induced breakdown spectroscopy (LIBS) coupled with univariate and multivariate regression methods. *Analytical Methods*. 2019. Vol. 11, iss. 23. P. 3006–3013. DOI: 10.1039/C9AY00890J.
9. Настенко Е., Павлов В., Бойко Г., Носовен О. Многокритериальный алгоритм шаговой регрессии. *Біомедична інженерія і технологія*, 2020. № 3. С. 48–53. DOI: 10.20535/2617-8974.2020.3.195661.
10. Babatunde G., Emmanuel A. A., Oluwaseun O. R., Bunmi O. B., Precious A. E. Impact of climatic change on agricultural product yield using k -means and multiple linear regressions. *International Journal of Education and Management Engineering (IJEME)*. 2019. Vol. 9, no. 3. P. 16–26. DOI: 10.5815/ijeme.2019.03.02.
11. Худсон Д. *Статистика для физиков: Лекции по теории вероятностей и элементарной статистике*. Москва: Мир, 1970. 296 с.
12. Pavlov A. A., Holovchenko M. N., Drozd V. V. Construction of a multivariate polynomial given by a redundant description in stochastic and deterministic formulations using an active experiment. *Вісник Нац. техн. ун-ту «ХПІ»: зб. наук. пр. Темат. вип.: Системний аналіз, управління та інформаційні технології*. Харків: НТУ «ХПІ», 2022. № 1 (7). С. 3–8. DOI: 10.20998/2079-0023.2022.01.01.
13. Pavlov A., Holovchenko M., Mukha I., Lishchuk K., Drozd V. A Modified Method and an Architecture of a Software for a Multivariate Polynomial Regression Building Based on the Results of a Conditional Active Experiment. *Advances in Computer Science for Engineering and Education VI (ICCSEE 2023)*. 2023. (у друці)
14. Pavlov A. A., Holovchenko M. N. Modified method of constructing a multivariate linear regression given by a redundant description. *Вісник Нац. техн. ун-ту «ХПІ»: зб. наук. пр. Темат. вип.: Системний аналіз, управління та інформаційні технології*. Харків: НТУ «ХПІ», 2022. № 2 (8). С. 3–8. DOI: 10.20998/2079-0023.2022.02.01.

References (transliterated)

1. Yu L. Using negative binomial regression analysis to predict software faults: a study of Apache Ant. *International Journal of Information Technology and Computer Science (IJITCS)*. 2012, vol. 4, no. 8, pp. 63–70. DOI: 10.5815/ijitcs.2012.08.08.
2. Shahrel M.Z., Mutalib S., Abdul-Rahman S. PriceCop – price monitor and prediction using linear regression and LSVM-ABC methods for e-commerce platform. *International Journal of Information Engineering and Electronic Business (IJIEEB)*. 2021, vol. 13, no. 1, pp. 1–14. DOI: 10.5815/ijieeb.2021.01.01.

3. Satter A., Ibtahaz N. A regression based sensor data prediction technique to analyze data trustworthiness in cyber-physical system. *International Journal of Information Engineering and Electronic Business (IJIEEB)*. 2018, vol. 10, no. 3, pp. 15–22. DOI: 10.5815/ijieeb.2018.03.03.
4. Isabona J., Ojuh D. O. Machine learning based on kernel function controlled gaussian process regression method for in-depth extrapolative analysis of Covid-19 daily cases drift rates. *International Journal of Mathematical Sciences and Computing (IJMSC)*. 2021, vol. 7, no. 2, pp. 14–23. DOI: 10.5815/ijmsc.2021.02.02.
5. Sinha P. Multivariate polynomial regression in data mining: methodology, problems and solutions. *International Journal of Scientific & Engineering Research*. 2013, vol. 4, iss. 12, pp. 962–965.
6. Kalivas J. H. Interrelationships of multivariate regression methods using eigenvector basis sets. *Journal of Chemometrics*. 1999, vol. 13 (2), pp. 111–132. DOI: 10.1002/(SICI)1099-128X(199903/04)13:2<111::AID-CEM532>3.0.CO;2-N.
7. Ortiz-Herrero L., Maguregui M. I., Bartolomé L. Multivariate (O)PLS regression methods in forensic dating. *TrAC Trends in Analytical Chemistry*. 2021, vol. 141, 116278. DOI: 10.1016/j.trac.2021.116278.
8. Guo G., Niu G., Shi Q. et al. Multi-element quantitative analysis of soils by laser induced breakdown spectroscopy (LIBS) coupled with univariate and multivariate regression methods. *Analytical Methods*. 2019, vol. 11, iss. 23, pp. 3006–3013. DOI: 10.1039/C9AY00890J.
9. Nastenko E., Pavlov V., Boyko G., Nosovets O. Mnogokriterial'nij algoritm shagovoy regressii. *Biomedychna inzheneriya i tekhnolohiya* [Biomedical engineering and technology]. 2020, no. 3, pp. 48–53. DOI: 10.20535/2617-8974.2020.3.195661.
10. Babatunde G., Emmanuel A. A., Oluwaseun O. R., Bunmi O. B., Precious A. E. Impact of climatic change on agricultural product yield using *k*-means and multiple linear regressions. *International Journal of Education and Management Engineering (IJEME)*. 2019, vol. 9, no. 3, pp. 16–26. DOI: 10.5815/ijeme.2019.03.02.
11. Hudson D. J. Statistics Lectures, Volume 2: Maximum Likelihood and Least Squares Theory. CERN Reports 64(18). Geneva, CERN, 1964. (Russ. ed.: Hudson D. *Statistika dlja fizikov: Lekcii po teorii veroyatnostej i jelementarnej statistike*. Moscow, Mir Publ., 1970. 296 p.). DOI: 10.5170/CERN-1964-018.
12. Pavlov A. A., Holovchenko M. N., Drozd V. V. Construction of a multivariate polynomial given by a redundant description in stochastic and deterministic formulations using an active experiment. *Visnyk Nats. tekhn. un-tu "KhPI": zb. nauk. pr. Temat. vyp.: Systemnyy analiz, upravlinnya ta informatsiyni tekhnolohiyi* [Bulletin of the National Technical University "KhPI": a collection of scientific papers. Thematic issue: System analysis, management and information technology]. Kharkov, NTU "KhPI" Publ., 2022, no. 1 (7), pp. 3–8. DOI: 10.20998/2079-0023.2022.01.01.
13. Pavlov A., Holovchenko M., Mukha I. et al. A Modified Method and an Architecture of a Software for a Multivariate Polynomial Regression Building Based on the Results of a Conditional Active Experiment. *Advances in Computer Science for Engineering and Education VI (ICCSEEA 2023)*. 2023. (to appear)
14. Pavlov A. A., Holovchenko M. N. Modified method of constructing a multivariate linear regression given by a redundant description. *Visnyk Nats. tekhn. un-tu "KhPI": zb. nauk. pr. Temat. vyp.: Systemnyy analiz, upravlinnya ta informatsiyni tekhnolohiyi* [Bulletin of the National Technical University "KhPI": a collection of scientific papers. Thematic issue: System analysis, management and information technology]. Kharkov, NTU "KhPI" Publ., 2022, no. 2 (8), pp. 3–8. DOI: 10.20998/2079-0023.2022.02.01.

Received 15.04.2023

УДК 004:519.24:681.3.06

О. А. ПАВЛОВ, доктор технічних наук, професор, Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», м. Київ, Україна, професор кафедри інформатики та програмної інженерії; e-mail: pavlov.fiot@gmail.com; ORCID: <https://orcid.org/0000-0002-6524-6410>

М. М. ГОЛОВЧЕНКО, Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», м. Київ, Україна, старший викладач кафедри інформатики та програмної інженерії; e-mail: ma4ete25@ukr.net; ORCID: <https://orcid.org/0000-0002-9575-8046>

В. В. ДРОЗД, Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», м. Київ, Україна, бакалавр кафедри інформатики та програмної інженерії; e-mail: drozdllera@gmail.com, ORCID: <https://orcid.org/0000-0003-0418-1139>

ОБґРУНТУВАННЯ ЕФЕКТИВНОСТІ СИНТЕТИЧНОГО МЕТОДУ ПОБУДОВИ БАГАТОВИМІРНОЇ ПОЛІНОМІАЛЬНОЇ РЕГРЕСІЇ, ЗАДАНОЇ НАДЛИШКОВИМ ОПИСОМ

Протягом останніх років автори в своїх публікаціях паралельно розвивали два різних підходи до побудови багатовимірних поліноміальних, зокрема, лінійних регресій, заданих надлишковим описом. Перший підхід дозволяв знаходження оцінок коефіцієнтів при нелінійних членах багатовимірної поліноміальної регресії зводити до побудови послідовності одновимірних поліноміальних регресій та розв'язання відповідних невідроджених систем лінійних рівнянь. Другий підхід був реалізований на прикладі багатовимірної лінійної регресії, заданої надлишковим описом, і привів до створення методу, названого авторами модифікованим методом групового урахування аргументів (МГУА), так як він є модифікацією широко відомого методу евристичної самоорганізації МГУА (автор МГУА – академік НАН України О. Г. Івахненко). Модифікація полягає в тому, що завдання багатовимірної лінійної регресії надлишковим описом дозволяє для побудови множини часткових описів, один з яких має структуру шуканої регресії, використовувати не багаторівневий селекційний алгоритм, а ефективний алгоритм розбиття коефіцієнтів багатовимірної лінійної регресії на два класи. Як і в класичному МГУА, розв'язок знаходиться за допомогою перевірконої послідовності даних. Цей метод легко поширюється на випадок багатовимірної поліноміальної регресії, так як невідомі коефіцієнти в багатовимірну поліноміальну регресію входять лінійно. Кожен з двох підходів має свої переваги і недоліки. Очевидним наступним кроком є поєднання обох підходів в один. Це призвело до створення синтетичного методу, який реалізує переваги обох підходів, частково компенсуючи їх недоліки. В цій роботі наведена агрегована алгоритмічна структура синтетичного методу, теоретичні властивості часткових випадків і, як наслідок, обґрунтування його ефективності в цілому.

Ключові слова: одновимірна поліноміальна регресія, багатовимірна поліноміальна регресія, надлишковий опис, метод найменших квадратів, перевірна послідовність, повторний експеримент.

Повні імена авторів / Author's full names

Автор 1 / Author 1: Павлов Олександр Анатолійович, Pavlov Alexander Anatolievich

Автор 2 / Author 2: Головченко Максим Миколайович, Holovchenko Maxim Nikolaevich

Автор 3 / Author 3: Дрозд Валерія Валеріївна, Drozd Valeriia Valeriivna

А. І. ЛЕВТЕРОВ, кандидат технічних наук, професор, Харківський національний автомобільно-дорожній університет, завідувач кафедри інформатики та прикладної математики; м. Харків, Україна, e-mail: lai@khadi.kharkov.ua, ORCID: <https://orcid.org/0000-0001-6586-1061>

Г. А. ПЛЕХОВА, кандидат технічних наук, доцент, Харківський національний автомобільно-дорожній університет, доцент кафедри інформатики та прикладної математики; м. Харків, Україна, e-mail: plehovaanna1@gmail.com, ORCID: <https://orcid.org/0000-0002-6912-6520>

М. В. КОСТИКОВА, кандидат технічних наук, доцент, Харківський національний автомобільно-дорожній університет, доцент кафедри інформатики та прикладної математики; м. Харків, Україна, e-mail: kmv_topaz@ukr.net, ORCID: <https://orcid.org/0000-0001-5197-7389>

Н. Г. БЕРЕЖНА, кандидат технічних наук, доцент, Державний біотехнологічний університет, доцент кафедри транспортних технологій і логістики; м. Харків, Україна, e-mail: bereg_nat@ukr.net, ORCID: <https://orcid.org/0000-0001-8740-3387>

А. О. ОКУНЬ, кандидат технічних наук, доцент, Національний технічний університет «Харківський політехнічний інститут», доцент кафедри комп'ютерного моделювання та інтегрованих технологій обробки тиском; м. Харків, Україна, e-mail: okunanton@gmail.com, ORCID: <https://orcid.org/0000-0002-6467-4229>

ДОСЛІДЖЕННЯ МЕТОДІВ БЕЗПЕЧНОЇ МАРШРУТИЗАЦІЇ У ПРОГРАМНО-КОНФІГУРОВАНИХ МЕРЕЖАХ

У сучасному світі безпека мережі є ключовим питанням інформаційної безпеки. Віртуальні мережі стали невід'ємною частиною сучасної IT-інфраструктури, що ставить перед нами виклики у сфері безпеки. Одним з рішень цієї проблеми є використання програмно-визначеної мережі (SDN), яка надає засоби контролю та керування мережевим трафіком. Однак, як і будь-яка технологія, SDN має свої вразливості, які необхідно враховувати під час її розгортання. Одним із інструментів, який допомагає врахувати вразливості мережевої інфраструктури, є стандарт Common Vulnerability Scoring System (CVSS). Це дозволяє кількісно визначити рівень уразливості інфраструктури, що забезпечує ефективний захист мережі. Аналіз стандарту CVSS є важливим етапом у розробці стратегії безпеки мережі. У цій статті аналізуються стандарти для побудови програмно-конфігурованих мереж. Зазначається, що SDN – це сучасний підхід до проектування, побудови та експлуатації інформаційних комунікаційних мереж. Використання SDN дає можливість безпосередньо програмувати та динамічно керувати мережею, а також абстрагувати функціональні можливості рівня інфраструктури. Однак зростання інтересу до SDN виявило недоліки їх застосування в боротьбі із загрозами кібербезпеці. Сама архітектура SDN, зовнішні шкідливі атаки, недостатній контроль доступу та засоби шифрування були визнані основними проблемами безпеки. Запропоновано використання інструментів безпечної маршрутизації на основі показників уразливості для підвищення рівня безпеки мережі площини даних SDN. Відповідно до проведеного аналізу вразливостей площини даних SDN та функціональності інструментів маршрутизації, автори рекомендують використовувати стандарт CVSS для кількісної оцінки рівня вразливості інфраструктури під час розробки та дослідження перспективних підходів до безпечної маршрутизації в площині даних програмно налаштованих мереж.

Ключові слова: SDN, CVSS, вразливість, NFV, площина даних, NVD, програмно-визначена мережа.

Вступ. Розгортання таких мережних архітектур, як програмно-конфігуровані мережі (Software-Defined Networking, SDN), стикається з новими загрозами кібербезпеці, які вимагають розробку та дослідження нових спеціалізованих рішень щодо підвищення рівня мережної безпеки. Незважаючи на високу відкритість і можливості програмованості, архітектура SDN замінює традиційну мережу, проте збільшує кількість потенційних мережних атак, що призводить до нових проблем безпеки.

Саме тому у статті проведений аналіз стандартів побудови програмно-конфігурованих мереж. Крім того, представлено аналіз вразливостей архітектур SDN.

Частина роботи присвячена аналізу вразливостей площини даних SDN і функціональних можливостей засобів маршрутизації щодо протидії можливим атакам. Приділено увагу специфіці еволюції площини даних архітектури SDN, а також існуючим технологіям і підходам захисту площини даних SDN. Показано перспективність використання засобів безпечної маршрутизації на основі базових метрик критичності

вразливостей для підвищення рівня мережної безпеки площини даних SDN.

Також проведений аналіз стандарту CVSS щодо кількісного розрахунку рівня вразливості мережного обладнання. Було розглянуто категорії вразливостей, використання національної бази даних вразливостей та загальної системи оцінки вразливостей, метрики загальної системи оцінки вразливостей.

Аналіз публікацій. Основні стандарти побудови програмно-конфігурованих мереж. Програмно-конфігуровані мережі є підходом до проектування, побудови та експлуатації інфокомунікаційних мереж шляхом розділення площин управління (control plane) та передавання даних (data plane). Такий розподіл надає мережі безпосередньої програмованості та динамічності, а також дозволяє абстрагувати функціональні можливості рівня інфраструктури.

Сам термін виник ще у середині 1990-х років. Однак лише у 2011 році була заснована Open Networking Foundation для просування SDN, що почала використовувати OpenFlow як ключовий протокол для опису того, як керувати мережею та вносити відповідні

© Левтеров А. І., Плехова Г. А., Костікова М. В., Бережна Н. Г., Окунь А. О., 2023



Дослідницька стаття: Цю статтю опубліковано видавництвом *НТУ «ХПІ»* у збірнику «Вісник Національного технічного університету "ХПІ" Серія: Системний аналіз, управління та інформаційні технології». Ця стаття поширюється за міжнародною ліцензією Creative Commons Attribution (CC BY 4.0). **Конфлікт інтересів:** Автор/и заявив/или про відсутність конфлікту.



зміни стандартизованим способом. Наразі OpenFlow – лише один з прикладів із множини протоколів взаємодії SDN контролерів і пристроїв площини даних, і для надання можливостей SDN можна використовувати інші механізми зв'язку [1].

Існують різні архітектури, що відокремлюють логіку управління від ресурсів поза пристроєм, але всі підходи до побудови SDN включають контролер SDN і відповідні прикладні програмні інтерфейси – Southbound API та Northbound API (рис. 1).

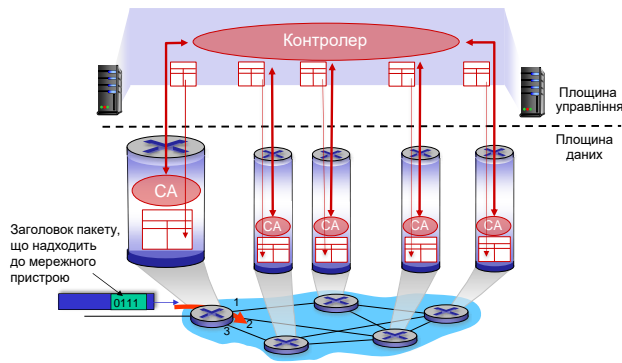


Рис. 1. Базова архітектура програмно-конфігурованої мережі [2]

Рис. 2. Концепція віртуалізації мережних функцій [1]

Зі свого боку віртуалізація мережних функцій (Network Functions Virtualization, NFV) є стандартизованим способом розробки, впровадження та керування мережними службами. Дана концепція передбачає заміну спеціальних пристроїв для мережної інфраструктури, таких як маршрутизатори та брандмауери, стандартними серверами, комутаторами, сховищем, хмарою або навіть туманною обчислювальною інфраструктурою [1]. NFV відокремлює такі функції мережі, як маршрутизація, комутація та безпека, від пропрієтарних або виділених апаратних пристроїв, щоб дозволити їм працювати в межах програмного забезпечення.

Головна мета використання NFV полягає в тому, щоб використовувати стандартні технології віртуалізації для консолідації апаратного забезпечення та віртуалізації мережних функцій у блоки, які можна об'єднувати для створення наскрізних комунікаційних послуг. Це може бути реалізовано для будь-якої функції площини управління або площини даних у середовищі як проводових, так і безпроводових мереж.

NFV з'явився в 2012 році, коли сім провідних глобальних постачальників мережних послуг об'єдналися під Групою галузевих специфікацій (Industry Specification Group, ISG) ETSI для NFV [1]. Це основна група для розробки вимог та архітектури для віртуалізації різних функцій в інфокомунікаційних мережах і для архітектурних стандартів NFV, зокрема NFV MANO (Management and Orchestration), який став стандартом де-факто для NFV.

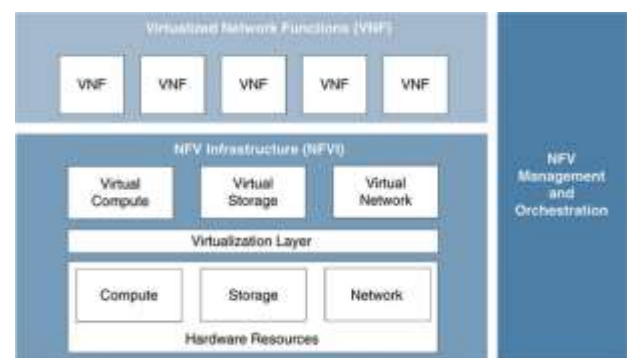
Отже, NFV складається з трьох основних компонентів [1]:

- Віртуалізовані мережні функції (Virtualized Network Functions, VNF) – мережні функції, які реалізовані в програмному забезпеченні та можуть бути розгорнуті в інфраструктурі віртуалізації мережних функцій (Network Functions Virtualization Infrastructure, NFVI).

- NFVI – загальні апаратні та програмні компоненти, де розгортаються VNF.

- Середовище керування та оркестровки NFV (Management and Orchestration, MANO) – архітектурний фреймворк, що містить функціональні блоки, сховища даних, що використовуються цими блоками, та інтерфейси, через які функціональні блоки обмінюються інформацією з метою управління та оркестрації NFVI та VNF.

З концепцією віртуалізації мережних функцій можна ознайомитися на рис. 2.



Нижче наведено добірку ключових груп розробників стандартів тісно пов'язаних передових технологічних областей NFV і SDN:

- Фокус група SDN і NFV Alliance for Telecommunications Industry Solutions (ATIS) було створено в січні 2014 року з метою об'єднання сервісів апаратного забезпечення/пристроїв NFV за допомогою концепції SDN, а також розробки API для інтерфейсів та OAM для зв'язування служб OpenStack та OpenDaylight [3].

- Broadband Forum (BBF) працює як з SDN, так і з NFV, зосереджуючись на наданні хмарного широко-смугового з'єднання та NFV. У 2017 році BBF уклав меморандум про взаєморозуміння (MoU) з SDN/NFV Industry Alliance [4].

- ETSI є головним глобальним рушієм стандартизації NFV [5].

- Інститут інженерів з електротехніки та електроніки (Institute of Electrical and Electronics Engineers, IEEE) займається стандартизацією SDN у межах дослідницьких груп, що вивчають можливості стандартизації в програмно-конфігурованих мережах, віртуалізації мережних функцій і суміжних областях [6].

- Міжнародна рада з великих електричних систем (International Council on Large Electrical Systems, CIGRE) має робочу групу, діяльність якої зосереджена на розробці стандарту IEEE P1915.1, що визначає структуру безпеки, моделей, аналітики та вимог для SDN/NFV [7]. CIGRE розглядає моделі безпеки, термінологію та аналітику (основні компоненти середовища

SDN і NFV), щоб забезпечити конфіденційність, цілісність і доступність.

- Сектор стандартизації телекомунікацій Міжнародного союзу електрозв'язку (International Telecommunication Union Telecommunication Standardization Sector, ITU-T) зосередився на SDN з листопада 2012 року. ITU-T прийняв резолюцію 77, щоб просувати стандартизацію в SDN, а також виконує роботу, орієнтовану на SDN, у групі WTSA-12 [8].

- Internet Engineering Taskforce (IETF) працює як над SDN (через RFC 7426 SDN), так і над новою групою стандартів IETF SDN (I2RS), зосередившись на протоколах програмування Southbound, NFV і мережних послуг [9].

- Internet Research Task Force (IRTF) має групи SDN, метою якої є надання переваг усім типам мереж, включаючи безпроводові, мобільні, домашні, корпоративні, центри обробки даних і глобальні мережі [10]. Дослідницька група програмно-визначених мереж (Software-Defined Networking Research Group, SDNRG) прагне визначити підходи, які можна розгорнути та використати найближчим часом, а також завдання майбутніх досліджень. Основні сфери інтересів включають масштабованість рішень, абстракції, мови програмування та парадигми, які особливо корисні в контексті SDN.

- Internet Society (ISOC) має групи IRTF і IETF, які зосереджуються на NFV і SDN і забезпечують архітектурний нагляд за Internet Architecture Board (IAB) [11].

- Metro Ethernet Forum (MEF), заснований у 2001 році, сприяє створенню нейтральних для промисловості середовищ для оркестровки послуг (OpenLSO) і служб підключення L2–L7 (OpenCS) на основі SDN і NFV [12].

- Альянс відкритих центрів даних (Open Data Centre Alliance, ODCA) був створений у жовтні 2010 року. ODCA керує об'єднаною хмарною архітектурою зі спільними стандартами як для апаратного, так і для програмного забезпечення [13].

- OpenDaylight Foundation була заснована в 2013 році і просуває платформу SDN з відкритим кодом для створення програмованих і гнучких мереж [14].

- Open Networking Foundation (ONF) – некомерційна організація, орієнтована на користувачів, яка прагне прискорити впровадження SDN і NFV. ONF дійсно стимулював рух SDN у 2011 році, а у 2017 році ONF оголосив Open Innovation Pipeline, щоб спрямувати галузь до наступного покоління SDN і NFV [15].

- Відкрита платформа для NFV (Open Platform for NFV, OPNFV), сформована в 2014 році фондом

Linux, сприяє розробці та еволюції компонентів NFV для екосистем з відкритим кодом [16].

- Optical Internetworking Forum (OIF) має робочу групу з архітектури та сигналізації, яка визначає інтерфейси для SDN [17].

Специфіка еволюції площини даних архітектури SDN. На сучасному етапі розвитку архітектур SDN, вітчизняні та іноземні дослідники вважають, що типові проблеми безпеки у програмно-конфігурованих мережах перш за все проявляються в таких аспектах: шкідливе програмне забезпечення, вразливість контролера, легітимність та узгодженість правил потоків, проблема стандартизації північного інтерфейсу, безпека комунікації в процесі використання південного інтерфейсу та інше. У межах роботи проаналізовано типові проблеми безпеки та об'єкти атак у межах площини даних – мережних пристроїв, що керуються контролером SDN. Результати аналізу наведено в табл. 1. Очевидно, що об'єктами атаки можуть бути пристрої різних рівнів мережі, і відповідно до чіткої багаторівневої архітектури SDN можна класифікувати загрози безпеці на різних рівнях з урахуванням підвищення.

Отже, площина даних складається з комутаторів та інших мережних пристроїв і головним чином відповідає за обробку даних, їх пересилання, відкидання, а також збір статистики. Функціонування площини даних відбувається на основі правил потоків, що надаються контролером мережі. Типові проблеми безпеки у площині даних розглядаються відповідно до наступного [18]:

- Авторизована автентифікація. По суті, на рівні площини даних відсутній ефективний механізм автентифікації між мережним обладнанням і контролером. Таким чином, можуть виникнути деякі проблеми, а саме видавання особи за іншу особу та незаконний доступ. Зловмисний SDN-комутатор може генерувати підроблений або фальшивий потік даних у мережі, маніпулювати або перевіряти вміст пакетів даних і відхиляти дозволені пакети даних. Це може призвести до порушення цілісності даних і вплинути на доступність площини даних. Крім того, якщо комутатор встановлює з'єднання з контролером без автентифікації, комутатором може керувати зловмисний контролер, що може призвести до фальсифікації інформації таблиці потоків, що призведе до витоку даних та інших проблем безпеки. Не відповідні керуючі інструкції можуть спричинити плутанину в таблицях потоків комутаторів та безпосередньо збільшити ризики порушення безпеки.

- Правомірність і узгодженість правил потоків є одними з головних проблем на рівні площини даних.

Таблиця 1 – Загальна класифікація метрик CVSS [23]

Проблема безпеки	Об'єкт атаки	Причина
Авторизована автентифікація	Мережне обладнання	Управління доступом
Правомірність правил потоків	Правила потоків	Управління доступом
Узгодженість правил потоків	Правила потоків	Архітектура SDN
DoS/DDoS атаки	Таблиці потоків	Архітектура SDN, зловмисна атака
Атака сторонніми каналами	Конфіденційність даних	Зловмисна атака

Правомірність правил потоків стосується зловмисного або невірнього впровадження правил потоків. Зі свого боку узгодження правил потоків в основному включає три аспекти [18]. Під час процесу генерації функціонування значної кількості застосунків може викликати конфлікти або перевизначення правил потоків. Під час процесу випуску правил потоків затримка передавання або зловмисне втручання може також спричинити їхню неузгодженість між контролером і комутаторами. Зі свого боку процес оновлення ініціює синхронізацію правил потоків між різними комутаторами. Слід зазначити, що до оновлення правил потоків в SDN призводять, наприклад, відмови вузлів мережі, передача службового навантаження або технічне обслуговування мережі, через що пакети бачитимуть неузгоджені уявлення мережі щодо її поточного стану. Таким чином, якщо пакети даних передаються відповідно до нових і старих правил потоків, то можуть виникнути такі проблеми, як «чорні діри», циклічні шляхи або перевантаження мережі.

- DoS/DDoS атаки. Простір таблиці потоків обмежений, проте за звичайних обставин розмірність таблиці потоків комутатора відповідає вимогам пересилання пакетів даних. Однак в умовах DoS/DDoS атак зловмисник створює низку незаконних доступів, і простір таблиці потоку переповнюється недійсними правилами трафіку [18]. У разі цього буде спожито значну кількість ресурсів таблиці потоків, а звичайні правила потоків не мають достатньо місця для обробки. Отже, DoS/DDoS-атаки можуть значно погіршити продуктивність мережі.

- Атака сторонніми каналами. SDN весь час переносить приватну та конфіденційну інформацію [18]. У SDN атрибути процесу (наприклад, атрибут часу) кожної дії виконання різні. Використовуючи атаки сторонніми каналами, зловмисник може отримати пов'язану з мережею інформацію про стан (наприклад, інформацію таблиці потоків), перевіряючи час виконання конкретного типу пакета даних. Тому таблиця потоків може спричинити проблеми з витоком даних. Хоча атаки через сторонні канали безпосередньо не впливають на доступність, конфіденційність або цілісність даних, вони можуть викликати подальші атаки.

Таким чином, основними причинами проблем безпеки є власне архітектура SDN, зовнішні шкідливі атаки, недостатність контролю доступу та засобів шифрування.

Мета та постановка задачі. Об'єктом дослідження є процес забезпечення мережної безпеки у програмно-конфігурованих мережах засобами маршрутизації.

Предмет дослідження – це стандарти побудови програмно-конфігурованих мереж (SDN), стандарт Common Vulnerability Scoring System (CVSS) щодо кількісного розрахунку рівня вразливості мережного обладнання, функціональні можливості протоколів IP-маршрутизації.

Мета роботи – аналіз стандартів побудови програмно-конфігурованих мереж, стандарту CVSS, функціональних можливостей протоколів IP-маршрутизації.

В роботі використані наступні методи досліджень: формалізація, порівняння.

Дана робота присвячена аналізу загроз, об'єктів атак і потенційних рішень щодо підвищення рівня мережної безпеки у площині даних SDN мереж засобами маршрутизації.

Аналіз вразливостей архітектур SDN. Завдяки зростаючому інтересу до SDN та широкому розгортанню програмно-конфігурованих мереж поступово виявляються їхні недоліки у боротьбі із загрозами кібербезпеці. Відповідно питання щодо безпеки тісно пов'язані з характеристиками SDN. Згідно з дослідженнями [18, 19], наступні аспекти роблять програмно-конфігуровані мережі вразливими до атак:

- SDN контролер, що відповідає за централізований контроль мережею, має недостатні механізми захисту, через що стає цілком зовнішніх зловмисних атак.

- У процесі складної взаємодії пов'язаних між собою різноманітних прикладних програм і мережних застосунків між ними часто виникають конфлікти у правилах передавання потоків у межах площини даних.

- Відсутність належних механізмів авторизації та автентифікації прикладних програм робить їх вразливими для атак з використанням зловмисного програмного забезпечення.

- Існує певна недостатність засобів безпеки та шифрування в процесі комунікації між площиною управління та площиною даних. Отже, правила передавання потоків (flow rules) вразливі щодо зловмисного втручання під час їхньої публікації.

У межах архітектури SDN реалізуються окремі рівні щодо площини управління та площини даних на відміну від традиційних мереж, де ці площини функціонують в єдиному пристрої. Площина управління формує правила потоків, а площина даних зі свого боку відповідає лише за пересилання пакетів даних відповідно до правил потоків. У той же час контролер мережі може отримати статус кожного пристрою через південний інтерфейс (Southbound Interface, SBI), щоб мати глобальне уявлення про стан мережі. Таке функціональне розділення в архітектурах SDN значно підвищує гнучкість мереж, але також створює нові загрози безпеці, які переважно відображаються в наступних двох аспектах [18]:

- Вибірковий дозвіл пристроїв захисту. SDN – це мережа, що керується правилами потоків, де фізичний пристрій захисту не має права прийняття рішень. Саме правила потоків визначають, чи будуть пакети даних передаватися через пристрій захисту та коли саме. Таким чином, в SDN зловмисник може обійти пристрій захисту, що призведе до збою заходів безпеки перед розгортанням мережі.

- Автоматичне отримання глобального уявлення про стан мережі. В SDN контролер, як командний і контрольний центр усієї мережі, може створити глобальне уявлення про мережу. Він може отримувати різноманітну інформацію про стан мережі в режимі реального часу. Таким чином, інформацію щодо стану безпеки мережі можна легко отримати від контролера. Отже, зловмисники можуть дістати глобальне уявлен-

ня про мережу безпосередньо від контролера, очікуючи найкращої нагоди для масштабної атаки.

Технології та підходи захисту площини даних SDN. Базуючись на дослідженнях типових проблем безпеки, стає очевидним, що централізоване керування та функції програмованості SDN надають зловмисникам потужні та зручні канали атак. Отже, поширене використання архітектур SDN на практиці робить питання безпеки все більш помітними. Основні існуючі технології захисту площини даних SDN наведені в табл. 2.

- Виявлення помилок у мережі може бути реалізовано за допомогою алгоритмів виявлення. Так, наприклад, NICE [18] надає тестову схему, засновану на символічній моделі виконання з метою перевірки, чи створює програма верхнього рівня неузгоджений стан мережі. Однак ці методи не є рішеннями в реальному часі, мають велику затримку обробки та високий рівень запитів. Такі рішення не можуть принципово усунути вплив конфігураційних конфліктів на мережу.

- Поділ правил потоків. Розподіл дозволів додатків на основі ролі та пріоритету об'єкта є основним методом вирішення правомірності та узгодженості правил потоків. Дозвіл правил потоків поділяється за такими методами, як цифровий підпис, розподіл ролей і класифікація функцій.

- Виявлення зв'язності може бути реалізовано різними засобами, наприклад, шляхом використання технології віртуалізованого обміну з метою розділення мережі на ізольовані фрагменти і таким чином запобігання конфлікту правил потоків у них; механізмів виявлення цілісності пакетів на основі поєднання правил потоків з номером версії тощо [18].

- Модуль автентифікації й авторизації може використовувати сервер RADIUS для автентифікації ідентифікатора хоста. Крім того, можливе застосування квантової криптографії для розробки надлегкої верифікації цілісності [18]. Зазначається, що квантові паролі є більш безпечними, ніж традиційні, проте їхнє впровадження може бути дорожчим.

- Захист від DoS/DDoS-атак у площині даних може бути трансформована в задачу оптимізації таблиці потоків.

Засоби безпечної маршрутизації для підвищення рівня мережної безпеки площини даних SDN. На сьогоднішній день важливе місце у комплексі засобів підвищення мережної безпеки, у тому числі мереж SDN, відводиться протоколам маршрутизації, які потребують системної та скоординованої взаємодії одночасно множини мережних елементів – SDN-

комутаторів, і контролерів мережі під час формування (розрахунку) шляхів і правил потоків, вздовж яких має забезпечуватися необхідний рівень безпеки за обраними показниками або критеріями.

Отже, у напрямку безпечної маршрутизації проведено значну кількість теоретичних досліджень, починаючи від найпростіших емпіричних варіантів рішень до системних оптимізаційних підходів.

Так, у роботі [20] розроблено та досліджено модель безпечної маршрутизації з балансуванням навантаження в мережах на основі SD-WAN. Технологічне завдання безпечної маршрутизації з балансуванням навантаження було сформульовано у формі оптимізаційної задачі з квадратичним критерієм оптимальності. Така форма критерію дозволяє збалансувати частки потоків, що передаються в мережі. Представлена модель безпечної маршрутизації з балансуванням навантаження з адитивною метрикою враховує продуктивність і безпеку мережі, дозволяє ефективніше використовувати наявні мережні ресурси, але також враховує ймовірність компрометації каналів зв'язку під час прийняття маршрутних рішень.

Тоді як у роботах [21, 22] пропонуються потокові моделі маршрутизації з врахуванням ризиків інформаційної безпеки за допомогою базових метрик критичності вразливостей. Основу моделі складають умови реалізації одно- та багатошляхової маршрутизації, збереження потоку та запобігання перевантаженню каналів зв'язку мережі, а задача безпечної маршрутизації також сформульована як оптимізаційна. У моделі [21] для розрахунку маршрутних метрик використовуються вирази, які характеризують ризик інформаційної безпеки в каналах зв'язку мережі та відповідно до рекомендацій NIST враховують збитки від порушення конфіденційності та цілісності інформації, доступності мережного ресурсу у випадку використання наявних вразливостей; показники складності використання вразливостей на вузлах мережі та отримання доступу до мережних елементів та мережі загалом внаслідок використання зазначених вразливостей. Запропонований авторами підхід до формування маршрутних метрик може бути використаний під час комплексного врахування в процесі розв'язання задач маршрутизації як показників мережної безпеки, так і показників якості обслуговування.

Аналіз стандарту CVSS щодо кількісного розрахунку рівня вразливості мережного обладнання. Виявлення й ідентифікація вразливостей – це процес виявлення вразливостей, які можуть бути використані загрозами для заподіяння шкоди активам. Вразливість – це слабе місце або недолік у процедурах безпеки

Таблиця 2 – Існуючі технології безпеки у площині даних SDN [18]

Технологія захисту	Проблема безпеки
Виявлення помилок у мережі	Помилка конфігурації
Класифікація правил потоків	Правомірність правил потоків
Формальний математичний аналіз	Правомірність правил потоків
Виявлення зв'язності	Узгодженість правил потоків
Модуль автентифікації й авторизації	Авторизована автентифікація

системи, проектуванні, реалізації або внутрішніх засобах контролю, які можуть бути випадково активовані або навмисно використані під час виявлення загрози. Отже, далі розглянуто категорії вразливостей та проаналізовано підходи до ідентифікації та документування вразливостей, а також обговорено використання національної бази даних вразливостей – National Vulnerability Database (NVD).

Категорії вразливостей. Узагальнена класифікація вразливостей може бути представлена наступним чином [23]:

- Технічні вразливості: недоліки в розробці, реалізації та/або конфігурації програмного забезпечення та/або апаратних компонентів, включно з прикладним програмним забезпеченням, системним програмним забезпеченням, комунікаційним програмним забезпеченням, обчислювальним обладнанням, комунікаційним обладнанням і вбудованими пристроями.

- Вразливості, спричинені діяльністю людини: залежність від певних осіб, прогалини в обізнаності та навчанні, прогалини в дисципліні та неправомірне обмеження доступу.

- Фізичні вразливості та вразливості середовища: недостатній контроль (фізичного) доступу, невіддалене розміщення обладнання, невідповідний контроль температури/вологості та неналежне кондиціонування у приміщеннях, де знаходиться обладнання.

- Операційні вразливості: відсутність керування змінами, неналежний розподіл обов'язків, відсутність контролю за встановленням програмного забезпечення, відсутність контролю над обробкою та зберіганням медіафайлів, відсутність контролю над системним зв'язком, неналежний контроль доступу або недоліки в процедурах контролю доступу, неналежний запис та/або перегляд записів системної діяльності, неадекватний контроль над ключами шифрування, неадекватне звітування, обробка та/або вирішення інцидентів безпеки, а також неадекватний моніторинг та оцінка ефективності засобів контролю безпеки.

- Вразливості безперервності бізнесу та дотримання нормативних вимог: недоречні, відсутні або невідповідні процеси для належного управління бізнес-ризиками; неадекватне планування безперервності бізнесу та дій у разі виникнення надзвичайних ситуацій; а також неадекватний моніторинг та оцінка відповідності керівним політикам і нормам.

У багатьох сферах, перерахованих тут, виявлення вразливості критично залежить від ініціативи керівництва та подальших дій. Такі методи, як інтерв'ю, анкетування, перегляд попередніх оцінок ризиків та аудиторських звітів, а також контрольні списки – усі вони сприяють створенню ефективного уявлення ландшафту вразливості.

Національна база даних вразливостей та загальна система оцінки вразливостей. Розглянемо більш детально сферу технічних вразливостей. Видатним ресурсом є національна база даних про вразливості NIST – National Vulnerability Database (NVD), і відповідна загальна система оцінки вразливостей – Common Vulnerability Scoring System (CVSS), описана в NISTIR

7946, Посібник із впровадження CVSS [21–25]. NVD – це вичерпний список відомих технічних вразливостей систем, апаратного та програмного забезпечення. CVSS забезпечує відкриту структуру для передачі характеристик вразливостей. CVSS визначає вразливість як помилку, недолік, слабкість або відкритість програми, системного пристрою чи сервісу, що може призвести до збою конфіденційності, цілісності чи доступності.

Отже, модель CVSS намагається забезпечити повторювані та точні вимірювання, одночасно дозволяючи користувачам переглядати базові характеристики вразливості, які використовуються для створення числових оцінок. CVSS надає загальну систему вимірювання для галузей промисловості, організацій та урядів, які вимагають точних і послідовних оцінок використання вразливостей та їхнього впливу.

Розуміння CVSS дозволяє оцінити широкий спектр вразливостей, які впливають на системи. Крім того, систематизована схема для оцінки вразливостей у CVSS є корисною для розробки подібного системного підходу до інших вразливостей, таких як ті, що пов'язані з організаційними питаннями, політикою та процедурами, а також фізичною інфраструктурою. На сьогоднішній день CVSS широко прийнятий і використовується для кількісного розрахунку рівня вразливості мережного обладнання [21, 22].

Кожен запис NVD містить наступну інформацію [20]:

- унікальний словниковий ідентифікатор вразливостей і ризиків – Common Vulnerabilities and Exposure (CVE);
- опис вразливості;
- посилання на веб-сайти та інші посилання з інформацією, пов'язаною з вразливістю;
- метрики CVSS.

Метрики загальної системи оцінки вразливостей – CVSS. Існує 14 метрик CVSS, розділених за трьома групами [23]. У табл. 3 перелічені окремі показники та показані рівні, визначені для кожного з них. У кожному випадку рівні вказані від найвищого до найнижчого. По суті, підрахунок балів здійснюється таким чином: для кожної виявленої вразливості NVD надає рівень для кожного показника в базовій групі на основі характеристик вразливості. Наприклад, метрика вектора атаки вказує, чи можна атаку запустити віддалено через мережу чи через Інтернет, запускати лише через мережу, до якої підключено як джерело атаки, так і цільову систему, має бути здійснено за допомогою локального входу, або вимагає фізичного доступу до машини. Чим віддаленіша атака, тим більше джерел атаки можливе, а отже, тим серйозніша вразливість. Ця інформація є безцінною, оскільки дозволяє користувачам зрозуміти характеристики вразливості.

Як показано в табл. 3, кожен рівень метрики має описову назву [23]. Крім того, CVSS призначає числове значення за шкалою від 0,0 до 10,0, де 10,0 є найсерйознішою проблемою безпеки. Числові оцінки для показників у групі базових показників поміщаються в

рівняння, визначене в CVSS, яке створює сукупний базовий показник безпеки в діапазоні від 0,0 до 10,0.

здатності вразливості в одному компоненті програмного забезпечення впливати на ресурси, що виходять за

Таблиця 3 – Загальна класифікація метрик CVSS [23]

Група базових метрик		Група часових метрик	Група базових метрик
Можливість використання	Вплив		
Вектор атаки: – Мережа – Прилеглий – Локальний – Фізичний Складність атаки: – Низька – Висока Необхідність привілеїв: – Немає – Низька – Висока Взаємодія з користувачем: – Немає – Вимагається Область застосування: – Без змін – Змінена	Вплив на конфіденційність: – Високий – Низький – Немає Вплив на цілісність: – Високий – Низький – Немає Вплив на доступність: – Високий – Низький – Немає	Зрілість коду експлоїту: – Не визначений – Високий – Функціональний – Доказ концепції – Недоведений Рівень виправлення: – Не визначений – Обхідний шлях – Тимчасове виправлення – Офіційне виправлення Достовірність звіту: – Не визначена – Підтверджена – Обгрунтована – Невідома	Вимоги конфіденційності: – Не визначені – Високі – Середні – Низькі Вимоги цілісності: – Не визначені – Високі – Середні – Низькі Вимоги доступності: – Не визначені – Високі – Середні – Низькі

Базова група метрик представляє внутрішні характеристики вразливості, які є незмінними протягом часу та серед користувачів. Він складається з трьох наборів показників [23]:

1. Можливість використання (Exploitability): ці показники відображають легкість і технічні засоби, за допомогою яких використовується вразливість. Показники:

- Вектор атаки, який вказує, наскільки віддаленим може бути зловмисник від вразливого компонента.
- Складність атаки передає рівень складності, необхідний зловмиснику для використання вразливості після ідентифікації цільового компонента. Складність оцінюється як висока, якщо зловмисник не може здійснити атаку за власним бажанням, але повинен докласти певних зусиль для підготовки або виконання.
- Необхідні привілеї характеризують доступ, потрібний зловмиснику для використання вразливості. Значення: «немає/none» (привілейований доступ не потрібен), «низький/low» (базові привілеї користувача) і «високий/high» (права адміністратора).
- Взаємодія з користувачем вказує, чи має брати участь інший користувач, крім зловмисника, для успішної атаки.

2. Вплив (Impact): ці показники вказують на ступінь впливу на основні цілі безпеки – конфіденційність, цілісність і доступність. У кожному з цих випадків оцінка відображає найгірший результат, якщо уражено більше ніж один компонент. Для кожної з трьох цілей вводяться аналогічні значення впливу: «високий/high» (повна втрата конфіденційності, цілісності або доступності), «низький/low» (певні втрати) і «немає/none» (відсутність впливу).

3. Сфера застосування (Scope): цей показник знаходиться в групі базових показників, хоча він є дещо незалежним від решти груп. Він стосується

межі його можливостей, або на привілеї. Прикладом є вразливість у віртуальній машині, яка дозволяє зловмиснику видалити файли в операційній системі хоста.

Зазвичай базові та часові метрики визначаються аналітиками бюлетенів вразливостей, постачальниками програмних засобів безпеки або програмного забезпечення, оскільки вони мають кращу інформацію щодо характеристик вразливостей, ніж користувачі. Однак показники, що стосуються середовища, визначаються користувачами, оскільки вони найкраще можуть оцінити потенційний вплив вразливості у своєму власному середовищі.

Група часових метрик представляє характеристики вразливості, які змінюються з часом, але не в середовищі користувача. Він складається з трьох показників. У випадку, коли значення такої метрики «не визначено», цей показник слід пропустити в рівнянні оцінки.

Зрілість коду експлоїту оцінює поточний стан методів експлоїту або доступність коду. Загальнодоступний простий у використанні код експлоїта збільшує кількість потенційних зловмисників, включаючи некваліфікованих осіб, тим самим підвищуючи серйозність вразливості. Рівні відображають ступінь доступності та придатності експлоїта для використання вразливості.

Рівень виправлення вимірює ступінь доступності виправлення.

Достовірність звіту вимірює ступінь впевненості в існуванні вразливості та достовірність відомих технічних деталей.

Група метрик середовища фіксує характеристики вразливості, пов'язані з IT-середовищем користувача. Це дає змогу аналітику налаштувати оцінку CVSS залежно від важливості ураженого IT-активу для

організації користувача, вимірюючого з точки зору конфіденційності, цілісності та доступності.

Висновки. Таким чином, у даній роботі було проаналізовано стандарти побудови програмно-конфігурованих мереж. Зазначено, що SDN є сучасним підходом до проектування, побудови та експлуатації інфокомунікаційних мереж шляхом розділення площин управління та передавання даних. Завдяки такому розподілу мережі набувають безпосередньої програмованості та динамічності, а також абстрагування функціональних можливостей рівня інфраструктури. Крім того, віртуалізація мережних функцій стала стандартизованим способом розробки, впровадження та керування мережними службами. Дана концепція передбачає заміну спеціальних пристроїв мережної інфраструктури, таких як маршрутизатори та брандмауери, стандартними серверами, комутаторами, сховищем, хмарою.

Проте, зростаючий інтерес до SDN та їхнього розгортання дозволили виявити їхні недоліки у боротьбі із загрозами кібербезпеці. На сучасному етапі розвитку архітектур SDN доведено, що типові проблеми безпеки у програмно-конфігурованих мережах перш за все проявляються в таких аспектах, як шкідливе програмне забезпечення, вразливість контролера, легітимність та узгодженість правил потоків, проблема стандартизації північного інтерфейсу, безпека комунікацій в процесі використання південного інтерфейсу. Таким чином, основними причинами проблем безпеки є власне архітектура SDN, зовнішні шкідливі атаки, недостатність контролю доступу та засобів шифрування. Базуючись на дослідженнях типових проблем безпеки, стає очевидним, що централізоване керування та функції програмованості SDN надають зловмисникам потужні та зручні канали атак.

Проведений аналіз вразливостей площини даних SDN і функціональних можливостей засобів маршрутизації щодо протидії можливим атакам показав перспективність використання засобів безпечної маршрутизації на основі базових метрик критичності вразливостей для підвищення рівня мережної безпеки площини даних SDN.

Аналіз стандарту CVSS щодо кількісного розрахунку рівня вразливості мережного обладнання довів доцільність його використання під час розробки та дослідження перспективних підходів до безпечної маршрутизації у площині даних програмно-конфігурованих мереж.

Список використаної літератури

1. Sabella A., Irons-Mclean R., Yannuzzi M. *Orchestrating and automating security for the internet of things: Delivering advanced security capabilities from edge to cloud for IoT*. Cisco Press, 2018. 1008 p.
2. Kurose J. F., Ross K. *Computer networking*. 8th Edition. Pearson, 2020 775 p.
3. *The Alliance for Telecommunications Industry Solutions (ATIS)*. URL: <http://www.atis.org/> (дата звернення: 03.05.2023).
4. *The Broadband Forum Member (BBF)*. URL: <https://www.broadband-forum.org/> (дата звернення: 03.05.2023).
5. *The European Telecommunications Standards Institute (ETSI)*. URL: <http://www.etsi.org/technologies-clusters/technologies/nfv> (дата звернення: 03.05.2023).

6. *The Institute of Electrical and Electronics Engineers (IEEE)*. URL: <https://sdn.ieee.org/> (дата звернення: 03.05.2023).
7. *The International Council on Large Electrical Systems (CIGRE)*. URL: <http://www.cigre.org/> (дата звернення: 03.05.2023).
8. *The International Telecommunication Union Telecommunication Standardization Sector (ITU-T)*. URL: <http://www.itu.int/en/ITU-T/sdn/Pages/default.aspx> (дата звернення: 03.05.2023).
9. *The Internet Engineering Taskforce (IETF)*. URL: <https://ietf.org> (дата звернення: 03.05.2023).
10. *The Internet Research Task Force (IRTF)*. URL: <https://irtf.org/concluded/sdnrg> (дата звернення: 03.05.2023).
11. *The Internet Society (ISOC)*. URL: <https://www.internetsociety.org/> (дата звернення: 03.05.2023).
12. *The Metro Ethernet Forum (MEF)*. URL: <https://mef.net/> (дата звернення: 03.05.2023).
13. *The Open Data Centre Alliance (ODCA)*. URL: <https://opendatacenteralliance.org/> (дата звернення: 03.05.2023).
14. *OpenDaylight*. URL: <https://www.opendaylight.org/> (дата звернення: 03.05.2023).
15. *The Open Networking Foundation (ONF)*. URL: <https://www.opennetworking.org/> (дата звернення: 03.05.2023).
16. *The Open Platform for NFV (OPNFV)*. URL: <https://www.opnfv.org> (дата звернення: 03.05.2023).
17. *The Optical Internetworking Forum (OIF)*. URL: <http://www.oiforum.com/> (дата звернення: 03.05.2023).
18. Liu Y., Zhao B., Zhao P., Fan P., Liu H. A survey: Typical security issues of software-defined networking. *China Communications*. 2019. Vol. 16(7). P. 13–31.
19. Sagare A. A., Khondoker R. Security analysis of SDN routing applications. *SDN and NFV Security. Lecture Notes in Networks and Systems*. Springer, Cham, 2018. Vol. 30. P. 1–17.
20. Yerenenko O., Persikov M., Lemeshko V., Altaki B. Research and development of the secure routing flow-based model with load balancing. *Проблеми телекомунікацій*. 2021. № 2(29). С. 3–14.
21. Свдодименко М. О., Шаповалова А. С., Шаповал М. М. Поточкова модель маршрутизації із врахуванням ризиків інформаційної безпеки за допомогою базових метрик критичності вразливостей. *Проблеми телекомунікацій*. 2020. № 1(26). С. 48–62.
22. Yevdokymenko M., Yerenenko O., Shapovalova A., Shapoval M., Porokhniak V., Rogovaya N. Investigation of the Secure Paths Set Calculation Approach Based on Vulnerability Assessment. *Workshop Proceedings of the MoMLeT+DS 2021: 3rd International Workshop on Modern Machine Learning Technologies and Data Science*. June 5, 2021. Lviv-Shatsk, Ukraine. P. 207-217.
23. Stallings W. *Effective Cybersecurity: Understanding and Using Standards and Best Practices*. Addison-Wesley, 2019. 800 p.
24. Common Vulnerability Scoring System v3.0: Examples, Forum of Incident Response and Security Teams. URL: <https://www.first.org/cvss/examples> (дата звернення: 03.05.2023).
25. NIST National Vulnerability Database, URL: <https://nvd.nist.gov> (дата звернення: 03.05.2023).

References (transliterated)

1. Sabella A., Irons-Mclean R., Yannuzzi M. *Orchestrating and automating security for the internet of things: Delivering advanced security capabilities from edge to cloud for IoT*. Cisco Press, 2018. 1008 p.
2. Kurose J. F., Ross K. *Computer networking*. 8th Edition. Pearson, 2020. 775 p.
3. *The Alliance for Telecommunications Industry Solutions (ATIS)*. URL: <http://www.atis.org/> (accessed 03.05.2023).
4. *The Broadband Forum Member (BBF)*. URL: <https://www.broadband-forum.org/> (accessed 03.05.2023).
5. *The European Telecommunications Standards Institute (ETSI)*. URL: <http://www.etsi.org/technologies-clusters/technologies/nfv> (accessed 03.05.2023).
6. *The Institute of Electrical and Electronics Engineers (IEEE)*. URL: <https://sdn.ieee.org/> (accessed 03.05.2023).
7. *The International Council on Large Electrical Systems (CIGRE)*. URL: <http://www.cigre.org/> (accessed 03.05.2023).
8. *The International Telecommunication Union Telecommunication Standardization Sector (ITU-T)*. URL: <http://www.itu.int/en/ITU-T/sdn/Pages/default.aspx> (accessed 03.05.2023).
9. *The Internet Engineering Taskforce (IETF)*. URL: <https://ietf.org>.

10. *The Internet Research Task Force (IRTF)*. URL: <https://irtf.org/concluded/sdnrg> (accessed 03.05.2023).
11. *The Internet Society (ISOC)*. URL: <https://www.internetsociety.org/> (accessed 03.05.2023).
12. *The Metro Ethernet Forum (MEF)*. URL: <https://mef.net/> (accessed 03.05.2023).
13. *The Open Data Centre Alliance (ODCA)*. URL: <https://opendatacenteralliance.org/> (accessed 03.05.2023).
14. *OpenDaylight*. URL: <https://www.opendaylight.org/> (accessed 03.05.2023).
15. *The Open Networking Foundation (ONF)*. URL: <https://www.opennetworking.org/> (accessed 03.05.2023).
16. *The Open Platform for NFV (OPNFV)*. URL: <https://www.opnfv.org>.
17. *The Optical Internetworking Forum (OIF)*. URL: <http://www.oiforum.com/> (accessed 03.05.2023).
18. Liu Y., Zhao B., Zhao P., Fan P., Liu H. A survey: Typical security issues of software-defined networking. *China Communications*. 2019, vol. 16(7), pp. 13–31.
19. Sagare A. A., Khondoker R. Security analysis of SDN routing applications. *SDN and NFV Security. Lecture Notes in Networks and Systems*. Springer, Cham, 2018, vol. 30, pp. 1–17.
20. Yeremenko O., Persik M., Lemeshko V., Altaki B. Research and development of the secure routing flow-based model with load balancing. *Problemy telekomunikatsii [Telecommunication Problems]*. 2021, no. 2(29), pp. 3–14.
21. Yevdokymenko M., Shapovalova A., Shapoval M. Potokova model marshrutyzatsii iz vrakhuvanniam ryzykiv informatsiinoi bezpeky za dopomohoiu bazovykh metryk krytychnosti vrazlyvosti [Flow model of routing taking into account information security risks using basic vulnerability criticality metrics]. *Problemy telekomunikatsii [Telecommunication Problems]*. 2020, no. 1(26), pp. 48–62.
22. Yevdokymenko M., Yeremenko O., Shapovalova A., Shapoval M., Porokhniak V., Rogovaya N. Investigation of the Secure Paths Set Calculation Approach Based on Vulnerability Assessment. *Workshop Proceedings of the MoMLeT+DS 2021: 3rd International Workshop on Modern Machine Learning Technologies and Data Science*, June 5, 2021, Lviv-Shatsk, Ukraine, pp. 207–217.
23. Stallings W. *Effective Cybersecurity: Understanding and using standards and best practices*. Addison-Wesley, 2019. 800 p.
24. Common Vulnerability Scoring System v3.0: Examples, Forum of Incident Response and Security Teams. URL: <https://www.first.org/cvss/examples> (accessed 03.05.2023).
25. NIST National Vulnerability Database, URL: <https://nvd.nist.gov> (accessed 03.05.2023).

Надійшла (received) 10.05.2023

UDC 004.056.5

A. I. LEVTEROV, Candidate of Technical Sciences, Professor, Kharkiv National Automobile and Highway University, Head of the Department of Informatics and Applied Mathematics, Kharkiv, Ukraine, e-mail: lai@khadi.kharkov.ua, ORCID: <https://orcid.org/0000-0001-6586-1061>

H. A. PLIEKHOVA, Candidate of Technical Sciences, Docent, Kharkiv National Automobile and Highway University, Associate Professor at the Department of Informatics and Applied Mathematics, Kharkiv, Ukraine, e mail: plehovaanna11@gmail.com, ORCID: <https://orcid.org/0000-0002-6912-6520>

M. V. KOSTIKOVA, Candidate of Technical Sciences, Docent, Kharkiv National Automobile and Highway University, Associate Professor at the Department of Informatics and Applied Mathematics, Kharkiv, Ukraine, e mail: kmv_topaz@ukr.net, ORCID: <https://orcid.org/0000-0001-5197-7389>

N. G. BEREZHNA, Candidate of Technical Sciences, Docent, State Biotechnological University, Associate Professor at the Department of Transport Technologies and Logistics, Kharkiv, Ukraine, e mail: bereg_nat@ukr.net, ORCID: <https://orcid.org/0000-0001-8740-3387>

A. O. OKUN, Candidate of Technical Sciences, Docent, National Technical University "Kharkiv Polytechnic Institute", Associate Professor at the Department of Computer Modeling and Integrated Forming Technologies, Kharkiv, Ukraine, e mail: okunanton@gmail.com, ORCID: <https://orcid.org/0000-0002-6467-4229>

ENHANCING SECURITY IN SOFTWARE-DEFINED NETWORKING THROUGH ROUTING TECHNIQUES EXPLORATION

In today's world, network security is a key issue of information security. Virtual Networks have become an integral part of modern IT infrastructure, which presents us with challenges in the field of security. One solution to this problem is the use of software-defined networking (SDN), which provides a means to control and manage network traffic. However, as with any technology, SDN has its vulnerabilities that must be considered when deploying it. One of the tools that helps to take into account the vulnerabilities of network infrastructure is the Common Vulnerability Scoring System (CVSS) standard. It allows you to quantify the level of vulnerability of the infrastructure, which enables effective network protection. Analysis of the CVSS standard is an important stage in the development of a network security strategy. This paper analyzes the standards for building software-configured networks. It is noted that SDN is a modern approach to the design, construction, and operation of information communication networks. Using SDN makes it possible to directly program and dynamically manage the network, as well as to abstract the functionality of the infrastructure layer. However, the growing interest in SDN has revealed the shortcomings of their application in the fight against cybersecurity threats. The SDN architecture itself, external malicious attacks, and insufficient access control and encryption tools were found to be the main security challenges. The use of secure routing tools based on vulnerability metrics is proposed to increase the level of SDN data plane network security. According to the conducted analysis of SDN data plane vulnerabilities and the functionality of routing tools, the authors recommend using the CVSS standard to quantify the level of infrastructure vulnerability during the development and research of promising approaches to secure routing in the data plane of software-configured networks.

Keywords: SDN, CVSS, vulnerability, NFV, data plane, NVD, Software-Defined Networking.

Повні імена авторів / Author's full names

Автор 1 / Author 1: Левтеров Андрій Іванович, Levterov Andrii Ivanovych

Автор 2 / Author 2: Пলেখова Ганна Анатоліївна, Pliekhova Hanna Anatoliivna

Автор 3 / Author 3: Костікова Марина Володимирівна, Kostikova Maryna Volodymyrivna

Автор 4 / Author 4: Бережна Наталія Георгіївна, Berezhna Nataliia Heorhiivna

Автор 5 / Author 5: Окунь Антон Олександрович, Okun Anton Oleksandrovych

УПРАВЛІННЯ В ТЕХНІЧНИХ СИСТЕМАХ

CONTROL IN TECHNICAL SYSTEMS

DOI: 10.20998/2079-0023.2023.01.03

УДК 681.515

О. А. ЖУЧЕНКО, доктор технічних наук, професор, Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», м. Київ, Україна; e-mail: azhuch@ukr.net; ORCID: <https://orcid.org/0000-0001-5611-6529>

А. П. КОРОТИНСЬКИЙ, кандидат технічних наук (PhD), старший викладач, Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», м. Київ, Україна; e-mail: ihfantkor@gmail.com; ORCID: <https://orcid.org/0000-0002-6309-5970>

А. О. АБРАМОВА, кандидат технічних наук (PhD), доцент, Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», м. Київ, Україна; e-mail: alla_abramova@ukr.net; ORCID: <https://orcid.org/0000-0003-3475-8584>

Д. Є. ЧЕПОВ, студент, Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», м. Київ, Україна; e-mail: pyroducttail@gmail.com; ORCID: <https://orcid.org/0009-0004-5458-8214>

РОЗРОБКА СИСТЕМИ КЕРУВАННЯ ВИПАРНИМ АПАРАТОМ У ПРОЦЕСІ ВИРОБНИЦТВА ОЦТОВОЇ КИСЛОТИ НА БАЗІ MPC-РЕГУЛЯТОРА

У більшості випадків очищення вихідного продукту процесу виробництва оцтової кислоти здійснюється шляхом ректифікації. Оскільки процес ректифікації полягає в розділенні сумішей на фази по температурі кипіння, то стає зрозуміло, що досить важливим технологічним параметром є температура та концентрація суміші, що обробляється. Зазвичай, дані технологічні параметри задаються кип'ятильником при ректифікаційній колоні, який ще зветься кубом колони. Саме тому існує потреба у якісному та ефективному керуванні кип'ятильником, як одним із важливих технологічних об'єктів. У роботі розроблена та проаналізована система керування кип'ятильником на базі MPC-регулятора. Робота MPC-регулятора базується на визначенні поведінки об'єкта та прогнозування майбутньої поведінки, після чого за рахунок оптимізації керування на певному проміжку часу здійснюється виведення його на бажану траєкторію. Проміжок часу на якому здійснюється розрахунок оптимальної траєкторії ще називається горизонтом. Визначення горизонту впливає на характер здійснюваного керування, тому його вибір повинен бути обраний виходячи з характеру динамічних властивостей об'єкту керування. Саме тому при розробці даних регуляторів потрібно досліджувати характер впливу даних параметрів на перебіг процесу. В результаті дослідження було отримано стратегії керування для різних значень горизонтів прогнозування та керування. Проаналізувавши результати роботи регулятора при різних налаштуваннях горизонтів прогнозування та керування можна дійти висновку, що при збільшенні горизонту прогнозування збільшується час перехідної характеристика, але також зменшується перерегулювання. Тому при виборі оптимального значення потрібно керуватися впливом перерегулювань на процес.

Ключові слова: система керування, горизонт, MPC-регулятор, ректифікація, керування, перехідна характеристика

Вступ. Оцтова кислота є дуже широко використовуваним продуктом в промисловості, проте у побуті її розчин використовують не рідше. Сфери застосування у неї надзвичайно широкі. В промисловості її застосовують починаючи з виробництва полімерів та закінчуючи харчовою галуззю. В побуті ж вона найчастіше використовується в якості харчової добавки чи розчинника. Тому нагальною стала потреба її виробництва у промислових масштабах. Для виробництва використовують цілий ряд технологічних схем. Виробництвом оцтової кислоти займалися ще у давнину, але тоді процес був досить тривіальним, і був представлений звичайною ферментацією продуктів виноробства. Таким методом користувались досить

довго, аж до початку активного розвитку промисловості, який прийшовся на XIX – XX ст., саме в той період почались дослідження стосовно того, які можливості збільшення обсягів виробництва та якості вихідного продукту можливо застосувати. Також це пов'язано з різноманіттям доступної в той час сировини, чого раніше не було. Саме завдяки цьому ми маємо зараз таке різноманіття технологій виробництва.

Разом з розвитком обчислювальної техніки з'являється можливість автоматизації в промисловості загалом та зокрема для виробництва оцтової кислоти. Це дозволило підвищити ефективність вже існуючих схем виробництва та впровадити нові змогливі до чіткого підтримання параметрів виробництва. З моменту впровадження автоматизації децю змінилась

© Жученко О. А., Коротинский А. П., Абрамова А. О., Чепов Д. Є., 2023



Дослідницька стаття: Цю статтю опубліковано видавництвом *НТУ «ХПІ»* у збірнику «Вісник Національного технічного університету "ХПІ" Серія: Системний аналіз, управління та інформаційні технології». Ця стаття поширюється за міжнародною ліцензією [Creative Commons Attribution \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/). Конфлікт інтересів: Автор/и заявив/или про відсутність конфлікту.



конструкція приладів, що її реалізують, але загальний принцип не зазнав значних змін. Сучасні мікроконтролери дозволяють впроваджувати більш комплексні закони керування, що засновані на сучасних методах, а також окрім цього, з'явилась можливість впровадження SCADA-систем для цілих виробничих ліній.

Завдяки значному зростанню обчислювальної потужності комп'ютерів, є змога досить складного моделювання технологічних об'єктів. Це дозволяє заощаджувати чималі кошти на розробленні нових технологічних апаратів, так як є можливість провести дослідження роботи апарату на комп'ютерній моделі, не витрачаючи кошти на виготовлення реального апарату.

Отже, розроблення автоматизованих систем керування та дослідження ефективності їх роботи на виробництві оцтової кислоти є актуальним напрямком дослідження.

Аналіз останніх досліджень і публікацій. У процесі виробництва оцтової кислоти важливу роль відіграє якість вихідного продукту, що у великій мірі залежить від ефективності процесу очищення (ректифікації). Ефективність процесу очищення від таких технологічних параметрів як температура та концентрація суміші, що задаються кип'ятильником при RECTИФІКАЦІЙНІЙ колоні. Виходячи із вищесказаного, є потреба у якісному та ефективному керуванні кип'ятильником, як одним із важливих технологічних об'єктів у виробництві оцтової кислоти. Це підтверджено у роботах авторів [1–2], а також розглянуто особливості, які потрібно врахувати для покращення параметрів процесу RECTИФІКАЦІЇ. У праці [3] досліджено роботу системи при наявності в ній нелінійних елементів та розподіленого характеру параметрів протікання процесу, що використовується при побудові регулятора об'єкту.

Постановка завдання. Метою даної роботи є розробка та дослідження ефективності автоматизованої системи керування на базі MPC-регулятора процесу виробництва оцтової кислоти та порівняння з класичним ПІД-регулятором. Дослідження проводяться у вигляді порівняльного аналізу роботи ПІД-, MPC-регуляторів.

Матеріали та результати дослідження. Для проведення порівняльної характеристики MPC-регулятора, пропонується реалізувати умовний базовий

рівень точності та ефективності роботи автоматизованої системи керування. Вирішення даної задачі базується на синтезі класичного ПІД-регулятора. *Синтез системи керування на базі ПІД-регулятора.* Найпростішим варіантом для налаштування регулятора є схема одноканального ПІД-регулятора [4]. В якості каналу керування використаємо витрату нагріваючої пари – температура парів оцтової кислоти на виході з апарату. Для налаштування будемо використовувати Matlab Simulink [5–7]. Для початку необхідно зібрати схему в програмному середовищі Matlab Simulink. Схему системи керування представлено на рис. 1. Для налаштування параметрів регулятора буде використано утиліту PID Tuner [8, 9]. Вікно налаштування контролера наведено на рис. 2. В результаті налаштування отримали наступні параметри регулятора: $P = 0.02319$, $I = 0.00878$, $D = -0.145037$, $N = 0.15989$.

Синтез системи керування на базі MPC-регулятора. Робота MPC регулятора базується на визначенні поведінки об'єкта та прогнозування майбутньої поведінки, після чого за рахунок оптимізації керування на певному проміжку часу здійснюється виведення його на бажану траєкторію. В основі алгоритму керування знаходиться задача мінімізації функції вартості. Зазвичай використовується лінійно-квадратичний функціонал якості [10]. Важливою особливістю даного методу керування є те, що розрахунок відбувається не одразу для всієї траєкторії, а лише для заданого проміжку часу. Ця особливість відрізняє MPC від LQR регулятора. Завдяки цьому LQR має кращу глобальну продуктивність, а MPC в свою чергу краще досягає локальних оптимумів. Проміжок часу на якому здійснюється розрахунок оптимальної траєкторії ще називається горизонтом. Визначення горизонту впливає на характер здійснюваного керування, тому його вибір повинен бути обраний виходячи з характеру динамічних властивостей об'єкту керування [11, 12].

Для синтезу MPC-регулятора будемо використовувати Matlab MPC Designer [5–7]. Для початку було зібрано систему керування в середовищі Simulink. Схема системи керування з використанням MPC регулятора в середовищі Simulink представлена на рис. 3. Для подальшого налаштування контролера необхідно задати його структуру. Для цього необхідно відкрити вікно MPC Structure. Вікно налаштування структури MPC контролера представлено на рис. 4.

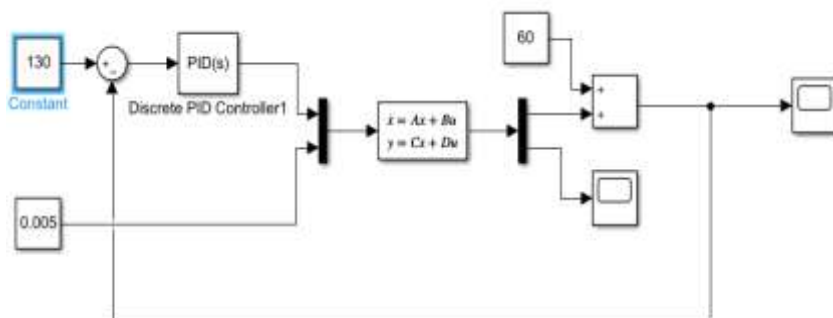


Рис. 1. Схема системи керування на основі ПІД-регулятора

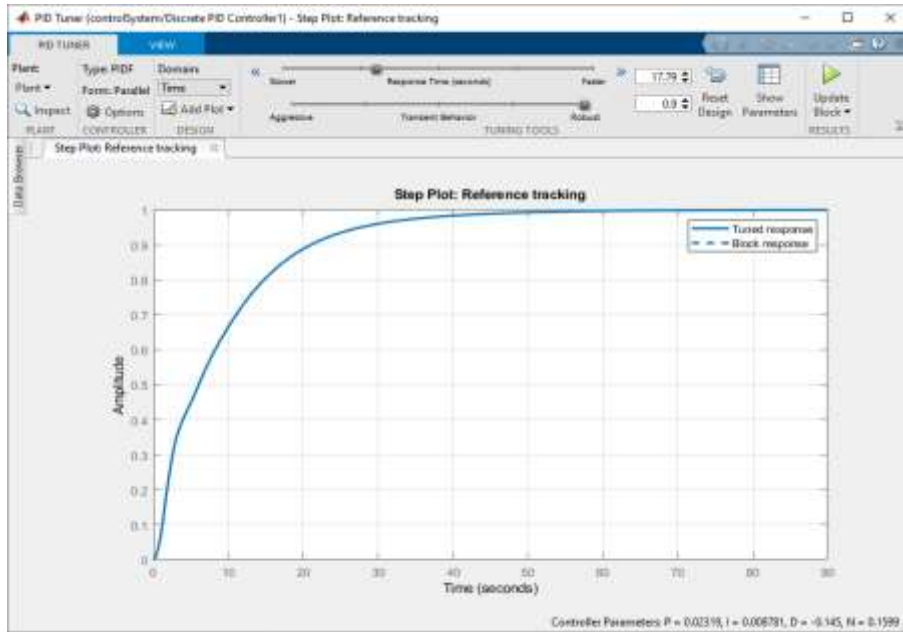


Рис. 2. Налаштування контролера

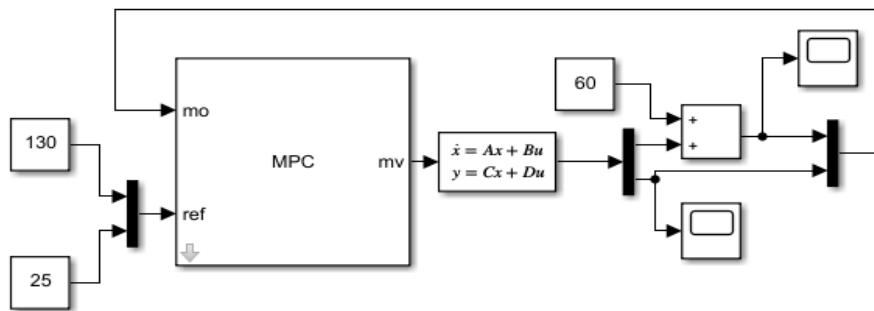


Рис. 3. Схема системи керування в Matlab Simulink

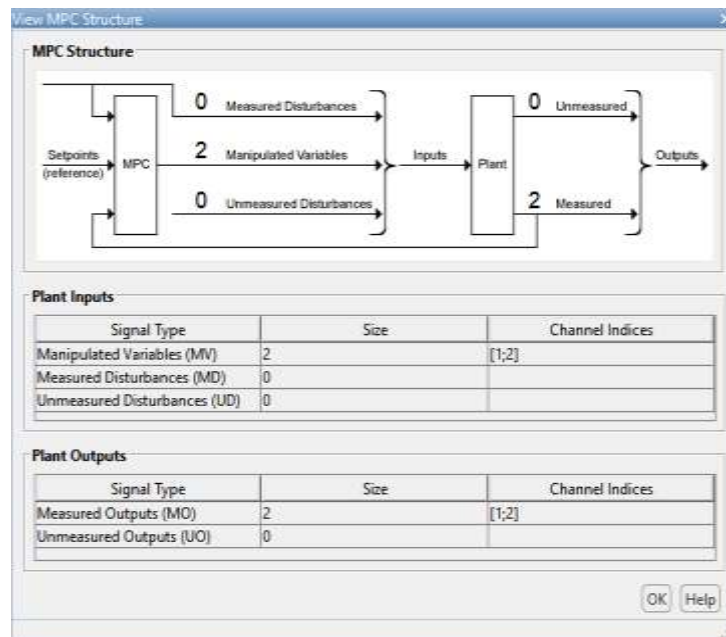


Рис. 4. Вікно налаштування структури регулятора

Налаштувавши структуру системи керування необхідно провести налаштування контролера. Параметрами, які впливають на роботу контролера, є горизонт управління, який визначає кількість кроків розрахунку оптимального керування, горизонт прогнозування, який визначає кількість кроків для розрахунку майбутньої поведінки.

Дослідження впливу горизонту прогнозування MPC-регулятора. Основною особливістю MPC-регулятора є розрахунок майбутньої поведінки на певному

горизонті прогнозування, тому його визначення сильно впливає на поведінку контролера, і правильний вибір горизонту прогнозування – важлива частина налаштування регулятора [12]. Для вибору буде використано експериментальний метод, тобто проведемо дослідження при різних налаштуваннях горизонту та оберемо найбільш відповідний нашому об'єкту варіант. Почнемо з горизонту прогнозування в 15 кроків, результат цього налаштування представлено на рис. 5. Наступ-

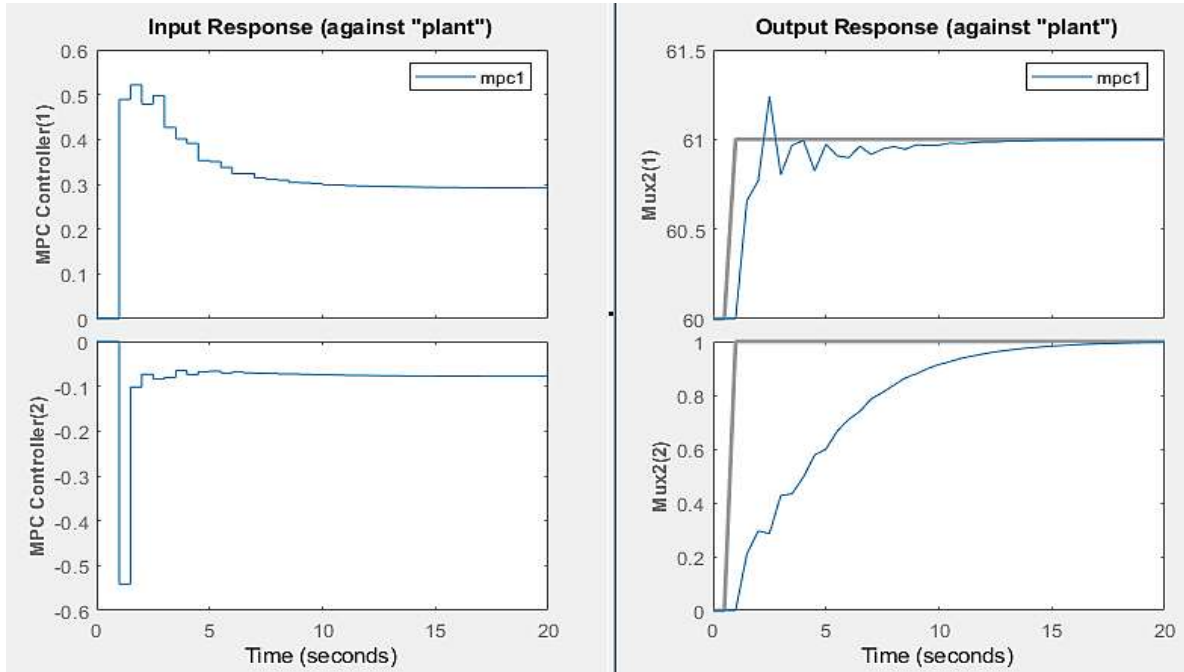


Рис. 5. Робота регулятора при горизонті прогнозування 15 кроків

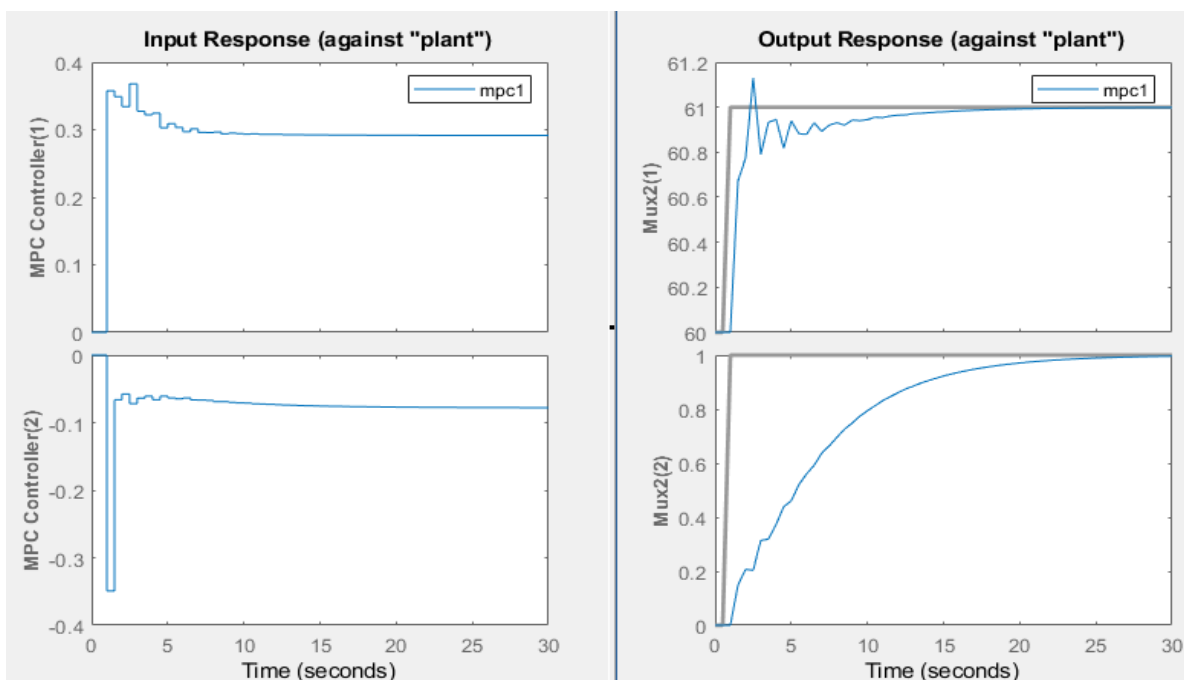


Рис. 6. Робота регулятора при горизонті прогнозування 25 кроків

ним значенням буде 25 кроків – рис. 6. Проаналізувавши результати роботи регулятора при різних налаштуваннях горизонту прогнозування можна дійти висновку, що при збільшенні горизонту прогнозування збільшується час перехідної характеристики, але також зменшується перерегулювання. Тому при виборі оптимального значення потрібно керуватися впливом перерегулювань на процес. У нашому випадку перерегулювання не є занадто критичними в певних межах, тому оптимальним для досліджуваного об'єкта буде

значення горизонту прогнозування 15 кроків.

Дослідження впливу горизонту управління MPC-регулятора. Ще одним параметром значення якого необхідно налаштувати у відповідності з необхідною поведінкою для технологічного об'єкта є горизонт управління. Горизонт управління визначає кількість кроків для розрахунку оптимального керування [12].

Для вибору цього параметру скористаємось також експериментальним методом. Проаналізуємо поведінку системи керування при різних налаштуваннях

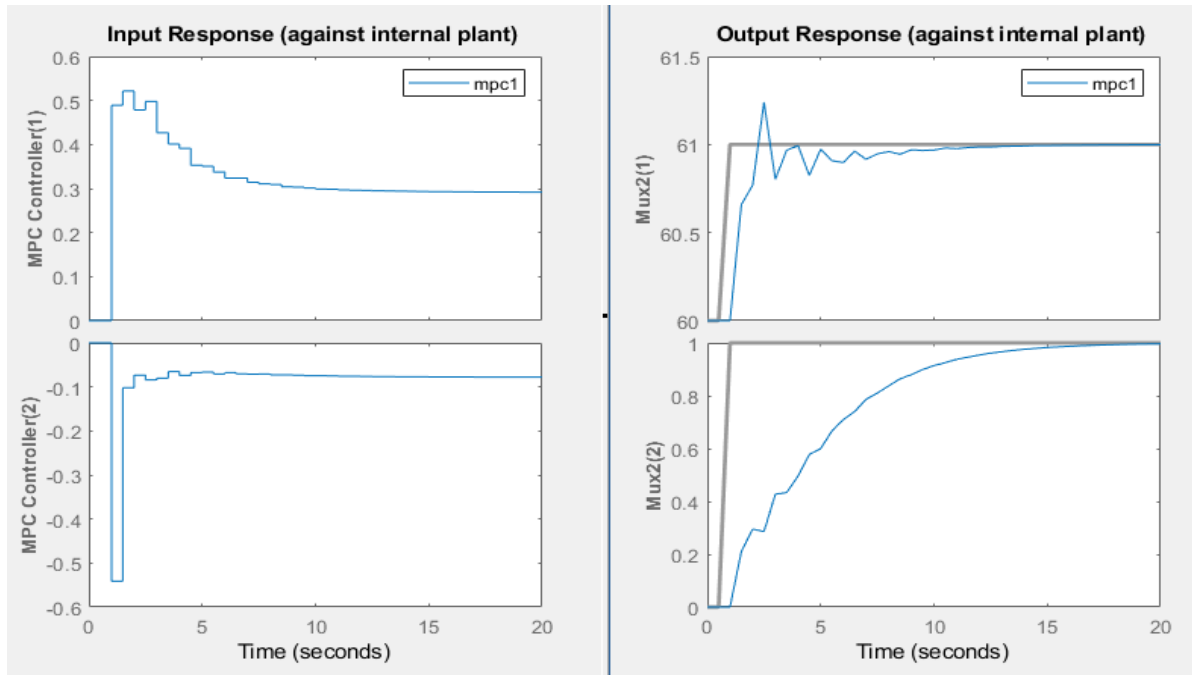


Рис. 7. Робота регулятора при горизонті управління 2 кроки

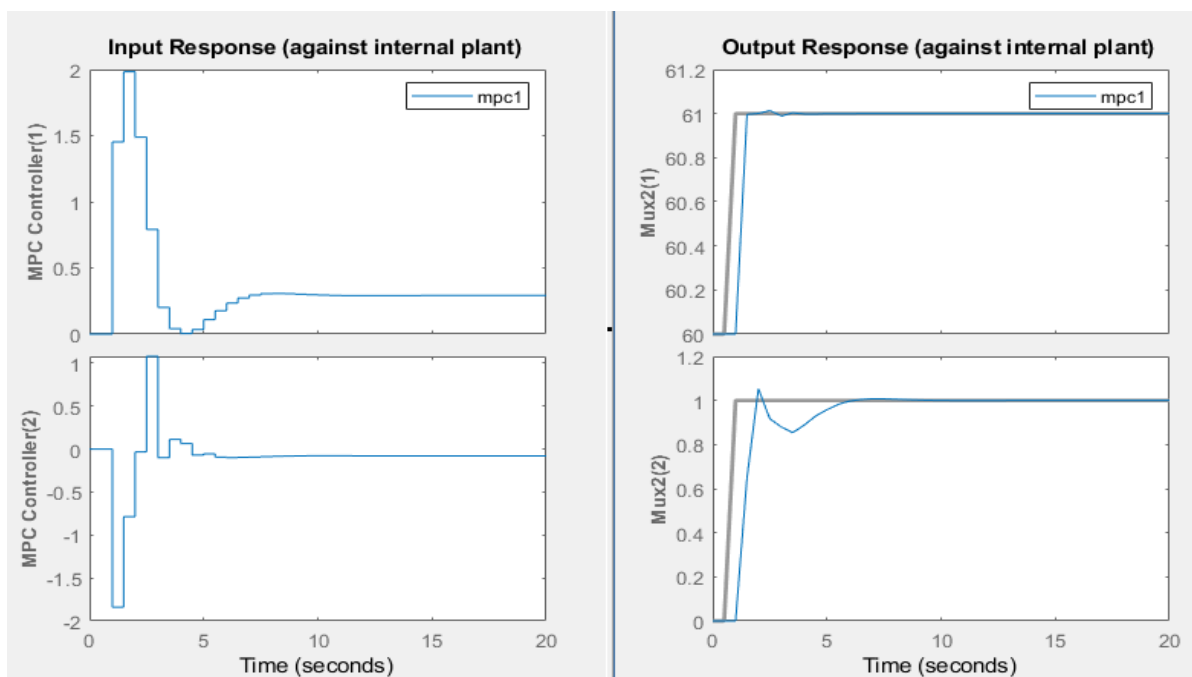


Рис. 8. Робота регулятора при горизонті управління 5 кроків

даного параметра і оберемо значення, що найкраще відповідає характеру нашого процесу.

Мінімальним значенням горизонту управління було обрано 2, поведінку контролера при цьому значенні представлено на рис. 7. Результат роботи системи при горизонті управління 5 кроків представлено на рис. 8. На рис. 9 наведено результати керування при горизонті управління 8 кроків. Максимальним значенням горизонту управління було обрано 10 кроків, поведінка системи при даному значенні зображена

на рис. 10.

Проаналізувавши отримані результати можна дійти висновку, що поведінка при значенні горизонту управління 8 кроків є найбільш прийнятною, так як дає найбільшу швидкість перехідного процесу та найменше значення перерегулювання.

Дослідження ефективності системи керування з MPC-регулятором. Порівняльна характеристика ПД-та MPC-регуляторів представлена на рис. 11.

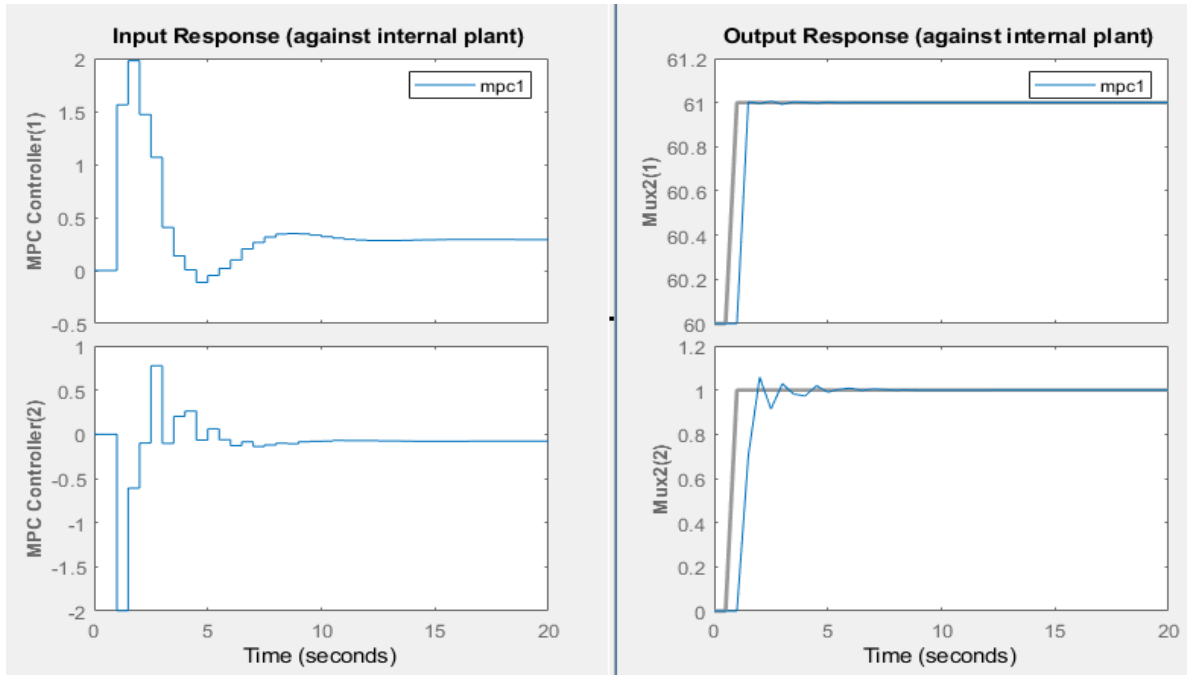


Рис. 9. Робота регулятора при горизонті управління 8 кроків

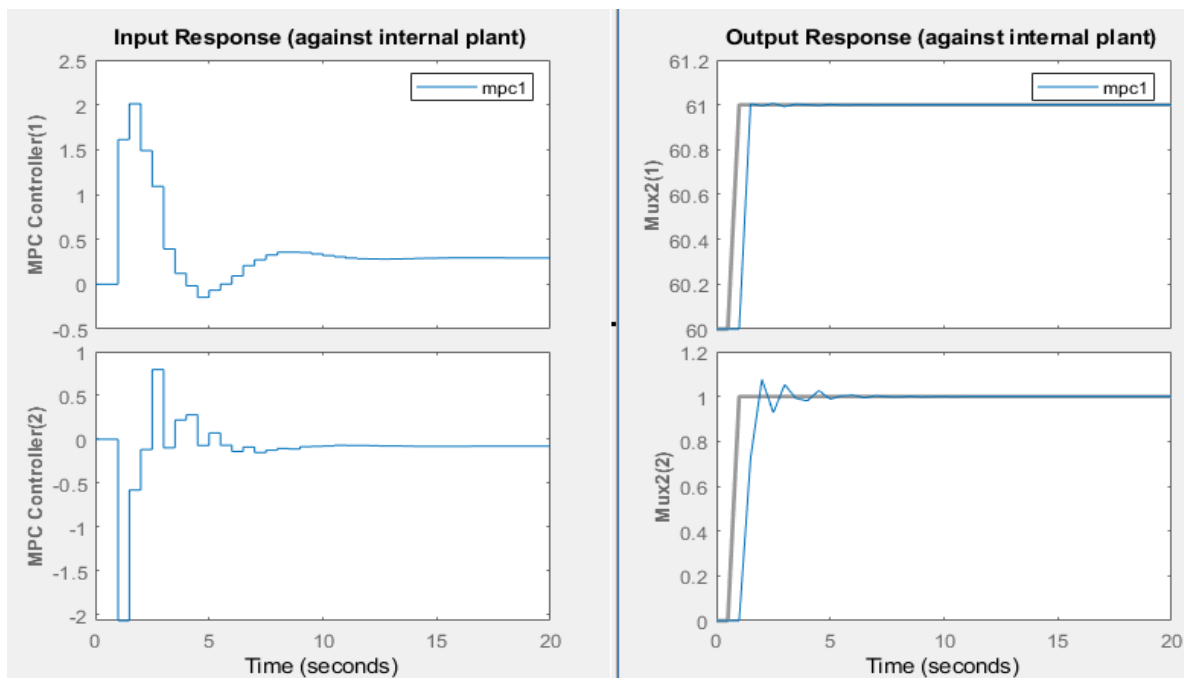


Рис. 10. Робота регулятора при горизонті управління 10 кроків

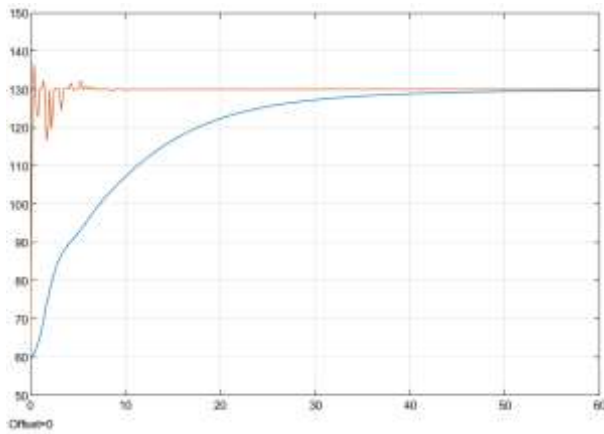


Рис. 11. Порівняльна характеристика регуляторів

За порівняльною характеристикою можна побачити, що MPC-регулятор дає на порядок швидше перехідний процес та з перерегулюванням, але перерегулювання становить менше 10%, тому воно не вплине на якість технологічного процесу. Отже для нашого апарату MPC-регулятор дає значно кращий результат.

Висновки.

Проведено налаштування параметрів одноканального ПІД-регулятора, отримано перехідні характеристики температури та витрати парів оцтової кислоти у виробництві оцтової кислоти.

Проведено дослідження впливу горизонту прогнозування MPC-регулятора. Особливістю такого регулятора є вплив горизонту прогнозування та горизонту управління на налаштування. Для правильного вибору горизонту прогнозування та управління використано експериментальний метод дослідження (кроки горизонту прогнозування 5, 10, 15, 20, 25, кроки горизонту управління 2, 5, 8, 10).

Проведено порівняльний аналіз ефективності керування заданими технологічними параметрами на основі ПІД- та MPC-регуляторів. Отримані результати підтверджують, що ефективнішим регулятором для керування необхідними технологічними параметрами у виробництві оцтової кислоти є MPC-регулятор. В результаті роботи було продемонстровано характер впливу горизонтів керування та прогнозування на характер керування технологічним об'єктом.

Список використаної літератури

1. Дударенко В. О. Система керування процесом ректифікації спирту. URL: <https://essuir.sumdu.edu.ua/bitstream-download/123456789/81324/1/Dudarenko%20V.O..pdf;jsessionid=0424F64A8EB9CC97B860E30C99DF9830> (дата звернення: 20.12.2022).
2. Jun-Li Gao *Research on Boiler Water Supply Control System Based on AT89C55 and Fractional order PID Algorithm* URL: <https://www.sciencedirect.com/science/article/pii/S1877050919307951> (дата звернення: 01.11.2022).
3. Lung Chien I. *Design and control of acetic acid dehydration system via heterogeneous azeotropic distillation* URL: https://www.researchgate.net/publication/244117428_Design_and_control_of_acetic_acid_dehydration_system_via_heterogeneous_azeotropic_distillation (дата звернення: 05.10.2022).
4. Гаєв Є. О., Нестеренко Б. М. Універсальний математичний пакет MATLAB і типові задачі обчислювальної математики. Навчальний посібник. Київ: НАУ, 2004. 176 с.

5. Гераймчук М. Д., Лазарев Ю. Ф., Толочко Т. О. *Модельовання систем у середовищі MATLAB-SIMULINK*. URL: <https://ela.kpi.ua/handle/123456789/30394> (дата звернення: 20.03.2023).
6. Qin S. J., Badgwell T. A. A survey of industrial model predictive control technology. *Control Engineering Practice*. 2003. Vol. 11, iss. 7. P. 733–764.
7. Parisio A, Rikos E, Glielmo L. Stochastic model predictive control for economic/environmental operation management of microgrids: An experimental case study [J]. *Journal of Process Control*. 2016. Vol. 43. P. 24–37.
8. Ricker N. L., Lee J. H. Nonlinear Model Predictive Control of the Tennessee Eastman Challenge Process, «*Computers & Chemical Engineering*», Vol. 19, no. 9, 2005. P. 961–981.
9. Alberto Bemporad, Manfred Morari, N. Lawrence Ricker Model Predictive Control Toolbox For Use with MATLAB. *The MathWorks Inc.* 3 Apple Hill Drive Natick, MA 01760-2098. 128 p.
10. Мовчан А. П., Степанець О. В. *Методи статичної оптимізації*. Навчальний посібник. Київ: НТУУ «КПІ», 2012. 138 с.
11. Zhuchenko O., Korotynskiy A. Investigation influence of predict horizon of MPC-regulator for control of the baking process. *Slovak international scientific journal*. 2020. Vol.2, no. 37., pp. 29–34.
12. Жученко О. А., Коротинський А. П. Дослідження впливу горизонту прогнозування MPC-регулятора при синтезі системи керування процесом випалювання. *Матеріали VI Міжнародної науково-технічної Інтернет-конференції «Сучасні методи, інформаційне, програмне та технічне забезпечення систем керування організаційно-технічними та технологічними комплексами», 20 листопада 2019*. Київ: НУХТ, 2019. С. 44–45.

References (transliterated)

1. Dudarenko V. A. Sistema upravleniya protsessom rektifikatsii spirta. [System for controlling the alcohol rectification process] Available at: <https://essuir.sumdu.edu.ua/bitstream-download/123456789/81324/1/Dudarenko%20V.O..pdf;jsessionid=0424F64A8EB9CC97B860E30C99DF9830> (accessed 20.12.2022).
2. Jun-Li Gao *Research on Boiler Water Supply Control System Based on AT89C55 and Fractional order PID Algorithm*. Available at: <https://www.sciencedirect.com/science/article/pii/S1877050919307951> (accessed 01.11.2022).
3. Lung Chien I. *Design and control of acetic acid dehydration system via heterogeneous azeotropic distillation* Available at: https://www.researchgate.net/publication/244117428_Design_and_control_of_acetic_acid_dehydration_system_via_heterogeneous_azeotropic_distillation (accessed 05.10.2022).
4. Hayev E. O., Nesterenko B. M. *Universalnyy matematichnyy paket MATLAB i tipovi zadachi obchysluvalnoyi matematyky*. [Universal mathematical package MATLAB and typical problems of computational mathematics]. Navchalnyy posibnyk. Kyiv, NAU Publ., 2004. 176 p.
5. Herayimchuk M. D., Lazaryev Y. F., Tolochko T. O. Modelyuvannya system u seredovyschi MATLAB-SIMULINK Available at: <https://ela.kpi.ua/handle/123456789/30394> (accessed 20.03.2023).
6. Qin S. J., Badgwell T. A. A survey of industrial model predictive control technology. *Control Engineering Practice*. 2003, vol. 11, iss. 7, pp. 733–764.
7. Parisio A, Rikos E, Glielmo L. Stochastic model predictive control for economic/environmental operation management of microgrids: An experimental case study [J]. *Journal of Process Control*. 2016, vol. 43, pp. 24–37.
8. Ricker N. L., Lee J. H. Nonlinear Model Predictive Control of the Tennessee Eastman Challenge Process, «*Computers & Chemical Engineering*». 2005, vol. 19, no. 9, pp. 961–981.
9. Alberto Bemporad, Manfred Morari, N. Lawrence Ricker Model Predictive Control Toolbox For Use with MATLAB. *The MathWorks Inc.* 3 Apple Hill Drive Natick, MA 01760-2098. 128 p.
10. Movchan A. P., Stepanets O. V. *Metody statychnoyi optymizatsiyi*. Navchalnyy posibnyk. [Methods of static optimization]. Kyiv, NTUU «КПІ» Publ., 2012. 138 p.
11. Zhuchenko O., Korotynskiy A. Investigation influence of predict horizon of MPC-regulator for control of the baking process. *Slovak international scientific journal*. 2020, vol. 2, no. 37, pp. 29–34.
12. Zhuchenko O. A., Korotynskiy A. P. Doslidzhennya vplyvu horyzontu prohnozuvannya MPC-rehulyatora pry syntezi systemy

keruvannya protsesom vypalyvannya. [Study of the influence of the prediction horizon of the MPC controller in the synthesis of the firing process control system] *Materialy VI Mizhnarodnoi naukovo-tekhnichnoi Internet-konferentsiyi «Suchasni metody, informatsiynе, prohramne ta tekhnichne zabezpechennya system keruvannya*

orhanizatsiyno-tekhnichnymy ta tekhnolohichnymy kompleksamy», 20 lystopada 2019. Kyiv, NUKHT, 2019, pp. 44–45.

Надійшла (received) 02.05.2023

UDC 681.515

O. A. ZHUCHENKO, Doctor of Technical Sciences, Full Professor, National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”; Kyiv, Ukraine, e-mail: azhuch@ukr.net; ORCID: <https://orcid.org/0000-0001-5611-6529>

A. P. KOROTYNSKYI, Candidate of Technical Sciences (PhD), Senior Lecture, National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”; Kyiv, Ukraine, e-mail: ihfantkor@gmail.com; ORCID: <https://orcid.org/0000-0002-6309-5970>

A. O. ABRAMOVA, Candidate of Technical Sciences (PhD), Docent, National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”; Kyiv, Ukraine, e-mail: alla_abramova@ukr.net; ORCID: <https://orcid.org/0000-0003-3475-8584>

D. E. CHEPOV, student, National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”; Kyiv, Ukraine, e-mail: pyroducktail@gmail.com; ORCID: <https://orcid.org/0009-0004-5458-8214>

DEVELOPMENT OF THE EVAPORATOR CONTROL SYSTEM IN THE ACETIC ACID PRODUCTION PROCESS BASED ON THE MPC-REGULATOR

In most cases, the initial product of the acetic acid production process is purified by rectification. Since the rectification process consists in separating the mixtures into phases based on the boiling point, it becomes clear that the temperature and concentration of the processed mixture are quite important technological parameters. Usually, these technological parameters are set by the boiler at the distillation column, which is also called the cube of the column. That is why there is a need for high-quality and effective management of the boiler as one of the important technological objects. A boiler control system based on the MPC regulator was developed and analyzed in this work. The operation of the MPC regulator is based on determining the behavior of the object and predicting its future behavior, after which it is brought to the desired trajectory by optimizing the control over a certain period of time. The period of time during which the calculation of the optimal trajectory is carried out is also called the horizon. The definition of the horizon affects the nature of the control performed, so its choice should be based on the nature of the dynamic properties of the control object. That is why, when developing these regulators, it is necessary to study the nature of the influence of these parameters on the course of the process. As a result of the study, management strategies were obtained for different values of forecasting horizons. Having analyzed the results of the controller at different settings of the forecasting horizon, it can be concluded that when the forecasting horizon is increased, the time of the transient characteristic increases, but overregulation also decreases. Therefore, when choosing the optimal value, it is necessary to be guided by the impact of adjustments on the process.

Keywords: control system, horizon, MPC regulator, rectification, control, transient characteristic.

Повні імена авторів / Author's full names

Автор 1 / Author 1: Жученко Олексій Анатолійович, Zhuchenko Oleksii Anatoliyovich

Автор 2 / Author 2: Коротинський Антон Петрович, Korotynskyi Anton Petrovych

Автор 3 / Author 3: Абрамова Алла Олександрівна, Abramova Alla Oleksandrivna

Автор 4 / Author 4: Чепов Дмитро Євгенійович, Chepov Dmytro Evgeniyovych

О. С. МАЗМАНІШВІЛІ, доктор фізико-математичних наук, професор, старший науковий співробітник ННЦ ХФТІ, м. Харків, Україна, e mail: mazmanishvili@gmail.com, ORCID: <https://orcid.org/0000-0003-0373-0626>

М. Г. РЕШЕТНЯК, кандидат фізико-математичних наук, старший науковий співробітник ННЦ ХФТІ, м. Харків; Україна, e mail: nreshetnyak@kpt.kharkov.ua, ORCID: <https://orcid.org/0000-0002-2345-6789>

Г. Ю. СИДОРЕНКО, кандидат технічних наук, доцент, доцент кафедри системного аналізу та інформаційно-аналітичних технологій НТУ «ХПІ», доцент кафедри моделювання систем і технологій ХНУ ім. В. Н. Каразіна, м. Харків, Україна, e mail: annsydorenko01@gmail.com, ORCID: <https://orcid.org/0000-0002-0761-2793>

ЕЛЕКТРОННІ ПУЧКИ У ГРАДІЄНТНОМУ МАГНІТНОМУ ПОЛІ: УПРАВЛІННЯ ПЕРЕТВОРЕННЯМ ПОЗДОВЖНЬОГО РУХУ У ПОПЕРЕЧНИЙ

В роботі розглянуто рух електронів у циліндричному магнітному полі з потенціалом градієнтного вигляду. Отримано, що у вибраному полі вихідний рух електронів уздовж поздовжньої осі перетворюється на радіальний рух. Визначено, що таке перетворення обумовлено впливом соленоїдального магнітного поля з великим поздовжнім градієнтом. Перетворення поздовжнього напрямку руху на поперечне виявилось стійким в діапазоні енергій 20...55 кеВ електронів і в інтервалі 5...50 мм радіальних розмірів пучка частинок. За допомогою програмного засобу вивчено основні залежності руху електронного пучка в заданому соленоїдальному магнітному полі. В даній роботі приведені результати чисельного моделювання траєкторій електронів у градієнтному магнітному полі зі вторинноemisійним катодом кругової форми, розташованим у середній системі. Для вивчення механізму стійкості по відношенню до магнітного поля використано два експериментально реалізовані магнітні поля. На основі цих двох полів чисельно синтезовано масиви додаткових 4 полів. Для сукупності з 6 названих полів чисельно вивчена робота гармати, коли частка відчуває стійке перетворення напрямку руху. Показано, що при заданій енергії електрона та фіксованому магнітному полі параметром, що визначає поворот частинок, є градієнт магнітного поля на межі ділянки вольту. Отримано, що ефект повороту має місце для розглянутого інтервалу радіальних розмірів пучка, що призводить до фокусування частинок. Показана можливість на основі регулювання поля в цілому керувати вертикальною координатою сфокусованого пучка, тим самим дано інтерпретацію порогової залежності реєстрації електронів на детекторі. Досліджено залежність формування підсумкового розподілу частинок від амплітуди та градієнта магнітного поля вздовж осі системи. Наводяться результати чисельного моделювання руху електронного потоку. На основі моделі руху електронного потоку розглянуто характеристики результуючого електронного пучка. В даній статті показано, що пучок, що має радіальні розміри 5...50 мм, фокусується по вертикалі на ділянці розміром 1 мм.

Ключові слова: електронний пучок, магнетронна гармата, динаміка електронів, перетворення напрямку руху, керування, градієнтне магнітне поле, математичне моделювання.

Вступ. Магнетронні гармати з холодними вторинноemisійними катодами перспективні для використання їх при створенні потужних НВЧ-приладів, прискорювальної техніки у зв'язку з високою щільністю струму, конструктивною простотою, стабільністю емісії катодів тощо [1–3]. Як об'єкт дослідження, магнетронна гармата має ту перевагу, що вона може бути модифікована для проведення експериментів різного типу [4–6]. Застосування інтенсивних магнітних полів з підвищеним градієнтом їхньої напруженості дає можливість розглянути нові явища в просторовій картині еволюції електронних пучків. Незважаючи на складність аналітичного дослідження [6–8], для отримання кількісних залежностей цілком можливе застосування обчислювальних методів та порівняння числових результатів із даними експерименту.

Раніше в ході чисельних експериментів [8–11] були виявлені випадки перетворення поздовжнього руху електронів на радіальне, які мали місце у градієнтному магнітному полі магнетронної гармати. У роботі на основі гамільтонового формалізму руху електронів у магнітному полі синтезовано програмний засіб [6, 7], що дозволяє провести чисельне моделювання динаміки трубчастих електронних потоків у циліндричному магнітному полі. Разом з досвідченими даними наводяться результати чисельного моделювання руху електронного потоку. Отримані результати моделювання свідчать про явище регульованого поперечного руху.

У ННЦ «ХФТІ» на основі магнетронної гармати було створено прискорювач електронів, який використовує осьовий електронний пучок для опромінення металевих мішеней [1, 3, 6] (див. рис. 1). Як об'єкт для досліджень магнетронна гармата має ту перевагу, що, з одного боку, вона може бути модифікована для проведення експериментів різного типу, з іншого боку, незважаючи на складність виконання аналітичного дослідження, для отримання кількісних залежностей є можливим використання обчислювальних методів та порівняння числових результатів із даними експерименту. Одним із можливих напрямів є комбінація магнітного поля в гарматі, створеного соленоїдами, та полем додаткових постійних магнітів [7–9, 12–15].

Об'єктом даної роботи є вивчення параметрів електронного пучка при його транспортуванні в інтенсивному градієнтному магнітному полі соленоїда і побудова обчислювальної моделі руху електронного потоку в області зростання поля.

Метою дослідження було створення моделі руху електронного пучка з енергією в десятки кеВ у значимому градієнтному магнітному полі, вивчення формування траєкторій електронів, залежності динаміки частинок від початкових умов та розподілу магнітного поля вздовж осі системи. Наводяться результати моделювання руху електронного потоку. У процесі моделювання є можливість в умовах експери-

© Мазманішвілі О. С., Решетняк М. Г., Сидоренко Г. Ю., 2023



Дослідницька стаття: Цю статтю опубліковано видавництвом НТУ «ХПІ» у збірнику «Вісник Національного технічного університету «ХПІ» Серія: Системний аналіз, управління та інформаційні технології». Ця стаття поширюється за міжнародною ліцензією [Creative Commons Attribution \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/). Конфлікт інтересів: Автор/и заявив/или про відсутність конфлікту.



менту встановити явище перетворення поздовжнього руху на радіальне.

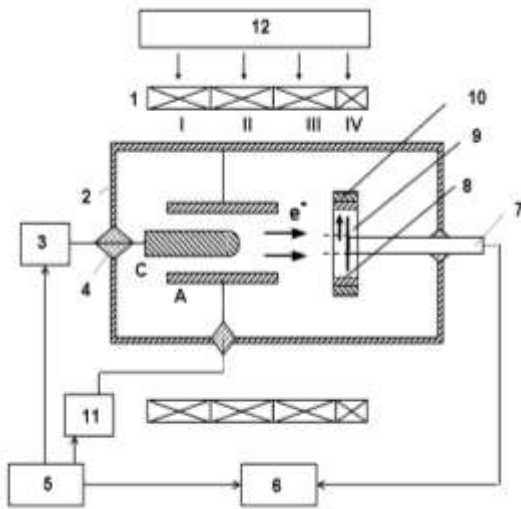


Рис. 1. Блок-схема експериментальної установки.

1 – секції соленоїда (I, II, III, IV), 2 – вакуумний об'єм, 3 – високовольтний імпульсний генератор, 4 – ізолятор, 5 – блок синхронізації, 6 – система, що вимірює, 7 – стрижень, що центрує, 8 – мідна вставка, 9 – циліндр Фарадея, 10 – постійний магніт, 11 – генератор, 12 – джерело струмів для управління магнітним полем соленоїда, А – анод, С – катод

1. Методика досліджень

Проведено дослідження по формуванню електронного пучка магнетронною гарматою (діаметр 40 мм, діаметр анода 78 мм) та вимірюванню його параметрів при напрузі на катоді 20...55 кВ.

В Табл. 1 наведено параметри електронного пучка в залежності від конфігурації магнітного поля. Проаналізувавши її, можна зробити висновок, що за певних умов пучок енергії 55 кеВ як ціле рухався по осі системи і реєструвався на циліндрі Фарадея. За інших умов пучок на циліндрі Фарадея не реєструвався, оскільки, рухаючись в радіальному напрямку, влучав на мідну вставку. При цьому спостерігалися випадки, коли зміна напрямку пучка та місце реєстрації мала стрибкоподібний характер.

Таблиця 1 – Типи магнітного поля та напрямки руху пучка

Магнітне поле (рис. 2)	Струм в осьовому напрямку, А	Струм в радіальному напрямку, А	Струм в чотирьох секціях соленоїда, А
B_1	30	0	35, 0, 18, 54
B_2	0	18	35, 0, 31, 0
B_3	0	17	-33, 0, -16, -54

На рис. 2 представлено розподіл поля B_1 (ромбики). Конфігурація цього поля, досить плавна, забезпечує проходження пучка вздовж вертикальної осі і реєстрацію на циліндрі Фарадея. На рис. 2 також наведені конфігурації полів B_2 (кружечки) і B_3 (хрестики), у яких напруженість (модулем) не перевищує 0.1 Т, при цьому градієнт поля досягав 9 Т/м. При синтезі цих полів у схемі рис. 1 був додатково використаний

постійний кільцевий магніт напруженості 0.21 Т, напрямок поля якого було протилежне напрямку поля соленоїдів. В результаті конфігурація магнітного поля була знакозмінною і мала ділянки з великим градієнтом. Рух частинок у полях таких конфігурацій якісно відрізнявся. Зокрема, пучок електронів як ціле повертався на 90° і рухався в радіальному напрямку, причому струм пучка на циліндрі Фарадея пороговим чином звертався в нуль.

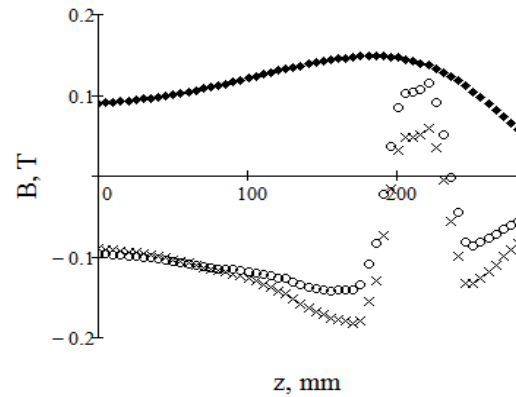


Рис. 2. Три магнітні поля B ; ромбики – поле B_1 , кола та хрестики – поля B_2 та B_3

Таким чином, виникає задача з вивчення властивостей електронного пучка при його транспортуванні в інтенсивному магнітному градієнтному полі і побудови обчислювальної моделі руху електронного потоку в області зростання поля.

2. Математичне моделювання. Конфігурації полів B_1 , B_2 і B_3 відрізняються від поля B_1 явно вираженою градієнтною властивістю в інтервалі від 170 мм до 210 мм. Після емісії пучок частинок, рухаючись уздовж осі системи, влітає в область з великим градієнтом поля, взаємодія з яким визначає траєкторію кожної частинки.

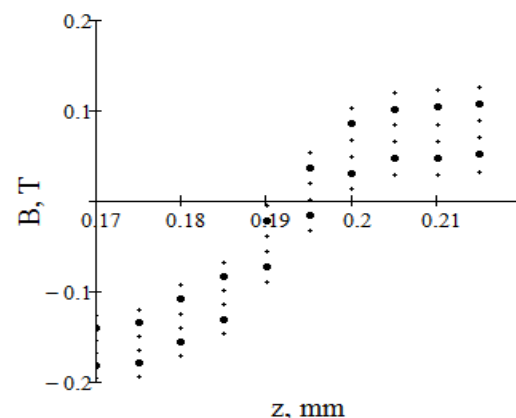


Рис. 3. Набір із 6 магнітних полів A_1, \dots, A_6 в області з великим градієнтом, кружальцями показані поля A_2 та A_5

Оскільки поля B_2 і B_3 створюються шляхом підібраних наборів варіантів струмів на обмотках секцій-соленоїдів (I, II, III, IV), оцифровані масиви цих

полів були взяті за основу подальшого моделювання. Програмний засіб [5] було використано для чисельного моделювання динаміки трубчастих електронних потоків у магнітному циліндричному полі з великим градієнтом. З цією метою було створено числові масиви A_1, \dots, A_6 з 6 додаткових магнітних полів:

$$A_k = B_2 + \frac{k-2}{3}(B_2 - B_3). \quad (1)$$

Таким чином, при $k = 2$ отримуємо поле $A_2 = B_2$ а при $k = 5$ отримуємо $A_5 = B_3$ (див. рис. 3).

При моделюванні був обраний пучок з 9 частинок, початкова умова на старті яких був рівномірний розподіл за радіусом від $r_0 = 28$ мм до $r_0 = 85$ мм у площині $(x_0; y_0)$.

Типовий приклад динаміки пучка з 9 частинок наведено на рис. 4. Можна бачити, що для всіх початкових радіусів траєкторії електронів схожі, а саме, в градієнтній області їхній рух перетворюється з поздовжнього на поперечне. Така зміна має місце всім розглянутих початкових радіусів. При цьому здійснюється фокусування частинок уздовж вертикальної осі системи.

Для деталізації типу руху та супутніх повороту та фокусування було виконано серію чисельних експериментів, у яких змінювалися як поля A_1, \dots, A_6 , так і початкові умови на площині $(x_0; y_0)$. Стійкість руху також перевірялася шляхом варіацій енергії частинок E , а також початкових умов вертикалі z_0 .

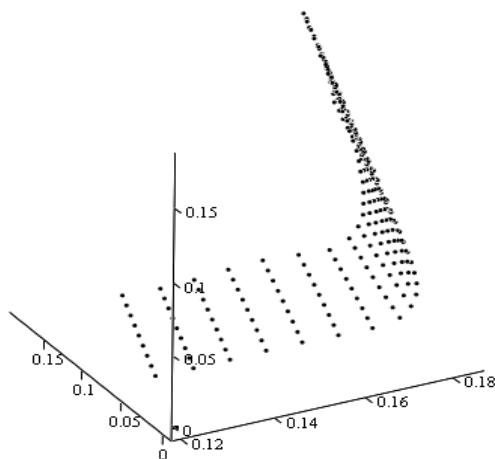


Рис. 4. Динаміка пучка (приклад із 9 частинками)

На рис. 5 наведено траєкторії частинок для двох полів градієнтного типу A_1 і A_6 , а також поля B_1 з рис. 2 зі слабо вираженою градієнтною властивістю. Можна бачити, що траєкторія частки у полі B_1 носить регулярний характер з невеликою зміною поточного радіусу $r = r(z)$ у прогнотному інтервалі від 170 мм до 210 мм. У тій же градієнтній області має місце поворот траєкторії з наступним фокусуванням, що реєструвалося для всіх шести полів A_1, \dots, A_6 .

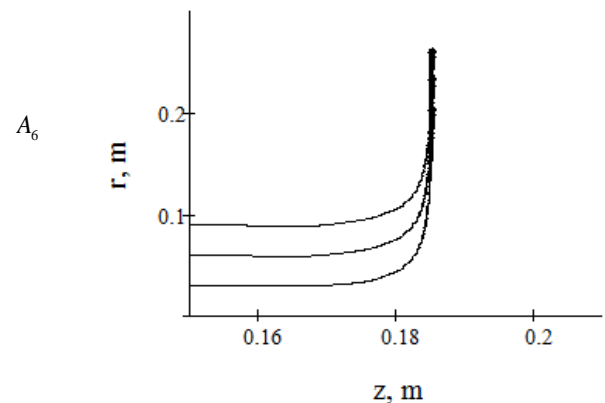
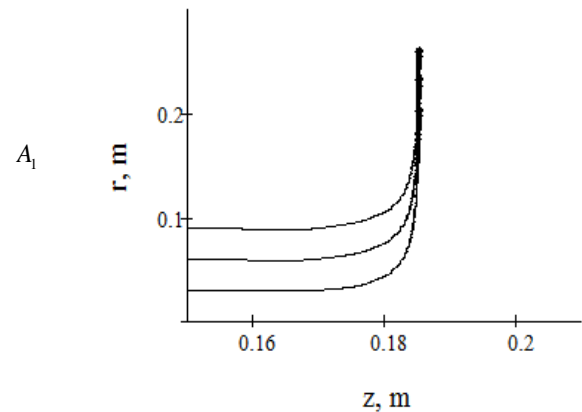
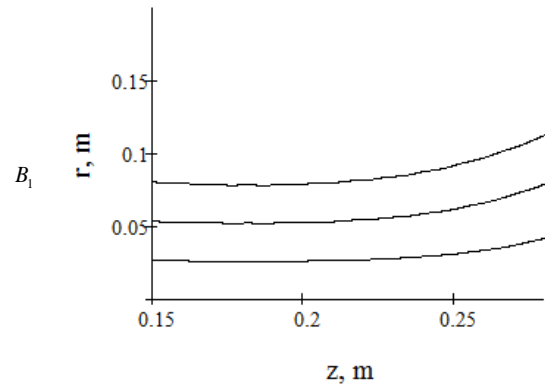


Рис. 5. Приклади траєкторій частинок

Зміщення Δz точки фокусування показано на рис. 6. При зміні параметра керування k в діапазоні параметра $0 \dots 7$ воно склало $180 \leq \Delta z \leq 197$ мм. Таким чином, область, що містить місце регульованого фокусування, має поздовжній розмір ≈ 17 мм. Сама ділянка фокусування частинки пучка, у свою чергу має розмір $\delta z \approx 2$ мм уздовж вертикальної осі, що утворений ларморовськими коливаннями електронів у циліндричному магнітному полі, чия амплітуда визначається початковим радіусом r_0 .

3. Математична модель та обговорення результатів. На основі гамільтонового формалізму руху електронів у магнітному полі побудовано програмний засіб, що дозволяє провести чисельне

модельовання динаміки електронних потоків у градієнтному магнітному полі [3, 4]. Скористаємося полярною системою координат (r, z, ϑ) та запишемо гамільтоніан електрона в магнітному полі в циліндричній системі [5, 6]

$$H = \frac{p_r^2 + p_z^2}{2m} + \frac{1}{2m} \left(\frac{p_\vartheta}{r} - e_0 A \right)^2, \quad (2)$$

де e_0 , m – заряд та маса спокою електрона, p_r , p_z , p_ϑ – канонічні імпульси, A – магнітний потенціал. Його з урахуванням азимутальної симетрії запишемо у вигляді $A = A(r, z) = rB(z)$, де $B(z)$ – напруженість магнітного поля в точці.

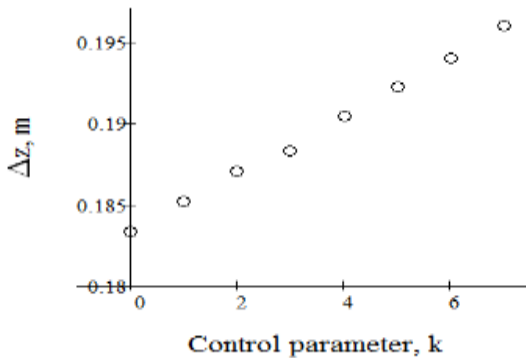


Рис. 6. Залежність керованого зміщення по вертикалі Δz від параметра вибору поля k

Рівняння руху для координат та імпульсів мають загальний вигляд

$$\left\{ \begin{array}{l} \dot{r} = \frac{p_r}{m}, \\ \dot{z} = \frac{p_z}{m}, \\ \dot{\vartheta} = \frac{1}{m} \left(\frac{p_\vartheta}{r} - e_0 r B(z) \right) \frac{1}{r}, \\ \dot{p}_r = -\frac{1}{m} \left(\frac{p_\vartheta}{r} - e_0 r B(z) \right) \left(\frac{p_\vartheta}{r^2} - e_0 B(z) \right), \\ \dot{p}_z = -\frac{1}{m} \left(\frac{p_\vartheta}{r} - e_0 r B(z) \right) \left(-e_0 r \frac{\partial}{\partial z} B(z) \right) \\ \dot{p}_\vartheta = 0. \end{array} \right. \quad (3)$$

Як видно з рівнянь (3), на формування траєкторії частинки впливає як розподіл магнітного поля B вздовж осі z , так і його похідна.

У системі (3) із 6 рівнянь розглянемо друге рівняння для z та п'яте для канонічного імпульсу p_z . З цих рівнянь випливає, що похідна dp_z/dt може зійтися до нуля, якщо

$$\left(p_\vartheta - e_0 r^2 B(z_F) \right) \frac{\partial}{\partial z} B(z) = 0. \quad (4)$$

У градієнтній області похідна $\frac{\partial}{\partial z} B(z)$ ніде не перетворюється на нуль (рис. 2, рис. 3). При $z = z_0$ імпульс p_ϑ пропорційний канонічному імпульсу p_z , тобто $(2E/E_0)^2$, де E – енергія електрона, $E_0 = 511$ кеВ – енергія спокою електрона. На початку руху імпульс p_ϑ змінюється незначно. Під час руху вздовж поздовжньої вісі електрон наближається ліворуч до градієнтної області, у якій функція $B(z)$ швидко зростає. При досить великій величині $B(z_F)$ в деякій точці $z = z_F$ значення круглої дужки в (4) обертається в нуль. Наближення частки до цієї точки z , однак, носить асимптотичний характер руху у радіальному напрямку. Сама координата z_F на осі z залежить від обраного магнітного поля (див. рис. 3) і для описаних вище полів A_0, \dots, A_7 зміщувалася від $z_F = 180$ мм при $k = 0$ до $z_F = 197$ мм при $k = 7$. Таким чином, збільшення керуючого параметра k відповідає загальному зменшенню амплітуди поля A згідно (1) і тому більш влучанню частинки в градієнтну область (див. рис. 6) при більшій координаті $z = z_F$. Можна сказати, що має місце «бульдозерний» ефект.

Розглянуто рух електронів у циліндричному магнітному полі з потенціалом градієнтного вигляду. Вивчено формування пучка з енергією 55 кеВ у такому магнітному полі. Отримано, що у вибраному полі вихідний рух електронів уздовж поздовжньої осі перетворюється на радіальний рух. Визначено, що таке перетворення обумовлено впливом соленоїдального магнітного поля з великим поздовжнім градієнтом. Перетворення поздовжнього напрямку руху в поперечне виявилось стійким у діапазоні енергій 20...55 кеВ електронів та інтервалі 5...50 мм радіальних розмірів пучка частинок. Показано, що при заданій енергії електрона та фіксованому магнітному полі параметром, що визначає поворот частинок, є градієнт магнітного поля на межі ділянки вльоту. Отримано, що ефект повороту має місце для інтервалу радіальних розмірів пучка, що призводить до фокусування частинки.

Висновки. Удосконалення магнетронної гармати, що полягає у застосуванні інтенсивних магнітних полів з підвищеним градієнтом їхньої напруженості, дає можливість розглянути нові явища в просторовій картині еволюції електронних пучків.

Показана можливість на основі регулювання поля як ціле керувати вертикальною координатою сфокусованого пучка, тим самим дано інтерпретацію порогової залежності реєстрації електронів на детекторі. Досліджено залежність формування підсумкового розподілу частинок від амплітуди та градієнта магнітного поля вздовж вісі системи. На основі чисельної моделі руху електронного потоку розглянуто характеристики результуючого електронного пучка. Показано, що пучок, що має вхідні радіальні розміри 5...50 мм, фокусується по вертикалі на ділянку 1 мм.

Список використаної літератури

- Ayzatsky M. I., Dovbnya A. N., Mazmanishvili A. S., Reshetnyak N. G., Romasko V. P., Chertishchev I. A. Studies on formation of the radially-directed electron beam generated by the magnetron gun with the secondary emission cathode. *Problems of Atomic Science and Technology. Series "Nuclear Physics Investigations"*. 2016. Issue 66, vol. 3(103). P. 78–94.
- Saveliev Y. M., Sibbet W., Parkes D. M. Self-excitation and characteristics of the crossed-fields emission electron source. *Review of Scientific Instruments*. 1999. Vol. 70, no. 12. P. 4502–4514.
- Ayzatsky N. I., Churyumov G. I., Dovbnya A. N., Zakutin V. V., Reshetnyak N. G., Starchevskiy Y. L. Generation and formation of axially symmetrical tubular electron beam in a cold metal secondary-emission cathode magnetron gun. Part 1. Experiment // *IEEE Transaction and electron device*, April, 2016. Vol. 63, no. 4. P. 1704–1709.
- Miyamoto K. *Fundamentals of Plasma Physics and Controlled Fusion*. Moscow: Fizmatlit. 2007. 424 p.
- Engelko V., Mueller G., Andreev A., Berutchev N., Komarov O., Petuchov A., Tkatchenko K. Pulsed electron beam facilities (GESA) for surface treatment of materials. *Proceedings of the 10th International Conference on Applied Charged Particle Accelerators in Medicine and Industry*, 2012. P. 412–417.
- Довбня А. Н., Закутин В. В., Решетняк Н. Г. и др. Исследование формирования пучка в ускорителе электронов с вторичноэмиссионным источником. *Вісник Харківського університету, серія фізична «Ядра, частинки, поля»*. 2006. № 732, вып. 2(30). С. 96–100.
- Mazmanishvili A. S., Reshetnyak N. G. Electron beam transversion management on exit of magnetic gun by gradient magnetic field. *Problems of Atomic Science and Technology, series "Nuclear Physics Investigations"*. 2019. Issue 72. Vol. 6(124). P. 106–113.
- Mazmanishvili A. S., Reshetnyak N. G., Shovkoplyas O. A. Beam and sector regimes of electron flows in cylindrical magnetic field of the magnetron gun. *Journal of Nano- and Electronic Physics*. 2020. Vol. 12, no. 3, 03001(5cc).
- Mazmanishvili A.S., Reshetnyak N.G. Transformation of data array of cylindrical magnetic field of the magnetron gun and problem of the radial motion of electrons. *Applied Problems of Mathematical Modeling*, 2020, vol. 3, no. 1. P. 108–116.
- Довбня А. Н., Закутин В. В., Решетняк Н. Г. и др. Исследование формирования пучка в ускорителе электронов с вторичноэмиссионным источником. *«Вісник Харківського університету», серія фізична «Ядра, частинки, поля»*. 2006. № 732, вып. 2(30). С. 96–100.
- Тузигов А. В., Александров В. С., Шевцов В. Ф. Моделирование динамики пучка в магнитном поле спиральной катушки. *Письма в ЭЧАЯ*. 2006. Т. 3, № 7 (136). С. 55–59.
- Маделунг Э. *Математический аппарат физики*. Москва, Физматгиз, 1961. 620 с.
- Мазманишвили А. С., Решетняк Н. Г., Сидоренко А. Ю. Динамика электронного пучка формируемого магнетронной пушкой с вторичноэмиссионным катодом, в спадающем магнитном поле соленоида: эксперимент и моделирование. *Вісник Національного технічного університету «ХПИ». Серія: Системний аналіз, управління та інформаційні технології*. Харків: НТУ «ХПИ», 2021. № 2 (6). С. 96–100.
- Глухих В. А., Печерский О. П., Энгелько В. И., Яценко Б. И. Модификация поверхностных свойств материалов с помощью мощных импульсных электронных пучков. *Сб. докладов X Международного совещания по применению ускорителей заряженных частиц в промышленности и медицине* (г. Санкт-Петербург, 1–4 октября, 2001). Москва: ЦНИИатоминформ, 2001. С. 445–447.
- Азарян Н. С., Баренгольц С. А., Доля С. Н., Месяц Г. А., Перельштейн Э. А. Азарян Н. С., Баренгольц С. А., Доля С. Н., Месяц Г. А., Перельштейн Э. А. Моделирование формирования электронного кольца с использованием пикосекундных электронных пучков в магнитной системе типа "касп". *Журнал технической физики*. 2010. Т. 80, вып. 4. С. 126–134.
- magnetron gun with the secondary emission cathode. *Problems of Atomic Science and Technology. Series "Nuclear Physics Investigations"*, 2016, issue 66, vol. 3(103), pp. 78–94.
- Saveliev Y. M., Sibbet W., Parkes D. M. Self-excitation and characteristics of the crossed-fields emission electron source. *Review of Scientific Instruments*, 1999, vol. 70, no. 12, pp. 4502–4514.
- Ayzatsky N. I., Churyumov G. I., Dovbnya A. N., Zakutin V. V., Reshetnyak N. G., Starchevskiy Y. L. Generation and formation of axially symmetrical tubular electron beam in a cold metal secondary-emission cathode magnetron gun. Part 1. Experiment. *IEEE Transaction and electron device*, April, 2016, vol. 63, no. 4, pp. 1704–1709.
- Miyamoto K. *Fundamentals of Plasma Physics and Controlled Fusion*. Moscow, Fizmatlit. 2007, 424 p.
- Engelko V., Mueller G., Andreev A., Berutchev N., Komarov O., Petuchov A., Tkatchenko K. Pulsed electron beam facilities (GESA) for surface treatment of materials. *Proceedings of the 10th International Conference on Applied Charged Particle Accelerators in Medicine and Industry*, 2012, pp. 412–417.
- Dovbnya A. N., Zakutin V. V., Reshetnyak N. G. etc. Issledovanie formirovaniya puchka v uskoritele elektronov vtorichnoemissionnym istochnikom. [Study of beam formation in an electron accelerator with a secondary emission source]. *Visnyk Kharkivs'kogo universytetu, seria fizychna "Yadra, chastynky, polya"* [Bulletin of Kharkiv University, physical series "Nucleus, particles, fields"], 2006, no. 732, issue. 2(30). pp. 96–100.
- Mazmanishvili A. S., Reshetnyak N. G. Electron beam transversion management on exit of magnetic gun by gradient magnetic field. *Problems of Atomic Science and Technology, series "Nuclear Physics Investigations"*, 2019, issue 72, vol. 6(124), pp. 106–113.
- Mazmanishvili A. S., Reshetnyak N. G., Shovkoplyas O. A. Beam and sector regimes of electron flows in cylindrical magnetic field of the magnetron gun. *Journal of Nano- and Electronic Physics*, 2020, vol. 12, no. 3, 03001(5cc).
- Mazmanishvili A.S., Reshetnyak N.G. Transformation of data array of cylindrical magnetic field of the magnetron gun and problem of the radial motion of electrons. *Applied Problems of Mathematical Modeling*, 2020, vol. 3, no. 1, pp. 108–116.
- Tusikov A. V., Aleksandrov V. S., Shevtsov V. F. Modelirovanie dinaiki puchka v magnitnom pole spiral'no' katushki. [Simulation of beam dynamics in the magnetic field of a spiral coil]. *Pis'ma v EChAYa* [The brief of ECAY]. 2006, vol. 3, no. 7 (136), pp. 55–59.
- Astrelin V. T., Kandaurov I. V., Sveshnikov V. M. Chislennoe modelirovanie transportirovki i szhatiya elektronogo puchka shodyatshim magnitnym polem pri ego ihzhekcii v mnogoprotochnuyu lovushku GOL-3 [Numerical simulation of the transport and compression of an electron beam by a converging magnetic field during its injection into a GOL-3 multi-mirror trap]. *Uspehi prikladnoy fiziki* [Advances in Applied Physics], 2013, vol. 1, no. 5, pp. 580–585.
- Madelung E. *Matematicheskiy apparat fiziki* [Mathematical machine of physics]. Moscow, Fizmatgiz, 1961, 620 p.
- Mazmanishvili A. S., Reshetnyak N. G., Sydorenko A. Yu., Dynamika elektronogo puchka formiruемого магнетронной пушкой с вторичноэмиссионным катодом, в спадающем магнитном поле соленоида: експеримент і моделювання [Dynamics of an electron beam formed by magnetron gun with the secondary emission cathode in the declining magnetic field of solenoid: experiment and simulation]. *Nats. tekhn. un-ta "KhPI": sb. nauch. tr. Temat. vyp.: Sistemnyy analiz, upravlenie i informatsionnye tekhnologii* [Bulletin of the National Technical University "KhPI": a collection of scientific papers. Thematic issue: System analysis, management and information technology]. Kharkiv, NTU "KhPI" Publ., 2021, no. 2 (6), pp. 96–100.
- Gluhiv V. A., Pecherskiy O. P., Engel'ko V. I., Yatsenko B. I. Modifikatsiya poverhnostnykh svoystv materialov s pomotshyu motshnykh impul'snykh ustanovok [Modification of surface properties of materials using high-power pulsed electron beams]. *Sb. dokladov X mezhdunarodnogo sovetshania po primineniyu sarzhzhennykh chastits v promishlennosti i medicine* [Collection of reports of the X International Meeting on the Application of Charged Particle Accelerators in Industry and Medicine] (sity Sankt-Peterburg, 1–4 oktober, 2001). Moscow: TsNIAtominform Publ., 2001, pp. 93–396.
- Asaryan N. S., Barengol'ts S. A., Dolya S. N., Mesyats G. A., Perel'shteyn E. A. Modelirovanie formirovaniya elektronogo kol'tsa

References (transliterated)

- Ayzatsky M. I., Dovbnya A. N., Mazmanishvili A. S., Reshetnyak N. G., Romasko V. P., Chertishchev I. A. Studies on formation of the radially-directed electron beam generated by the

s ispol'zovaniem pikosekundnykh elektronnykh putchkov v magnitnoy sisteme tipa "kasp" [Simulation of the formation of an electron ring using picosecond electron beams in a "cusp" magnetic system].

Gurnal tehniceskoy fiziki. [Journal of Technical Physics], 2010, issue. 80, no. 4, pp. 126–134.

Надійшла (received) 05.03.2023

UDC 519.2

O. S. MAZMANISHVILI, Doctor of Physical and Mathematical Sciences, Full Professor, senior staff scientist, National Science Center "Kharkov Institute of Physics and Technology", Kharkov, Ukraine, e-mail: mazmanishvili@gmail.com, ORCID: <https://orcid.org/0000-0003-0373-0626>

M. G. RESHETNYAK, Candidate of Physical and Mathematical Sciences (PhD), senior staff scientist, National Science Center "Kharkov Institute of Physics and Technology", Kharkiv, Ukraine, e-mail: nreshetnyak@kipt.kharkov.ua, ORCID: <https://orcid.org/0000-0002-2345-6789>

G. YU. SYDORENKO, Candidate of Technical Sciences (PhD), Docent, National Technical University "Kharkiv Polytechnic Institute", Associate Professor at the Department of system analysis and information-analytical technologies, Kharkiv, Ukraine, e-mail: annsydorenko01@gmail.com, ORCID: <https://orcid.org/0000-0002-0761-2793>

ELECTRON BEAMS IN THE GRADIENT MAGNETIC FIELD: CONTROL FOR CONVERTING LONGITUDINAL MOTION INTO TRANSVERSAL

The motion of electrons in a cylindrical magnetic field with the gradient-type potential is considered. The motion of electrons in the cylindrical magnetic field with the gradient-type potential is considered. It is found that in the selected field, the initial motion of electrons along the longitudinal axis is converted into motion along the radius. It is determined that such the transformation is due to the action of a solenoidal magnetic field with large longitudinal gradient. The transformation of the longitudinal direction of motion into the transverse one turned out to be stable in the energy range of 20...55 keV of electrons and in the range of 5...50 mm of radial dimensions of the particle beam. The main dependences of the motion of the electron beam in the given solenoidal magnetic field are studied with the help of the software tool. The results of numerical simulation of electron trajectories in the gradient magnetic field with the circular secondary emission cathode located in the middle of the system are presented. To study the mechanism of stability with respect to magnetic field, two experimentally realized magnetic fields were used. Based on these two fields, arrays of additional 4 fields are numerically synthesized. For the set of 6 named fields, the operation of the gun, in which the particle undergoes the stable transformation of the direction of motion, is numerically studied. It is shown that for the given electron energy and the fixed magnetic field, the parameter that determines the rotation of the particles is the magnetic field gradient at the boundary of the entry region. It is found that the rotation effect takes place for the considered range of radial beam sizes, which leads to particle focusing. The possibility is shown to control the vertical coordinate of the focused beam on the basis of the field adjustment, thereby giving the interpretation of the pore dependence of the registration of electrons on the detector. The dependence of the formation of the final distribution of particles on the amplitude and gradient of the magnetic field along the axis of the system is studied. The results of numerical simulation on the motion of the electron beam are presented. Based on the model of electron flow motion, the characteristics of the resulting electron beam are considered. It is shown that the beam, having radial dimensions of 5...50 mm, is transformed and focused vertically on the area of 1 mm.

Keywords: electron beam, magnetron gun, electron dynamics, motion direction transformation, gradient magnetic field, control, mathematical modeling.

Повні імена авторів / Author's full names

Автор 1 / Author 1: Мазманішвілі Олександр Сергійович, Mazmanishvili Oleksandr Serhiyovych

Автор 2 / Author 2: Решетняк Микола Григорович, Reshetnyak Mykola Grygorovych

Автор 3 / Author 3: Сидоренко Ганна Юріївна, Sydorenko Ganna Yurijivna

УПРАВЛІННЯ В ОРГАНІЗАЦІЙНИХ СИСТЕМАХ

MANAGEMENT IN ORGANIZATIONAL SYSTEMS

DOI: 10.20998/2079-0023.2023.01.05

УДК 004:[004.415:004.77](477.87)

Н. Е. КУНАНЕЦЬ, доктор наук із соціальних комунікацій, кандидат історичних наук, професор, професор кафедри інформаційних систем та мереж, Національний університет «Львівська політехніка», м. Львів, Україна, e-mail: nek.lviv@gmail.com, ORCID: <https://orcid.org/0000-0003-3007-2462>

П. П. ФЕДОРКА, аспірант, асистент кафедри програмного забезпечення систем, ДВНЗ «Ужгородський національний університет», м. Ужгород, Україна, e-mail: pavlo.fedorka@uzhnu.edu.ua, ORCID: <https://orcid.org/0000-0002-9242-5588>

В. І. КУТ, кандидат технічних наук, доцент, завідувач кафедри інформатики та фізико-математичних дисциплін, ДВНЗ «Ужгородський національний університет», м. Ужгород, Україна, e-mail: vasilij.kut@uzhnu.edu.ua, ORCID: <https://orcid.org/0000-0001-5267-331X>

ФОРМУВАННЯ РЕКОМЕНДАЦІЙНОЇ СИСТЕМИ ДЛЯ «РОЗУМНОГО РЕГІОНУ» З МЕТОЮ ОБРАННЯ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ ТА ЇХ РЕАЛІЗАЦІЙ ПРИ СТВОРЕННІ ЗАСТОСУНКІВ

Проведено аналіз можливості втілення найперспективніших методологій, що спрямовані на створення інфраструктури «розумного» регіону. Детальне опрацювання релевантних досліджень наукових розвідок дозволяє виділити кращі напрацювання. Багато дослідників зосереджуються тільки на теоретичному аспекті питання. Автори статті проаналізували засади створення інформаційно-технологічної платформи, що покликана полегшити процес взаємодії мешканця регіону з оточуючою інфраструктурою. Формування «розумного регіону» покликане забезпечити активну комунікацію представників різних сфер агломерації, тобто органів місцевого управління та різних галузей функціонування регіону. Саме кооперація цих важливих ланок життєзабезпечення конгломерації населених пунктів, дозволяє розвивати стратегію націлену на покращення умов проживання та ефективного використання ресурсів. На сьогодні, вже є ряд успішних прикладів, щодо формування певного базису і закладки успішних підвалин для розвитку «розумних регіонів». Саме аналіз кращих методик з метою стає основою аплікації на реалізацію концепції «Розумний регіон Закарпаття» та розроблення власної новаторської інформаційно-технологічної платформи для впровадження технологій «розумного» середовища. Саме через розроблення якісної інформаційно-технологічної платформи здійснюється управління розумним регіоном, щоб покращити якість послуг, що пропонуються громадянам, і зробити усі процеси управління більш ефективними. Для віднайдення найкращого шляху проектування та забезпечення вимог громади розробляється рекомендаційна система обрання інформаційних технологій та їх реалізацій, і як результат вихідними даними, що маємо отримати збалансовану, оптимальну модель застосунку для інформаційно-технологічної платформи «Розумний регіон Закарпаття».

Ключові слова: інфраструктура, рекомендаційна система, інформаційно-технологічна платформа, веб-застосунок, дані, розумне місто, розумний регіон, алгоритм.

Вступ. У сучасному суспільстві для формування розумного регіону генерується набір технологічних інструментів і рішень, спрямованих на забезпечення комфортного проживання його мешканців, якості їх життя, ефективної комунікації, розвитку інфраструктури, транспорту, економіки, науки, технології тощо. Розумні ініціативи спрямовані на покращення систем і структур регіону. Важливим напрямом формування розумного регіону є створення системно виваженої інформаційно-технологічної платформи, що сприятиме покращенню інтеграції всіх учасників та їх взаємодії в багатьох сферах. Розумний регіон розглядається як екосистема і фокусується на забезпеченні соціальних благ, економічного зростання та створенні нових можливостей, ґрунтуючись на використанні інформаційних технологій, оскільки саме вони можуть забезпечити надійними рішеннями, які приносять користь

громадянам. Інформаційно-технологічна платформа (ІТ-платформа) – це зручне середовище для створення, розгортання та управління різноманітними програмними застосунками, яке об'єднує у собі апаратне забезпечення, операційну систему та різні компоненти програмного забезпечення.

ІТ-платформи можуть включати в себе різноманітні технології, такі як віртуалізація, облікові записи користувачів, безпеку даних, аналітику даних, розподілений доступ до даних, управління процесами у різних галузях діяльності розумного регіону. Вони можуть бути розроблені для підтримки певних видів програмного забезпечення, наприклад, веб-застосунків, мобільних застосунків або інформаційних систем.

Саме з допомогою інформаційно-технологічної платформи здійснюватиметься управління «Розумним регіоном Закарпаття», щоб покращити якість послуг,

© Кунанець Н. Е., Федорка П. П., Кут В. І., 2023



Дослідницька стаття: Цю статтю опубліковано видавництвом *НТУ «ХПІ»* у збірнику «Вісник Національного технічного університету "ХПІ" Серія: Системний аналіз, управління та інформаційні технології». Ця стаття поширюється за міжнародною ліцензією [Creative Commons Attribution \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/). Конфлікт інтересів: Автор/и заявив/или про відсутність конфлікту.



що пропонуються громадянам, і зробити усі процеси управління більш ефективними.

Мета даної статті – проаналізувати особливості створення рекомендаційної системи вибору інформаційних технологій та їх реалізацій з метою розроблення застосунку для інформаційно-технологічної платформи «Розумний регіон Закарпаття».

Аналіз останніх досліджень. Корінна Моранді, Андреа Роландо, Стефано Ді Віта презентують результати дослідження взаємозв'язку між інформаційними та комунікаційними технологіями і просторовим плануванням, розвиваючи концепцію «розумності міста» до регіонального виміру. Автори дослідили використання інформаційних та комунікаційних технологій для представлення, просування, управління та поширення інтегрованої системи послуг у розумному місті; просторовий вплив цифрових послуг на розвиток у різних масштабах (регіональний, міський, місцевий) [1].

Г. Вдова, Й. Кнілінг відзначають, що у Нижній Саксонії застосовують новий підхід до управління з метою підвищення ефективності розвитку та впровадження інновацій під назвою «розумний регіон». Автори аналізують фактори, що сприяють успішності таких проєктів, а також проблеми, що виникають в процесі їх реалізації, досліджують роль співпраці у визначенні інноваційних пріоритетів та налагодженні інноваційних процесів. У статті відзначаються мережі, кластери та платформи як важливі формати співпраці для передачі знань та просування інновацій. Автори відзначають, що процес стратегічного планування «знизу вгору» у «розумному регіоні» підтриманий при визначенні пріоритетних напрямків економічного розвитку та технологічної спеціалізації в регіоні Гамбург, що у Південній Ельбі.

Автори наголошують на доцільності формування регіональних проєктів. В регіоні Гамбург розроблений Генеральний план реалізації інноваційних проєктів, що враховує стратегічні пріоритети спеціалізації регіону. Формування «розумного регіону» покликане забезпечувати активну комунікацію представників органів місцевого управління, економічної галузі та наукових кіл.

Автори стверджують, що пов'язані з інформаційними та комунікаційними технологіями інновації сприяють формуванню розумного регіону, у якому активно розвиваються університети, великі компанії, організації і тематичні кластери. Результати проведеного дослідження підтверджують, що співпраця з інноваційними суб'єктами, розташованими поблизу столичних міст Гамбурга, Бремена та Ганновера сприяє стимулюванню інновацій у регіоні Гамбург.

Автори відзначають, що інноваційні підходи до управління «розумними регіонами» сприяють сталому регіональному розвитку та територіальній єдності, такий підхід забезпечує часткове вирішення місцевих економічних, соціальних чи екологічних проблем [2].

Фабріціо Арнеодо; Роберта Кастеллі; Данило Ботта аналізують досвід регіону П'ємонт, в якому проведено реорганізацію парадигми «розумного міста» на концепцію «розумного регіону». Різні проєкти та технології, задіяні в цій перспективній зміні, описані

через генезис єдиного регіонального центру обслуговування, втілення підходу мобільності як послуги. Розумний регіон є не лише початковою точкою, але й відправною точкою нових змін [3].

Концепт «інформаційно-технологічна платформа». Для створення комфортного комунікативного середовища «розумного регіону» доцільно розробляти інформаційно-технологічну платформу, що поєднуватиме різні програмні застосунки, інформаційні технології та відкриті інформаційні ресурси.

«Розумний регіон» за аналогією «розумного міста» можемо визначити як відкриту, складну та адаптовану систему, що ґрунтується на комп'ютерній мережі, інформаційно-технологічній платформі та інформаційних ресурсах, які сприяють створенню віртуального цифрового середовища регіону.

ІТ-платформи дозволяють створювати, тестувати, впроваджувати та керувати програмними застосунками в зручний спосіб, що сприяє підвищенню продуктивності та зниженню витрат на розробку та експлуатацію програмного забезпечення. Також вони можуть допомогти забезпечити стандартизацію технологій, що дозволяє зменшити складність та збільшити ефективність розроблення програмного забезпечення.

Інформаційно-комунікаційна платформа формується із окремих програмних застосунків, інформаційних технологій та інформаційних ресурсів. Таким чином концепт «інформаційно-технологічна платформа», тісно пов'язаний з інформаційними технологіями, хоча охоплює множину компонентів. Такі платформи набувають все більшого інноваційного значення та використовуються в багатьох сферах функціонування регіону. Разом з тим цей концепт не отримав чіткого визначення, а ґрунтується на загальному розумінні, ним послуговуються у наукових публікаціях різних галузей знань.

Р. Фішман відзначає, що в процесі розвитку інформаційних технологій концепт «інформаційно-технологічна платформа» все частіше розглядається як технологічний тренд, який забезпечує основу для розгортання значної кількості застосунків та бізнес-процесів [4]. Низка дослідників імплементувала цей концепт у ряд академічних дисциплін. Дослідники Масачусетського університету Т. Eisenmann та ін. [5] вважають платформу засобом, що дозволяє об'єднати функціональні можливості розробленого застосунку із цільовою метою застосунків, представлених в множині застосунків, що розташовані на платформі, яка забезпечує спільне їх використання. Такий підхід сприяє економії часу та обсягу хмарних сховищ.

Дослідники розглядають підходи до розгортання платформ, серед яких сформулювали стратегію, яку назвали охопленням платформи. Її суть полягає в тому, що у випадку, коли у провайдерів платформ, які обслуговують різні сегменти, іноді збігаються бази користувачів, що потребують подібних компонентів, доцільно застосовувати технологію консолідації, об'єднавши функціональність своєї власної платформи з цільовою, щоб забезпечити спільні потреби користувачів у загальних компонентах [6].

М. Кусумано [7] вважає, що технологічна платформа для екосистеми – це набір спільних основних технологій і технологічних стандартів, що лежать в основі організаційної сфери і підтримують спільне створення цінностей через спеціалізацію та додаткові пропозиції. На думку дослідника – це найширокомасштабніший і неоднорідний потік, який спирається на різноманітні теоретичні точки зору, включаючи промислове співтовариство, економічні ефекти та ресурси [8].

Л. Томас, Е. Давід, М. Ганн [9] вважають, що технологічні платформи – це не стільки продукти, скільки процеси чи послуги, які сприяють зростанню продуктивності економіки. Дослідники трактують концепт платформа як сукупність продуктів, процесів і послуг.

Р. Мангемантин трактують концепт «технологічна платформа» як продукт кумулятивних інвестицій у науково-дослідні розробки, які створюють сімейства технологічних варіантів [10]. Такий підхід перегукується з трактуванням Т. Куна [11], а також Б. Кіма, який вважає, що набір основних технологій дозволяє впроваджувати та стимулювати інновації [12]. Б. Кім вважає, що при формуванні технологічної платформи використовується декілька поколінь технологій. На думку дослідника, велика кількість продуктів і послуг може бути отримано з кожного покоління технологій, які проходять життєвий цикл і з часом будуть замінені наступним поколінням. Як старі, так і нові технології, на думку авторів, доцільно розташовувати на одній технологічній платформі [12]. Дослідники аналізують, як керувати переходом поколінь технологій з точки зору макrorівня, розробляючи аналітичні моделі та тестуючи їх за допомогою інформаційного моделювання. Результати показують, як ключові визначальні фактори взаємодіють один з одним, зокрема, який оптимальний час переходу між поколіннями технологій, як його можна визначити, оцінюють вартість переходу від одного покоління до іншого та користь, яку економіка отримує від використання технологічної платформи.

С. Велврайгт, К. Кларк та ін. [13] вважали проєкти платформи орієнтованими на потреби основної групи клієнтів, призначеними для легкої модифікації шляхом додавання, заміни та видалення функцій. Таке визначення концепту не пояснює, як відбувається модифікація продуктів, а концепцію платформи було застосовано до різних типів модифікацій.

Т. Сімпсон, Ю. Маєр, Ф. Містре [14] розподілили концепт «технологічна платформа» на декілька типів:

До першого типу віднесено визначення Т. Сімпсон, Ю. Маєр, Ф. Містре, яке переконує, що одним із типів є масштабована продуктова платформа, яка має компоненти, які мають однакову функцію, але з різними можливостями [15].

Другий тип – це визначення технологічної платформи як платформи продуктів, і розглядає життєвий цикл продукту як основу для швидкого розвитку наступного покоління продуктів, що базується на відмінності між життєвим циклом платформи продукту,

життєвим циклом дизайну продукту та життєвим циклом варіанту продукту Г. Вольфмана та А. Албласа [16].

Третім, типом трактування платформ М. Маєром та Ю. Унтербаком є розгляд їх як модульних платформ, де варіанти продуктів в середині сімейства продуктів створюються шляхом заміни модулів [17].

Розроблення рекомендаційної системи. Одним із елементів інформаційно-технологічної платформи «Розумний регіон Закарпаття» є програмний застосунок – рекомендаційна система, яка допомагає обирати ефективні інформаційні технології, їх реалізації для створення інформаційних систем та застосунків з врахуванням функціональних завдань, які з їх допомогою вирішуватимуться. Дані щодо вимог до майбутнього програмного застосунку вносяться як вхідні дані для рекомендаційної системи. Далі проводиться їх аналіз і на його основі генеруються рекомендації щодо обрання інформаційних технологій, а в межах кожного класу інформаційних технологій кращі реалізації, для розроблення програмного продукту. Для визначення доцільного використання тієї чи іншої інформаційно-технологічної реалізації при розробленні застосунку у рекомендаційній системі використовується метод аналізу ієрархій. Результати видаються у формі таблиць (табл. 2 та табл. 3) із подальшою візуалізацією у вигляді діаграм (рис. 3 та рис. 4).

Метод аналізу ієрархій розроблений Томасом Л. Сааті [18] і від тоді вдосконалювався. Метод аналізу ієрархій передбачає виділення трьох складових: кінцевої мети або проблеми, яку необхідно вирішити, можливі рішення, які називаються альтернативами, і критерії, за якими відбуватиметься оцінювання альтернатив. Метод забезпечує раціональну структуру для прийняття рішення шляхом кількісного визначення критеріїв та альтернативних варіантів.

Експерти порівнюють важливість критеріїв шляхом попарного порівняння. Оцінювання може бути вербальним, яке згодом перетворюється в вагові коефіцієнти, які можна порівнювати. Ця кількісна оцінювання відрізняє метод аналізу ієрархій від інших методів, які дозволяють генерувати рекомендації для прийняття рішень.

На останньому етапі процесу обчислюються числові пріоритети для кожного з альтернативних варіантів.

Користувачами рекомендаційної системи переважно будуть 5 акторів:

- власник продукту;
- проєктний менеджер;
- інженер із забезпечення якості;
- DevOps-інженер;
- розробник.

Робота рекомендаційної системи відбувається за наступним алгоритмом:

Крок 1. Реєстрація користувачів, яка дозволяє користувачам створювати облікові записи з унікальними ідентифікаторами та паролями для авторизації.

Крок 2. Авторизація та автентифікація користувачів, які дозволяють користувачам авторизуватися на

сайті та проводити автентифікацію з використанням засобів, таких як токени, сесійні ключі тощо.

Крок 3. Внесення даних про сформовані замовником вимоги до застосунку.

Крок 4. Підбір інформаційної технології для розроблення застосунку.

Крок 5. Підбір реалізацій обраної інформаційної технології.

Крок 5. Проведення експертного оцінювання на основі аналізу коментарів на офіційних сайтах розробників застосунків.

Крок 6. Генерування рекомендацій з використанням методу аналізу ієрархій.

Крок 7. Візуалізація результатів.

Розглянемо генерування рекомендацій системою на прикладі реалізацій класу хмарних технологій. Переваги кожної реалізації оцінюються за різними критеріями.

Досягнення мети полягає у виборі однієї з альтернатив на основі множини сформульованих критеріїв. Вибір альтернативи виконується обчисленням елементів вектора пріоритетів, які поставлено у відповідність кожній альтернативі. Альтернатива з найбільшим значенням такого елемента вважається прийнятим рішенням.

При проведенні попарних порівнянь реалізацій хмарних технологій визначається, який із двох порівнюваних об'єктів важливіший. Міра переваги об'єкта A_i над об'єктом A_j виражається елементом матриці a_{ij} . Матриця парних порівнянь A має властивість зворотної симетрії і виконується умова:

$$a_{ij} = \frac{1}{a_{ji}}$$

Використанню методу аналізу ієрархій передують експертне оцінювання реалізацій за низкою критеріїв. Для порівняння обраної реалізації хмарних технологій AWS, Azure, Google.

Реалізація Amazon Web Services (AWS) є провідним провайдером хмарних технологій. AWS пропонує широкий спектр послуг, включаючи обчислювальні, мережеві, зберігання даних, бази даних, інструменти аналітики, штучний інтелект, машинне навчання та інші.

AWS має безліч переваг, включаючи високу доступність, масштабованість та безпеку. Клієнти можуть використовувати різні послуги AWS, щоб забезпечити виконання своїх бізнес-потреб, знизити витрати та покращити ефективність.

AWS також пропонує безліч інструментів управління та моніторингу, таких як Amazon CloudWatch, AWS Management Console та AWS Command Line Interface (CLI), які дозволяють користувачам з легкістю контролювати та відлагоджувати свої хмарні сервіси.

Загалом, AWS є досить потужним та розширеним інструментом, який може допомогти підприємствам будь-якого розміру збільшувати свою продуктивність та знижувати витрати.

Реалізація Azure – це набір послуг, що надаються Microsoft Azure для забезпечення хмарних обчислень. Azure є однією з провідних платформ хмарних обчислень, яка пропонує різноманітні сервіси:

Віртуальні машини: Це можливість запускати віртуальні машини в хмарі, де можна налаштувати різноманітні операційні системи та програмні засоби, які потрібні для запуску застосунків.

Azure App Service: Це можливість розгортання та керування веб-застосунками на Azure, що дозволяє зосередитися на розробці програми, а не на керуванні інфраструктурою.

Azure SQL Database: Це послуга, яка дозволяє створювати та керувати базами даних в Azure.

Azure Cosmos DB: Це глобально розподілена база даних, яка дозволяє швидкий доступ до даних в будь-якому місці світу.

Azure Storage: Це набір послуг для зберігання даних в хмарі, включаючи блок-зберігання, файлове зберігання та інші.

Azure Virtual Network: Це послуга, яка дозволяє налаштовувати приватні мережі в хмарі та забезпечувати безпеку між різними компонентами застосунку.

Azure Active Directory: Це ідентифікаційна та управлінська послуга, яка дозволяє керувати доступом користувачів до різних сервісів та додатків в хмарі.

Ця реалізація дозволяє розробникам та підприємствам швидко та ефективно створювати, розгортати та керувати застосунками.

Реалізація Google – це набір послуг, які надаються Google Cloud Platform (GCP) для забезпечення хмарних обчислень. GCP є однією з провідних платформ хмарних обчислень, яка надає різноманітні сервіси, такі як обчислення, зберігання даних, штучний інтелект та інші. Деякі з хмарних послуг Google, які надаються Google Cloud Platform, включають:

Google Compute Engine: Це послуга віртуальних машин, яка дозволяє запускати віртуальні машини з різними операційними системами та налаштуваннями.

Google Kubernetes Engine: Це послуга контейнеризації, яка дозволяє керувати та оркеструвати контейнери з застосунками.

Google Cloud Storage: Це набір послуг для зберігання даних в хмарі, включаючи об'єктне зберігання, файлове зберігання та інші.

Google BigQuery: Це послуга опрацювання та аналізу великих обсягів даних з використанням SQL-запитів.

Google Cloud AI Platform: Це набір інструментів для створення та навчання моделей штучного інтелекту.

Google Cloud Functions: Це послуга, яка дозволяє запускати код без необхідності керування віртуальними машинами.

Google Cloud Identity and Access Management: Це ідентифікаційна та управлінська послуга, яка дозволяє керувати доступом користувачів до різних сервісів та застосунків в хмарі.

Ця реалізація дозволяє розробникам та підприємствам швидко та ефективно створювати, розгортати

та керувати застосунками та послугами в хмарі, забезпечуючи високу продуктивність та масштабованість.

Шкала експертних оцінок або ступенів важливості – для парних порівнянь при оцінюванні переваги одного об’єкта над іншим використовуватимемо в межах від 1 до 9.

У табл. 1 наведено результати експертного оцінювання за 10-бальною шкалою.

Табля 1 – Результати експертного оцінювання

	Функціональні і можливості	Безпека	Надійність
AWS	9	8	8
Azure	7	8	7
Google	6	7	6

Базуючись на результати експертного оцінювання визначимо найкращу реалізацію хмарних сервісів із застосуванням методу Сааті. Перший крок МАІ полягає в декомпозиції та представленні задачі в ієрархічній формі. На рис. 1 наведено побудовану ієрархію цілей для вибору кращої реалізації хмарних технологій з метою розроблення застосунку для інформаційно-технологічної платформи «Розумний регіон».

Ми розглядаємо домінантні ієрархії, які будуються з вершини через проміжні рівні (критерії, від яких залежать наступні рівні) до найнижчого рівня, який складає перелік альтернатив. Ієрархія вважається повною, якщо кожен елемент заданого рівня функціонує як критерій для всіх елементів рівня, що стоїть нижче. Закон ієрархічної безперервності вимагає, щоб елементи нижчого рівня були попарно зрівняні відносно елементів наступного рівня і т. д. до вершини ієрархії.

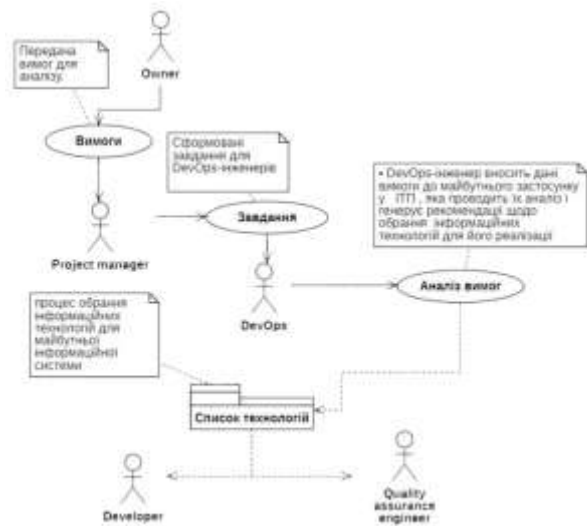


Рис. 1. Діаграма варіантів використання рекомендаційної системи

Застосовуючи метод аналізу ієрархій сформувавши тривірневу структуру. Верхній рівень займає ціль, яка складатиме основу рекомендації, що генеруватиме

рекомендаційна система, другий рівень формує множина критеріїв, за якими здійснюється вибір альтернативних реалізацій хмарних технологій для розроблення застосунку. Самі ж альтернативи реалізацій утворюють нижній рівень ієрархічної структури. Дерево цілей подано на рис. 2. Генерування рекомендації полягає у виборі однієї з можливих альтернатив на основі побудови вектора пріоритетів. Пріоритет оцінюється як дійсне число, яке ставиться у відповідність кожній альтернативі. Альтернатива з найвищим пріоритетом і буде основою рекомендації, яку генеруватиме система.

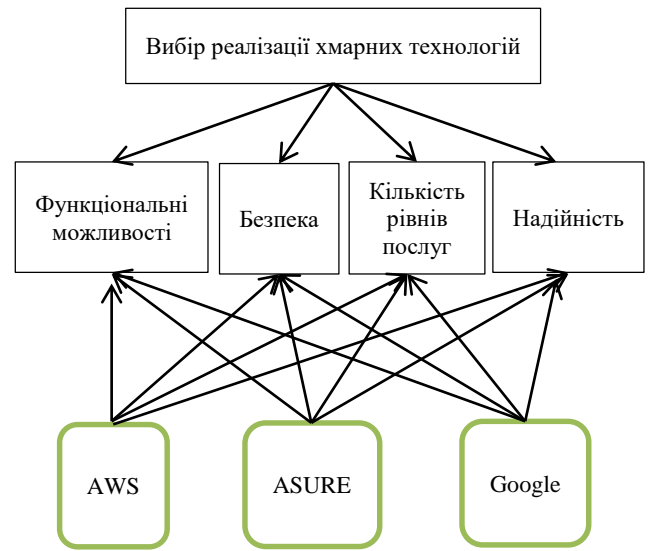


Рис. 2. Ієрархія для вирішення проблеми вибору кращої реалізації хмарних сервісів

Під альтернативами розуміємо реалізації хмарних технологій, аналіз яких здійснюється за обраними критеріями, за якими відбуватиметься оцінювання. Важливим фактором у формулюванні критеріїв, за якими буде обрано реалізації є їх можливості щодо створення інформаційної системи з функціоналом, визначеним у вимогах замовника.

Для матриць попарних порівнянь, побудованої за критерієм «Функціональні можливості», обчислено такі параметри:

- оцінка найбільшого власного значення, яка обчислена за формулою:

$$\lambda_{\max} = \sum_{i=1}^n W_i S_i,$$

де W_i – вага альтернативи з номером i , S_i – сума елементів стовпця з номером i матриці попарних порівнянь, n – кількість альтернатив;

- індекс узгодженості:

$$CI = \frac{\lambda_{\max} - n}{n - 1};$$

- індекс послідовності співвідношень:

$$CR = \frac{CI}{RI}.$$

Тут і далі $RI = 1.24$ – це випадковий індекс для $n = 6$, значення якого є однаковим для всіх подальших обчислень ваг альтернатив.

Аналогічні розрахунки проводяться за кожним з критеріїв.

У табл. 2 наведено нормовану матрицю попарних порівнянь альтернатив відносно критерію «Функціональні можливості». Також на рис. 3 додано це графічне відображення у діаграмі.

Таблиця 2 – Нормована матриця попарних порівнянь альтернатив відносно критерію «Функціональні можливості»

Функціональні можливості	Azure	Google	AWS	Вектор локальних пріоритетів	Нормований вектор
Azure	1	7	9	3,979	0,765
Google	1/7	1	1/8	0,261	0,050
AWS	1/9	8	1	0,961	0,185
$\Sigma =$	1,25	16,00	10,13	5,202	1

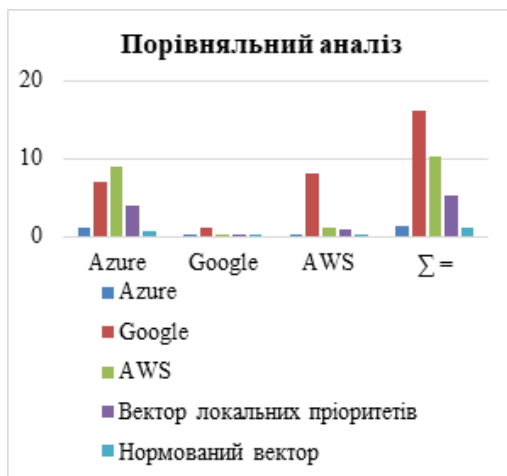


Рис. 3. Порівняльний аналіз реалізацій хмарних технологій за критерієм «Функціональні можливості»

У табл. 3 наведено нормовану матрицю попарних порівнянь альтернатив відносно критерію «Безпека». Також на рис. 4 додано це графічне відображення у діаграмі.

Після обчислення даних для матриці попарних порівнянь, побудованої за критерієм «Функціональні можливості», параметри набувають таких значень:

$$a_{\max} = 3.635,$$

- індекс узгодженості $IY = \frac{a_{\max} - n}{n - 1} = -0.12$;
- індекс послідовності співвідношень $BY = \frac{IY}{RI} = -0.14$.

Оскільки $BY = -0.14 < 1$, то матрицю попарних порівнянь за критерієм «Функціональні можливості» вважаємо узгодженою.

Таблиця 3 – Нормована матриця попарних порівнянь альтернатив відносно критерію «Безпека»

Безпека	Azure	Google	AWS	Вектор локальних пріоритетів	Нормований вектор
Azure	1	9	8	4,160	0,775
Google	1/9	1	1/7	0,251	0,047
AWS	1/8	7	1	0,956	0,178
$\Sigma =$	1,24	17,00	9,14	5,368	1



Рис. 4. Порівняльний аналіз реалізацій хмарних сервісів за критерієм «Безпека»

Після обчислення для матриці попарних порівнянь, побудованої за критерієм «Безпека», вказані параметри набувають таких значень:

$$a_{\max} = 3.383,$$

- індекс узгодженості $IY = \frac{a_{\max} - n}{n - 1} = 0.19$;
- індекс послідовності співвідношень $BY = \frac{IY}{RI} = 0.58$.

Оскільки $BY = 0.58 < 1$, то матрицю попарних порівнянь за критерієм «Функціональні можливості» вважаємо узгодженою.

В результаті дослідження можемо констатувати, що реалізація Azure – переможець, помітно слабшою є AWS та найменшу оцінку отримала реалізація Google.

Висновки. Дане дослідження, вищеописаних теорій та практик, дозволяє скласти комплексний алгоритм дій для розуміння правильної дорожньої карти на шляху до впровадження широкого спектру «Smart Technology», для міста чи регіону. Разом з цим, показані аналітичні дані та базові напрацювання у використанні перших проявів «розумних» програмних рішень, тобто «хмарні» дані, та робота з допомогою хмарних сервісів та відповідних корпорацій, що розвивають цей напрямок. На сьогодні не існує єдино правильного варіанту побудови «розумного» регіону. Кожне місто чи регіон, агломерація має свої власні спеціальні потреби, запити громади, рівень соціального та управлінського контролю, тому при проектуванні певних систем чи інформаційно-технологічних платформ, варто брати широкий аспект потреб, саме конкретного округу і старатися виділити кілька основних напрямків впровадження, оскільки метод ієрархій допомагає побудувати стійку структуру від меншого до більшого та побудувати певні закономірності, щодо задоволення якомога більше потреб «розумного регіону». Грунтовний аналіз показує наскільки важливе використання кращих напрацювань при впровадженні «розумних» технологій. Сучасне життя громадян насичене величезними масивами інформації, яку генерує сам мешканець і яку потребує він чи вона для свого повсякденного задоволення потреб. Тому слід наголосити, що розвиток самодостатніх, соціально орієнтованих регіонів неможливий без поступу новітніх інформаційних технологій та розроблення комплексних заходів щодо впровадження інформаційно-технологічних платформ. Тільки комплексний підхід до вирішення цих питань та розробки відповідного програмного, може забезпечити розвиток та добробут регіонів і загалом держави. Це особливо актуально в сучасних реаліях економічної нестабільності, потреби у соціальному захисті населення та необхідності допомоги у взаємодії людини і регіональної інфраструктури. Саме соціально орієнтована стратегія розвитку регіонів має бути центральною та об'єднуючою ланкою між всіма дотичними сферами життєдіяльності міста. Цей опрацьований матеріал мав на меті показати доцільність та пряму залежність економічного і соціального добробуту мешканців «розумних регіонів» від використання інноваційних інформаційних технологій.

Список використаної літератури

1. Morandi C., Rolando A., Di Vita S. From Smart City to Smart Region. Digital Services for an Internet of Places. Milano, 2016. 103 p.
2. Vladova G., Knieling J. "Smart Region" governance for innovation. *Rural-Urban Linkages for Sustainable Development*. Hamburg, 2020. P. 176-190. DOI: 10.4324/9780429288111-11.
3. Arneodo F., Castelli R., Botta D. Towards a "Smart Region" paradigm: Beyond Smart Cities borders: Piedmont Region experience. *Proc. International Conference of Electrical and Electronic Technologies for Automotive. (15-16 June 2017, Torino)*. Torino, 2017. P. 1-8, DOI: 10.23919/eeta.2017.7993225.
4. Fichman R. G. Real Options and IT Platform Adoption: Implications for Theory and Practice. *Information Systems Research*. 2004. Vol. 15, no.2. P. 132-154. DOI: 10.1287/isre.1040.0021.

5. Eisenmann T., Parker G., Van Alstyne M. Platform Envelopment. *Strategic Management Journal*. 2011. Vol. 32, no. 12. P. 1270-1285. DOI: 10.1002/imj.935.
6. Cusumano M. A. The evolution of platform thinking. *Communications of the ACM*. 2010. Vol. 53, no. 1. P. 32-34.
7. Cusumano M.A., Gawer A. The elements of platform leadership. *MIT Sloan Management Review*. 2002. P. 51-58 DOI: 10.1109/emr.2003.1201437.
8. Llewellyn D. W. T., Autio E., Gann D. M. Architectural Leverage: Putting Platforms in Context. *Academy of Management Perspectives*. 2014. Vol. 28, no. 2. P. 198-219. DOI: 10.5465/amp.2011.0105.
9. Robinson D., Rip A., Mangematin V. Technological agglomeration and the emergence of clusters and networks in nanotechnology. *Research Policy*. 2007. Vol. 36(6). P. 871-879.
10. Kuhn T. S. *The Structure Of Scientific Revolutions*. Chicago, University of Chicago Press, 1970. 210 p.
11. Kim B. Managing the transition of technology life cycle. *Technovation*. 2003. Vol. 23, issue 5. P. 371-381
12. Wheelwright S.C., Clark K.B. Creating project plans to focus product development. *Harvard Business Review*. 1992. Vol. 70, no. 2. P. 70-82.
13. Simpson T. W., Maier J. R., Mistree F. Product platform design: Method and application. *Research in Engineering Design*. 2001. Vol. 13. P. 2-22. DOI:10.1007/s001630100002.
14. Wortmann H., Alblas A. Product platform life cycles: a multiple case study. *International Journal of Technology Management*. 2009. Vol.48, no. 2. P. 188-201. DOI:10.1504/ijtm.2009.024915.
15. Meyer M. H., Utterback J. M. *The product family and the dynamics of core capability*. Cambridge, Massachusetts Institute of Technology, 1993. 32 p.
16. Saaty T. L. *The Analytic Hierarchy Process: Planning, Priority Setting, Resource Allocation*. New York, McGraw-Hill, 1980. 287 p.

References (transliterated)

1. Morandi C., Rolando A., Di Vita S. From Smart City to Smart Region. Digital Services for an Internet of Places. Milano, 2016. 103 p.
2. Vladova G., Knieling J. "Smart Region" governance for innovation. *Rural-Urban Linkages for Sustainable Development*. Hamburg, 2020, pp. 176-190. DOI: 10.4324/9780429288111-11.
3. Arneodo F., Castelli R., Botta D. Towards a "Smart Region" paradigm: Beyond Smart Cities borders: Piedmont Region experience. *Proc. International Conference of Electrical and Electronic Technologies for Automotive. (15-16 June 2017, Torino)*. Torino, 2017, pp. 1-8. DOI: 10.23919/eeta.2017.7993225.
4. Fichman R. G. Real Options and IT Platform Adoption: Implications for Theory and Practice. *Information Systems Research*. 2004, vol. 15, no. 2, pp. 132-154. DOI: 10.1287/isre.1040.0021.
5. Eisenmann T., Parker G., Van Alstyne M. Platform Envelopment. *Strategic Management Journal*. 2011, vol. 32, no. 12, pp. 1270-1285. DOI: 10.1002/imj.935.
6. Cusumano M.A. The evolution of platform thinking. *Communications of the ACM*. 2010, vol. 53, no. 1, pp. 32-34.
7. Cusumano M. A., Gawer A. The elements of platform leadership. *MIT Sloan Management Review*. 2002, pp. 51-58. DOI: 10.1109/emr.2003.1201437.
8. Llewellyn D. W. T., Autio E., Gann D. M. Architectural Leverage: Putting Platforms in Context. *Academy of Management Perspectives*. 2014, vol. 28, no.2, pp.198-219. DOI:10.5465/amp.2011.0105.
9. Robinson D., Rip A., Mangematin V. Technological agglomeration and the emergence of clusters and networks in nanotechnology. *Research Policy*. 2007, vol. 36(6), pp. 871-879.
10. Kuhn T. S. *The Structure Of Scientific Revolutions*. Chicago, University of Chicago Press, 1970. 210 p.
11. Kim B. Managing the transition of technology life cycle. *Technovation*. 2003, vol. 23, issue 5, pp.371-381.
12. Wheelwright S. C., Clark K. B. Creating project plans to focus product development. *Harvard Business Review*. 1992, vol. 70, no. 2, pp. 70-82.
13. Simpson T. W., Maier J. R., Mistree F. Product platform design: Method and application. *Research in Engineering Design*. 2001, vol. 13, pp. 2-22. DOI:10.1007/s001630100002.
14. Wortmann H., Alblas A. Product platform life cycles: a multiple case study. *International Journal of Technology Management*. 2009, vol. 48, no. 2, pp. 188-201. DOI:10.1504/ijtm.2009.024915.

15. Meyer M.H., Utterback J.M. *The product family and the dynamics of core capability*. Cambridge, Massachusetts Institute of Technology, 1993. 32 p.
16. Saaty T. L. *The Analytic Hierarchy Process: Planning, Priority Setting, Resource Allocation*. New York, McGraw-Hill, 1980. 287 p.

Надійшла (received) 05.05.2023

UDC 004:[004.415:004.77](477.87)

N. E. KUNANETS, Doctor of Social Communications, Candidate of Historical Sciences, Professor, Professor of the Department of Information Systems and Networks, Lviv Polytechnic National University, Lviv, Ukraine, e-mail: nek.lviv@gmail.com, ORCID: <https://orcid.org/0000-0003-3007-2462>

P. P. FEDORKA, graduate student, assistant at the Department of Systems Software, Uzhhorod National University, Uzhhorod, Ukraine, e-mail: pavlo.fedorka@uzhnu.edu.ua, ORCID: <https://orcid.org/0000-0002-9242-5588>

V. I. KUT, Candidate of Technical Sciences, Associate Professor, Head of the Department of Informatics and Physical and Mathematical Sciences, Uzhgorod National University, Uzhhorod, Ukraine, e-mail: vasilij.kut@uzhnu.edu.ua, ORCID: <https://orcid.org/0000-0001-5267-331X>

FORMATION OF A RECOMMENDATION SYSTEM FOR THE "SMART REGION" FOR THE PURPOSE OF CHOOSING INFORMATION TECHNOLOGIES AND THEIR IMPLEMENTATION WHEN CREATING APPLICATIONS

An analysis of the possibility of implementing the most promising methodologies aimed at creating the infrastructure of a "smart" region was carried out. A detailed study of research-relevant scientific intelligence allows us to highlight the best developments. Many researchers focus only on the theoretical aspect of the issue. The authors of the article analyzed the principles of creating an information technology platform designed to facilitate the process of interaction of a resident of the region with the surrounding infrastructure. The formation of a "smart region" is designed to ensure active communication between representatives of various spheres of the agglomeration, i.e. local government bodies and various branches of the region's functioning. It is the cooperation of these important life support links of the conglomeration of settlements that allows the development of a strategy aimed at improving living conditions and efficient use of resources. Today, there are already a number of successful examples of the formation of a certain basis and the laying of successful foundations for the development of "smart regions". It is the analysis of the best methods with the goal that becomes the basis of the application for the implementation of the "Smart Region of Zakarpattia" concept and the development of its own innovative information technology platform for the implementation of "smart" environment technologies. It is through the development of a high-quality information technology platform that the smart region is managed in order to improve the quality of services offered to citizens and make all management processes more efficient. In order to find the most optimal way of designing and meeting the community's requirements, a recommendation system for selecting information technologies and their implementations is being developed, and as a result, we should get a balanced, optimal application model for the "Smart Region of Transcarpathia" information technology platform.

Keywords: infrastructure, recommendation system, information technology platform, web application, data, smart city, smart region, algorithm.

Повні імена авторів / Author's full names

Автор 1 / Author 1: Кунанець Наталія Едуардівна, Kunanets Nataliia Eduardivna

Автор 2 / Author 2: Федорка Павло Павлович, Fedorka Pavlo Pavlovich

Автор 3 / Author 3: Кут Василь Іванович, Kut Vasyl Ivanovich

М. Д. ГОДЛЕВСЬКИЙ, доктор технічних наук, професор, Національний технічний університет «Харківський політехнічний інститут», директор інституту комп'ютерних наук та інформаційних технологій; м. Харків, Україна; e-mail: god_asu@kpi.kharkov.ua; ORCID: <https://orcid.org/0000-0003-2872-0598>

Г. О. БУРЛАКОВ, Національний технічний університет «Харківський політехнічний інститут», аспірант, м. Харків, Україна; e-mail: george@sheerchain.com; ORCID: <https://orcid.org/0000-0002-0423-6024>

ВЕРБАЛЬНИЙ ОПИС ТЕХНОЛОГІЇ ПЛАНУВАННЯ ПІДВИЩЕННЯ ЯКОСТІ ПІДМНОЖИНИ ПРОЦЕСІВ ЕТАЛОННОЇ МОДЕЛІ ЗРІЛОСТІ SPICE

Зазначено, що для вирішення задачі розробки інформаційної технології планування підвищення рівня зрілості підмножини процесів еталонної моделі зрілості SPICE, попередньо це питання необхідно розглянути на вербальному рівні як технологію (послідовність етапів) реалізації поставленої проблеми. На першому етапі проведено формалізацію структури еталонної моделі SPICE, яка складається з множини окремих процесів. Ця множина представлена як ієрархічна структура: перший рівень – множина процесів і підпроцесів; другий рівень – множина груп процесів; третій рівень – множина категорій процесів. На другому етапі розглянуто метод оцінки рівня можливості окремого процесу/підпроцесу еталонної моделі зрілості SPICE, який наведено з погляду двох вимірів моделі SPICE: призначення процесів; атрибути процесів (вимірні характеристики, необхідні для управління процесом і підвищення можливості його виконання). Третій етап технології присвячено синтезу моделі планування розвитку підмножини процесів моделі SPICE, яка визначає рівень якості окремої складової процесу розробки програмного забезпечення (ПР ПЗ), що в свою чергу позитивно впливає на рівень розробки програмних систем. Оцінка та планування рівня можливості підмножини процесів реалізується на деякому плановому періоді в умовах обмежених ресурсів на основі їх оптимального розподілу з урахуванням важливості окремих процесів та їх практик упродовж планового періоду. На четвертому етапі технології розглядається алгоритм планування розвитку підмножини процесів еталонної моделі зрілості SPICE на основі методу послідовного аналізу варіантів. Це обумовлено в першу чергу тим, що модель оптимізації має адитивну цільову функцію. Метод дозволяє під час його використання відкидати ті допустимі рішення, які не містять оптимальних. У подальшому при конкретизації моделі планується вибрати алгоритм, який відноситься до цього методу і адаптувати його до розв'язання поставленої задачі. П'ятий етап присвячено інформаційній технології реалізації розробленої моделі та алгоритму. На цьому етапі вирішується наступна множина задач. Аналіз бізнес-процесів технології покращення якості ПР ПЗ. Визначення вимог до ПЗ. Формування діаграми варіантів використання. Розробка моделей даних. Обґрунтування вибору інструментарію для розробки ПЗ. Вибір еталонної системної архітектури. Далі кодування і тестування ПЗ. На шостому етапі на основі інформаційної технології формується множина варіантів плану підвищення якості підмножини процесів моделі зрілості SPICE з метою підтримки прийняття рішення керівництвом IT-компанії. Для цього попередньо визначається множина ефективних рішень, яка пропонується для визначення остаточного варіанта, який реалізується у подальшому.

Ключові слова: рівень зрілості, підмножина процесів, процес розробки програмного забезпечення, технологія, модель, метод послідовного аналізу варіантів, множина ефективних рішень.

Вступ. Розробка програмного забезпечення стала складним процесом, який різні автори у сфері програмної інженерії визначають у свій спосіб, намагаючись підкреслити найбільш важливі, на їхню думку, аспекти, етапи і суть даного поняття. У розділі «Інформаційні технології» глосарію «Gather» надається таке визначення процесу розробки програмного забезпечення (ПР ПЗ): управління специфікаціями, проектуванням, програмуванням, тестуванням, інсталяцією та навчанням, пов'язаними з конкретним проектом розробки додатків будь-якого розміру [1]. Це визначення відображає одне з найбільш поширених точок зору ПР ПЗ, поділеного на чіткі етапи. Іншими словами, ПР ПЗ – це структурований підхід до розробки програмного забезпечення для системи чи проекту [2]. Процес розробки програмного забезпечення також можна визначити як «набір дій, методів, практик і перетворень, які люди використовують для розробки та підтримки програмного забезпечення та пов'язаних продуктів». Згідно з Institute of Electrical and Electronics Engineers (IEEE), процес розробки програмного забезпечення – це процес, за допомогою якого потреби користувача перетворюються на програмні продукти. Для розробки програмних продуктів були прийняті різні методи розробки програмного забезпечення, такі як: модель

водоспаду, ітеративна та інкрементна модель, спіральна модель, V-модель, швидка розробка програм, модель прототипування, гнучка модель та гібридна спіральна модель [3]. Як бачимо, ПР ПЗ став сукупним поняттям, яке охоплює всі складові, які свідчать про постійне ускладнення як самого процесу розробки, так і способів контролю його якості.

Спираючись на вищевказані визначення можна прийти до висновку, що сьогодні управління розробкою є складним і комплексним завданням, яке вимагає відповідного підходу для розробки відповідної програмної системи. Почнемо з того, що навіть такий простий раніше процес як визначення вимог до кінцевого програмного продукту на даний момент потребує залучення спеціалістів і відіграє важливу роль, про що зазначено у праці [4]. Робота [5] присвячена впливу критичних факторів на успіх покращення процесу розробки. Також складність програмного забезпечення впливає і на те, що зараз навіть складно уявити, щоб процес розробки дійсно складних програмних систем відбувався за участі лише однієї команди або лише в одному місці. Тому проводяться різноманітні дослідження з питань розробки ПЗ глобально розподіленими командами, проектами та компаніями [6], [7] та виявлення бар'єрів, які обмежують обмін неявними знан-



нями в командах [8]. І це лише один з прикладів розширення ПР ПЗ додатковими процесами, про які раніше не велася й мова.

Таким чином, ці дослідження свідчать, що ПР ПЗ відіграє важливу роль у розробці якісного програмного продукту і питання його покращення актуальне.

Постановка та мета задачі дослідження. Хоча програмне забезпечення є важливим у всіх сферах сучасного світу, сама розробка програмного забезпечення ще не стала ідеальним процесом. Незважаючи на спроби застосувати методології розробки програмного забезпечення, розробка ПЗ не була постійно успішною, про що свідчить висока кількість відкладених, залишених або відхилених проєктів програмного забезпечення. Велика кількість досліджень показує, що фактори успіху проєкту можуть бути пов'язані з вартістю, часом доставки, масштабом і якістю. Таким чином, відсутність якості безпосередньо сприяє провалу проєкту. Гнучкі методології були поширені, щоб покращити якість програмного забезпечення та легше реагувати на зміни. На жаль, результати підтверджують, що цієї мети ще не досягнуто, і відсутність якості програмного забезпечення продовжує викликати занепокоєння [9].

Багато дослідників застосовували концепцію вдосконалення програмного процесу (Software Process Improvement, SPI), щоб зосередитися на якості програмного забезпечення. SPI розглядається як життєво важливий аспект оптимізації процесу розробки програмного забезпечення, особливо для малих і середніх організацій. Однією з найбільших проблем індустрії програмного забезпечення є розробка ініціатив щодо впровадження SPI, які ефективно допоможуть малим і середнім організаціям. Наразі дослідницькі зусилля зосереджені на впровадженні стандартів/фреймворків SPI для підвищення якості програмного забезпечення та підвищення продуктивності [10].

Поняття управління якістю за своєю суттю є невловимим, оскільки воно має різне значення для різних людей, на різних ролях, у різних ситуаціях і в різний час. Нижче наведено низку стандартів, які використовуються для розробки програмного забезпечення. ISO 9001 визначає набір операційних процесів і пропонує проєктування, документування, впровадження, моніторинг і постійне вдосконалення цих операційних процесів. ISO/IEC 9126: цей стандарт містить модель якості програмного забезпечення та набір показників для підтримки моделі. ISO/IEC 25010: у 2011 році цей стандарт замінив ISO 9126. Він містить модернізований набір атрибутів якості, і, зокрема, атрибутам, пов'язаним з безпекою, приділено більше уваги. ISO/IEC 15504: цей стандарт, який іноді називають SPICE, охоплює широкий набір процесів, пов'язаних із придбанням, розробкою, експлуатацією, постачанням, обслуговуванням і підтримкою програмного забезпечення. CMM/CMMI: термін CMM розшифровується як Capability Maturity Model і визначає процеси розробки програмного забезпечення. Він структурований таким чином, що організацію розробки можна визначити як таку, що належить до одного з набору рівнів зрілості, залежно від процесів, які вона має.

Моделі зрілості особливо користуються популярністю серед підходів, які підвищення якості ПР ПЗ. Тому їх використовують для управління не лише якістю, а й людськими ресурсами [11]. Але існує низка проблем, які відносяться до моделей зрілості.

Першою проблемою є відсутність формалізованого опису моделей зрілості. Хоч згідно [12] 52,17% (48 досліджень) вибраних досліджень назвали фактор «план впровадження SPI» як ефективний SSF, який має позитивний вплив на галузь малого програмного забезпечення, та все ж впровадження вдосконалення якості ПР ПЗ досі викликають труднощі у компаніях. Перш за все це пов'язано з тим, що не існує формалізованого опису для моделей зрілості. Тому дослідження в даному напрямку залишаються дуже актуальними.

Друга проблема представляє собою те, що існуючі формалізовані моделі зрілості надають оцінки якості ПР ПЗ без врахування особливостей ІТ-компанії. Для CMMI було формалізовано опис та створено програмне забезпечення [13-16], яке дозволяє компаніям легко впроваджувати дану модель зрілості, однак хоч CMMI реалізована у двох варіантах (безперервному та дискретному) та все ж має такий недолік як відсутність врахування особливостей ІТ-компанії. Вона не враховує специфічні технології та моделі життєвих циклів, що використовує конкретна компанія, не враховує особливості предметної області та цілі організації. З іншого боку, модель SPICE дозволяє компенсувати вищезгаданий недолік моделі CMMI, однак не має змоги надавати оцінку ПР ПЗ в цілому, оскільки реалізований лише в безперервному виді, що свідчить про третю проблему. Окремі моделі описують не весь ПР ПЗ, а лише окремі його процеси. Звідси витікає те, що модель SPICE може оцінювати рівні можливості окремих процесів, але для неї не сформовані поняття рівня зрілості всього ПР ПЗ ІТ-компанії.

І четвертою проблемою є відсутність моделей та методів, які описують не лише оцінку якості, а й процес планування розвитку ПР ПЗ і керування цим процесом. Таким чином, можна зробити висновок, що незважаючи на велику кількість досліджень, усі наявні методи та моделі не задовольняють потребу ІТ-компаній у наданні загальної оцінки її ПР ПЗ та оцінки окремо всіх його підпроцесів, серед яких є й ті, що відображають особливості розробки у даній ІТ-компанії, у тому числі з можливістю планування подальшого розвитку й покращення якості кожного підпроцесу, властивому цій компанії, з метою підняття загального рівня зрілості всього ПР ПЗ.

Отже метою роботи є вербальний опис технології планування підвищення якості підмножини процесів еталонної моделі зрілості SPICE, який є першою стадією розробки відповідної інформаційної технології.

Вербальний опис технології планування підвищення якості підмножини процесів еталонної моделі зрілості SPICE. Для вирішення задачі розробки моделей, методів та інформаційної технології планування підвищення рівня зрілості підмножини процесів еталонної моделі зрілості розглянемо це питання на

вербальному рівні у вигляді технології реалізації проблеми (послідовності етапів). На рис. 1 наведена у наочній формі така технологія. Розглянемо докладно її етапи реалізації з метою у подальшому формалізації кожного з них.



Рис. 1. Технологія підвищення якості підмножини процесів моделі SPICE

Етап 1. Формалізація структури еталонної моделі SPICE. Еталонна модель SPICE на відміну від моделі CMMI має тільки безперервний варіант розвитку стосовно окремих процесів та підпроцесів моделі. Кожний з них має своє призначення при реалізації процесу розробки програмного забезпечення ІТ-компанії. Враховуючи те, що метою подальшого наукового дослідження є формалізація вербального опису окремих процесів моделі зрілості SPICE, а також розробка моделі та алгоритму планування підвищення рівня зрілості деякої підмножини процесів, яким приділяється особлива увага керівництвом ІТ-компанії, в умовах обмежених ресурсів виникає проблема формалізації опису всієї множини процесів моделі SPICE. Ця множина може бути представлена у вигляді ієрархічної структури. До першого рівня ієрархії відносяться вся множина процесів і підпроцесів еталонної моделі зрілості SPICE. Другому рівню відповідає множина груп процесів. Таких груп три: основні процеси; допоміжні процеси; організаційні процеси. В свою чергу, кожна група процесів складається з певної множини категорій процесів, які відповідають третьому рівню ієрархічної системи.

З огляду на еталонну модель SPICE основні процеси містять в собі наступні категорії: процеси, що безпосередньо пов'язані зі споживачем; процеси, що встановлюють вимоги до системи і програмного

продукту, процеси реалізації та супроводу. Допоміжні процеси мають одну категорію. Це процеси, які забезпечують і підвищують продуктивність інших процесів проекту. Організаційні процеси мають наступні категорії: процеси запуску проекту та управління його ресурсами; процеси, що визначають бізнес-цілі організації і дозволяють їх досягти. На четвертому рівні знаходяться окремі процеси і підпроцеси. На п'ятому рівні – практики, на основі яких реалізуються процеси і підпроцеси.

Отже, вербально наведена структура еталонної моделі SPICE, яку у подальшому буде формалізовано з метою використання при синтезі динамічної моделі планування розвитком підмножини її процесів.

Етап 2. Метод оцінки рівня можливості окремого процесу еталонної моделі зрілості SPICE. Одним з ключових питань є формалізація проблеми оцінки рівня можливості окремого процесу/підпроцесу еталонної моделі зрілості SPICE. Фактично стоїть задача розробки методу оцінки. Відповідно до вербального опису моделі SPICE, вона має два виміри:

- «процес» – містить твердження про призначення процесів;
- «можливість» – містить набір атрибутів, які придатні до будь-якого процесу, що представляють вимірні характеристики, необхідні для управління процесом/підпроцесом і підвищення можливості його виконання.

Атрибути процесів це їх риси, які оцінюються за деякою шкалою, на якій визначається ступінь володіння процесом/підпроцесом відповідною рисою. З метою візуалізації цього питання можна провести порівняння з функцією приналежності у теорії нечітких множин (рис. 2)



Рис. 2. Наочна інтерпретація

де: *N* – не володіє; *P* – володіє частково; *L* – володіє в основному; *F* – володіє повністю.

У еталонній моделі SPICE використовується дев'ять атрибутів, які визначають рівень можливості процесу або підпроцесу. Це наступні атрибути:

- 1) виконання процесу;
- 2) управління виконанням;
- 3) управління робочими продуктами;
- 4) визначення процесу;
- 5) забезпечення процесу ресурсами;
- 6) вимірювання;
- 7) кількісне управління процесом;
- 8) зміна процесу;
- 9) безперервне удосконалення.

На основі володіння рисами, завдяки вербальному опису моделі SPICE, визначається рівень можливості окремих процесів/підпроцесів наступним чином. Процес має перший рівень можливості, якщо перший атрибут має ступінь володіння своєю рисою L/F . Другий рівень відповідає ступеню володіння рисою першого атрибута F , а другого та третього L/F . Процес/підпроцес досягає третього рівня можливості, коли перший, другий та третій атрибути мають ступінь володіння рисою F , а четвертий та п'ятий – L/F . Далі процес/підпроцес мають четвертий рівень можливості, коли атрибути з першого по п'ятий мають ступінь володіння рисою F , а шостий та сьомий – L/F . І під кінець п'ятий рівень можливості процесу/підпроцесу відповідає ступеню володіння рисою атрибутів з першого по сьомий F , а восьмого та дев'ятого – L/F .

Отже, згідно з таким вербальним описом рівня можливості окремого процесу/підпроцесу стоїть задача формалізації поняття «оцінка рівня можливості».

Етап 3. Синтез моделі планування розвитку підмножини процесів моделі SPICE. Оцінка рівня можливості окремого процесу/підпроцесу або деякої множини процесів еталонної моделі SPICE визначає рівень якості окремої складової процесу розробки програмного забезпечення, що в свою чергу впливає на рівень розробки програмних систем, підвищує ймовірність того, що IT-проект буде виконано вчасно на запланованому рівні якості та виконання наперед заданих функціональних і нефункціональних вимог. Крім цього, зменшується ризик провальних проектів. Але оцінка рівня можливості процесу є більш актуальною і важливою, якщо вона реалізується на деякому плановому періоді на множині окремих його відрізків. Отже, можна стверджувати, що така інформація на порядок більш вагома ніж статичний варіант. Набагато більш актуальною є розв'язання задачі управління (планування) такими оцінками в умовах обмежених ресурсів на основі наперед заданого критерія або множини критеріїв, кожний з яких має свої вагові коефіцієнти. Для множини процесів така задача вирішується за рахунок оптимального розподілу ресурсів між ними з урахуванням важливості кожного з них протягом планового періоду. Це дозволяє з деякою ймовірністю наперед оцінювати якість програмних продуктів, яка напряму залежить від якості процесу розробки програмного забезпечення.

Отже, виникає задача формування динамічної моделі планування підвищення рівня можливості підмножини процесів/підпроцесів моделі SPICE. Таку модель будемо синтезувати як дискретний процес на плановому періоді $[0, T]$, тривалість якого і кількість підперіодів планування визначаються керівництвом IT-компанії залежно від тих задач, які перед нею стоять. Якщо ми розглядаємо підмножину процесів моделі SPICE, то будемо вважати, що вона на кожному підперіоді планування характеризується деяким станом \bar{X}_t . Цей стан дозволяє провести оцінку рівня можливості кожного процесу і їх множини. Будемо

вважати, що $\bar{X}_0 = \bar{X}^0$ визначає початковий стан системи з погляду планового періоду. У результаті можна запропонувати наступну залежність стану \bar{X}_t від стану \bar{X}_{t-1}

$$\bar{X}_t = \bar{F}_t(\bar{X}_{t-1}, r_t), \quad t \in [1, T],$$

де r_t – варіант керуючого впливу, який залежить від фінансових ресурсів. Вище було підкреслено, що процес, який розглядається, реалізується в умовах обмежених ресурсів. Нехай $r_t \in R_t$, де R_t – множина варіантів розвитку процесу, яка залежить від виділених на t -му підперіоді фінансових ресурсів. Далі будемо вважати, що на основі стану \bar{X}_{t-1} і керуючого впливу r_t може бути проведена оцінка рівня можливості всіх процесів, які розглядаються

$$\Phi_t(\bar{X}_{t-1}, r_t), \quad t \in [1, T].$$

У результаті стоїть задача максимізувати суму таких оцінок на плановому періоді за допомогою множини управляючих впливів $\{r_t\}$. Така задача називається динамічною задачею дискретної оптимізації та може бути сформована у такий спосіб

$$\begin{aligned} \max_{\{r_t\}} \sum_{t=1}^T \Phi_t(\bar{X}_{t-1}, r_t) + \Phi_0(\bar{X}_0), \quad \bar{X}_t = \\ = \bar{F}_t(\bar{X}_{t-1}, r_t), \quad r_t \in R_t, \quad t \in [1, T]. \end{aligned}$$

Етап 4. Розробка алгоритму планування розвитку підмножини процесів еталонної моделі зрілості SPICE на основі методу послідовного аналізу варіантів. При виборі методу, а на його основі алгоритму оптимізації моделі, необхідно ретельно вивчити структуру моделі, її характеристики. Як відомо, всі моделі можна поділити на аналітичні та алгоритмічні. У цьому разі ми будемо розглядати аналітичну модель. В свою чергу, такі моделі можна поділити на моделі з безперервними змінними і дискретні моделі. Як варіант – гібридні моделі, які мають дискретні і безперервні змінні. У подальшому ми будемо розглядати модель з дискретними змінними, на цільову функцію якої накладаються ресурсні обмеження. Крім цього, необхідно підкреслити, що модель має адитивну цільову функцію.

На основі такого аналізу можна зробити висновок, що одним з найбільш придатних методів для оптимізації моделі є метод послідовного аналізу варіантів. Він дозволяє під час його використання відкидати ті допустимі рішення, які не містять оптимальних. По іншому можна сказати, що це процедура спрямованого перебору варіантів, під час якого стає відомо, що деякі з них не входять до оптимального рішення.

Одними з перших праць, в яких було запропоновано такий підхід до оптимізації моделей були наукові напрацювання таких видатних вчених як Беллман Р., Вальд А., Айзекс Р. У подальшому при

конкретизації моделі планування розвитку підмножини процесів еталонної моделі зрілості SPICE необхідно буде вибрати деякий алгоритм, який відноситься до цього методу, і адаптувати його до розв'язання поставленої задачі.

Етап 5. Інформаційна технологія планування підвищення якості підмножин процесів еталонної моделі зрілості SPICE. На цьому етапі вирішується наступна множина задач. Аналіз бізнес-процесів технології покращення якості ПР ПЗ. Визначення вимог до ПЗ. Формування діаграми варіантів використання. Розробка моделей даних. Обґрунтування вибору інструментарію для розробки ПЗ. Вибір еталонної системної архітектури. Далі кодування і тестування ПЗ.

Етап 6. Формування множини варіантів плану підвищення якості підмножини процесів моделі зрілості SPICE з метою підтримки прийняття рішень ОПР.

Розглянемо задачу підтримки прийняття рішень безпосередньо до формування плану підвищення якості підмножини процесів моделі зрілості SPICE. Необхідно підкреслити, що на п'ятому етапі технології, яку представлено на рис. 1, інформаційна технологія дозволяє визначити деякий конкретний план підвищення якості процесів моделі SPICE, але маючи тільки один варіант плану неможливо прийняти рішення з метою його реалізації на практиці. Необхідно сформулювати декілька альтернативних планів на основі варіювання множини різних параметрів і на основі теорії прийняття рішень визначити деякий компромісний варіант. Розглянемо це питання у загальному вигляді з метою його реалізації при дослідженні моделі етапу 3. По-перше, необхідно визначити, якого типу процес прийняття рішень розглядається. Можуть бути представлені два варіанта: колективне прийняття рішень і одноосібне. Звісно, що у процесі прийняття рішень приймає участь три типи осіб: експерти; особи, що формують рішення (ОФР); особи, що приймають рішення (ОПР) [17, 18]. ОФР це фактично аналітики, які розробляють модем, алгоритми і технології формування компромісних (конструктивних) рішень за згодою з ОПР. Тому фактично поняття колективного та одноосібного прийняття рішень відноситься до експертів і ОПР. У нашому разі ОПР це керівники ІТ-компанії або господарі компанії. Множина експертів формується з кваліфікованого складу працівників ІТ-компанії, керівників відповідних підрозділів. В деяких випадках залучаються фахівці з інших компаній як незалежні експерти. Перейдемо до розгляду основних етапів розв'язання задачі прийняття рішень.

Перший етап – визначення мети задачі. Будемо вважати, що мета визначена і як це було розглянуто вище – підвищення рівня зрілості ПР ПЗ.

Другий етап – формування основних умов, припущень, базових основ, вихідних даних при побудові моделі об'єкта дослідження, який реалізує поставлену мету. Будемо вважати, що на основі групи осіб, що приймають рішення за допомогою відповідних експертів, прийнято рішення, що така модель буде створена на основі еталонної моделі зрілості SPICE. А метою є планування підвищення зрілості (якості) деякої підмножини процесів моделі SPICE, яка з

погляду ОПР є найбільш важливою для ІТ-компанії на даному етапі її функціонування.

Третій етап – безпосередня побудова моделі на основі всіх допущень, які було зроблено на попередньому етапі. До формування такої моделі долучається команда ОФР, яка визначається ОПР або групою ОПР. Ця задача реалізується на основі етапів 2, 3, 4 технології підвищення якості підмножини процесів моделі SPICE (рис. 1). Будемо вважати, що і алгоритм оптимізації моделі також реалізується цією командою ОФР.

Четвертий етап – визначення основних показників (критеріїв), на основі яких у подальшому припускається аналіз множини можливих варіантів плану розвитку підмножини процесів моделі SPICE з метою визначення деякого компромісного варіанта.

Необхідно підкреслити, що при розв'язанні багатокритеріальних задач одним з підходів є визначення основного (головного) критерію, на основі якого проводиться оптимізація, а всі інші критерії повинні бути не менше (не більше) деяких встановлених величин t_i , які називаються пороговими значеннями

$$f_i(x) \rightarrow \max \quad (1)$$

$$x \in \bar{X}, f_i(x) \geq t_i, i = 1, 2, \dots, m; i \neq l \quad (2)$$

де m – кількість критеріїв, які всі максимізуються.

Оптимальним вважається будь-яке рішення x^0 цієї задачі. Таке рішення завжди є слабо ефективним, а якщо воно єдине, то і ефективним. Питання про вибір головного критерію слід вирішувати так, щоб полегшити призначення величини t_i . Практично вказується серія наборів $\{t_i\}$ і для кожного набору вирішується задача (1), (2). Якщо значення t_i великі, то обмеження можуть виявитись несумісними і задача не буде мати рішення. Тому це питання необхідно враховувати при формуванні серії наборів $\{t_i\}$ [19].

Перейдемо безпосередньо від загального підходу з розв'язання задачі багатокритеріальної оптимізації до розгляду нашої конкретної проблеми. Будемо вважати, що розглядається двокритеріальна задача, в якій першим критерієм є приріст рівня зрілості (якості) ПР ПЗ, а другим критерій – фінансові ресурси, необхідні для розв'язання цієї задачі. В свою чергу, у якості обмежень будемо використовувати фінансові ресурси, які виділяються на кожному підперіоді планування.

П'ятий етап – на основі визначених критеріїв формування множини варіантів плану підвищення якості підмножини процесів моделі SPICE. Як було визначено, на четвертому етапі задача прийняття рішень розглядається з погляду двох критеріїв. При цьому один критерій – підвищення якості ПР ПЗ – вважається головним, а другий критерій – фінансові ресурси – визначаємо як обмеження, для якого встановлюємо ряд порогових значень, кожне з яких визначає деякий план розвитку підмножини процесів моделі SPICE при фіксованих параметрах моделі. Будемо вважати, що деякі параметри можна варіювати в межах заданих інтервалів. Такими параметрами є:

- вагові коефіцієнти важливості окремих практик і процесів моделі SPICE;
- вагові коефіцієнти важливості окремих підперіодів планування;
- верхні межі обмежень на фінансові ресурси по кожному підперіоду планування при збереженні загальних фінансових ресурсів на всьому періоді планування.

На рис. 3 представлена наочна інтерпретація множини варіантів плану підвищення якості ПР ПЗ залежно від виділених ресурсів і варіювання параметрів. З огляду на фундаментальні властивості системи при фіксованих параметрах моделі можна вважати, що залежність ефективності (якості, корисності, зрілості) від вкладених ресурсів для будь-якої складної системи, наприклад ПР ПЗ, на всьому інтервалі життєвого циклу може бути якісно описана логістичною кривою, яка має три відрізки (рис. 4). Перший відрізок $[0, S_1]$ опуклий вниз. На цьому рисунку ефективність системи дуже кволо зростає від наданих ресурсів і градієнт зростання збільшується поки опукла вниз залежність не перетворюється у лінійну. На другому відрізку $[S_1, S_2]$ залежність ефективності від ресурсів наближається до лінійного характеру. На третьому відрізку $[S_2, +\infty]$ залежність ефективності від ресурсів опукла вгору. Це означає, що ресурси дають менший ефект, ніж на лінійному відрізку, і з деякого моменту нема сенсу надалі вкладати ресурси у підвищення якості системи.

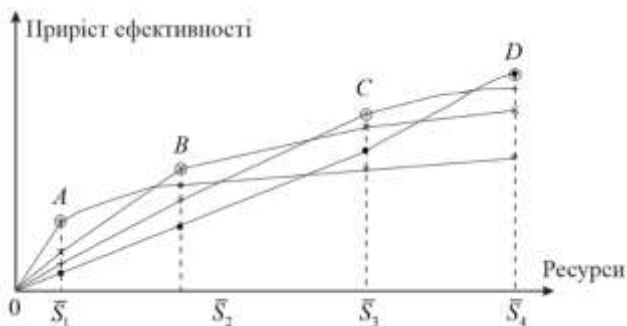


Рис. 3. Множина варіантів плану

Шостий етап – визначення множини ефективних рішень. На початку планування ми маємо деякий вихідний варіант ПР ПЗ з погляду якості процесів моделі SPICE і будемо вважати, що вкладання ресурсів починається на лінійному відрізку залежності, представленої на рис. 4. На рис. 3 визначені варіанти плану для чотирьох варіантів множини параметрів моделі, які позначені наступними символами: «·», «+», «×», «*». Кожний варіант параметрів моделі використано при оптимізації моделі для чотирьох варіантів обмежень на загальні ресурси: $\bar{S}_1, \bar{S}_2, \bar{S}_3, \bar{S}_4$. Враховуючи те, що критерій «Ресурси» мінімізується, а критерій «Приріст ефективності» максимізується, на рис. 3 визначена множина ефективних рішень. Це плани A, B, C, D, які представляються ОПП для вибору остаточного варіанта, який реалізується у подальшому.

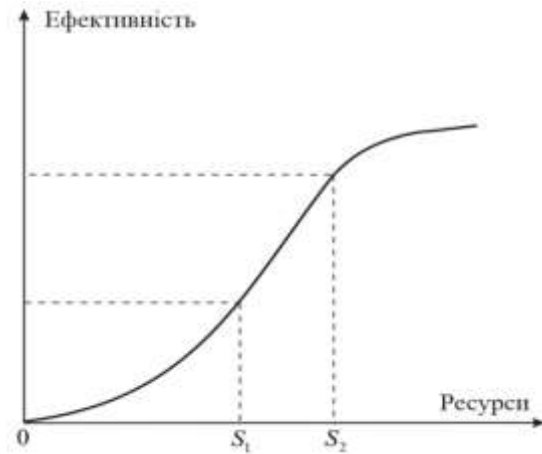


Рис. 4. Фундаментальна властивість системи

Висновки, шляхи подальших досліджень. У роботі зазначено, що для вирішення задачі розробки інформаційної технології планування підвищення рівня зрілості підмножини процесів еталонної моделі зрілості SPICE, попередньо це питання необхідно розглянути на вербальному рівні як технологію (послідовність етапів) реалізації поставленої проблеми. На першому етапі проведено формалізацію структури еталонної моделі SPICE, яка складається з множини окремих процесів. На другому етапі розглянуто метод оцінки рівня можливості окремого процесу/підпроцесу еталонної моделі зрілості SPICE, який наведено з погляду двох вимірів моделі SPICE. Третій етап технології присвячено синтезу моделі планування розвитку підмножини процесів моделі SPICE, яка визначає рівень якості окремої складової процесу розробки програмного забезпечення. На четвертому етапі технології розглядається алгоритм планування розвитку підмножини процесів еталонної моделі зрілості SPICE на основі методу послідовного аналізу варіантів. П'ятий етап присвячено інформаційній технології реалізації розробленої моделі та алгоритму. На шостому етапі на основі інформаційної технології формується множина варіантів плану підвищення якості підмножини процесів моделі зрілості SPICE з метою підтримки прийняття рішення керівництвом ІТ-компанії.

Подальші дослідження будуть проводитись у наступних напрямках:

- формалізація методу оцінки рівня можливості окремого процесу моделі зрілості SPICE;
- синтез моделі планування розвитку підмножини процесів моделі зрілості SPICE;
- розробка алгоритму планування розвитку підмножини процесів моделі зрілості SPICE;
- розробка інформаційної технології планування якості підмножини процесів моделі зрілості SPICE.

Список використаної літератури

1. *Gartner Glossary-Information Technology Glossary-S-Software Development [Electronic resource]: IT part of glossary created by company "Gartner". Electronic text. URL: <https://www.gartner.com/en/information-technology/glossary/software-development> (дата звернення 23.04.2023).*

2. *AcqNotes Program Management Tool for Aerospace: article about software development process*. URL: <https://acqnotes.com/acqnote/careerfields/software-development-process> (дата звернення 23.04.2023).
3. Singh B., Gautam S. The Impact of Software Development Process on Software Quality: A Review. *2016 8th International Conference on Computational Intelligence and Communication Networks (CICN)*. URL: https://www.researchgate.net/publication/320748110_The_Impact_of_Software_Development_Process_on_Software_Quality_A_Review (дата звернення 23.04.2023).
4. Franch X., Palomares C., Quer C. et al. *The state-of-practice in requirements specification: an extended interview study at 12 companies*. Requirements Eng. URL: <https://doi.org/10.1007/s00766-023-00399-7> (дата звернення 25.04.2023).
5. Alsawat Mohammed, Qureshi M. Rizwan. The Impact of Critical Factors for the Successful Software Process Improvement // *International Journal of Computer Applications*. 2022. Vol. 184. P. 25–33.
6. Kuhrmann C. Ebert M., Prikladnicki R. Global software engineering: evolution and trends. *2016 IEEE 11th International Conference on Global Software (ICGSE)*. 2016. P. 144–153.
7. Khan Arif, Keung Jacky. Systematic Review of success factors and barriers for Software Process Improvement in Global Software Development. *IET Software*. 2016. Vol. 10. Pp. 125–135.
8. Mtsweni Emmanuel, Gorejena Koga. Team Barriers to Tacit Knowledge Sharing in Software Development Project Teams. *Electronic Journal of Knowledge Management*. Vol. 21. 2023. P. 59–72.
9. Arcos-Medina Gloria, David Mauricio. The Influence of the Application of Agile Practices in Software Quality Based on ISO/IEC 25010 Standard: Research Anthology on Agile Software, Software Development, and Testing, edited by Information Resources Management Association. *IGI Global*. 2022. P. 1416–1443.
10. Keshta I. *Approaches to software process improvement: A state-of-the-art review*. *Journal of Software*. 2019. Vol. 14. P. 519–529.
11. Moser Patricia Cristina, de Moura Hermano Perrelli. Maturity Models for Managing People in Software Development Teams: A Systematic Literature Review. *International Institute for Science, Technology and Education (IISTE): E-Journals*. 2022. Vol. 12.
12. Balogun Abdullateef, Almomani Malek A., Basri Shuib, Almomani Omar, Capretz Luiz, Khan Arif. Towards the sustainability of small and medium software enterprises through the implementation of software process improvement: Empirical investigation. *Journal of Software: Evolution and Process*. 2022. Vol. 34. 18 p.
13. Годлевский М. Д., Брагинский И. Л. Динамическая модель и алгоритм управления качеством процесса разработки программных систем на основе модели зрелости. *Проблемы информационных технологий*. Херсон: ОЛДИ-Плюс, 2012. С. 6–13.
14. Годлевский М. Д., Брагинский И. Л. Информационная технология управления качеством процесса разработки программного обеспечения. *Східно-Європейський журнал передових технологій*. Харків, 2013, № 2/9 (62). С. 63–67.
15. Годлевский М. Д., Голоскокова А. А. Синтез статических моделей планирования улучшения качества процесса разработки программного обеспечения. *Східно-Європейський журнал передових технологій*. Харків, 2015. № 3/2 (75). С. 23–29.
16. Годлевский М. Д., Рубин Э. Е., Голоскокова А. А. Динамическая модель планирования улучшения качества процесса разработки программного обеспечения. *Вісник Національного технічного університету «ХПІ»*. Серія: Системний аналіз, управління та інформаційні технології. Харків: НТУ «ХПІ», 2015. № 58 (1167). С. 3–6.
17. Гнатієнко Г. М., Снитюк В. С. *Експертні технології прийняття рішень: монографія*. Київ: ТОВ «Маклаут», 2008. 444 с.
18. Волошин О. Ф., Машенко С. О *Теорія прийняття рішень: навч. посіб.* Київ: ВПЦ «Київський університет», 2006. 304 с.
19. Зайченко Ю. П. *Теорія прийняття рішень: підручник*. Київ: НТУУ «КПІ», 2014. 412 с.
20. *Gartner Glossary-Information Technology Glossary-S-Software Development* [Electronic resource]: IT part of glossary created by company "Gather". Electronic text. Available at: <https://www.gartner.com/en/information-technology/glossary/software-development> (accessed 23.04.2023).
21. *AcqNotes Program Management Tool for Aerospace: article about software development process*. Available at: <https://acqnotes.com/acqnote/careerfields/software-development-process> (accessed 23.04.2023).
22. Singh B., Gautam S. The Impact of Software Development Process on Software Quality: A Review. *2016 8th International Conference on Computational Intelligence and Communication Networks (CICN)*. Available at: https://www.researchgate.net/publication/320748110_The_Impact_of_Software_Development_Process_on_Software_Quality_A_Review (accessed 23.04.2023).
23. Franch X., Palomares C., Quer C. et al. *The state-of-practice in requirements specification: an extended interview study at 12 companies*. Available at: <https://doi.org/10.1007/s00766-023-00399-7> (accessed 25.04.2023).
24. Alsawat Mohammed, Qureshi M. Rizwan. (2022). The Impact of Critical Factors for the Successful Software Process Improvement. *International Journal of Computer Applications*. 2022, vol. 184, pp. 25–33
25. Kuhrmann C. Ebert M., Prikladnicki R.. Global software engineering: evolution and trends. [Глобальна програмна інженерія: еволюція та тенденції]. *2016 IEEE 11th International Conference on Global Software (ICGSE)*. 2016, pp. 144–153.
26. Khan Arif, Keung Jacky. Systematic Review of success factors and barriers for Software Process Improvement in Global Software Development. *IET Software*. 2023, vol. 21, pp. 59–72.
27. Mtsweni Emmanuel, Gorejena Koga. Team Barriers to Tacit Knowledge Sharing in Software Development Project Teams. *Electronic Journal of Knowledge Management*. 2023, vol. 21, pp. 59–72.
28. Arcos-Medina Gloria, David Mauricio. The Influence of the Application of Agile Practices in Software Quality Based on ISO/IEC 25010 Standard: Research Anthology on Agile Software, Software Development, and Testing, edited by Information Resources Management Association. *IGI Global*. 2022, pp. 1416–1443.
29. Keshta I. Approaches to software process improvement: A state-of-the-art review. *Journal of Software*. 2019, vol. 14, pp. 519–529.
30. Patricia Cristina Moser, Hermano Perrelli de Moura. Maturity Models for Managing People in Software Development Teams: A Systematic Literature Review. *International Institute for Science, Technology and Education (IISTE): E-Journals*. 2022, vol. 12.
31. Balogun Abdullateef, Almomani Malek A., Basri Shuib, Almomani Omar, Capretz Luiz, Khan Arif. Towards the sustainability of small and medium software enterprises through the implementation of software process improvement: Empirical investigation. *Journal of Software: Evolution and Process*. 2022, vol. 34. 18 p.
32. Hodlevskyy M. D., Brahynskyy Y. L. Dynamicheskaya model' y alhorytm upravlenyya kachestvom protsesssa razrabotky prohrammnykh system na osnove modely zrelosty. *Problemy ynformatsyonnykh tekhnolohyy*. Kherson: OLDY-Plyus Publ., 2012. pp. 6–13.
33. Hodlevskyy M. D., Brahynskyy Y. L. Ynformatsyonnaya tekhnolohyya upravlenyya kachestvom protsesssa razrabotky prohrammnoho obespechenyya [Information technology for quality management of the software development process]. *Vostochno-Evropeyskyy zhurnal peredovykh tekhnolohyy* [Eastern-European Journal of Enterprise Technologies]. Kharkiv, 2013, no. 2/9 (62). pp. 63–67.
34. Hodlevskyy M. D., Holoskokova A. A. Syntez statycheskykh modeley planirovaniya uluchsheniya kachestva protsesssa razrabotky prohrammnoho obespechenyya [Synthesis of static planning models for improving the quality of the software development process]. *Skhidno-Yevropeys'kyy zhurnal peredovykh tekhnolohiy* [Eastern-European Journal of Enterprise Technologies]. Kharkiv, 2015, no. 3/2 (75), pp. 23–29.
35. Hodlevskyy M. D., Rubyn E. E., Holoskokova A. A. Dynamicheskaya model' planirovaniya uluchsheniya kachestva protsesssa razrabotky prohrammnoho obespechenyya [Dynamic planning model for improving the quality of the software development process]. *Visnyk Natsional'noho tekhnichnoho universytetu "KhPI"*. Seriya: Systemnyy analiz, upravlinnya ta informatsiyini tekhnolohiyi [Bulletin of NTU "KhPI". Series: System analysis, control and information technology]. Kharkiv, NTU "KhPI" Publ., 2015, no. 58 (1167), pp. 3–6.

References (transliterated)

1. *Gartner Glossary-Information Technology Glossary-S-Software Development* [Electronic resource]: IT part of glossary created by company "Gather". Electronic text. Available at:

17. Hnatiyenko H. M., Snytyuk V. Ye. *Ekspertni tekhnolohiyi pryynyattya rishen': monohrafiya* [Expert decision-making technologies: monograph]. Kyiv, TOV «Maklout» Publ., 2008. 444 p.
18. Voloshyn O. F., Mashchenko S. O. *Teoriya pryynyattya rishen': navch. posib* [Theory of decision-making: education. manual]. Kyiv, VPTs «Kyivivs'kyi universytet» Publ., 2006. 304 p.
19. Zaychenko Yu. P. *Teoriya pryynyattya rishen': pidruchnyk* [Theory of decision-making: a textbook]. Kyiv, NTUU «KPI» Publ., 2014. 412 p.

Надійшла (received) 11.05.2023

UDC 004.4: 519.816

M. D. GODLEVSKYI, Doctor of Technical Sciences, Full Professor, National Technical University "Kharkiv Polytechnic Institute", Director of the Institute of Computer Science and Information Technology, Kharkiv, Ukraine; e-mail: god_asu@kpi.kharkov.ua, ORCID: <https://orcid.org/0000-0003-2872-0598>

G. O. BURLAKOV, National Technical University "Kharkiv Polytechnic Institute", Graduate Student, e-mail: george@sheerchain.com, ORCID: <https://orcid.org/0000-0002-0423-6024>

VERBAL DESCRIPTION OF THE TECHNOLOGY FOR PLANNING THE QUALITY IMPROVEMENT OF A SUBSET OF THE PROCESSES OF THE SPICE MATURITY REFERENCE MODEL

It is noted that to solve the problem of developing information technology for planning the increase of maturity level of a subset of the processes of the SPICE reference maturity model, this issue must first be considered at the verbal level as a technology (sequence of stages) for the implementation of the given problem. At the first stage, the structure of the SPICE reference model, which consists of a number of separate processes, was formalized. This set is presented as a hierarchical structure: the first level is a set of processes and subprocesses; the second level is a set of process groups; the third level is a set of process categories. At the second stage, the method of assessing the level of possibility of a separate process/sub-process of the reference maturity model SPICE is considered. It is given from the point of view of two dimensions of the SPICE model: the purpose of the processes; the attributes of the processes (measurable characteristics necessary for managing the process and increasing the possibility of its execution). The third stage of the technology is focused on the synthesis of planning model of the subset development processes for the SPICE model, which determines the quality level of a separate component of the software development process (SDP), which in turn has a positive effect on the level of development of software systems. Assessment and planning of the possibility level of a subset of processes is implemented during a certain planning period under conditions of limited resources based on their optimal distribution, taking into account the importance of individual processes and their practices during the planning period. At the fourth stage of the technology, an algorithm for planning the development of a subset of processes of the reference maturity model SPICE is considered based on the method of sequential analysis of options. This is primarily due to the fact that the optimization model has an additive objective function. The method allows discarding those appropriate solutions that do not contain optimal solutions. In the future, when specifying the model, it is planned to choose an algorithm that belongs to this method and adapt it to the solution of the given problem. The fifth stage is devoted to information technology implementation of the developed model and algorithm. At this stage, the following set of problems is solved. Analysis of business processes of the technology of improving the quality of software development process. Definition of software requirements. Forming a diagram of use cases. Development of data models. Justification of the choice of tools for software development. Selection of reference system architecture. Next, software coding and testing. At the sixth stage, based on information technology, a number of variants of the plan for improving the quality of a subset of the processes of the SPICE maturity model are formed in order to support decision-making by the management of the IT company. For this, a set of effective solutions is preliminarily determined, which is proposed to determine the final option, which is implemented later.

Keywords: maturity level, subset of processes, software development process, technology, model, method of sequential analysis of options, set of effective solutions.

Повні імена авторів / Author's full names

Автор 1 / Author 1: Годлевський Михайло Дмитрович, Godlevskiy Mykhaylo Dmytrovych

Автор 2 / Author 2: Бурлаков Георгій Олександрович, Burlakov Georgii Oleksandrovych

МАТЕМАТИЧНЕ І КОМП'ЮТЕРНЕ МОДЕЛЮВАННЯ

MATHEMATICAL AND COMPUTER MODELING

DOI: 10.20998/2079-0023.2023.01.07

УДК 004.42:519.8

О. Ю. МЕЛЬНИКОВ, кандидат технічних наук (PhD), доцент, Донбаська державна машинобудівна академія, доцент кафедри інтелектуальних систем прийняття рішень; м. Краматорськ, Україна; e-mail: alexandr@melnikov.in.ua; ORCID: <https://orcid.org/0000-0003-2701-8051>

О. Ю. ЗАКАБУЛА, студент, Донбаська державна машинобудівна академія; м. Краматорськ, Україна; e-mail: superuser254@gmail.com; ORCID: <https://orcid.org/0009-0008-4139-2326>

МОДЕЛЮВАННЯ РІВНЯ НЕЗАДОВОЛЕННЯ ПОТРЕБ МЕШКАНЦІВ МАЛИХ МІСТ У СИСТЕМІ ПІДТРИМКИ ПРИЙНЯТТЯ РІШЕНЬ ДЛЯ ВОДОПОСТАЧАННЯ В ЕКСТРЕМАЛЬНИХ ВИПАДКАХ

Розглядається проблема забезпечення водопостачання малих населених пунктів в екстремальних випадках. Наведено, що автори раніше сформулювали та розв'язали задачу створення системи підтримки прийняття рішень, яка дозволяє при наявності даних про кількість мешканців у кожному районі та відстанях між районами розрахувати оптимальний маршрут пересування цистерни з водою. З використанням низки параметрів (середній час обслуговування, об'єм цистерни, середній відсоток населення, що виходить за водою, обмеження на обсяг видаваної води тощо) система дозволяє скласти розклад (графік) руху цистерни, а також дати рекомендації щодо збільшення кількості цих цистерн і оптимального їхнього розподілу по районах, визначити місця їх найкращого розташування в кожному районі для максимального задоволення всіх його мешканців. Наразі розроблено математичну модель для розрахунку спеціального коефіцієнта, який дозволяє оцінити рівень незадоволення потреб мешканців в забезпеченні питною водою. Пропонований коефіцієнт містить три складові частини, а саме – віднесення об'єму рекомендованої видачі максимально можливої кількості літрів на одну людину до розрахункової; віднесення реального «часу у дорозі» до оптимально розрахованого; середня відносна віддаленість від розрахованого центру району. Створена модель додана як додатковий модуль до наявної системи підтримки прийняття рішень, наведено приклади роботи системи під час розрахунку маршруту та положення цистерн, що забезпечують м. Торезьк Донецької області. Приклади свідчать, що причиною найбільшого незадоволення є недостатня кількість машин, на другому місці – їх не дуже вдале розташування; наявний час майже не впливає на результат. Збільшення кількості машин до двох знижує першу складову коефіцієнту до одиниці. Результати свідчать про можливість зниження показника з 1,305 до 1,087.

Ключові слова: забезпечення водою, оптимальний маршрут, гугл-карти, центри кластерів, задоволення потреб, математична модель.

Вступ. Система водопостачання, яка діє у більшості невеликих українських міст, може бути порушена в результаті екстремальних подій, і доставляння води споживачеві буде здійснюватися за допомогою спеціалізованого автотранспорту. У районах (мікрорайонах, окремих кварталах) міста розташовуються тимчасові пункти розливу питної води з автоцистерн у тару споживача [1–5].

Постановка задачі та огляд літератури. Автори сформулювали [6] та розв'язали задачу створення системи підтримки прийняття рішень, яка дозволяє при наявності даних про кількість мешканців у кожному районі та відстанях між районами розрахувати оптимальний маршрут пересування цистерни з водою. З використанням таких параметрів, як середній час обслуговування, об'єм цистерни, середній відсоток населення, що виходить за водою, і обмеження на обсяг видаваної води, система дозволяє скласти розклад (графік) руху цистерни, а також дати рекомендації

щодо збільшення кількості цих цистерн і оптимального їхнього розподілу по районах [7]. Також було поставлено та вирішено [8] наступне завдання: визначити в кожному районі таке місце розміщення цистерни, щоб воно було рівновіддаленим від усіх прилеглих будинків з урахуванням кількості мешканців. Ще одна можливість, яку надав новий модуль користувачам – це визначення точних координат розташування цистерни на місцевості за допомогою гугл-карт. У роботі [9] було сформульоване нове завдання: за наявності статистичних даних про аварії, що сталися раніше на водопроводі даного міста, і тривалості ремонтних робіт у кожному випадку потрібно передбачити (тобто спрогнозувати) тривалість ремонтних робіт з відновлення водопостачання міста.

Формулювання нової математичної моделі. Було поставлене завдання визначення нового коефіцієнта, що дозволив би оцінити рівень незадоволення потреб населення в забезпеченні питною водою. Відомі

© Мельников О. Ю., Закабула О. Ю., 2023



Дослідницька стаття: Цю статтю опубліковано видавництвом *НТУ «ХПІ»* у збірнику «Вісник Національного технічного університету «ХПІ» Серія: Системний аналіз, управління та інформаційні технології». Ця стаття поширюється за міжнародною ліцензією [Creative Commons Attribution \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/). **Конфлікт інтересів:** Автор/и заявив/или про відсутність конфлікту.



рішення розрахунку (найпоширеніший спосіб – індекс споживчої задоволеності) базуються на опитуванні споживачів [10], тому неприйнятні.

Визначимо такі дані:

N – кількість факторів, що впливають на якість виконання завдання;

d_{iReal} , $i = 1 \dots N$ – значення i -го фактора, розрахованого для реальних даних;

$d_{iOptimal}$, $i = 1 \dots N$ – значення i -го фактора, розрахованого для оптимальних даних;

k_i , $i = 1 \dots N$ – коефіцієнт перевищення значення фактора, який можна розрахувати як $d_{iReal} / d_{iOptimal}$ (приймаємо $k_{iOptimal} = 1$).

Таким чином, коефіцієнт оцінки рівня задоволення потреб населення в забезпеченні питною водою в загальній постановці можна записати так (зважаючи на його прагнення до оптимального значення, тобто до одиниці):

$$K = \frac{\sum_{i=1}^N k_i}{N} \geq 1. \quad (1)$$

У [6–8] було визначено трійку впливових факторів, а саме:

– кількість спеціалізованих машин (автоцистерн), які доставляють воду;

– маршрут руху спеціалізованої техніки;

– місце розташування цистерни у районі (таке, щоб воно було рівновіддаленим від усіх прилеглих будинків та максимально зручним для мешканців).

Бажана кількість спеціалізованих машин розраховується в [7], тому першим фактором приймаємо віднесення об'єму рекомендованої видачі максимально можливої кількості літрів на одну людину до розрахункової:

$$k_1 = \begin{cases} V_{rec} / V_{calc}, & \text{якщо } > 1 \\ 1, & \text{інакше} \end{cases}, \quad (2)$$

$$V_{calc} = \frac{N_c \cdot V_{auto}}{S_{hab} \cdot Out}, \quad (3)$$

де V_{rec} – рекомендована видача максимально можливої кількості літрів на одну людину;

V_{calc} – розрахована видача максимально можливої кількості літрів на одну людину;

N_c – кількість цистерн;

V_{auto} – місткість однієї цистерни;

S_{hab} – кількість населення, яке мешкає в означених районах;

Out – доля населення, яке виходить за водою.

Параметри V_{rec} та Out визначаються заздалегідь та є константами, інші мають бути розраховані.

Розглянемо наш приклад [6–8]: $V_{rec} = 18$ літрів, $V_{auto} = 4000$ літрів, $Out = 0,01$ (приймаємо як константи), $V_{calc} = 4000 / (34378 \cdot 0,01) = 11,64$ літра, $k_1 = 18 / 11,64 = 1,55$. Якщо взяти до уваги рекомендацію збільшення кількості цистерн хоча б до двох, то на кожну цистерну

прийдеться менша кількість населення, і V_{calc} прийме значення від 23,27 літра, а $k_1 = 1$.

Розрахунок оптимального маршруту руху спеціалізованої техніки та складання його графіку наведено в [2–3]; також там є список M районів міста із зазначенням чисельності населення S_i в кожному (всього в

місті $S_{sum} = \sum_{i=1}^M S_i$ мешканців) та таблиця відстаней

між районами з урахуванням можливості або заборони прямого проїзду між кожною парою районів:

$$F(X) = \sum_{i=1}^M \sum_{j=1}^M C_{ij} X_{ij} \rightarrow \min, \quad (4)$$

де M – кількість мікрорайонів з зупинками цистерни для споживачів води;

C_{ij} , $i, j = 1 \dots M$ – «матриця витрат», C_{ij} – «витрати на перехід» з i -го мікрорайону в j -й, тобто відстань між цими мікрорайонами;

X_{ij} – матриця переходів з компонентами: $X_{ij} = 1$, якщо цистерна робить переїзд з i -го мікрорайону в j -й, $X_{ij} = 0$, якщо цистерна не виконує переїзд; $i, j = 1 \dots M$, $i \neq j$.

Маємо обмеження:

$$\sum_{i=1}^M X_{ij} = 1, j = 1 \dots M, \quad (5)$$

$$\sum_{j=1}^M X_{ij} = 1, i = 1 \dots M, \quad (6)$$

$$U_i - U_j + M \cdot X_{ij} \leq M - 1, i, j = 1 \dots M, i \neq j. \quad (7)$$

Умова (5) означає, що цистерна кожний район залишає тільки один раз; умова (6) – входить в кожен район тільки один раз; умова (7) забезпечує закритість маршруту, що містить мікрорайони M , і не має закритих внутрішніх петель. Оскільки кількість мікрорайонів у місті Торезьку дорівнює 6 [6–8], для вирішення завдання можна використовувати метод повного перебору, що вимагає перебору максимум $(n - 1)!$ варіантів ($5! = 120$).

Для розрахунку другої складової коефіцієнту можна використати «час у дорозі», тобто скільки часу потрібно цистерні для переміщення та видачі води:

$$k_2 = \frac{t_{real}}{t_{optm}}, \quad (8)$$

де t_{real} – наявний час у дорозі;

t_{optm} – час, який було розраховано у [7].

У нашому прикладі ці параметри майже однакові: $t_{real} = 540$ хвилин, $t_{optm} = 526$ хвилин, $k_1 = 540 / 526 = 1,03$.

Останній пункт математично описано так [8]. У кожному з M районів є H_i будинків ($i = 1 \dots M$), кожен

будинку D_j ($j = 1..H_i$) має умовні координати D_{jx} та D_{jy} , Необхідно визначити такі C_{ix} та C_{iy} , щоб середня відстань від будинку до центру збору (розташування машини) була мінімальною:

$$F_i(C_{ix}, C_{iy}) = \frac{\sum_{j=1}^{H_i} \sqrt{(D_{jx} - C_{ix})^2 + (D_{jy} - C_{iy})^2}}{M_i} \rightarrow \min, \quad (9)$$

$i = 1..M$.

Знаходження центрів сукупностей будинків здійснюватимемо в такий спосіб. Спочатку знаходимо попередні координати:

$$C_{ix} = \frac{\sum_{j=1}^{M_i} D_{jx}}{H_i}, C_{iy} = \frac{\sum_{j=1}^{M_i} D_{jy}}{H_i}, i = 1..M. \quad (10)$$

Потім уточнюємо ці координати одним із методів [11–13]. Також використовуємо такі припущення: координатами будинку вважаємо його «геометричний центр» (кількість будинків «особливої форми» невелика, і їх не приймаємо); оскільки в кожному районі розташовані будинки приблизно однакової поверховості, кількість мешканців у кожному будинку розраховуємо пропорційно до площі цього будинку. Центри сукупностей будинків можуть бути розраховані трьома способами [8]: «звичайний» геометричний центр кластера; центр, що враховує кількість будинків; центр, що враховує кількість мешканців.

Для розрахунку третьої складової нашого коефіцієнта використаємо середню відносну віддаленість від розрахованого центру, тобто (для приведення даних) – відносні координати:

$$k_3 = 1 + \left(\sum_{i=1}^M \left(\frac{|R_{ix} - C_{ix}|}{C_{ix}} + \frac{|R_{iy} - C_{iy}|}{C_{iy}} \right) / 2 \right) / M, \quad (11)$$

де C_{ix}, C_{iy} – розраховані координати оптимального центру i -го району, $i = 1..M$;

R_{ix}, R_{iy} – реальні координати розміщення цистерни в i -му районі.

Остаточно формула коефіцієнта оцінки рівня задоволення потреб населення під час постачання жителям малих міст питної води в екстремальних випадках буде виглядати таким чином:

$$K = \begin{cases} V_{\text{rec}} / \left(\frac{N_c \cdot V_{\text{auto}}}{S_{\text{hab}} \cdot Out} \right), \text{ якщо } > 1 \\ 1, \text{ інакше} \end{cases} + \frac{t_{\text{real}}}{t_{\text{optm}}} + 1 + \left(\sum_{i=1}^M \left(\frac{|R_{ix} - C_{ix}|}{C_{ix}} + \frac{|R_{iy} - C_{iy}|}{C_{iy}} \right) / 2 \right) / M \geq 1. \quad (12)$$

Результати розрахунків. Далі потрібно внести зміни в наявний додаток для можливості здійснення розрахунків.

Перший варіант для міста Торецьк [14] наведено на рис. 1. Він свідчить, що причиною найбільшого незадоволення є недостатня кількість машин, на другому місці – їх не дуже вдале розташування. Найвний час майже не впливає на результат.

Далі збільшимо кількість машин до двох. Це знижує першу складову коефіцієнту до одиниці (рис. 2).

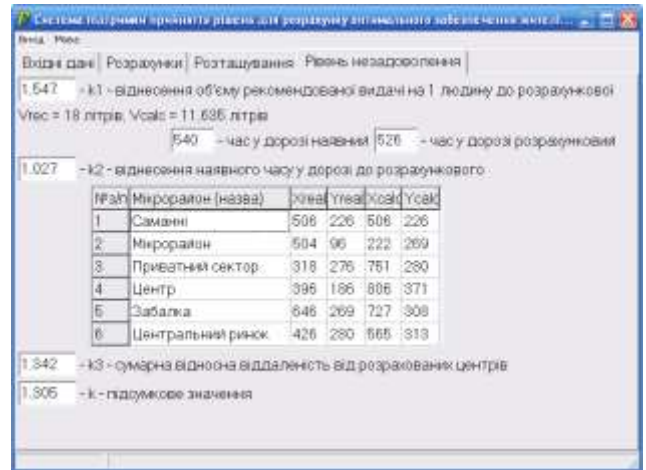


Рис. 1. Розрахунки у системі, варіант 1

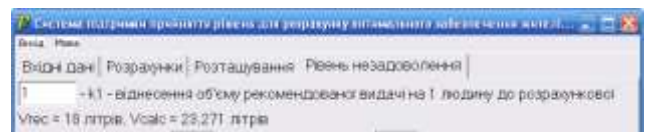


Рис. 2. Розрахунки у системі, варіант 2

Для корегування розташування машин здійснимо переміщення червоного кола найближче до синього (приклад двох районів наведено на рис. 3 та рис. 4, інші за аналогією).

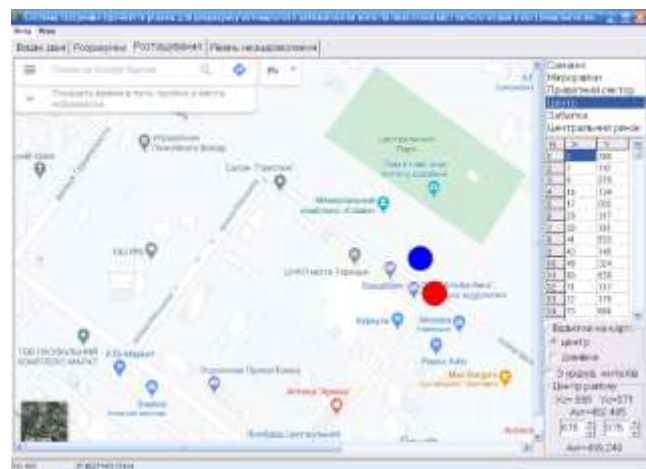


Рис. 3. Зміна розташування машини у районі Центр

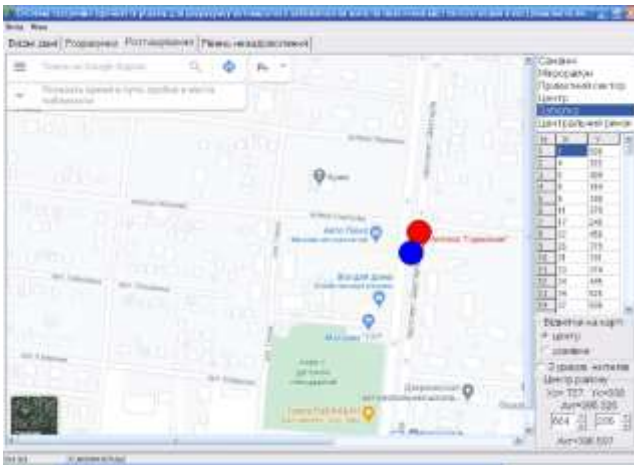


Рис. 4. Зміна розташування машини у районі Забалка

Остаточний варіант розрахунків див. на рис. 5.

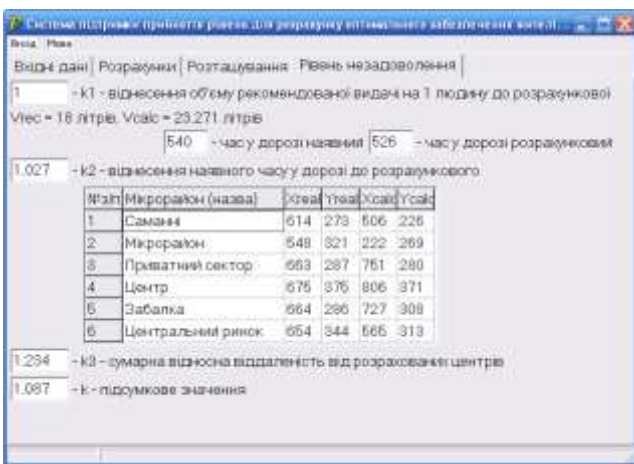


Рис. 5. Розрахунки в системі, остаточний варіант

Висновки. Наведено математичну модель розрахунку спеціального коефіцієнта, який дозволяє оцінити рівень незадоволення потреб мешканців. Приклад використання розробленої моделі для розрахунку маршруту та положення цистерн, що забезпечують м. Торезьк Донецької області, свідчить про можливість зниження показника з 1,305 до 1,087.

Список використаної літератури

1. Про питну воду, питне водопостачання та водовідведення: Закон України від 10.01.2002 р. № 2918-III. URL: <https://zakon.rada.gov.ua/laws/show/2918-14> (дата звернення: 21.10.2022).
2. Про затвердження Порядку визначення потреб населення адміністративно-територіальної одиниці у соціальних послугах: Наказ Міністерства соціальної політики України від 20.01.2014 р. №28. URL: <https://zakon.rada.gov.ua/laws/show/z0253-14> (дата звернення: 21.03.2023).
3. Свинцов А. П., Малов А. Н., Масри Г. Х. Реализация водопроводно-канализационной продукции на сегментированном рынке для жилищного фонда. *Водоснабжение и санитарная техника*. 2009. № 2. С. 23–27.
4. Синиси Л., Эртгеертс Р. *Руководство по вопросам водоснабжения, канализации и санитарно-профилактических мероприятий при чрезвычайных погодных явлениях*. URL: <https://www.google.com.ua/maps/place/Торезьк,+Донецкая+область,+85200> (дата звернення: 21.03.2023).

5. https://unece.org/DAM/env/water/whmop2/Guidance_on_water_supply_sanitation_final_ru.pdf (дата звернення: 21.03.2023).
6. De Graaf R. E., Van de Ven F. H. M. Transitions to more sustainable concepts of urban water management and water supply. *10th International Conference on Urban Drainage*. Copenhagen, Denmark, 21–26 August, 2005. URL: https://www.academia.edu/2825827/10_th_International_Conference_on_Urban_Drainage (дата звернення: 21.03.2023).
7. Закабула О. Ю. Задача розрахунку оптимального забезпечення жителів невеликих міст питною водою в екстремальних випадках. *Наукові записки молодих учених*. 2020. № 6. URL: <https://phm.cuspu.edu.ua/ojs/index.php/SNYS/article/view/1749> (дата звернення: 21.02.2023).
8. Zakabula Olexij, Melynykov Oleksandr. Decision Support System for calculating the Optimal Provision of Residents of Small Towns with Drinking Water in Extreme Cases. *Central European Researchers Journal*. CERES, 2021, volume 7, issue 73, pp. 89–94. URL: http://ceres-journal.eu/download.php?file=2021_01_10.pdf (дата звернення: 21.02.2023).
9. Мельников О. Ю., Закабула О. Ю. Модуль визначення розташування цистерн в системі підтримки прийняття рішень для оптимального забезпечення жителів невеликих міст питною водою в екстремальних випадках. *Сучасні інформаційні технології, засоби автоматизації та електропривод: матеріали V Всеукраїнської науково-технічної конференції*. Краматорськ: ДДМА, 2021. URL: <http://dSPACE.dgma.donetsk.ua/handle/DSEA/799> (дата звернення: 21.03.2023).
10. Мельников О. Ю., Закабула О. Ю. Застосування нейронних мереж для прогнозування тривалості ремонтних робіт під час пошкодження водопроводу в невеликому місті. *Нейромережні технології та їх застосування НМТіЗ-2021: збірник наукових праць XX Міжнародної наукової конференції «Нейромережні технології та їх застосування НМТіЗ-2021»*. Краматорськ: ДДМА, 2021. С. 99–102.
11. Бест Роджер. *Маркетинг от потребителя*. Москва: Манн, Иванов і Фербер, 2019. 696 с.
12. Григорків В. С., Григорків М. В. *Оптимізаційні методи та моделі: підручник*. Чернівці: Чернівецький нап. ун-т, 2016. 400 с.
13. Зайченко Ю. П. *Дослідження операцій*. Київ: ЗАТ «ВІПОЛ», 2000. 688 с.
14. Наконечний С. І., Савіна С. С. *Математичне програмування: навчальний посібник*. Київ: КНЕУ, 2003. 452 с.
15. Місто Торезьк. Google Maps. URL: <https://www.google.com.ua/maps/place/Торезьк,+Донецкая+область,+85200> (дата звернення: 21.03.2023).

References (transliterated)

1. *Pro pytnu vodu, pytne vodopostachannia ta vodovidvedennia: Zakon Ukrainy vid 10.01.2002 r. #2918-III* [On drinking water, drinking water supply and drainage: Law of Ukraine of 01.04.2014 r. #2918-III]. Available at: <https://zakon.rada.gov.ua/laws/show/2918-14> (accessed 21.10.2022).
2. *Pro zatverdzhennia Poriadku vyznachennia potreb naseleння адміністративно-територіальної одиниці у соціальних послугах: Nakaz Ministerstva sotsialnoi polityky Ukrainy vid 20.01.2014r. #28* [On approval of the Procedure for determining the needs of the population of an administrative-territorial unit in social services: Order of the Ministry of Social Policy of Ukraine from 20.01.2014 # 28]. Available at: <https://zakon.rada.gov.ua/laws/show/z0253-14> (accessed 21.03.2023).
3. Svynstov A. P., Malov A. N., Masry H. Kh. Realyzatsiya vodoprovodno-kanalizatsyonnoi produktsyy na sehmentyrovannom rynke dlia zhylyshchnoho fonda. [Sales of water and sewer products on a segmented housing market]. *Vodosnabzhenie y sanytarnaia tekhnika* [Water supply and sanitary engineering]. 2009, no 2, pp. 23–27.
4. Synysy L., Ertheerts R. *Rukovodstvo po voprosam vodosnabzhenija, kanalizacii i sanitarno-profilakticheskikh meroprijatij pri chrezvychajnykh pogodnykh javlenijah*. [Guidance on Water Supply, Sanitation and Sanitation in Extreme Weather Events]. Available at: https://unece.org/DAM/env/water/whmop2/Guidance_on_water_supply_sanitation_final_ru.pdf (accessed 21.03.2023).

5. De Graaf R. E., Van de Ven F. H. M. Transitions to more sustainable concepts of urban water management and water supply. *10th International Conference on Urban Drainage*. Copenhagen, Denmark, 21–26 August, 2005. URI: https://www.academia.edu/2825827/10_th_International_Conference_on_Urban_Drainage (accessed 21.03.2023).
6. Zakabula O. Yu. Zadacha rozrakhunku optimalnogo zabezpechennia zhyteliv nevelykykh mist pytnoiu vodoiu v ekstremalnykh vypadkakh. [The task of calculating the optimal provision of drinking water for residents of small towns in extreme cases]. *Naukovi zapysky molodykh uchenykh*. [Scientific notes of young scientists]. 2020, no 6. Available at: <https://phm.cuspu.edu.ua/ojs/index.php/SNYS/article/view/1749> (accessed 21.02.2023).
7. Zakabula Olexij, Melnykov Oleksandr. Decision Support System for calculating the Optimal Provision of Residents of Small Towns with Drinking Water in Extreme Cases. *Central European Researchers Journal*. CERES, 2021, volume 7, issue 73, pp. 89–94. Available at: http://ceres-journal.eu/download.php?file=2021_01_10.pdf (accessed 21.02.2023).
8. Melnykov O. Yu., Zakabula O. Yu. Modul vyznachennia roztashuvannia tsystem v systemi pidtrymky pryiniattia rishen dlia optimalnogo zabezpechennia zhyteliv nevelykykh mist pytnoiu vodoiu v ekstremalnykh vypadkakh. [The module for determining the location of tanks in the decision support system for optimal provision of drinking water to residents of small towns in extreme cases]. *Suchasni informatsiini tekhnologii, zasoby avtomatyzatsii ta elektropryvod: materialy V Vseukrainskoi nauково-tekhnichnoi konferentsii* [Proc. of the Int. Conf. "Modern information technologies, automation tools and electric drive"]. Kramatorsk, DDMA Publ., 2021. Available at: <http://dspace.dgma.donetsk.ua/handle/DSEA/799> (accessed 21.03.2023).
9. Melnykov O. Yu., Zakabula O. Yu. Zastosuvannia neironnykh merezh dlia prohnozuvannia tryvalosti remontnykh robit pid chas poshkodzhennia vodoprovodu v nevelykomu misti. [Application of neural networks for predicting the duration of repair work during water main damage in a small city]. *Neiromerezni tekhnologii ta yikh zastosuvannia NMTiZ-2021: zbirnyk naukovykh prats XX Mizhnarodnoi naukovoï konferentsii "Neiromerezni tekhnologii ta yikh zastosuvannia NMTiZ-2021"*. [Proc. of the Int. Conf. "Neural network technologies and their applications"]. Kramatorsk, DDMA Publ., 2021, pp. 99–102.
10. Best Roger. *Marketing ot potrebitelja*. [Marketing from the consumer]. Moscow, Mann, Ivanov i Ferber Publ., 2019. 696 p.
11. Hryhorkiv V. S., Hryhorkiv M. V. *Optimizatsiini metody ta modeli: pidruchnyk*. [Optimization methods and models: a textbook]. Chernivtsi, Chernivetskyi nats. un-t Publ., 2016. 400 p.
12. Zaichenko Yu. P. *Doslidzhennia operatsii*. [Operations Research]. Kyiv, ZAT "VIPOL" Publ., 2000. 688 p.
13. Nakonechnyi S. I., Savina S. S. *Matematychni prohramuvannia: navchalnyi posibnyk*. [Mathematical programming: a study guide]. Kyiv, KNEU Publ., 2003. 452 p.
14. *Misto Toretsk. GoogleMaps*. [The city of Toretsk]. Available at: <https://www.google.com.ua/maps/place/Торецк,+Донецкая+область,+85200> (accessed 21.03.2023).

Hadziuua (received) 27.03.2023

UDC 004.42:519.8

O. YU. MELNYKOV, Candidate of Technical Sciences (PhD), docent, Donbas State Engineering Academy, Associate Professor at the Department of Intelligent Decision Making Systems; Kramatorsk, Ukraine; e-mail: alexandr@melnikov.in.ua; ORCID: <https://orcid.org/0000-0003-2701-8051>

O. JU. ZAKABULA, student, Donbas State Engineering Academy; Kramatorsk, Ukraine; e-mail: superuser254@gmail.com; ORCID: <https://orcid.org/0009-0008-4139-2326>

MODELING THE LEVEL OF DISSATISFACTION OF THE NEEDS OF RESIDENTS OF SMALL TOWNS IN THE DECISION SUPPORT SYSTEM FOR WATER SUPPLY IN EXTREME CASES

The problem of ensuring water supply to small settlements in extreme cases is considered. It is stated that the authors previously formulated and solved the problem of creating a decision support system, which allows to calculate the optimal route of movement of a water tank with data on the number of residents in each district and the distances between districts. Using a number of parameters (average service time, tank volume, average percentage of the population going out for water, restrictions on the volume of water dispensed, etc.), the system allows you to make a schedule (schedule) of the movement of the tank, as well as give recommendations on increasing the number of these tanks and their optimal distribution by districts, to determine the places of their best location in each district for the maximum satisfaction of all its residents. Currently, a mathematical model has been developed for calculating a special coefficient that allows to assess the level of dissatisfaction of residents' needs in the provision of drinking water. The proposed coefficient contains three components, namely, the ratio of the volume of the recommended dispensing of the maximum possible number of liters per person to the calculated one; assignment of real "time on the road" to optimally calculated; the average relative distance from the calculated center of the district. The created model is added as an additional module to the existing decision-making support system, examples of the system's operation during route calculation and the position of tanks supplying the city of Toretsk, Donetsk region are given. The examples show that the reason for the biggest dissatisfaction is the insufficient number of cars, in second place is their not very good location; available time has almost no effect on the result. Increasing the number of machines to two reduces the first component of the coefficient to unity. The results indicate the possibility of reducing the indicator from 1.305 to 1.087.

Keywords: water supply, optimal route, Google maps, cluster centers, satisfaction of needs, mathematical model.

Повні імена авторів / Author's full names

Автор 1 / Author 1: Мельников Олександр Юрійович, Melnykov Oleksandr Yuriyovych

Автор 2 / Author 2: Закабула Олексій Юрійович, Zakabula Oleksij Yuriyovych

A. M. KOPP, Doctor of Philosophy (PhD), Docent, National Technical University "Kharkiv Polytechnic Institute", Associate Professor at the Department of Software Engineering and Management Intelligent Technologies, Kharkiv, Ukraine, e mail: kopp93@gmail.com, ORCID: <https://orcid.org/0000-0002-3189-5623>

D. L. ORLOVSKYI, Candidate of Technical Sciences (PhD), Docent, National Technical University "Kharkiv Polytechnic Institute", Associate Professor at the Department of Software Engineering and Management Intelligent Technologies, Kharkiv, Ukraine, e mail: orlovskiy.dm@gmail.com, ORCID: <https://orcid.org/0000-0002-8261-2988>

AN ALGORITHM FOR NLP-BASED SIMILARITY MEASUREMENT OF ACTIVITY LABELS IN A DATABASE OF BUSINESS PROCESS MODELS

Business process modeling is an important part of organizational management since it enables companies to obtain insights into their operational workflows and find opportunities for development. However, evaluating and quantifying the similarity of multiple business process models can be difficult because these models frequently differ greatly in terms of structure and nomenclature. This study offers an approach that uses natural language processing techniques to evaluate the similarity of business process models in order to address this issue. The algorithm uses the activity labels given in the business process models as input to produce textual descriptions of the associated business processes. The algorithm includes various preprocessing stages to guarantee that the textual descriptions are correct and consistent. First, single words are retrieved and transformed to lower case from the resulting textual descriptions. After that, all non-alphabetic and stop words are removed from the retrieved words. The remaining words are then stemmed, which includes reducing them to their base form. The algorithm evaluates the similarity of distinct business process models using similarity measures, including Jaccard, Sorensen – Dice, overlap, and simple matching coefficients, after the textual descriptions have been prepared and preprocessed. These metrics provide a more detailed understanding of the similarities and differences across various business process models, which can then be used to influence decision-making and business process improvement initiatives. The software implementation of the proposed algorithm demonstrates its usage for similarity measurement in a database of business process models. Experiments show that the developed algorithm is 31% faster than a search based on the SQL LIKE clause and allows finding 18% more similar models in the business process model database.

Keywords: business process model, database of business process models, natural language processing, similarity measurement algorithm, activity labels, software implementation of the algorithm.

Introduction. Business process modeling is the baseline technique of the Business Process Management (BPM) approach. It focuses on the depiction of organizational workflows in the form of visual diagrams similar to workflows but focused on business activities rather than programming tasks. Business process modeling helps to describe activities visually to train new employees, detect inefficient spots in the company operations, capture requirements to enterprise information systems, design new business processes, etc.

Today BPMN is the de-facto standard for business process modeling maintained by the Object Management Group (OMG) since 2005 and then updated to the BPMN 2.0 in 2011 [1]. This modeling notation has been extended to the XML-based (eXtensible Markup Language) language suitable not only for visual depiction of business process scenarios but also to exchange created diagram files between heterogeneous BPM suites and execute depicted workflows using BPM engines that can automate routine process scenarios.

Therefore, organizations at the higher levels of BPM maturity tend to continuously improve their business activities using BPMN modeling techniques. However, it does not mean enterprises should deal with BPM decisions only based on their resources and experience. Most organizations used so-called “reference models” – collections of typical business processes, generic or industry-specific. Such reference models accumulate proven industry standards, and best practices of multiple successful companies, and could be customized according to particular business needs [2]. Some of the most widely-

spread collections of reference business process models are Process Classification Framework (PCF) by American Productivity & Quality Center (APQC) and Supply-Chain Operations Reference (SCOR). While APQC’s PCF is the cross-industry taxonomy of business processes [3], SCOR focuses on logistics process areas, such as supply, manufactory, and delivery [4].

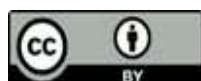
Hence, many organizations may face the problem of searching for similar business process models in collections of business process models, such as APQC’s PCF or SCOR. The solution should provide the capability to find similar BPMN models to a given model or only by the textual description of a business process if an organization does not have a BPMN model yet.

Literature review and problem statement. The similarity measurement between business process models has been studied in many papers. Some of the most relevant studies ([5]; [6]; [7]; [8]; [9]; [10]) are described below.

Paper [5] introduces the similarity search problem, where the objects in a collection and a query object are business process models. The authors consider the similarity search as the comparison of the query object against a collection of objects to identify ones that are close to the query object [5].

Study [6] proposes three business process model similarity metrics: (i) “node matching” metric based on labels comparison, (ii) “structural similarity” metric based on topology comparison, and (iii) “behavioral similarity” based on causal relations [6].

In paper [7], authors mention that comparing business process models is a complex problem, performed mostly



manually. This is why the authors propose an approach to measure the semantic similarity between business process models in an automated manner [7].

The authors of [8] have identified a linear search of similar business process models when a query model is compared to each model in a collection as inefficient and computationally complex. Hence, this paper suggests a fast similarity search algorithm based on the comparison of business process model features [8].

In paper [9] authors propose the “behavioral-based” comparison of business process models based on the causal footprints captured formally using Petri nets and informally using Event-driven Process Chain (EPC) notation [9].

A previous study in this field proposes the business process model similarity metric based on the graph structural characteristics [10]. It allows comparing models described using different notations and standards and takes into account not only workflow elements but also the secondary objects given in business process models (e.g., organizational units, data objects, information systems, etc.) [10].

According to reviewed studies ([5]; [6]; [7]; [8]; [9]; [10]), business process models were mostly compared by their structure or behavior, but the label comparison is the less elaborated approach. Due to the lack of studies that measure similarity between business process models using NLP (Natural Language Processing) techniques, this study should propose the respective approach to bridge the gap.

Research objective and tasks. The objective of this paper is the improvement of a search process for similar business process models within and across organizational collections and reference process libraries.

Therefore, the following tasks are considered:

- to propose the NLP-based algorithm to measure business process model similarity using activity labels given in business process models;
- to use the proposed algorithm to compare business process models given in different notations;
- to use the proposed algorithm compare textual descriptions to business process models.

However, compared business process models should be machine-readable, e.g. based on eXtensible Markup Language (XML), JavaScript Object Notation (JSON), or Yet Another Markup Language (YAML) formats.

NLP-based similarity measurement of business process models. The proposed algorithm for NLP-based similarity measurement of business process models includes the following steps:

- take two BPMN business process models A and B as the input and extract activity labels from these two models to obtain two collections L_A and L_B :

$$\begin{aligned} L_A &= \{l_{Ai}, i = \overline{1, n}\}, \\ L_B &= \{l_{Bj}, j = \overline{1, m}\}, \end{aligned} \quad (1)$$

where l_{Ai} – the i -th activity label extracted from the business process model A ;

l_{Bj} – the j -th activity label extracted from the business process model B ;

n – is the number of activity labels extracted from the business process model A ;

m – is the number of activity labels extracted from the business process model B ;

- split labels (1) into words and change obtained words to lower case – two sets of words W_A and W_B will be obtained as the result:

$$\begin{aligned} W_A &= \{w_{Ai}, i = \overline{1, p}\}, \\ W_B &= \{w_{Bj}, j = \overline{1, q}\}, \end{aligned} \quad (2)$$

where w_{Ai} – the i -th word extracted from labels of the business process model A ;

w_{Bj} – the j -th word extracted from labels of the business process model B ;

p – is the number of words extracted from labels of the business process model A ;

q – is the number of words extracted from labels of the business process model B ;

• remove non-alphabetic and stop words from the previously obtained sets of words (2) to get cleansed sets of words C_A and C_B respectively:

$$\begin{aligned} C_A &= \{c_{Ai}, i = \overline{1, r}\}, \\ C_B &= \{c_{Bj}, j = \overline{1, s}\}, \end{aligned} \quad (3)$$

where c_{Ai} – the i -th meaningful word that describes the business process model A ;

c_{Bj} – the j -th meaningful word that describes the business process model B ;

r – is the number of meaningful words that describe the business process model A ;

s – is the number of meaningful words that describe the business process model B ;

- stem remaining words (3) to finally obtain sets of words U_A and U_B :

$$\begin{aligned} U_A &= \{u_{Ai}, i = \overline{1, x}\}, \\ U_B &= \{u_{Bj}, j = \overline{1, y}\}, \end{aligned} \quad (4)$$

where u_{Ai} – the i -th stemmed word that describes the business process model A ;

u_{Bj} – the j -th stemmed word that describes the business process model B ;

x – is the number of stemmed words that describe the business process model A ;

y – is the number of stemmed words that describe the business process model B ;

- measure similarity between these two sets of words U_A and U_B (4) using Jaccard $J(U_A, U_B)$, Sorensen – Dice $SDC(U_A, U_B)$, overlap $overlap(U_A, U_B)$, and simple matching $SMC(U_A, U_B)$ coefficients [11]:

$$J(U_A, U_B) = \frac{|U_A \cap U_B|}{|U_A \cup U_B|}, \quad (5)$$

$$SDC(U_A, U_B) = \frac{2|U_A \cap U_B|}{|U_A| + |U_B|}, \quad (6)$$

$$overlap(U_A, U_B) = \frac{2|U_A \cap U_B|}{\min\{|U_A|, |U_B|\}}, \quad (7)$$

$$SMC(U_A, U_B) = \frac{|U_A \cap U_B|}{|U_A| + |U_B|}. \quad (8)$$

This algorithm is the improved algorithm for semantic quality analysis of business process models proposed earlier [12]. Whereas the earlier proposed algorithm was supposed to measure the closeness of a business process model to the textual description of a real business process, now the elaborated algorithm considers the comparison of business process models based on the semantic closeness of their activities (5) – (8).

An alternative algorithm takes a textual business process description *A* and a BPMN model *B* if the business process model that should be used as the search query does not exist. In this case, activity labels should be extracted only from the BPMN model, while the given textual description could be immediately split into a set of words turned into the lower case style (4).

The flowchart of the proposed algorithm is given below in fig. 1.

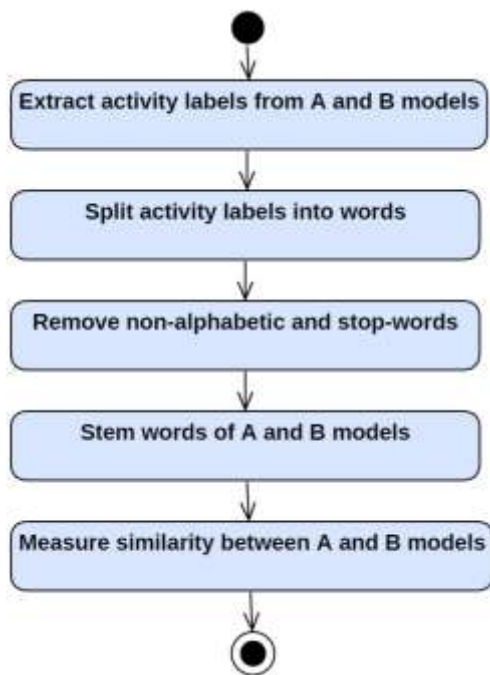


Fig. 1. The algorithm for NLP-based similarity measurement of business process models

Now it is necessary to verify the proposed algorithm. It can be implemented as a software component using the Python programming language [13] and Natural Language Toolkit (NLTK) [14]. The NLTK software platform is a

leading solution for building Python programs that handle natural language processing [14].

Experimental usage of the proposed algorithm. In this section we demonstrate the experimental usage of the proposed algorithm (fig. 1).

The products supply process [4] according to the SCOR model involves scheduling product deliveries with the supplier, receiving the products at a specified location, verifying the goods to ensure they meet requirements, transferring the goods to the appropriate storage location, and finally authorizing payment to the supplier after the goods have been successfully delivered and inspected.

The BPMN model of described supply process [4] is demonstrated in fig. 2.

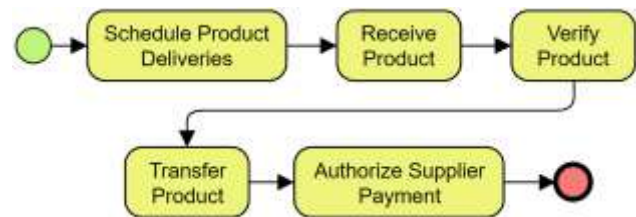


Fig. 2. The supply SCOR business process model [4]

The products delivery process [4] according to the SCOR model includes such steps as receiving, entering, and verifying an order, reserving inventory and setting a delivery date, preparing the products for delivery and loading them onto a vehicle, shipping the products, and invoicing the customer.

The BPMN model of described delivery process [4] is demonstrated in fig. 3.

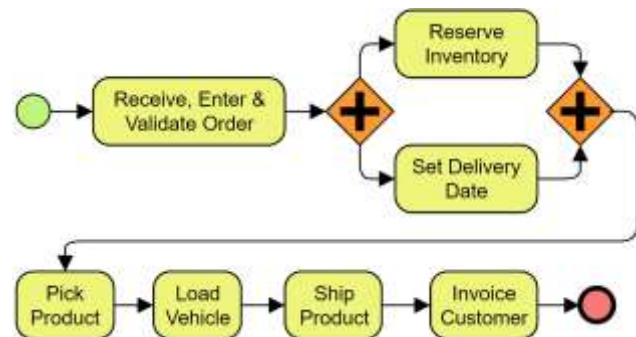


Fig. 3. The delivery SCOR business process model [4]

Two example models of supply (fig. 2) and delivery (fig. 3) business processes may be evaluated to illustrate the efficiency of the proposed algorithm in determining the similarity between such business process models.

The algorithm can construct and compare their relative textual descriptions by importing the activity labels from BPMN files that correspond these two business process models, providing for a more comprehensive understanding of their similarities and differences.

This data may be further used to discover areas for improvement and to enhance delivery and supply chain procedures for increased efficiency and profitability if comparing real company workflows toward reference models, such as SCOR [4] or APQC's PCF [3].

The original activity labels and obtained sets of words for supply and delivery business process models offered by the SCOR model [4] are demonstrated in Table 1.

Table 1 – The sets of words obtained for supply and delivery SCOR business process models

Business process activity labels	Sets of words (4)
Schedule Product Deliveries	schedul, product, deliveri, receiv, verifi, transfer, author, supplier, payment
Receive Product	
Verify Product	
Transfer Product	
Authorize Supplier Payment	receiv, enter, valid, order, reserv, inventori, set, deliveri, date, pick, product, load, vehicl, ship, invoic, custom
Receive, Enter & Validate Order	
Reserve Inventory & Set Delivery Date	
Pick Product	
Load Vehicle	
Ship Product	
Invoice Customer	

The similarity measurement between the two sets of words demonstrated in Table 1 using coefficients (5) – (8) and the proposed algorithm (fig. 1) allowed us to obtain the following values:

- 0.14 using the Jaccard coefficient;
- 0.24 using the Sorencen – Dice coefficient;
- 0.33 using the overlap coefficient;
- 0.14 using the simple matching coefficient.

The comparison histogram of business process model similarities calculated using coefficients (5) – (8) is shown in fig. 4.

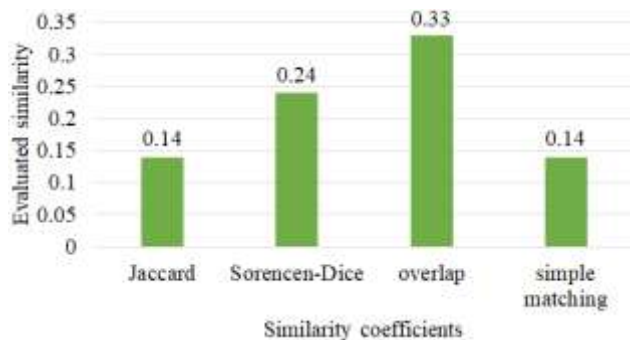


Fig. 4. The comparison of business process model similarities calculated using coefficients (5) – (8)

According to the obtained results, Jaccard and simple matching coefficients demonstrate equal similarity values of 0.14, since in this particular case the mutual absence of words in U_A and U_B is impossible. Sorensen – Dice and overlap coefficients show relatively low similarity values of 0.24 and 0.33 respectively for the given business process activity labels. Thus, further study in this direction may omit the simple matching coefficient, while focusing on the remaining ones.

The software tool for similarity measurement in a database of business process models. The software tool that implements the proposed algorithm and a database (DB) of business process (BP) models is demonstrated in fig. 5. The software tool, which implements the proposed algorithm, is a Python command-line application [13] that uses NLTK [14] and MySQL Connector [15]. Hence, the

database of business process models is a relational schema created using MySQL database management system [16].

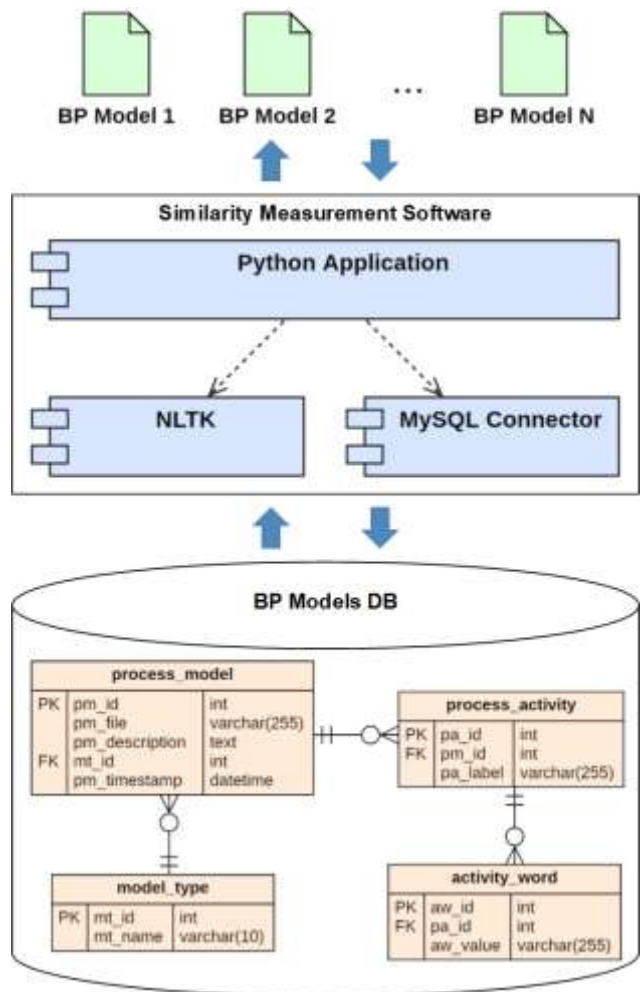


Fig. 5. Business process models database and software components that implement the proposed algorithm

According to fig. 5, the database of business process models stores text descriptions (i.e. the “pm_description” attribute) built from activity labels, activity label values (i.e. the “pa_label” attribute), and words extracted from activity labels (i.e. the “aw_value” attribute) for similarity measurement according to the proposed algorithm (fig. 1).

The software component “Python Application” that implements the proposed algorithm (fig. 1) uses Jaccard coefficient (5) for similarity measurement.

Experimental usage of the software tool. First of all, we have create the view “test_similarity” for querying the DB of BP models. The first query, which uses SQL LIKE clause, is shown in fig. 6.

```

1 sql = r"""SELECT
2     DISTINCT file_name
3     FROM
4     test_similarity
5     WHERE
6     description LIKE
7     '%invite logistic company%'"""

```

Fig. 6. The search query using SQL LIKE clause

The second query (fig. 7) uses prepared words for BP models according to the proposed algorithm (fig. 1) and the stemmed search words (we are using the Porter stemming algorithm [17]).

```

1 sql = """SELECT
2     file_name,
3     COUNT(file_name) / 3 AS similarity
4 FROM
5     test_similarity
6 WHERE
7     word IN (%s, %s, %s)
8 GROUP BY
9     file_name
10 HAVING
11     similarity = 1
12 ORDER BY
13     similarity DESC"""
14
15 val = ("invite", "logistic", "company")
16 val = tuple([porter_stemmer.stem(word)
17             for word in val])

```

Fig. 7. The search query using the proposed algorithm

According to fig. 7, the similarity degree is calculated using Jaccard index (5) and then used to filter only 100% matches. Both queries in fig. 6 and 7 were looking for the BP models containing “invite logistic company” activities or similar.

The comparison of search results is given in Table 2.

Table 2 – The comparison of search results

Similarity search	Seconds (average time)	Found BP models
SQL LIKE	0.0124	2 (3%)
Proposed algorithm	0.0085	14 (21%)

According to Table 2, the proposed algorithm is 31% faster than SQL LIKE clause (we measured the average time of 10 executions), while allowing to find 18% more similar BP models (14 against 2, from a collection of 68 models of a goods dispatch process [18]).

These results are compared visually in fig. 8.

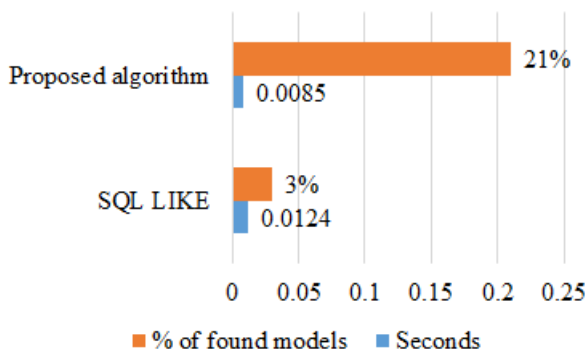


Fig. 8. The histogram of querying performance comparison

Conclusions. In this study, the algorithm for NLP-based similarity measurement of business process models is proposed.

1. The proposed algorithm (fig. 1) is based on the natural language processing techniques (such as

tokenization, stemming, stop words elimination) used to compare textual activity labels of business process models or business process descriptions to measure their similarity.

2. This algorithm was verified using supply (fig. 2) and delivery (fig. 3) BPMN process models based on the SCOR reference model. Obtained results (fig. 4) demonstrate Jaccard and simple matching coefficients give the same values when comparing two sets with the impossible mutual absence of elements. Thus it is proposed to use Jaccard or simple matching coefficient in the further search for similar business process models.
3. The proposed algorithm is implemented using Python and NLTK to measure similarity in the database of BP models created using MySQL. Experimental results demonstrate that proposed algorithm is 31% faster than the SQL LIKE clause and allows to find 18% more similar BP models than the SQL LIKE clause.
4. However, the limitation of the proposed approach is the necessity of preliminary process of business process models to apply the proposed algorithm. This requires processing of large collections of BP models, which may require significant computing resources. Nevertheless, such a pre-processing of BP models takes place much less frequently than search for similar business process models.

Future work includes elaboration of computational techniques and software solutions for efficient similarity search in large collections of BP models.

References

1. Geiger M. et al. *BPMN 2.0: The state of support and implementation*. URL: <https://doi.org/10.1016/j.future.2017.01.006> (access date: 01.04.2023).
2. Fettke P. et al. *Business Process Reference Models: Survey and Classification*. URL: https://doi.org/10.1007/11678564_44 (access date: 01.04.2023).
3. *APQC Process Classification Framework*. URL: <https://www.signavio.com/reference-models/apqc-framework/> (access date: 01.04.2023).
4. *SCOR Model*. URL: <https://scor.ascm.org/> (access date: 01.04.2023).
5. Dumas M. et al. *Similarity Search of Business Process Models*. URL: <http://sites.computer.org/debull/A09sept/marlon.pdf> (access date: 02.04.2023).
6. Dijkman R. *Similarity of business process models: Metrics and evaluation*. URL: <https://doi.org/10.1016/j.is.2010.09.006> (access date: 02.04.2023).
7. Humm B. G., Fengel J. *Semantics-Based Business Process Model Similarity*. URL: https://doi.org/10.1007/978-3-642-30359-3_4 (access date: 02.04.2023).
8. Yan Z., Dijkman R. *Fast business process similarity search*. URL: <https://doi.org/10.1007/s10619-012-7089-z> (access date: 02.04.2023).
9. van Dongen B. et al. *Measuring Similarity between Business Process Models*. URL: https://doi.org/10.1007/978-3-540-69534-9_34 (access date: 02.04.2023).
10. Kopp A. M., Orlovskiy D. L. *Estimation and analysis of business process models similarity in enterprise continuum repository*. URL: <https://doi.org/10.20535/SRIT.2308-8893.2018.4.04> (access date: 02.04.2023).
11. Verma V., Aggarwal R. K. *A comparative analysis of similarity measures akin to the Jaccard index in collaborative recommendations: empirical and theoretical perspective*. URL: <https://doi.org/10.1007/s13278-020-00660-9> (access date: 04.04.2023).

12. Kopp A., Orlovskiy D. *The approach and the software tool to calculate semantic quality measures of business process models*. URL: <http://dx.doi.org/10.20998/2079-0023.2022.02.12> (access date: 04.04.2023).
13. *Python*. URL: <https://www.python.org/> (access date: 06.04.2023).
14. *NLTK*. URL: <https://www.nltk.org/> (access date: 06.04.2023).
15. *MySQL Connector/Python Developer Guide*. URL: <https://dev.mysql.com/doc/connector-python/en/> (access date: 06.04.2023).
16. *MySQL*. URL: <https://www.mysql.com/> (access date: 07.04.2023).
17. *Porter Stemmer*. URL: <https://tartarus.org/martin/PorterStemmer/> (access date: 08.04.2023).
18. *BPMN for research*. URL: <https://github.com/camunda/bpmn-for-research> (access date: 10.04.2023).
7. Humm B. G., Fengel J. *Semantics-Based Business Process Model Similarity*. Available at: https://doi.org/10.1007/978-3-642-30359-3_4 (accessed 02.04.2023).
8. Yan Z., Dijkman R. *Fast business process similarity search*. Available at: <https://doi.org/10.1007/s10619-012-7089-z> (accessed 02.04.2023).
9. van Dongen B. et al. *Measuring Similarity between Business Process Models*. Available at: https://doi.org/10.1007/978-3-540-69534-9_34 (accessed 02.04.2023).
10. Kopp A. M., Orlovskiy D. L. *Estimation and analysis of business process models similarity in enterprise continuum repository*. Available at: <https://doi.org/10.20535/SRIT.2308-8893.2018.4.04> (accessed 02.04.2023).
11. Verma V., Aggarwal R. K. *A comparative analysis of similarity measures akin to the Jaccard index in collaborative recommendations: empirical and theoretical perspective*. Available at: <https://doi.org/10.1007/s13278-020-00660-9> (accessed 04.04.2023).

References (transliterated)

1. Geiger M. et al. *BPMN 2.0: The state of support and implementation*. Available at: <https://doi.org/10.1016/j.future.2017.01.006> (accessed 01.04.2023).
2. Fettke P. et al. *Business Process Reference Models: Survey and Classification*. Available at: https://doi.org/10.1007/11678564_44 (accessed 01.04.2023).
3. *APQC Process Classification Framework*. Available at: <https://www.signavio.com/reference-models/apqc-framework/> (accessed 01.04.2023).
4. *SCOR Model*. Available at: <https://scor.ascm.org/> (accessed 01.04.2023).
5. Dumas M. et al. *Similarity Search of Business Process Models*. Available at: <http://sites.computer.org/debull/A09sept/marlon.pdf> (accessed 02.04.2023).
6. Dijkman R. *Similarity of business process models: Metrics and evaluation*. Available at: <https://doi.org/10.1016/j.is.2010.09.006> (accessed 02.04.2023).
12. Kopp A., Orlovskiy D. *The approach and the software tool to calculate semantic quality measures of business process models*. Available at: <http://dx.doi.org/10.20998/2079-0023.2022.02.12> (accessed 04.04.2023).
13. *Python*. Available at: <https://www.python.org/> (accessed 06.04.2023).
14. *NLTK*. Available at: <https://www.nltk.org/> (accessed 06.04.2023).
15. *MySQL Connector/Python Developer Guide*. Available at: <https://dev.mysql.com/doc/connector-python/en/> (accessed 06.04.2023).
16. *MySQL*. Available at: <https://www.mysql.com/> (accessed 07.04.2023).
17. *Porter Stemmer*. Available at: <https://tartarus.org/martin/PorterStemmer/> (accessed 08.04.2023).
18. *BPMN for research*. Available at: <https://github.com/camunda/bpmn-for-research> (accessed 10.04.2023).

Received 05.05.2023

УДК 004.94

А. М. КОПП, доктор філософії (PhD), доцент, Національний технічний університет «Харківський політехнічний інститут», доцент кафедри програмної інженерії та інтелектуальних технологій управління, м. Харків, Україна, e-mail: kopp93@gmail.com, ORCID: <https://orcid.org/0000-0002-3189-5623>

Д. Л. ОРЛОВСЬКИЙ, кандидат технічних наук (PhD), доцент, Національний технічний університет «Харківський політехнічний інститут», доцент кафедри програмної інженерії та інтелектуальних технологій управління, м. Харків, Україна, e-mail: orlovskiy.dm@gmail.com, ORCID: <https://orcid.org/0000-0002-8261-2988>

АЛГОРИТМ ВИМІРЮВАННЯ ПОДІБНОСТІ МІТОК ДІЯЛЬНОСТЕЙ НА ОСНОВІ NLP У БАЗІ ДАНИХ МОДЕЛЕЙ БІЗНЕС-ПРОЦЕСІВ

Моделювання бізнес-процесів є важливою частиною організаційного управління, оскільки дозволяє компаніям отримати уявлення про свої операційні бізнес-процеси та знайти можливості для розвитку. Однак оцінити та кількісно виміряти схожість декількох моделей бізнес-процесів може бути складно, оскільки ці моделі часто сильно відрізняються за структурою та номенклатурою. Це дослідження пропонує підхід, який використовує методи обробки природної мови для оцінки схожості моделей бізнес-процесів, для розв'язку цієї задачі. Алгоритм використовує мітки діяльності, наведені в моделях бізнес-процесів, як вхідні дані для створення текстових описів пов'язаних бізнес-процесів. Алгоритм включає декілька етапів попередньої обробки, щоб гарантувати, що текстові описи є коректними і послідовними. Спочатку з отриманих текстових описів вилучаються окремі слова і представляються у нижньому регістрі. Після цього з отриманих слів видаляються всі нелітерні та стоп-слова. Потім слова, що залишилися, піддаються стемінгу, тобто приведенню до їхньої базової форми. Після підготовки та попередньої обробки текстових описів алгоритм оцінює схожість різних моделей бізнес-процесів за допомогою мір схожості, включаючи коефіцієнти Жаккара, Соренсена – Дайса, перетину та простого коефіцієнту відповідності. Ці метрики забезпечують більш детальне розуміння подібності і відмінності між різними моделями бізнес-процесів, які потім можуть бути використані для впливу на прийняття рішень та ініціатив щодо вдосконалення бізнес-процесів. Програмна реалізація запропонованого алгоритму демонструє його використання для вимірювання подібності в базі даних моделей бізнес-процесів. Експерименти демонструють, що розроблений алгоритм є на 31 % швидшим за пошук на основі виразу SQL LIKE, а також дозволяє знайти на 18 % більше подібних моделей у базі даних моделей бізнес-процесів.

Ключові слова: модель бізнес-процесу, база даних моделей бізнес-процесів, обробка природної мови, алгоритм вимірювання подібності, мітки діяльності, програмна реалізація алгоритму.

Повні імена авторів / Author's full names

Автор 1 / Author 1: Копп Андрій Михайлович, Kopp Andrii Mykhailovych

Автор 2 / Author 2: Орловський Дмитро Леонідович, Orlovskiy Dmytro Leonidovych

А. І. ЛЕВТЕРОВ, кандидат технічних наук, професор, Харківський національний автомобільно-дорожній університет, завідувач кафедри інформатики та прикладної математики; м. Харків, Україна, e-mail: lai@khadi.kharkov.ua, ORCID: <https://orcid.org/0000-0001-6586-1061>

Г. А. ПЛЕХОВА, кандидат технічних наук, доцент, Харківський національний автомобільно-дорожній університет, доцент кафедри інформатики та прикладної математики; м. Харків, Україна, e-mail: plehovaanna1@gmail.com, ORCID: <https://orcid.org/0000-0002-6912-6520>

М. В. КОСТИКОВА, кандидат технічних наук, доцент, Харківський національний автомобільно-дорожній університет, доцент кафедри інформатики та прикладної математики; м. Харків, Україна, e-mail: kmv_topaz@ukr.net, ORCID: <https://orcid.org/0000-0001-5197-7389>

А. О. ОКУНЬ, кандидат технічних наук, доцент, Національний технічний університет «Харківський політехнічний інститут», доцент кафедри комп'ютерного моделювання та інтегрованих технологій обробки тиском; м. Харків, Україна, e-mail: okunanton@gmail.com, ORCID: <https://orcid.org/0000-0002-6467-4229>

ГЕОМЕТРИЧНЕ МОДЕЛЮВАННЯ ТРАС І ПОТОКІВ

Досліджено та розроблено математичні моделі для вирішення задач оптимізації з'єднання в невідносних областях за типових технологічних обмежень на геометричні та топологічні параметри трас, насамперед, на кривизну та кількість вигинів. Моделі пов'язані з існуючими та перспективними топогеодезичними моделями полігональних зображень територій. Розв'язання задач зв'язку передбачає пошук оптимальних траєкторій маршрутів і сіток у межах необмежених геометричних форм. Для цього потрібна розробка безлічі загальних моделей як полів, де здійснюються зв'язки. Сполучення можуть бути різних типів, таких як гнучкі, манхеттенські, рівні, тверді, а також маршрути інших типів. Смеляков та Алісейко (Плехова Г. А.) зауважують, що глобальне та локальне регулювання геометричних зв'язків для розв'язання задач зв'язків можна представити як загальну оптимізаційну задачу зв'язку, яка визначається як задача вибору $\langle \Omega, R \rangle$, де R – набір альтернатив, Ω – принцип оптимальності. При цьому набір можна представити як сукупність фазового простору ϕ та обмежень Q , які застосовуються до параметрів фазового простору ϕ . У свою чергу, доцільно уявити, що фазовий простір ϕ є декартовим добутком $\phi = X * Y * Z * U$ вихідних даних X , збурень Y , параметрів керування U та результатів Z . Аналіз задачі свідчить про те, що насамперед ефективність моделювання фазового простору ϕ пов'язана з описом вихідних даних X про площу F і простір L можливих магістралей в F . Питання досліджується як розробка побудови структур моделей та методології їх використання, які б уможливили конструктивне та ефективне (в обчислювальній техніці) моделювання та дослідження різноманітних моделей та алгоритмів, які зберігають геометричність та інваріантність моделей, які необхідні для їх конкретного використання в умовах прийнятності використання різних вихідних структур даних. Дане дослідження присвячене розв'язанню задачі розробки моделі для задач зв'язку в рамках геометричного проектування.

Ключові слова: математична модель, задача оптимізації, обмеження, топологічний параметр, норма і правило побудови, гомотопія, точність.

Постановка проблеми та її актуальність. Об'єктом дослідження є математичні моделі задач пошуку оптимальних мереж і маршрутів, що виникають при автоматизації проектування та управління з урахуванням ландшафту при обмеженнях на форму, взаємне розташування та інші параметри зв'язку. Предметом дослідження є розробка математичної моделі для вирішення задачі пошуку оптимальних маршрутів і сполучних мереж у неоднозв'язних областях при обмеженнях на кривизну, кількість вигинів та інших геометричних топологічних параметрів зв'язку, що відображають типові технологічні вимоги. У цій статті використовуються такі методи дослідження: методи оптимізації, обчислювальні підходи, обчислювальна геометрія, математичне моделювання та підходи комбінаторної топології. Що стосується топологічних параметрів, то в даній роботі представлено їх декомпозицію як систему основних і стандартних задач; поставлено основні задачі оптимізації.

Аналіз останніх досліджень і публікацій. Найбільш ефективний підхід до вирішення аналогічних задач [1–5], які є NP-повними навіть для випадку вигинів у опуклій області, розроблено в роботах Ю. Г. Стояна та С. В. Смелякова [6–9]. Це синтез

комбінаторних і варіаційних методів оптимізації в рамках ієрархічної системи моделей. Низькорівневі моделі орієнтовані на вирішення базових задач пошуку оптимального маршруту за допомогою варіаційних методів, які були розроблені М. М. Моїсєєвим, С. В. Смеляковим і Н. З. Шором. Високорівневі моделі орієнтовані на дослідження гомотопних сімей магістралей і мереж на основі методів дискретної оптимізації, які були розроблені в роботах В. І. Михалевича, І. В. Сергієнка, Ю. Г. Стояна, С. В. Яковлева та ін. [10–12]. Неможливість повної формалізації вимог до будівельних норм і правил (БНП) (як наслідок їх неузгодженості та не повної строгості формул) та підпорядкований характер проблем зв'язку по відношенню до проблем нормалізації областей зумовлює необхідність їх вирішення у двох режимах – оптимізації та імітації, коли особа, яка приймає рішення, залучається до процесу інтерактивного вирішення проблем зв'язку.

Мета дослідження. Метою даної роботи є розробка математичної моделі для вирішення оптимізаційних задач з'єднання в неоднозв'язних областях при типових технологічних обмеженнях на геометричні та топологічні параметри траси, перш за все, на кривизну та кількість вигинів. Модель повинна

© Левтеров А. І., Плехова Г. А., Костікова М. В., Окунь А. О., 2023



Дослідницька стаття: Цю статтю опубліковано видавництвом *НТУ «ХПИ»* у збірнику «Вісник Національного технічного університету «ХПИ» Серія: Системний аналіз, управління та інформаційні технології». Ця стаття поширюється за міжнародною ліцензією [Creative Commons Attribution \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/). **Конфлікт інтересів:** Автор/и заявив/или про відсутність конфлікту.



відповідати існуючим і перспективним топогеодезичним моделям полігонального зображення території; методи оптимізації повинні гарантувати ефективне вирішення основних класів прикладних задач і допускати природну інтеграцію в існуючі та перспективні системи проектування та управління.

Для досягнення мети необхідно вирішити такі завдання:

- відсортувати основні критерії та обмеження, які пов'язані з геометричними та топологічними параметрами траси, які будуть розглядатися при проектуванні з'єднань, і на цій основі поставити основну задачу оптимізації;

- розробити загальну модель проблем зв'язку, що забезпечує постановку основних типів задач оптимізації та моделювання маршрутів з використанням функціональних класів ліній та дозволяє нарізати та інтегрувати в існуючі системи.

Побудова та опис моделі. Основні стандарти побудови програмно-визначених мереж. Відповідно до вимог, що висуваються до точності одержуваних розв'язків та обчислювальної ефективності методів їх побудови, а також до існуючих світових тенденцій використання полігональних моделей рельєфу в системах топогеодезичного забезпечення, в моделях областей і методах оптимізації для вирішення задач зв'язку слід використовувати полігональні моделі неоднорозв'язної області, яка вказана для з'єднань маршрутів, а не сіткові моделі, які використовуються в сучасних системах, таких як ReCAD і CREDO. Отже, область розраховується наступним чином:

$$F = Cl \left[\frac{F_0}{\left(\bigcup_{i=1}^{n_F} F_i \right)} \right], \quad (1)$$

де Cl – операція закриття; F_0 – загальна область, яка використовується для вивчення проблем моделювання; F_i , ($i = 1, 2, \dots, n_F$) – однорозв'язні області, які взаємно перетинаються («заборонені зони» для маршрутизації) в однорозв'язній області F_0 на площині.

Загальною математичною особливістю парних маршрутів (автомагістралей, залізничних і трамвайних колій) є вимога безперервності та обмеженості кривизни. Вимога повинна бути виконана для забезпечення безпеки руху транспорту та відсутності ударів. У зв'язку з цим у будівельних нормах і стандартах зазначено, що осьові лінії автомагістралей будуть представлені сплайнами

$$p = S_1 \eta_1 C_1 R_1 S_2 r_2 C_2 R_2 \dots S_m r_m C_m R_m S_{m+1}, \quad (2)$$

де S_i – катети; C_i – дуги кіл; r_i, R_i – сполучні криві, які можуть бути представлені фрагментами кривої та

кубічної параболи; m – кількість фрагментів відповідних кривих.

Аналіз будівельних норм і правил показує, що більшість досліджуваних технологічних обмежень для транспортних та інженерних мереж можна звести до наступних основних класів геометричних і топологічних обмежень: Q_1 – обмеження на максимальну довжину прямих катетів; Q_2 – обмеження на кут повороту α , $\alpha \in (-90^\circ, 90^\circ)$, у вершинах; Q_3 – обмеження на функціональний клас гладких кривих: \tilde{W}, SC, SKC, SPC ; Q_4 – умови на торці (для визначення дотичних та їх довжини в точках A і B); Q_5 – примикання з допустимими перетинами; Q_6 – вузли для узгодження потоків за напрямком; Q_7 – розв'язки для забезпечення їх досяжності; Q_8 – примикання до меж та маршрутів; Q_9 – топологічна структура шуканої мережі (зв'язність, цикли).

Відповідно до національних будівельних норм, вибір траси трубопроводу, магістралі тощо повинен здійснюватися за допомогою математичних методів на основі одного або кількох оптимальних критеріїв (наприклад, криві повинні проектуватися з такими великими радіусами, як можливо). При цьому критерії, як правило, виражаються геометричними параметрами маршрутів і топологічними параметрами мереж, причому деякі з них можуть бути неадитивними. Серед основних геометричних параметрів траси p необхідно вказати довжину $l(p)$, кількість $m(p)$ вставок кіл та їх радіуси кривизни $\{p_i\}_{i=1, m(p)}$ (або, для стрічкової лінії, кількість вигинів $n(p)$ і кути повороту $\{\phi_i\}_{i=1, m(p)}$), розташування маршруту p в області F .

Вони впливають на вартість будівництва $c(p)$, витрати на технічне обслуговування $e(p)$, технологічність $t(p)$ і надійність $b(p)$, що є одним із найважливіших і продукт можливих аналогів надійності

$$\begin{cases} b_m(p) = \prod_{i=1}^m \left(1 - \frac{1}{p_i}\right), & (p_i \geq 1) \text{ for } m > 0; \\ b_d(p) = \prod_{j=1}^d (1 - g_j)^{k_j}; \\ b_l(p) = \frac{1}{l(p)}, & (l(p) \geq 1). \end{cases} \quad (3)$$

Підсумовуючи, загальну задачу оптимізації з'єднань GPOC (General problem to optimize connections) з урахуванням введених критеріїв і обмежень можна поставити наступним чином.

Загальна проблема оптимізації з'єднань. Дана область F має як набір точок $\{A_i\}_{i=1, N}$, так і кілька мереж $\{S_j\}_{j=1, M}$. Необхідно з'єднати ці точки та мережі сполучною мережею s^* , яка складається з ліній заданого функціонального класу $S \in C$, щоб відповідати заданим обмеженням Q типу Q_1 – Q_9 , і вона буде найбільш ефективною за значенням заданого принципу оптимуму $R(s)$.

Враховуючи вищезазначену ефективність декомпозиції GPOC, яка базується на зведенні цієї задачі до системи базових задач на континуальних сімействах маршрутів, що дозволяють отримати точний розв'язок, можна мати основну типову задачу оптимізації MOP (Main optimization problem) для пошуку оптимального шляху в класі еквівалентності маршруту $[\tau]$ на заданому функціональному класі ліній $P(A, B)$ із заданими кінцями A і B при відповідному зниженні обмежень Q^* :

Основна проблема оптимізації MOP. Обчислити

$$\arg \text{extr}_{p \in P_{IQ}(A, B)} R(p). \quad (4)$$

Ця стаття також висуває інформаційні та обчислювальні вимоги до моделей для розрахунку зазначених проблем.

Розглянуто задачу глобальної та локальної декомпозиції та регуляризації GPOC на основі її зведення до системи однорідних базових задач типу (4). Це гарантує конструктивну регуляризацію та ефективний обчислювальний підхід для вирішення загальної проблеми для оптимізації з'єднань, використовуючи дискретні та безперервні моделі та методи. Для цього використовується дворівнева FL модель Ю. Стояна та С. Смелякова. На верхньому топологічному рівні структура цієї моделі визначається алгебраїчним розшаруванням ліній на класи еквівалентності маршрутів і мереж, а на геометричному рівні – розглядом необхідного функціонального класу ліній у кожному класі. Обмеження можуть бути накладені на компоненти на обох рівнях. Оскільки многовид F у формі (1) має гомотопічний характер типу диска з $n_F \geq 0$ отворами, структура гомотопічної моделі ліній складається з вільної групи $C_{(n_F)}$, яка дискретизує неперервні сімейства маршрутів.

Мережа $\Gamma = \{\gamma_i\}_{i=1, m}$ в області F являє собою набір випрямляних кривих, кількість яких обмежена. Вони не перетинаються між собою і не мають спільних точок, крім кінців. Огляд зв'язних комбінацій типів мереж, графів тощо приводить до задач трьох класів: пошуку оптимальних мереж на топологічному рівні без урахування геометрії області F ; оптимізація внесків абстрактної моделі в області F і структура оптимальної мережі в області F . Щоб вирішити ці проблеми, ми можемо ввести концепцію абстрактної мережі $s^* = (V^*, R^*)$ як абстрактного симплиціального комплексу з кінцевим зв'язком $K = (V^*, R^*)$. Реалізація абстрактної мережі геометричним комплексом $s = (V, R)$ є геометричною мережею. Мережі $s_1 = (V_1, R_1)$ і $s_2 = (V_2, R_2)$ є гомотопними в F , якщо їх абстрактні аналоги мають ізоморфізм $\omega: s_1 \rightarrow s_2$, при якому відповідність 0-симплексів означає комбінацію вершин у F , а відповідність 1-симплексів означає їх приналежність до одного класу еквівалентності маршрутів у F ; мережі s_1 і s_2 є вільними

гомотопними, якщо вони гомотопні з точністю до зсуву 0-симплексів у F і деформаційно еквівалентними, якщо вони ізоморфні та гомотопні, а їх різні ребра можуть мати спільні точки лише в основних вершинах, які вони з'єднують. Слід зазначити, що мережевий ізоморфізм зберігає алгебраїчні інваріанти для базових вершин, але не враховує структуру області F з обмеженнями в ній. Гомотопія зберігає гомотопію мереж, але не сприяє їх ізоморфізму через появу додаткових вершин. Деформація мережі зберігає гомотопічні інваріанти.

Геометрична модель маршрутів – це функціональні класи ліній $\Lambda = \{S, SC, SKC, SPC, W, \tilde{W}\}$ в області F , які мають сплайн-образ (2). До них також можна застосовувати топологічні обмеження. Введення моделей мереж і ліній дозволяє звести GPOC до системи базових задач типу (2) на різні функціональні класи ліній і типових задач на ті ж класи, що відображають поєднання обмежень, характерних для прикладних задач і зводяться до базових задач.

Висновки. Розроблено ієрархічну математичну модель загальної задачі зв'язку. Суть її полягає в пошуку оптимальних зв'язків (маршрутів і з'єднувальних мереж) в неоднотипних регіонах. Проблема виникає при проектуванні транспортних та інженерних мереж і регулюванні руху автотранспорту на пересіченій місцевості. У рамках моделі ця задача зводиться до системи базових і типових задач для пошуку оптимальних маршрутів. Використання цієї ієрархічної структури моделей у системах прийняття рішень дозволяє вирішити задачу адекватного моделювання зв'язків у неоднотипних областях з точки зору точності, обчислювальної ефективності та відсутності інформаційної надлишковості для всіх функціональних класів вигнутих та рівних ліній $\{S, SC, SKC, SPC\}$, які регулюються національними будівельними нормами, з обмеженнями на кривизну та інші геометричні та топологічні параметри трас.

Список використаної літератури

1. Upadhyay S., Ratnoo A. On existence and synthesis of smooth four parameter logistic paths inside annular passages. *IEEE Robotics and Automation Letters*. 2018. Vol. 3, Iss. 4, P. 4375–4382.
2. Cowlagi R., Tsiotras P. Curvature-bounded traversability analysis in motion planning for mobile robots. *IEEE Transactions on Robotics*. 2014. P. 1011–1019.
3. Bakolas E., Tsiotras P. Kinodynamic trajectory generation through rectangular channels using path and motion primitives. *47th IEEE Conference on Decision and Control*. 2008. P. 3725–3730.
4. Laumond J.-P., Jacobs P., Taix M., Murray R. M. A motion planner for nonholonomic mobile robots. *IEEE Transactions on Robotics and Automation*. 1994. Vol. 10, Iss. 5, P. 577–593.
5. Agarwal P., Biedl T., Lazard S., Robbins S., Suri S., Whitesides S. Curvature-constrained shortest paths in a convex polygon. *Proceedings 14th Annual ACM Symposium on Computational Geometry*. 1998. P. 392–401.
6. Плехова А. А., Смеляков С. В. Моделирование коммуникационных сетей с учетом ландшафта при разнородных критериях и ограничениях. *Информационные системы*. 1998. № 3(11). С. 143–146.
7. Stoyan, Y. G., Eschenko V. G., Vinarsky V. Y. Mathematical Methods of Geometric Design in Artificial Intelligence System. *IFAC Proceedings Volumes*. 1992. Vol. 25, Iss. 28, P. 101–105.
8. Smelyakov S. V., Stoyan Y. G. Modelling of the space of paths in problems of constructing optimal trajectories. *USSR Computational*

- Mathematics and Mathematical Physics*. 1993. Vol. 23, Iss. 1. P. 50–55.
- Smelyakov S. V. Construction of shortest line of restricted curvature in a non-singly-connected polygonal area. *Inclusion methods for nonlinear problems. Computing supplementa*. 2003. Vol. 16. P. 237–244.
 - Sergienko I. V., Shylo V. P. Problems of discrete optimization: Challenges and main approaches to solve them. *Cybernetics and Systems Analysis*. 2006. T. 42. P. 465–482.
 - Плехова А. Модель и метод решения задачи поиска оптимального соединения при ограничении кривизны. *Радіоелектроніка та інформатика*. 1998. № 3. С. 56–59.
 - Плехова А. Построение оптимальной трассы ограниченной кривизны в неодносвязной области. *Радіоелектроніка та інформатика*. 1999. № 3. С. 22–23.
 - Plekhoa A.A., Smelyakov S.V., Modeliuvannia komunikatsiinykh merezh z urakhuvanniam landshaftu za riznorodnykh umov ta obmezhen [Modelling of communication networks taking into account the landscape under heterogeneous criteria and restrictions], *Informatsionnye sistemy [Information Systems]*, vol. 3, no. 11, 1998, pp. 143-146.
 - Stoyan, Y.G., Eschenko V.G., Vinarsky V.Y. Mathematical methods of geometric design in artificial intelligence system. *IFAC Proceedings Volumes*. 1992, vol. 25, iss. 28, pp. 101–105.
 - Smelyakov S. V., Stoyan Y. G. Modelling of the space of paths in problems of constructing optimal trajectories. *USSR Computational Mathematics and Mathematical Physics*. 1983, vol. 23, iss. 1, pp. 50–55.
 - Smelyakov S. V. Construction of shortest line of restricted curvature in a non-singly-connected polygonal area. *Inclusion methods for nonlinear problems. Computing supplementa*. 2003, vol. 16, pp. 237–244.
 - Sergienko I. V., Shylo V. P. Problems of discrete optimization: Challenges and main approaches to solve them. *Cybernetics and Systems Analysis*. 2006, vol. 42, pp. 465–482.
 - Plekhoa A. Model i metod rozv'iazannia zadachi poshuku optymalnoho poiednannia pry obmezheni kryvlyzny [Model and methods for solving problems of searching for certain connections with a curvature constraint]. *Radioelektronika i informatika [Radioelectronics and Informatics]*. 1998, vol. 3, 1998, pp. 56–59.
 - Plekhoa A. Pobudova optymalnoi trasy obmezhenoi kryvlyzny u neodnozviaznii oblasti [Construction of an optimal path of bounded curvature in a non-simply connected domain]. *Radioelektronika i informatika [Radioelectronics and Informatics]*. 1999, vol. 3, pp. 22–23.

References (transliterated)

- Upadhyay S., Ratnoo A. On existence and synthesis of smooth four parameter logistic paths inside annular passages. *IEEE Robotics and Automation Letters*. 2018, vol. 3, iss. 4, pp. 4375–4382.
- Cowlagi R., Tsiotras P. Curvature-bounded traversability analysis in motion planning for mobile robots. *IEEE Transactions on Robotics*. 2014, pp. 1011–1019.
- Bakolas E., Tsiotras P. Kinodynamic trajectory generation through rectangular channels using path and motion primitives. *47th IEEE Conference on Decision and Control*. 2008, pp. 3725–3730.
- Laumond J.-P., Jacobs P., Taix M., Murray R. M. A motion planner for nonholonomic mobile robots. *IEEE Transactions on Robotics and Automation*. 1994, vol. 10, iss. 5, pp. 577–593.
- Agarwal P., Biedl T., Lazard S., Robbins S., Suri S., Whitesides S. Curvature-constrained shortest paths in a convex polygon. *Proceedings 14th Annual ACM Symposium on Computational Geometry*. 1998, pp. 392–401.

Надійшло (received) 15.05.2023

UDC 624

A. I. LEVTEROV, Candidate of Technical Sciences, Professor, Kharkiv National Automobile and Highway University, Head of the Department of Informatics and Applied Mathematics, Kharkiv, Ukraine, e-mail: lai@khadi.kharkov.ua, ORCID: <https://orcid.org/0000-0001-6586-1061>

H. A. PLIEKHOVA, Candidate of Technical Sciences, Docent, Kharkiv National Automobile and Highway University, Associate Professor at the Department of Informatics and Applied Mathematics, Kharkiv, Ukraine, e mail: plehovaanna1@gmail.com, ORCID: <https://orcid.org/0000-0002-6912-6520>

M. V. KOSTIKOVA, Candidate of Technical Sciences, Docent, Kharkiv National Automobile and Highway University, Associate Professor at the Department of Informatics and Applied Mathematics, Kharkiv, Ukraine, e mail: kmv_topaz@ukr.net, ORCID: <https://orcid.org/0000-0001-5197-7389>

A. O. OKUN, Candidate of Technical Sciences, Docent, National Technical University "Kharkiv Polytechnic Institute", Associate Professor at the Department of Computer Modeling and Integrated Forming Technologies, Kharkiv, Ukraine, e mail: okunanton@gmail.com, ORCID: <https://orcid.org/0000-0002-6467-4229>

GEOMETRIC MODELING: TRACKS AND FLOWS

Mathematical models to solve optimization connection problems in nonsimply connected regions under typical technological restrictions on geometric and topological parameters of routes, first of all, on curvature and the number of bends, have been investigated and developed. The models are linked with the extant and prospective topogeodesic models of the territory polygonal images. The solution of connection problems involves search for optimum trajectories of routes and nets within unrestricted geometric shape areas. It needs the development of a plethora of general models as fields where connections are carried out. The connections can be of various types such as bendy, Manhattan, even, solid as well as routes of other types. Smeliakov and Pliekhoa observe that the global and local regulation of geometric connections to solve connection problems can be presented as the general optimization connection problem that is defined as the problem of the choice of c , where Ω is a set of alternatives, R is a principle of optimality. In so doing, the set Ω can be presented as the totality of the phase space ϕ and the restrictions Q that are applied to the parameters of the phase space ϕ . In turn, it is expedient to imagine that the phase space ϕ is the Cartesian product $\phi = X*Y*Z*U$ of the output data X , disturbances Y , control parameters U and results Z . The analysis of problem indicates that first and foremost the effectiveness of the modelling of the phase space ϕ is linked with the description of the output data X on the area F and space L of possible highways in F . This research is devoted to the solution of the problem to develop a model for connection tasks within the framework of geometric design.

Keywords: mathematical model, optimization problem, restriction, topological parameter, construction norm and rule, homotopy, accuracy.

Повні імена авторів / Author's full names

Автор 1 / Author 1: Левтеров Андрій Іванович, Levterov Andrii Ivanovych

Автор 2 / Author 2: Плехова Ганна Анатоліївна, Pliekhoa Hanna Anatoliivna

Автор 3 / Author 3: Костікова Марина Володимирівна, Kostikova Maryna Volodymyrivna

Автор 4 / Author 4: Окунь Антон Олександрович, Okun Anton Oleksandrovych

DOI: 10.20998/2079-0023.2023.01.10
УДК 519.2

О. М. НИКУЛІНА, д-р техн. наук, професор, завідувачка кафедри інформаційних систем та технологій Національного технічного університету «Харківський політехнічний інститут», Харків, Україна; e mail: elniknik02@gmail.com; ORCID: <https://orcid.org/0000-0003-2938-4215>

В. П. СЕВЕРИН, д-р техн. наук, професор, професор кафедри системного аналізу та інформаційно-аналітичних технологій Національного технічного університету «Харківський політехнічний інститут», Харків, Україна; e mail: valerii.severyn@khp.edu.ua; ORCID: <https://orcid.org/0000-0002-2969-6780>

В. О. ШАРОВ, аспірант кафедри інформаційних систем та технологій Національного технічного університету «Харківський політехнічний інститут», Харків, Україна; e mail: wycrpiy@gmail.com; ORCID: 0000-0003-3152-0650

МОДЕЛЮВАННЯ ТА АНАЛІЗ КОДЕРІВ ЗАВАДОСТІЙКИХ КАСКАДНИХ КОДІВ ДЛЯ ДИНАМІЧНИХ СИСТЕМ

Для різних каналів передачі інформації, що можуть використовуватись у локальних системах управління широкого спектру процесів, необхідні доступні, здатні до масштабування й не витратні методи передачі, що дозволяють передавати необхідну інформацію достовірно без помилок. Помилки, що трапляються у дискретних безперервних каналах зв'язку – це основна перепона для достовірної передачі даних. Основні причини виникнення помилок – загасання сигналу, шум та різні перешкоди. Як результат аналізу статистики помилок у дискретних каналах передачі даних, було зроблено висновок, що однократні та двократні помилки трапляються у абсолютній більшості випадків. Запропоновано боротись із помилками шляхом використання завадостійкого кодування. Серед завадостійких кодів виділені надлишкові блокові роздільні систематичні коди, до яких відноситься код Хеммінга. В основу моделі завадостійкого кодування запропоновано покласти використання систематичного коду Хеммінга з подальшим каскадним кодуванням шляхом додавання біту перевірки на парність. Використана модель надала можливість боротись з усіма найбільш ймовірними випадками виникнення помилок. Оскільки дана модель кодування повинна гнучко використовуватись в різних системах, необхідним критерієм моделі є її універсальність. Тому виконане тестування моделі на вибірці з кодовими комбінаціями різної кількості бітів. Метою даної статті є аналіз залежності часу виконання процедури кодування для кодерів різного рівня: окремого кодера першого ступеню, а також кодера першого ступеню з додаванням кодера другого ступеню. Проведено експерименти з великою вибіркою, які були у подальшому проаналізовані та інтерпретовані. Оскільки модель у результаті експериментів виявила себе гнучкою, простою, стабільною та ефективною, а також процес боротьби із завадами шляхом використання каскадних кодів добре себе зарекомендував у світовій практиці, модель рекомендовано для використання у різних інформаційних управляючих системах. Наведена реалізація кодера для завадостійкої передачі даних.

Ключові слова: завадостійкість, біт, код, кодова комбінація, каскадні коди, кодування, кодер, перевірка на парність.

Вступ. У сучасних комунікаційних технологіях широко розповсюджені стандартні принципи передачі даних. Спочатку на одному пристрої інформація генерується, потім вона кодується, передається по каналах, декодується і використовується споживачем вже на іншому пристрої.

Завадостійка передача даних дуже важлива для інформаційної технології оптимізації управління динамічними системами [1]. Це особливо необхідно для надійного функціонування інформаційних управляючих систем в ядерній енергетиці [2]. Моделі завадостійких кодів повинні бути враховані при вдосконаленні інформаційної технології оптимізації управління складними динамічними системами [3].

У сучасній теорії передачі даних існує низка способів для покращення якості даних, збільшення достовірності та зменшення експлуатаційних витрат [4]. Тому ідея створення нових моделей кодування, а також покращення вже існуючих моделей є концептуальною однією з найважливіших.

У теперішній час ведуться активні розробки по створенню нових засобів забезпечення завадостійкості у каналах передачі даних з широким впровадженням та практичним використанням [4]. Ще з часів Клоду Шеннону та його праці «Математична теорія зв'язку» 1948 року концепція завадостійкого каналу передачі даних була однією з найбільш цікавою для досліджень, що дозволяли економити багато ресурсів та знижувати

собівартість інформаційних систем [1]. Оскільки з тих часів інформаційних систем і технологій стало набагато більше, то і дослідження за даним напрямом стали складнішими. Сучасні дослідники передачі даних, серед яких такі вчені як Фаді Аладжі, Тіме Ніколас, Альфорд Веслі, Бенджамін Флекер, Джон М. Бегс та інші, генерували нові ідеї в теорії інформації, пропонували нові підходи та вдосконалили попередні технології [5, 6]. На основі сучасної теорії та практики кодування актуальною є задача вдосконалення існуючих моделей і алгоритмів кодування, їх дослідження в сучасних інформаційних системах.

Мета та задачі дослідження. Мета даної статті полягає в розробці моделі завадостійкої передачі даних для інформаційної технології оптимізації управління динамічними системами шляхом аналізу можливостей кодерів завадостійких каскадних кодів.

Для цього поставлені задачі:

- 1) обрати завадостійкий код для моделі;
- 2) реалізувати модель на мові програмування;
- 3) провести експерименти з кодування;
- 4) проаналізувати результати експериментів.

Завадостійкі коди. Під завадостійкими кодами розуміють коди, що дозволяють виявляти або виправляти та виправляти помилки, які виникають у результаті впливу завад. Здатність коду до виявлення та виправлення помилок базується на введенні надмірності в кодову інформаційну комбінацію. Надмірні символи

© Нікуліна О. М., Северин В. П., Шаров В. О., 2023



Дослідницька стаття: Цю статтю опубліковано видавництвом *НТУ «ХПИ»* у збірнику «Вісник Національного технічного університету «ХПИ» Серія: Системний аналіз, управління та інформаційні технології». Ця стаття поширюється за міжнародною ліцензією [Creative Commons Attribution \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/). **Конфлікт інтересів:** Автор/и заявив/или про відсутність конфлікту.



формується за певними правилами і називаються перевірочними або контрольними. Збільшення числа таких символів в кодовій комбінації підвищує виявляючу і виправляючу здатності коду, але призводить до зниження швидкості передачі інформації.

Надлишковість коду розділяється на абсолютну та відносну. Абсолютна надлишковість визначається кількістю додаткових розрядів, які вводяться,

$$k = n - m, \quad (1)$$

де k – абсолютна надлишковість або кількість перевірочних елементів;

n – загальна кількість елементів у кодовій комбінації, довжина кодової комбінації;

m – загальна кількість інформаційних елементів у кодовій комбінації.

Властивості коригування надлишкових кодів є залежними від кодування або побудови самих кодів та їх параметрів: числа розрядів, надмірності, тривалості символів й інших. Розглянуті загальні принципи використання надлишковості.

Нехай на вхід кодуємого пристрою надійшла деяка послідовність m інформаційних двійковий розрядів. На виході їй буде відповідати послідовність з n двійкових символів, де $n > m$. Усього може бути 2^m різних послідовностей з 2^n , які є дозволеними кодівими комбінаціями. Інші $2^n - 2^m$ з можливих вихідних послідовностей для передачі не використовуються, тому вони називаються забороненими [5, 6, 7].

Існує велика кількість завадостійких кодів. На рис. 1 наведена часткова ієрархія видів завадостійких кодів [5].



Рис. 1. Види завадостійких кодів

Ураховуючи та аналізуючи усі переваги та недоліки різних видів кодів, розглянуті завадостійкі коди Хеммінґа, які є надлишковими блоковими роздільними систематичними кодами.

Коди Хеммінґа. Коди Хеммінґа надають можливість гарантовано виправляти одноразові помилки. Досліджується використання кодів Хеммінґа з додаванням до них біту перевірки на парність, що дозволяє

виправляти одноразові помилки та виявляти дворазові помилки.

Коди Хеммінґа формуються стандартним для систематичних кодів матричним способом. Перевірочна матриця $H_{(n-m) \times n}$ формується з одиничної матриці O_{n-m} та транспонованої перевірочної підматриці $P_{m \times (n-m)}^T$ і має вигляд

$$H_{(n-m) \times n} = (O_{n-m} | P_{m \times (n-m)}^T). \quad (2)$$

Коди Хеммінґа є одними з найпростіших представників блокових систематичних кодів. Принцип їх формування аналогічний загальному принципу побудови систематичних кодів. Основною ідеєю кодів Хеммінґа є чітке використання певних розрядів у комбінації, задля досягнення можливості виправляти однократну помилку [7].

Довжина коду n вибирається з умови

$$2^m \leq \frac{2^n}{1+n}, \quad (3)$$

що означає виконання умови

$$2^k < m + k + 1. \quad (4)$$

Код будується таким чином, щоб в результаті $p = n - k$ перевірок отримати p -розрядне двійкове число, що вказує номер спотвореної позиції в кодовій комбінації. Для цього перевірочні символи повинні перебувати на тих номерах позицій, які відповідають ступеням двійки $2^0, 2^1, 2^2, 2^3, \dots, 2^{k-1}$. Кожен з перевірочних символів входить тільки в одне з перевірочних рівнянь. Таким чином, якщо нумерувати позиції зліва направо, то контрольні символи повинні знаходитися на першій, другій, четвертій і т. д. позиціях.

Результат першої перевірки дає цифру молодшого розряду синдрому, тобто номеру спотвореного біту, в двійковому записі. Якщо в результаті цієї перевірки отримано значення 1, то один із символів перевіреної групи спотворений. Таким чином, першою перевіркою повинні бути охоплені символи з номерами, що містять в двійковій формі запису одиниці в першому розряді: 1, 3, 5, 7, 9 і т. д.

Результат другої перевірки дає цифру другого розряду синдрому. Отже, другою перевіркою повинні бути охоплені символи з номерами, що містять в двійковій формі запису одиниці в другому розряді: 2, 3, 6, 7, 10 і т. д.

Аналогічно, при третій перевірці повинні перевірятися символи, номери яких в двійковій формі запису містять одиниці в третьому розряді: 4, 5, 6, 7, 12 і так далі [6].

Принцип розташування розрядів для кодової комбінації Хеммінґа 7/4 наведено в табл. 1 [8].

У табл. 1 використані позначення:

№ – номер розряду у комбінації;

F – вказує на місця розрядів за призначенням;

I_i – інформаційні розряди;

K_i – перевірочні розряди;

S_i – розряди, за які відповідає при перевірці i -й перевірочний розряд.

X – розряди, які підсумовуються за модулем 2.

Таблиця 1 – Розташування розрядів коду Хеммінга 7/4

№	1	2	3	4	5	6	7
F	K_1	K_2	I_1	K_3	I_2	I_3	I_4
S_3				X	X	X	X
S_2		X	X			X	X
S_1	X		X		X		X

Таким чином, перевірочні групи повинні мати вигляд:

$$S_4 = k_8 \oplus k_9 \oplus k_{10} \oplus k_{11} \oplus k_{12} \oplus \dots$$

$$S_3 = k_4 \oplus k_5 \oplus k_6 \oplus k_7 \oplus k_{12} \oplus \dots$$

$$S_2 = k_2 \oplus k_3 \oplus k_6 \oplus k_7 \oplus k_{10} \oplus \dots$$

$$S_1 = k_1 \oplus k_3 \oplus k_5 \oplus k_7 \oplus k_9 \oplus \dots$$

де \oplus – підсумовування за модулем 2.

При додаванні до коду Хеммінга біту перевірки на парність проходить перевірка на парність усієї кодової послідовності, тобто формується каскадний код [9].

Каскадні коди. Каскадний код, або турбо-код – паралельний, блоковий, систематичний код, який здатен виправляти помилки, що виникають при передачі цифрової інформації каналом зв'язку з шумами. В основному, каскадні коди складаються з двох або більше, менших за розміром і простіших кодів. Основна ідея каскадного коду полягає у тому, що повідомлення одноразово кодується у кодері 1, потім це закодоване повідомлення кодується поверх кодером 2 і так далі. При декодуванні процес декодування виконується у зворотному порядку.

Можливість провести процес декодування повторно робить каскадні коди дуже зручними. Однією з ідей також є те, що другий декодер зберігає у собі адреси помилок, які zostалися від першого декодера, а при зверненні до першого декодера, у ньому зберігаються адреси помилок другого, тобто ці декодери пов'язані й збірні. Загальна схема каскадних кодів зображена на рис. 2 [8].

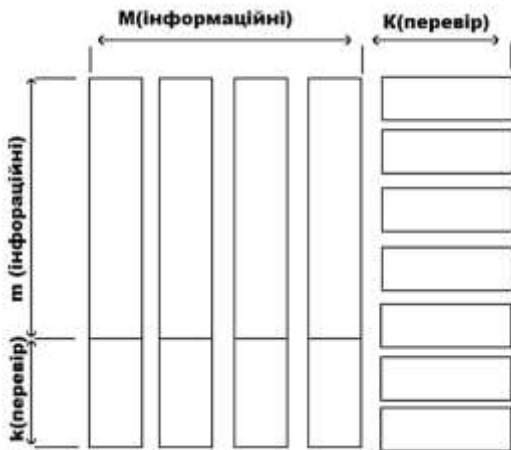


Рис. 2. Загальна схема каскадного коду

Загальна схема кодування та декодування каскадним кодом з додаванням біту перевірки на парність наведена на рис. 3 [9, 10].

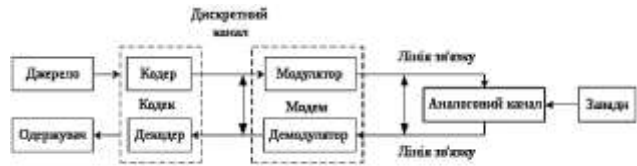


Рис. 3. Загальна схема кодування та декодування

У даній статті розглядається перший етап передачі інформації, а саме кодування кодом першого ступеню, після чого виконується кодування кодом другого ступеню. Загальна схема кодування каскадним кодом показана на рис. 4 [11, 12].



Рис. 4. Загальна схема кодування каскадним кодом

Принцип розташування розрядів для кодової комбінації Хеммінга 8/4 надано в табл. 2.

Таблиця 2 – Розташування розрядів коду Хеммінга 8/4

№	1	2	3	4	5	6	7	8
F	K_1	K_2	I_1	K_3	I_2	I_3	I_4	K_n
S_4	X	X	X	X	X	X	X	X
S_3				X	X	X	X	
S_2		X	X			X	X	
S_1	X		X		X		X	

У табл. 2 використано позначення K_n – місце розташування біту перевірки на парність.

Для коду Хеммінга додавання біту перевірки на парність означає, що після кодування кодом першого ступеню проходить підсумовування за модулем 2 для усіх елементів комбінації [13, 14]. Потім результат такого підсумовування записується у біт перевірки на парність, «закриваючи» байт.

Для комбінацій з більшою довжиною повідомлення принцип такий самий, як і в табл. 2. Оскільки при кодуванні кодом Хеммінга при будь-якій довжині інформаційного повідомлення завжди залишається один незайнятий біт, тому його доцільно використати для перевірки на парність [15].

Аналіз моделей кодерів коду Хеммінга. Кодування інформаційної комбінації кодом Хеммінга без додавання біту перевірки на парність займає менше часу, ніж кодування кодом Хеммінга з додаванням біту перевірки на парність.

Дослідження цієї різниці виконано з використанням різного програмного забезпечення. У даній статті використовується мова програмування Java (language version 8 або вище) без використання сторонніх фреймворків.

При дослідженні часу виконання програми кодера виявлені наступні закономірності. На рис. 4. наведено залежність часу кодування комбінацій кодом Хеммінга без перевірки на парність від кількості символів

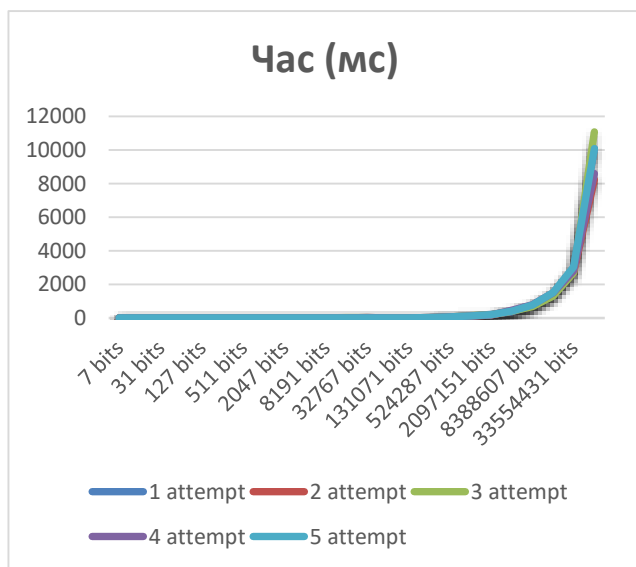


Рис. 4. Час кодування без перевірки на парність

Для полегшення роботи з вибіркою даних за основу початкової комбінації обирались біти, близькі до повних байтів, кількість яких подвоювалась кожного разу.

Необхідно відмітити, що до 8 388 607 бітів, що являється еквівалентом 2 097 152 байт, швидкість кодування кодером першого ступеню не перевищувала 3 секунд, для меншої кількості бітів час кодування був незначним, а для комбінацій до 8192 бітів швидкість була майже миттєвою.

На рис. 5 наведено час кодування комбінацій кодом Хеммінга з перевіркою на парність комбінацій залежно від кількості символів.

Схожі результати демонструє кодування кодом першого ступеню з додаванням до нього коду другого ступеню, а саме додавання біту перевірки на парність.

Додавання такого біту дозволяє повністю вмістити бітові комбінації у байти без пропусків. Це є ще однією перевагою моделі, бо навіть з додаванням біту перевірки на парність, кінцева кількість байтів не змінюється, що надає переваги над іншими каскадними кодами.

На рис. 6. наведено час кодування комбінацій кодом Хеммінга без перевірки на парність та з перевіркою на парність залежно від кількості символів.

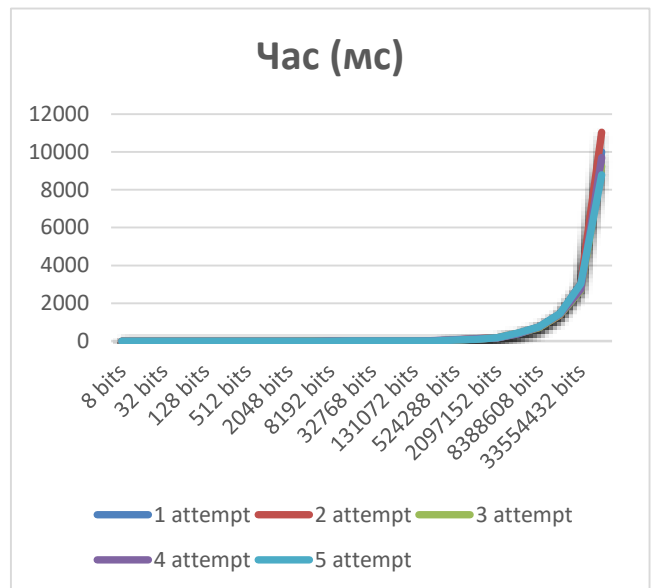


Рис. 5. Час кодування з перевіркою на парність

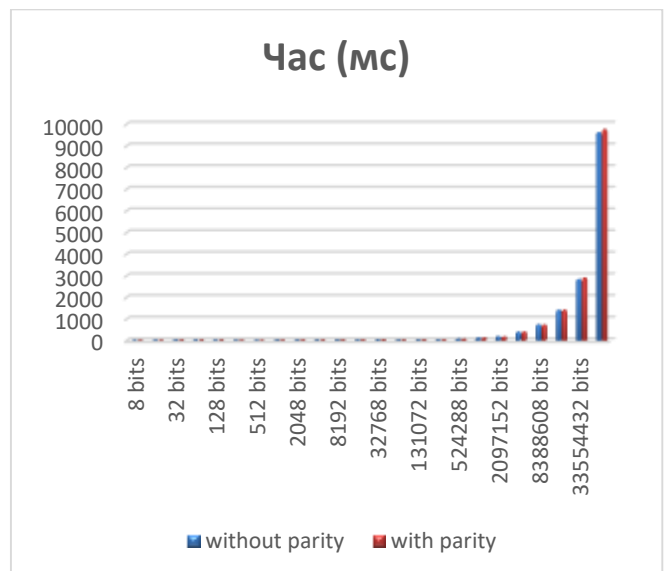


Рис. 6. Час кодування без перевірки та з перевіркою на парність

Аналіз результатів кодування. Аналіз отриманих результатів показує, що час виконання процедури кодування збільшується у геометричній прогресії відносно кількості елементів кодування.

Різне зростання часу обробки також може бути пов'язане з тим, що на кодуєчому пристрої, у якості якого використовувався ноутбук ASUS TUF F15, в якого наявні 8 Гб оперативної пам'яті, закінчилась вільна оперативна пам'ять, тому у повному обсязі про нелінійне зростання часу виконання процедури кодування висновки робити зарано. Для більшої надійності, необхідне тестування на більш потужних кодуєчих пристроях.

Порівнюючи час кодування без додавання біту перевірки на парність та з додаванням біту перевірки

на парність, стає очевидним, що час, який використовується для додавання біту перевірки на парність, не є значним.

Перевага додавання біту перевірки на парність у вигляді виявлення дворазових помилок є значною, бо без суттєвого ускладнення системи і без зміни розміру повідомлень, додавання біту перевірки на парність надає змогу боротись з основною кількістю ймовірних помилкових ситуацій.

Висновки. В результаті аналізу і дослідження наукових праць різних авторів, опираючись на сучасну теорію передачі інформації, у статті проводився аналіз моделі каскадного коду. Проаналізована ефективність кодування звичайного коду Хеммінга та каскадного коду Хеммінга з додаванням до нього біту перевірки на парність. Експериментально визначений час кодування для наведених кодів. У результаті співставлення експериментальних даних з'ясовано, що вплив додавання біту перевірки на парність є незначним. Для комбінацій невеликої довжини додатковий час не перевищує 7% від загального часу при додаванні біту перевірки на парність, що дозволяє економити час при передачі, а тому й збільшувати пропускну здатність. Необхідно відмітити, що додавання біту перевірки на парність не потребує додавання додаткових байтів при передачі у складі пакетів даних. Це не збільшує довжину пакету з даними, а тому не ускладнює протоколи передачі даних мережевого рівня. Тому каскадний код Хеммінга надає можливість працювати з усіма протоколами, які використовуються при передачі даних у всіх світових каналах, без змін.

Список використаної літератури

1. Нікуліна О. М., Северин В. П., Шаров В. О. Розробка моделі завадостійкої передачі даних для інформаційної технології оптимізації управління динамічними системами. *Вісник Національного технічного університету «Харківський політехнічний інститут»: сб. наук. пр. Темат. вип.: Системний аналіз, управління та інформаційні технології.* Харків: НТУ «ХПІ», 2022. № 2 (8). С. 57–62.
2. Нікуліна О. М., Северин В. П., Коцюба Н. В., Бубнов А. І. Моделювання теплових процесів парогенератора АЕС для інформаційної технології оптимізації управління. *Вісник Національного технічного університету «Харківський політехнічний інститут»: сб. наук. пр. Темат. вип.: Системний аналіз, управління та інформаційні технології.* Харків: НТУ «ХПІ», 2021. № 1 (5). С. 56–61.
3. Нікуліна О. М., Северин В. П., Коцюба Н. В. Розробка інформаційної технології оптимізації управління складними динамічними системами. *Вісник Національного технічного університету «Харківський політехнічний інститут»: сб. наук. пр. Темат. вип.: Системний аналіз, управління та інформаційні технології.* Харків: НТУ «ХПІ», 2020. № 2 (4). С. 63–69.
4. Захарченко Н. В., Горохов С. М., Кочетков А. В. *Інформаційні параметри позиційних кодів.* Одеса: ОНАС, 2018. 212 с.
5. Eklund J. E., Arvidsson R. A multiple sampling, single A/D conversion technique for I/Q demodulation. *IEEE Journal of Solid-State Circuits.* 1996, vol. 31, iss. 12, p. 1987–1994.
6. Blahut R. *Theory and Practice of Error Control Codes.* Addison-Wesley, 1983. 500 p.
7. Лосев Ю. І., Шматков С. І. *Основи теорії передачі інформації.* Харків: ХНУ ім. В. Н. Каразіна, 2013. 290 с.
8. Yue Tang, Tian Mao, Bing Jiang, Design and Experiment of Multi-resolution Composite Digital Array Antenna. *Journal of Radars.* 2016. Vol. 5, iss. 3. P. 265–270.
9. Банкет В. Л., Івашенко П. В., Іщенко М. О. *Завадостійке кодування в телекомунікаційних системах.* Одеса: ОНАЗ ім. О. С. Попова, 2011. 100 с.
10. Жураковський Ю. П., Полторак В. П. *Теорія інформації та кодування.* Київ: Вища шк., 2001. 255 с.
11. Кожевников В. Л., Кожевников В. Л. *Теорія інформації та кодування.* Дніпропетровськ: Національний гірничий університет, 2011. 108 с.
12. Jones G. A., Jones J. M. *Information and Coding Theory.* London: Springer, 2000. 210 p.
13. Шаров В. О., Бердніков А. Г. Моделювання коригувального каскадного коду в каналах передачі даних системи управління. *Матеріали міжнародної науково-технічної конференції «Комп'ютерне моделювання у наукоємних технологіях».* Харків: ХНУ ім. В. Н. Каразіна, 2021. С. 298–302.
14. Yurong Guan, Zhihua Hu, Chen Chen, Xiaoming Zhu, An anti-noise transmission algorithm for 5G mobile data based on constellation selection and channel joint mapping. *Alexandria Engineering Journal,* 2021. Vol. 60, iss. P. 3153–3160.

References (transliterated)

1. Nikulina O. M., Severyn V. P., Sharov V. O. Rozrobka modeli zavadostykoji peredachi danyh dlja informatsiynoi tekhnologii optimizatsii upravlinnja dynamichnymy sistemamy [Development of a data transfer model for information technology optimization of dynamic systems control]. *Vestnik Nats. tekhn. un-ta "KhPI": sb. nauch. tr. Temat. vyp.: Sistemnyy analiz, upravlenie i informatsionnye tekhnologii* [Bulletin of the National Technical University "KhPI": a collection of scientific papers. Thematic issue: System analysis, management and information technology]. Kharkiv, NTU "KhPI", Publ., 2022, no.2 (8), pp. 57–62.
2. Nikulina E. N., Severyn V. P., Kotsiuba N. V., Bubnov A. I. Modeluvannya teplovykh procesiv parogeneratora AES dlya informatsiynoi tekhnologii optimizatsii upravlinnja [Modeling of thermal processes of NPP steam generator for information technology of control optimization]. *Vestnik Nats. tekhn. un-ta "KhPI": sb. nauch. tr. Temat. vyp.: Sistemnyy analiz, upravlenie i informatsionnye tekhnologii* [Bulletin of the National Technical University "KhPI": a collection of scientific papers. Thematic issue: System analysis, management and information technology]. Kharkov, NTU "KhPI" Publ., 2021, no. 1 (5), pp. 56–61.
3. Nikulina E. N., Severyn V. P., Kotsiuba N. V. Rozrobka informatsiynoi tekhnologii optimizatsii upravlinnja skladnymy dynamichnymy sistemamy [Development of information technology for optimizing the control of complex dynamic systems]. *Vestnik Nats. tekhn. un-ta "KhPI": sb. nauch. tr. Temat. vyp.: Sistemnyy analiz, upravlenie i informatsionnye tekhnologii* [Bulletin of the National Technical University "KhPI": a collection of scientific papers. Thematic issue: System analysis, management and information technology]. Kharkiv, NTU "KhPI" Publ., 2020, no. 2 (4), pp. 63–69.
4. Zakharchenko N. V., Gorokhov S. M., Kochetkov A. V. *Informatsijni parametry pozitsijnykh kodiv* [Information parameters of positional codes]. Odessa, ONAS Publ., 2018. 212 p.
5. Eklund J. E., Arvidsson R. A multiple sampling, single A/D conversion technique for I/Q demodulation. *IEEE Journal of Solid-State Circuits.* 1996, vol. 31, iss. 12, pp. 1987–1994.
6. Blahut R. *Theory and Practice of Error Control Codes.* Addison-Wesley, 1983. 500 p.
7. Losev Yu. I., Shmatkov S. I. *Osnovy teotiji peredachi informatsiji* [Fundamentals of the theory of information transfer]. Kharkiv, KhNU named V. N. Karazina Publ., 2013. 290 p.
8. Yue Tang, Tian Mao, Bing Jiang Design and Experiment of Multi-resolution Composite Digital Array Antenna. *Journal of Radars.* 2016, vol. 5, iss. 3, p. 265–270.
9. Banket V. L., Ivashchenko P. V., Ishchenko M. O. *Zavadostijke koduvannja v telekomunatsijnyh sistemah* [Interference-resistant coding in telecommunication systems]. Odessa, ONAZ named O. S. Popova Publ., 2011. 100 p.
10. Zhurakovsky Yu. P., Poltorak V. P. *Teorija informatsiji ta koduvannja* [Theory of information and coding]. Kyiv, Vyshcha shk. Publ., 2001. 255 p.
11. Kozhevnikov V. L., Kozhevnikov V. L. *Teorija informatsiji ta koduvannja* [Theory of information and coding]. Dnipropetrovsk, National Mining University Publ., 2011. 108 p.
12. Jones G. A., Jones J. M. *Information and Coding Theory.* London, Springer, 2000. 223 p.

13. Sharov V. O., Berdnikov A. G. Modeluvannja koruguvalnogo kaskadnogo kodu v kanalah peredachi danyh systemy upravlinnja [Simulation of the correcting cascade code in data transmission channels of the control system]. *International scientific and technical conference "Computer simulation in scientific technologies"* [Materials of the international scientific and technical conference "Computer modeling in knowledge-intensive technologies"]. Kharkiv, KhNU named V. N. Karazina Publ., 2021, pp. 298–302.
14. Yurong Guan, Zhihua Hu, Chen Chen, Xiaoming Zhu, An anti-noise transmission algorithm for 5G mobile data based on constellation selection and channel joint mapping. *Alexandria Engineering Journal*. 2021, vol. 60, iss. 3, pp. 3153–3160.

Надійшла (received) 10.05.2023

UDC 519.2

O. M. NIKULINA, Doctor of Technical Sciences, Associate Professor, Head of Department Information Systems and Technologies National Technical University «Kharkiv Polytechnic Institute», Kharkiv, Ukraine; ORCID: <https://orcid.org/0000-0003-2938-4215>; e mail: elniknik02@gmail.com

V. P. SEVERYN, Doctor of Technical Sciences, Professor, Professor of Department System Analysis and Information-Analytical Technologies National Technical University «Kharkiv Polytechnic Institute», Kharkiv, Ukraine; ORCID: <https://orcid.org/0000-0002-2969-6780>; e mail: valerii.severyn@khp.edu.ua

V. O. SHAROV, Postgraduate of Department Information Systems and Technologies National Technical University «Kharkiv Polytechnic Institute», Kharkiv, Ukraine; ORCID: <https://orcid.org/0000-0003-3152-0650>; e mail: wycptiy@gmail.com

SIMULATION AND ANALYSIS OF ENCODERS OF INTERFERENCE-RESISTANT CASCADE CODES FOR DYNAMIC SYSTEMS

For various channels of information transmission that can be used in local control systems of a wide range of processes, affordable, scalable, and inexpensive transmission methods are needed that allow the necessary information to be transmitted reliably without errors. Errors that occur in discrete, continuous communication channels are the main obstacle to the reliability of transmitted data. The main causes of errors are signal attenuation, noise and various interferences. As a result of the analysis of error statistics in discrete data transmission channels, it was concluded that single and double errors occur in the absolute majority of cases. It is proposed to combat errors by using tamper-resistant coding. Among the interference-resistant codes, redundant block-separated systematic codes are highlighted, among which is the Hamming code. The basis of the interference-resistant coding model is proposed to be the use of a systematic Hamming code followed by cascade coding by adding a parity check bit. The used model made it possible to deal with all the most likely cases of errors. Since this coding model must be flexibly used in different systems, a necessary criterion of the model is its universality. Therefore, the model was tested on a sample with code combinations of different numbers of bits. The purpose of this article is to analyze the dependence of the coding procedure execution time for encoders of different levels: a separate first-stage encoder, as well as a first-stage encoder with the addition of a second-stage encoder. Experiments were conducted with a large sample, which were subsequently analyzed and interpreted. Since the model proved to be flexible, simple, stable and effective as a result of experiments, and the process of combating interference by using cascade codes has proven itself well in world practice, it is recommended for use in various information management systems. The implementation of the coder for interference-resistant data transmission is given.

Keywords: immunity, bit, code, code combination, cascade codes, coding, coder, parity check.

Повні імена авторів / Author's full names

Автор 1 / Author 1: Нікуліна Олена Миколаївна, Nikulina Olena Mykolaivna

Автор 2 / Author 2: Северин Валерій Петрович, Severyn Valerii Petrovych

Автор 3 / Author 3: Шаров Владислав Олегович, Sharov Vladyslav Olegovych

DOI: 10.20998/2079-0023.2023.01.11
УДК 004.891.3

С. Ф. ЧАЛИЙ, доктор технічних наук, професор, Харківський національний університет радіоелектроніки, професор кафедри інформаційних управляючих систем, м. Харків, Україна; e-mail: serhii.chalyi@nure.ua, ORCID: <https://orcid.org/0000-0002-9982-9091>

І. О. ЛЕЩИНСЬКА, кандидат технічних наук (PhD), доцент, Харківський національний університет радіоелектроніки, доцент кафедри програмної інженерії; м. Харків, Україна; e-mail: iryna.leshchynska@nure.ua, ORCID: <https://orcid.org/0000-0002-8737-4595>

КОНЦЕПТУАЛЬНА МЕНТАЛЬНА МОДЕЛЬ ПОЯСНЕННЯ В СИСТЕМІ ШТУЧНОГО ІНТЕЛЕКТУ

Предметом дослідження є процеси формування пояснень щодо отриманих в системах штучного інтелекту рішень. Для вирішення проблеми непрозорості прийняття рішень в таких системах користувачі мають отримати пояснення щодо отриманих рішень. Пояснення дозволяє довіряти цим рішенням та забезпечити їх використання на практиці. Мета роботи полягає у розробці концептуальної ментальної моделі пояснення для визначення базових залежностей, що визначають зв'язок між вхідними даними, а також діями з отримання результату в інтелектуальній системі, та її кінцевим рішенням. Для досягнення мети вирішуються такі задачі: структуризація підходів до побудови ментальних моделей пояснень; побудова концептуальної ментальної моделі пояснення на основі об'єднаного представлення знань користувача. Висновки. Виконано структуризацію підходів до побудови ментальних моделей пояснень в інтелектуальних системах. Ментальні моделі призначені для відображення сприйняття пояснення користувачем. Виділено каузальний, статистичний, семантичний та концептуальний підходи до побудови ментальних моделей пояснення. Показано, що концептуальна модель задає узагальнені схеми та принципи щодо процесу функціонування інтелектуальної системи. Її подальша деталізація виконується на основі каузального підходу у випадку побудови пояснення для процесів, статистичного підходу при побудові пояснення щодо результату роботи системи, а також семантичного при узгодженні пояснення із базовими знаннями користувача. Запропоновано тривірневу концептуальну ментальну модель пояснення, що містить рівні концепції щодо базових принципів функціонування системи штучного інтелекту, пояснення, що деталізує цю концепцію у прийнятному та зрозумілому для користувача вигляді, а також базових знань про предметну область, які є основою для формування пояснення. У практичному аспекті запропонована модель створює умови для побудови та упорядкування множини узгоджених пояснень, які описують процес та результат роботи інтелектуальної системи з урахуванням можливості їх сприйняття користувачем.

Ключові слова: : пояснення, система штучного інтелекту, зрозумілий штучний інтелект, залежності, ментальна модель, каузальні залежності.

Вступ. У процесі створення сучасних інтелектуальних інформаційних систем знайшли широке застосування методи машинного навчання, які дають можливість сформулювати моделі прийняття рішень на основі виявлення закономірностей у великих масивах даних [1]. Отримані складні моделі, що є «ядром» таких систем, стають непрозорими й незрозумілими для користувачів. Тому користувачі можуть не довіряти рішенням, запропонованим такими інтелектуальними системами, що зменшує ефективність використання отриманих результатів [2].

Для вирішення проблеми непрозорості прийняття рішень в системах штучного інтелекту користувачі мають отримати пояснення щодо отриманих рішень. Пояснення дозволяє довіряти цим рішенням та забезпечити їх використання на практиці [3].

Наведені положення вказують на важливість інтерпретації для користувача процесу формування рішень та отриманого в інформаційній системі результату.

Для вирішення цієї проблеми застосовуються два підходи:

- використання моделей, що можуть бути інтерпретовані безпосередньо;
- використання пояснень для «непрозорих» алгоритмів та моделей, що лежать в основі функціонування інтелектуальної системи.

Проблема інтерпретації процесу прийняття рішень в рамках першого підходу виникає лише у

випадку комерційних обмежень на інформацію щодо структури моделі.

В рамках другого підходу підсистема пояснень вбудовується на етапі розробки інтелектуальної системи або ж розробляється окремо для вже існуючої системи. При побудові пояснень, згідно з концепцією зрозумілого (пояснювального) штучного інтелекту [3, 4], вирішуються три базових задачі:

- побудова ментальних моделей пояснень на основі узагальнення та розширення психологічних теорій пояснень;
- розробка нових методів машинного навчання та методів побудови тлумачень для створення ефективних пояснень;
- оцінка ефективності методів побудови пояснень при вирішенні задач аналітики даних та підтримки автономних агентів.

Вирішення першої задачі дає можливість формалізувати сприйняття пояснення людиною-користувачем інтелектуальної системи. Також реалізація першої задачі створює умови для розробки нових методів побудови пояснень та оцінки останніх при вирішенні задач аналізу даних та розробки автономних агентів.

Використання ментальних моделей забезпечує умови для представлення пояснення у формі, зрозумілій для користувача, з урахування рівня його підготовки, а також його знань [5]. Зазначене свідчить про актуальність даного напрямку досліджень.

© С. Ф. Чалий, І. О. Лещинська, 2023



Дослідницька стаття: Цю статтю опубліковано видавництвом *НТУ «ХПИ»* у збірнику «Вісник Національного технічного університету «ХПИ» Серія: Системний аналіз, управління та інформаційні технології». Ця стаття поширюється за міжнародною ліцензією [Creative Common Attribution \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/). **Конфлікт інтересів:** Автор/и заявив/или про відсутність конфлікту.



Аналіз останніх досліджень і публікацій.

Дослідження щодо формування пояснень в сучасних системах штучного інтелекту були суттєво інтенсифіковані в рамках програми зрозумілого штучного інтелекту (XAI) [3], де було виділено як окремий напрямок досліджень задачу побудови ментальних моделей пояснень [4]. Можливості моделювання пояснень з урахуванням когнітивних аспектів, що суттєво впливає на сприйняття цих тлумачень людьми, було розглянуто в роботі [6]. Психологічні аспекти тлумачень, які визначають можливість побудови концептуальної ментальної моделі пояснень, розглянуті в роботах [7–9]. В цих роботах виділено структуру та функції пояснень. Переваги побудови пояснень з урахуванням як каузального, так і темпорального аспектів були розглянуті в роботах [10–12].

Виконаний аналіз показав, що на сьогодні головна увага при вирішенні задачі побудови ментальних моделей пояснень приділялась побудові описових психологічних моделей пояснень, а також спеціалізованих ментальних моделей, які відображають сприйняття пояснення щодо окремих задач та процесів в інтелектуальній системі.

Однак для розробки методів побудови тлумачень необхідно також розробити концептуальну модель пояснень, яка може бути доповнена локальними моделями. Така модель дає можливість користувачеві зрозуміти ключові принципи побудови пояснень, що створює для ефективного використання користувачем рішень інтелектуальної системи і свідчить про важливість вирішення задачі побудови такої моделі.

Мета та задачі дослідження. Мета роботи полягає у розробці концептуальної ментальної моделі пояснення для визначення базових залежностей, що визначають зв'язок між вхідними даними, а також діями з отримання результату в інтелектуальній системі, та її кінцевим рішенням.

Для досягнення мети вирішуються такі задачі:

- структуризація підходів до побудови ментальних моделей пояснень;
- побудова концептуальної ментальної моделі пояснення на основі об'єднаного представлення знань користувача.

Структуризація підходів до побудови ментальної моделі пояснень.

Мета розробки ментальної моделі пояснення – забезпечити розуміння користувачем внутрішніх процесів системи штучного інтелекту. Забезпечуючи прозоре та інтерпретоване пояснення, ментальна модель має створювати умови для підвищення довіри користувача, а також забезпечувати користувачам можливість оцінити релевантність рішень системи штучного інтелекту. Модель має сприяти ефективному використанню отриманих в інтелектуальній системі результатів.

Ментальна модель пояснень визначає концептуальне представлення, яке використовує інтелектуальна система для тлумачення своїх рішень або дій з отримання цих рішень. Це адаптоване до задачі побудови пояснень представлення знань щодо внутрішніх процесів та послідовності виводу в системі штучного

інтелекту, яке може бути інтерпретовано з тим, щоб допомогти користувачам або зацікавленим сторонам зрозуміти, чому інтелектуальна система прийняла певне рішення.

Ключові підходи до побудови ментальної моделі пояснень базуються на підходах до побудови представлення знань і містять у собі: каузальний; статистичний; семантичний; концептуальний підходи.

Порівняльну характеристику розглянутих підходів наведено у табл. 1.

Вибір форми представлення ментальної моделі залежить від задач, які вирішує система штучного інтелекту, а також контексту вирішення цих задач.

Підхід до побудови ментальної моделі пояснення, оснований на каузальних залежностях (наприклад, правилах), забезпечує формування тлумачень рішень системи штучного інтелекту на основі набору попередньо визначених правил: обмежень або умов. Відмінності між ними полягають у тому, що обмеження є істинними для всіх варіантів процесу формування рішення, а умови – лише для підмножини таких альтернативних варіантів.

Таблиця 1 – Підходи до побудови ментальної моделі пояснень

Підхід	Особливості
1. Каузальний	Використовуються каузальні залежності. Можливість надати пояснення як результату роботи інтелектуальної системи, так і процесу отримання даного результату.
2. Статистичний	Використовуються патерни та залежності, отримані на основі аналізу даних інтелектуальної системи. Орієнтований в першу чергу на пояснення рішення системи штучного інтелекту.
3. Семантичний	Використовуються знання з предметної області (а тому числі темпоральні) щодо процесу прийняття рішення та результату інтелектуальної системи. Можливість адаптації до рівня знань користувача.
4. Концептуальний	Використовуються знання щодо загальних принципів, схеми, технологій та процесу прийняття рішень у предметній області. Можливість побудови високорівневого пояснення, що є узагальненим для різних контекстів прийняття рішень у предметній області. Можливість деталізації на основі каузального, статистичного та семантичного підходів.

Вказані залежності мають охоплювати та узагальнювати весь процес прийняття рішень інтелектуальною системою з тим, щоб надати прозоре пояснення як окремих дій, так і їх послідовностей, що забезпечують досягнення результуючого рішення.

У каузальній ментальній моделі пояснення виводяться шляхом оцінки вхідних та проміжних даних з використанням набору логічних правил. Кожне правило складається з умови або набору умов, які перевіряють значення певних змінних, і пов'язаного пояснення або подальшої дії. Коли умови правила задовольняються, відповідне пояснення надається як результат. Або ж виконується дія із уточнення можливого пояснення. Послідовність таких правил може надати пояснення як для процесу, так і для результату, отриманого в системі штучного інтелекту.

Фактично, правила визначають граничні умови, які впливають на прийняття рішення.

Перевага даного підходу до побудови ментальної моделі полягає у можливості її безпосередньої інтерпретації. Такі пояснення є інтуїтивно зрозумілими, оскільки вони базуються на детермінованих залежностях, які спрощують процес прийняття рішень в системі штучного інтелекту.

Недолік підходу пов'язаний із обмеженнями в представленні складних зв'язків між вхідними даними і отриманим результатом, а також невизначеності яка може виникати в результаті зовнішніх впливів на інтелектуальну систему. Зазвичай такий підхід підходить для випадків, коли можуть бути визначені чіткі умови та обмеження у процесі прийняття рішення.

Побудова ментальної моделі на основі правил передбачає визначення правил, вибір функцій оцінки правил для відбору альтернатив пояснення, а також постійне вдосконалення множини правил з урахуванням змін і доповнень у процесі прийняття рішення у системі штучного інтелекту.

Статистичний підхід до побудови ментальної моделі пояснення в системі штучного інтелекту орієнтований на визначення ключових особливостей або факторів, які вплинули на отримане рішення. Пояснення виводяться на основі аналізу патернів та закономірностей, наявних у даних, які використовує система штучного інтелекту. Ці закономірності можуть бути отримані методами машинного навчання. На відміну від попереднього підходу, в даному випадку ключова увага приділяється встановленню зв'язків між вхідними даними та кінцевим результатом інтелектуальної системи.

Перевага даного підходу полягає у здатності зафіксувати складні нелінійні зв'язки та враховувати невизначеність в даних. Також враховується контекст прийняття рішення.

Недолік підходу полягає базується на його ймовірнісній природі і полягає у відсутності визначення детермінованих каузальних залежностей.

Семантичний підхід до побудови ментальної моделі пояснення базується на використанні семантики предметної області. В даному випадку можуть бути використані онтології, семантичні мережі, графи знань. Розробка представлення знань може бути виконана як за допомогою інженерів знань, так і з використанням методів автоматизованої побудови бази знань.

Перевага даного підходу полягає у можливості надавати чіткі, інтерпретовані та зрозумілі пояснення, спираючись на зрозумілі людині представлення. Такий

підхід може відображати складні відношення, в тому числі каузальні та темпоральні, дозволяючи користувачам пояснити розгортання процесу прийняття рішень у часі.

Недолік підходу полягає пов'язаний із обмеженням при формуванні знань та при поясненні зовнішніх впливів, оскільки вони не завжди відображені у відповідній базі знань.

Підхід до побудови абстрактної ментальної моделі пояснення в системі штучного інтелекту базується на використанні загальних принципів, концепцій та моделей, що лежать в основі процесу прийняття рішень. Цей підхід має на меті забезпечити концептуальне розуміння поведінки інтелектуальної замість пояснення конкретних процесів або результатів.

Концептуальна ментальна модель пояснення містить знання щодо структури та властивостей предметної області. Така модель базується на психологічних теоріях пояснення. Останні використовуються при аналізі навчання дітей у людському суспільстві. Наприклад, концептуальне пояснення щодо розпізнавання зображень має охоплювати принципи розпізнавання образів, виділення ознак, класифікації, а також концепції згорткових нейронних мереж, тощо. В даному випадку замість правил або патернів пояснень щодо конкретного результату модель підкреслює загальну схему, процес та принципи розпізнавання зображень у конкретній інтелектуальній системі.

У практичному аспекті концептуальне представлення може визначати ключові елементи та принципи обробки вхідних даних, які інтелектуальна система використовує для прийняття рішення.

Наприклад, для пояснення щодо рішення «людина на зображенні біжить», прийнятого системою розпізнавання зображень, можуть бути виділені зони зображення, що відображають відповідні фрагменти тіла людини у русі. В даному випадку концепція базується на визначенні підмножини фрагментів зображення, які є ключовим для прийнятого в інтелектуальній системі рішення.

Аналогічно, концептуальним поясненням щодо кредитного рішення в інтелектуальній системі, що має відповідати сприйняттю людини, може виступати залежність між кредитним рейтингом та доходом клієнта.

Перевага даного підходу полягає в тому, що концептуальна ментальна модель забезпечує розуміння поведінки інтелектуальної системи на більш високому рівні абстракції, що дає можливість узагальнити контекст прийняття рішення в рамках пояснення. Такий підхід дозволяє користувачам зрозуміти базові принципи та концепції прийняття рішень системою штучного інтелекту, що дає можливість сформулювати умови та обмеження щодо застосування цих рішень, визначити семантику цих рішень, а також передбачити додаткові можливості їх використання.

Недолік концептуального підходу до побудови ментальної моделі пояснень полягає у відсутності конкретних деталей щодо конкретних рішень інтелектуальної системи. Тому даний підхід потребує доповнення можливостями інших підходів..

Зв'язок підходів до побудови ментальних моделей пояснення представлено на рис. 1. Даний зв'язок відображає загальну схему побудови системи ментальних моделей в залежності від вимог до пояснень в системі штучного інтелекту.

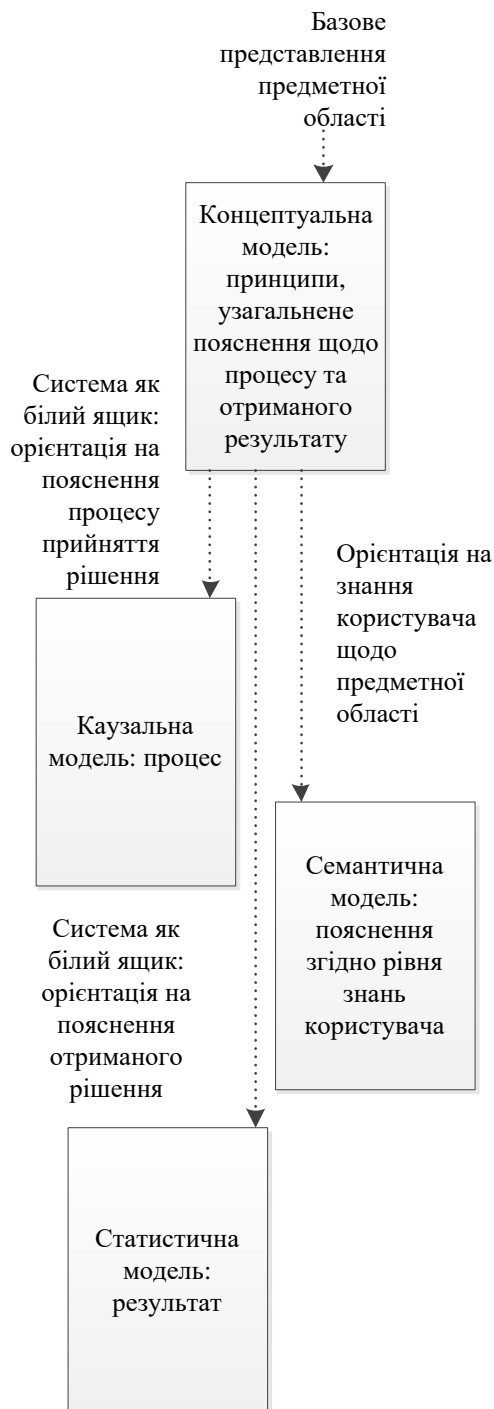


Рис. 1. Послідовність та умови реалізації підходів до побудови ментальних моделей пояснень

Порівняння представлених підходів до побудови ментальної моделі пояснень дає можливість зробити висновок про те, що концептуальна ментальна модель пояснення має розроблятися в першу чергу і в подальшому розширюватись з використанням розглянутих каузального, статистичного та семантичного підходів.

Статистичний підхід має переваги у випадку вирішення задачі пояснення результату роботи інтелектуальної системи, каузальний – для тлумачення процесу отримання результату. Семантичний підхід дає можливість адаптувати пояснення до рівня підготовки користувача та створює умови для пояснення щодо способу використання отриманого в системі результату.

Концептуальна ментальна модель пояснення.

Запропонована концептуальна ментальна модель складається із елементів, пов'язаних структурними та темпоральними відношеннями, враховує онтологію понять в предметній області, а також може якісно описувати впливи на процес прийняття рішень в системі штучного інтелекту.

Ключова особливість концептуальної моделі полягає в тому, що вказані елементи можуть бути неузгодженими. Це створює умови для побудови альтернативних варіантів пояснень. В подальшому кожен варіант пояснення може бути ймовірно оцінений.

Найбільш ймовірний варіант розглядається як ключове пояснення. Отримане пояснення має бути пов'язано із концепцією, що узагальнює ситуацію та умови прийняття рішення. В результаті базовим елементом концептуальної моделі є трійка (знання, що деталізують певну концепцію щодо прийняття рішення; безпосередньо концепція; пояснення).

Знання щодо предметної області визначають властивості певної концепції прийняття рішень в інтелектуальній системі і можуть містити:

- опис структури предметної області;
- опис базових залежностей, на яких базується процес прийняття рішень у системі штучного інтелекту;
- опис зовнішніх впливів на процес прийняття рішення в системі;
- показник оцінки пояснень, що дає можливість упорядкувати їх за відповідністю концепції.

Тобто концепція як узагальнене представлення щодо певної категорії або класу об'єктів та процесів, а також їх властивостей і взаємозв'язків між ними пояснюється через ієрархічну структуру предметної області та процесу використання цієї структури для отримання рішення.

Узагальнену структуру концептуальної моделі пояснення наведено на рис. 2.

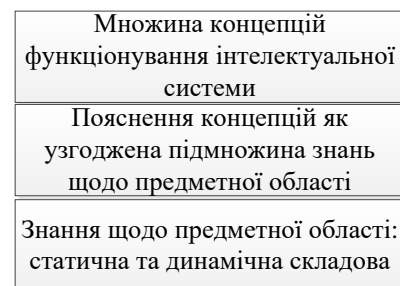


Рис. 2. Структура концептуальної моделі пояснення

Згідно представленої структури, концептуальна модель пояснення для користувача M складається із множини концепцій $C = \{c_i\}$, пояснень $Expl = \{\text{expl}_{i,j}\}$, а також знань для формування пояснень $K = \{k_m\}$ за умови локальної узгодженості концепції та пояснення:

$$M = \langle C, Expl, K \mid Expl \subset K \rangle. \quad (1)$$

З урахуванням можливих протиріч у знаннях щодо процесів прийняття рішень в системі штучного інтелекту, запропонована модель використовує принцип локальної узгодженості знань.

Згідно даного принципу кожна пара (концепція; пояснення) має бути узгодженою, тобто має характеризуватись властивостями повноти та несуперечливості.

Іншими словами, із множини знань K щодо предметної області для пояснення $\text{expl}_{i,j}$ відбираються лише узгоджені з концепцією знання. Узгодженість пояснення і концепції $c_i \sim \text{expl}_{i,j}$ визначається наступним чином:

$$\begin{aligned} c_i \sim \text{expl}_{i,j} &\equiv (\forall i) \{ \text{expl}_{i,j} \} \neq \emptyset, \\ (\forall i, j, m) &(\text{expl}_{i,j} \neg \text{expl}_{i,m}) = \text{false}. \end{aligned} \quad (2)$$

Концепція в рамках моделі пов'язує об'єкти o_i, o_m відношенням r_m^l , що дає можливість визначити високорівневі зв'язки у предметній області. Відповідно, концепція визначається трійкою:

$$o_i r_m^l o_m, (o_i, o_m) \in O, r_m^l \in R. \quad (3)$$

Представлення концепції у вигляді (3) дає можливість визначити як ключові складові ієрархічної структури предметної області, що впливають на прийняття рішення, так і задати ключові залежності у самому процесі прийняття рішення.

В першому випадку може бути використаний онтологічний підхід, а в другому – каузальні або темпоральні залежності.

Концепція у вигляді (3) формує узагальнені запитання для пояснення щодо результату роботи інтелектуальної системи: Чому об'єкти o_i, o_m мають між собою зв'язок виду r_m^l ? Запитання щодо процесу прийняття рішення має вигляд: Чому за об'єктом (або дією) o_i при прийнятті рішення було використано об'єкт o_m способом r_m^l ? Концептуальні запитання щодо семантики прийняття рішення можуть мати вигляд: Який сенс (можливості використання) має отриманий як рішення об'єкт o_m для вхідного об'єкту o_i з урахуванням контексту прийняття рішення r_m^l ?

У наведеному вище прикладі щодо аналізу зображень в якості об'єкту o_i концепції виступає фрагмент зображення людини (зокрема, зображення

ніг), який свідчить про її рух, в якості об'єкту o_m виступає характеристика руху людини (біжить, стоїть, йде), а залежність r_m^l визначає принцип розпізнавання руху на зображенні: за станом певних фрагментів тіла людини.

Кожний об'єкт концепції o_i задається через множину його властивостей $v_{i,n}$: $o_i = \{v_{i,n}\}$. Таке представлення об'єкту дає можливість деталізувати концепцію через залежності між окремими властивостями об'єктів. Такі залежності $\rho_{m,q}^{l,n}$ можуть виступати в якості основи для пояснень $\text{expl}_{i,j}$ щодо концепції c_i :

$$\text{expl}_{i,j} = v_{i,n} \rho_{m,q}^{l,n} v_{m,q}. \quad (4)$$

Висновки. Виконано структуризацію підходів до побудови ментальних моделей пояснень в інтелектуальних системах. Ментальні моделі призначені для відображення сприйняття пояснення користувачем. Виділено каузальний, статистичний, семантичний та концептуальний підходи до побудови ментальних моделей пояснення. Показано, що концептуальна модель задає узагальнені схеми та принципи щодо процесу функціонування інтелектуальної системи. Її подальша деталізація виконується на основі каузального підходу у випадку побудови пояснення для процесів, статистичного підходу при побудові пояснення щодо результату роботи системи, а також семантичного при узгодженні пояснення із базовими знаннями користувача.

Запропоновано трирівневу концептуальну ментальну модель пояснення, що містить рівні концепції щодо базових принципів функціонування системи штучного інтелекту, пояснення, що деталізує цю концепцію у прийнятному та зрозумілому для користувача вигляді, а також базових знань про предметну область, які є основою для формування пояснення. У практичному аспекті запропонована модель створює умови для побудови та упорядкування множини узгоджених пояснень, які описують процес та результат роботи інтелектуальної системи з урахуванням можливості їх сприйняття користувачем.

Список використаної літератури

- Engelbrecht Andries P. *Computational Intelligence: An Introduction*. New York: John Wiley & Sons, 2007. 632 p.
- Tintarev N., Masthoff J. A survey of explanations in recommender systems. *The 3rd international workshop on web personalisation, recommender systems and intelligent user interfaces (WPRSIUI'07)*. 2007. P. 801–810.
- Gunning D., Aha D. DARPA's Explainable Artificial Intelligence (XAI) Program. *AI Magazine*. 2019. Vol. 40(2). P. 44–58.
- Gunning D., Vorm E., Wang J., Turek M. DARPA's explainable AI (XAI) program: A retrospective. *Applied AI Letters*. Vol. 2, no. 4. 2021. DOI: <https://doi.org/10.1002/ail2.61/>
- Gilpin L. H., Bau D., Yuan B. Z., Bajwa A., Specter M., Kagal L. Explaining Explanations: An Overview of Interpretability of Machine Learning. *arXiv:1806.00069*. 2018.
- Miller T. Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*. 2019. Vol. 267 P. 1–38.
- Chi M., de Leeuw N., Chiu M., Lavancher C. Eliciting self-explanations improves understanding. *Cognitive Science*. 1994. Vol. 18. P. 439–477.

8. Carey S. *The origin of concepts*. New York: Oxford University Press, 2009. 608 p.
9. Holyoak Keith J., Morrison Robert G. *The Oxford Handbook of Thinking and Reasoning*. Oxford University Press, 2012. 864 p.
10. Чалий С.Ф., Лещинський В.О., Лещинська І.О. Декларативно-темпоральний підхід до побудови пояснень в інтелектуальних інформаційних системах. *Вісник Нац. техн. ун-ту "ХПІ": зб. наук. пр. Темат. вип. Системний аналіз, управління та інформаційні технології*. Харків: НТУ «ХПІ», 2020. № 2(4). С. 51–56.
11. Halpern J. Y., Pearl J. *Causes and explanations: A structural-model approach. Part II: Explanations*. URL: <https://arxiv.org/pdf/cs/0208034.pdf> (дата звернення: 11.05.202).
12. Chalyi S., Leshchynskyi V. Temporal representation of causality in the construction of explanations in intelligent systems. *Advanced Information Systems*. Kharkiv: NTU "KhPI", 2020. Vol. 4, № 3. P. 113–117.
5. Gilpin L. H., Bau D., Yuan B. Z., Bajwa A., Specter M., Kagal L. Explaining Explanations: An Overview of Interpretability of Machine Learning. *arXiv:1806.00069*. 2018.
6. Miller T. Explanation in artificial intelligence: Insights from the social sciences, *Artificial Intelligence*. 2019. Vol. 267 P. 1–38.
7. Chi M., de Leeuw N., Chiu M., Lavancher C. Eliciting self-explanations improves understanding. *Cognitive Science*. 1994. Vol.18. P. 439–477.
8. Carey S. *The origin of concepts*. New York: Oxford University Press, 2009. 608 p.
9. Holyoak Keith J., Morrison Robert G. *The Oxford Handbook of Thinking and Reasoning*. Oxford University Press, 2012. 864 p.
10. Chalyi S., Leshchynskyi V., Leshchynska I. Deklaratyvno-temporalnyi pidkhid do pobudovy poiasnen v intelektualnykh informatsiynykh systemakh [Declarative-temporal approach to the construction of explanations in intelligent information systems]. *Visnyk Nats. tekhn. un-tu "KhPI": zb. nauk. pr. Temat. vyp. Systemnyi analiz, upravlinnia ta informatsiini tekhnologii* [Bulletin of the National Technical University "KhPI": a collection of scientific papers. Thematic issue: System analysis, management and information technology]. Kharkov, NTU "KhPI" Publ., 2020, no. 2(4), pp. 51–56.

References (transliterated)

1. Engelbrecht Andries P. *Computational Intelligence: An Introduction*. NJ: John Wiley & Sons, 2007. 632 p.
2. Castelvocchi D. Can we open the black box of AI? *Nature News* 2016. Vol. 538 (7623). P. 20.
3. Tintarev N., Masthoff J. A survey of explanations in recommender systems. *The 3rd International workshop on web personalisation, recommender systems and intelligent user interfaces (WPRSUI07)*. 2007, pp. 801–810.
4. Gunning D., Vorm E., Wang J., Turek M. DARPA's explainable AI (XAI) program: A retrospective. *Applied AI Letters*. Vol. 2, no. 4, 2021. DOI: <https://doi.org/10.1002/ail2.61>.
11. Halpern J. Y., Pearl J. *Causes and explanations: A structural-model approach. Part II: Explanations*. Available at: <https://arxiv.org/pdf/cs/0208034.pdf> (accessed 11.05.202).
12. Chalyi S., Leshchynskyi V. Temporal representation of causality in the construction of explanations in intelligent systems. *Advanced Information Systems*. 2020, vol. 4, no 3, pp. 113–117.

Надійшло (received) 14.05.2023

UDC 004.891.3

S. F. CHALYI, Doctor of Technical Sciences, Full Professor, Kharkiv National University of Radio Electronics, Professor of the Department of Information Control System, Kharkiv; e mail: serhii.chalyi@nure.ua, ORCID: <https://orcid.org/0000-0002-9982-9091>

I. O. LESHCHYNSKA, Candidate of Technical Sciences (PhD), Associate Professor, Kharkiv National University of Radio Electronics, Associate Professor at the Department of Software Engineering доцент кафедри програмної інженерії, Kharkiv; e-mail: iryna.leshchynska@nure.ua, ORCID: <https://orcid.org/0000-0002-8737-4595>

THE CONCEPTUAL MENTAL MODEL OF EXPLANATION IN AN ARTIFICIAL INTELLIGENCE SYSTEM

The subject of research is the process of formation of explanations in artificial intelligence systems. To solve the problem of the opacity of decision-making in artificial intelligence systems, users should receive an explanation of the decisions made. The explanation allows you to trust these solutions and ensure their use in practice. The purpose of the work is to develop a conceptual mental model of explanation to determine the basic dependencies that determine the relationship between input data, as well as actions to obtain a result in an intelligent system, and its final solution. To achieve the goal, the following tasks are solved: structuring approaches to building mental models of explanations; construction of a conceptual mental model of explanation based on a unified representation of the user's knowledge. Conclusions. The structuring of approaches to the construction of mental models of explanations in intelligent systems has been carried out. Mental models are designed to reflect the user's perception of an explanation. Causal, statistical, semantic, and conceptual approaches to the construction of mental models of explanation are distinguished. It is shown that the conceptual model sets generalized schemes and principles regarding the process of functioning of the intellectual system. Its further detailing is carried out on the basis of a causal approach in the case of constructing an explanation for processes, a statistical approach when constructing an explanation about the result of the system's work, as well as a semantic approach when harmonizing the explanation with the user's basic knowledge. A three-level conceptual mental model of the explanation is proposed, containing levels of concepts regarding the basic principles of the functioning of the artificial intelligence system, an explanation that details this concept in an acceptable and understandable way for the user, as well as basic knowledge about the subject area, which is the basis for the formation of the explanation. In a practical aspect, the proposed model creates conditions for building and organizing a set of agreed explanations that describe the process and result of the intelligent system, considering the possibility of their perception by the user.

Keywords: explanation; artificial intelligence system; understandable artificial intelligence; dependencies; mental model; causal dependence.

Повні імена авторів / Author's full names

Автор 1 / Author 1: Чалий Сергій Федорович, Chalyi Serhii Fedorovych

Автор 2 / Author 2: Лещинська Ірина Олександрівна, Leshchynska Irina Oleksandrivna

ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ

INFORMATION TECHNOLOGY

DOI: 10.20998/2079-0023.2023.01.12

UDC 519.2

K. S. YAMKOVYI, National Technical University "Kharkiv Polytechnic Institute", Assistant Professor of Computer Mathematics and Data Analysis Department, Kharkiv, Ukraine, e-mail: klym.yamkovyi@cs.khpi.edu.ua; ORCID: <https://orcid.org/0000-0001-9512-4150>

ADAPTATION OF LAMBDMART MODEL TO SEMI-SUPERVISED LEARNING

The problem of information searching is very common in the age of the internet and Big Data. Usually, there are huge collections of documents and only multiple percent of them are relevant. In this setup brute-force methods are useless. Search engines help to solve this problem optimally. Most engines are based on learning to rank methods, i.e. first of all algorithms produce scores for documents based on their feature and after that sorts them according to the score in an appropriate order. There are a lot of algorithms in this area, but one of the most fastest and a robust algorithm for ranking is LambdaMART. This algorithm is based on boosting and developed only for supervised learning, where each document in the collection has a rank estimated by an expert. But usually, in this area, collections contain tons of documents and their annotation requires a lot of resources like time, money, experts, etc. In this case, semi-supervised learning is a powerful approach. Semi-supervised learning is an approach to machine learning that combines a small amount of labeled data with a large amount of unlabeled data during training. Unlabeled data, when used in combination with a small quantity of labeled data, can produce significant improvement in learning accuracy. This paper is dedicated to the adaptation of LambdaMART to semi-supervised learning. The author proposes to add different weights for labeled and unlabeled data during the training procedure to achieve higher robustness and accuracy. The proposed algorithm was implemented using Python programming language and LightGBM framework that already has supervised the implementation of LambdaMART. For testing purposes, multiple datasets were used. One synthetic 2D dataset for a visual explanation of results and two real-world datasets MSLR-WEB10K by Microsoft and Yahoo LTRC.

Keywords: learning to rank, information retrieval, semi-supervised learning, pairwise ranking, LambdaMART, pseudo labeling, NDCG.

Introduction. A modern person searches for useful information or some stuff every day. Usually, we sort a list of items to find the best option. But today in the Big Data age almost all information stored on the internet and information retrieval systems like search engines can help us to handle this. Search engines based on learning to rank (LR) methods. LR is the application of machine learning that constructs ranking models for information retrieval systems. The ranking model proposes to rank, i.e. producing a permutation of items in new, unseen lists in a similar way to rankings in the training data.

One of the naive and simple to understand approaches in LR is pairwise comparison. In this case, we take two objects and try to classify which of them is more preferable than another. For example, bubble sort works with the same rules. The main advantages of this approach are simplicity and the ability to describe an algorithm's behavior. The pairwise approach is popular in the state-of-the-art ranking methods, i.e. LambdaMART [1]. This method was developed more than ten years ago, but still one of the most efficient and faster.

As we say above, today we are dealing with a huge dataset. And not always there is an opportunity to label all data, because of cost or complexity. So when we can't mark

the whole dataset we can use some compromise in the form of semi-supervised learning: train model on labeled and unlabeled data at the same time. In our research, we've developed a semi-supervised version of the LambdaMART algorithm.

This paper is organized as follows: Section 2 describes the problem statement. Section 3 provides a short review of related literature in the area of the research. Section 4 describes the proposed semi-supervised learning to rank approach. After that, we'll show the accuracy of the proposed algorithm on multiple datasets in Section 5. Finally, Section 6 contains conclusions.

Problem statement. Suppose that $D = \{d_1, \dots, d_N\}$ – set of documents, where N – denotes number of documents, $Q = \{q_1, \dots, q_M\}$ – set of queries, where M – number of queries and $N > M$. $R = \{r_1, \dots, r_K\}$ – ordinal set of ranks, where K denotes the number of ranks. Usually in Information Retrieval (IR) $K \in [2, 5]$ e.g. $R = \{\text{irrelevant, relevant}\}$ in binary case and if $K = 5$ then $R = \{\text{bad, fair, good, excellent, perfect}\}$ [2]. There exists a total order between the ranks $r_L \succ r_{L-1} \succ \dots \succ r_1$, where " \succ " denotes a preference relationship.

© Yamkovyi K. S., 2023



Research Article: This article was published by the publishing house of NTU "KhPI" in the collection "Bulletin of the National Technical University "KhPI" Series: System analysis, management and information technologies." This article is distributed under a Creative Commons Attribution (CC BY 4.0). **Conflict of Interest:** The author/s declared no conflict of



Each query q_i related with a subset of retrieved documents $D_i = \{d_{i1}, \dots, d_{im(q_i)}\}$, where $n(q_i)$ – size of D_i and list of labels $Y_i = \{y_{i1}, \dots, y_{im(q_i)}\}$, where $y_{ij} \in Y$ and denotes rank of document d_{ij} , $n(q_i) > m(q_i)$ – some documents haven't labels as a part of the query. Feature vector $\mathbf{x} \in X$ represent the "query – document" pair $\mathbf{x}_{ij} = \Psi(q_i, d_{ij})$ [3]. Usually, only query – document features are available, queries and documents feature not disclosed because it's a commercial secret of search engines. Thus training data can be formally represent as subset of labeled query – document pairs $L = \{(\mathbf{x}_{ij}, y_{ij})\}$, $|L| = l$ and subset of unlabeled pairs, $U = \{\mathbf{x}_{ij}\}$, $|U| = u$ the set U will be used in an unsupervised manner.

In IR information about labels can take several forms such as absolute labels, preference relations, complete orderings, or partial ordering. We'll use pairwise preference relations of the form document "i is preferred to j", and denoted by $i \succ j$. This information can be obtained by comparison of documents' labels.

Our goal is to estimate a function $f(\mathbf{x}, \mathbf{w})$, where \mathbf{w} – the function's parameters, which compute the score of a document related to a query: $o_i = f(\mathbf{x}_i, \mathbf{w})$. For any pair of items \mathbf{x}_i and \mathbf{x}_j we'll compute the score o_i and o_j . The two scores are mapped to a learned probability that x_i should be ranked higher than x_j using a probabilistic model $P(\mathbf{x}_i \succ \mathbf{x}_j)$. The one of the widely used in IR is Bradley – Terry model:

$$P(i \succ j) = \frac{o_i}{o_i + o_j}. \tag{1}$$

Let $O = \{i_1 \succ j_1, i_2 \succ j_2, \dots\}$ be a set of observed pairwise preferences. We can estimate parameters \mathbf{w} by maximizing likelihood.

$$C = \sum_{(i,j) \in L} I_{i \succ j} \log P(i \succ j) + \sum_{(i,j) \in U} I_{i \succ j} \log P(i \succ j) \rightarrow \max_{\mathbf{w}} \tag{2}$$

where $I_{i \succ j}$ – the indicator that denotes the observed pairwise preference, $I_{i \succ j} = 1$ when $i \succ j$, 0 otherwise [4];

α – regularization coefficient.

Related works. During this research, we relied on the different groups of papers. First of all is an overview of ranking algorithms, especially pairwise methods. Christopher J.C. Burges showed in his work evolution from RankNet – and neural networks-based algorithm to LambdaMART boosting-based approach for ranking. And nowadays LambdaMART is one of the state-of-the-art algorithms in LR. According to him the key observation of LambdaRank is thus that to train a model, we don't need the costs themselves: we only need the gradients (of the costs

with respect to the model scores). The arrows (λ 's) mentioned on fig. 1 are exactly those gradients. Note that this does not mean that the gradients are not gradients of a cost. The model was designed to learn Normalized Discounted Cumulative Gain (NDCG). LambdaMART combines MART (multiple additive regression trees) and LambdaRank [5].



Fig. 1. An example of RankNet gradients (black) and LambdaRank gradients (red)

In [6] the authors covered clustering. They compare unsupervised and semi-supervised clustering, describing a typology of methods. The main difference of semi-supervised clustering is the availability of some external knowledge. But they also consider that the available knowledge is too far from being representative of a target classification of the items, so that supervised learning is not possible, even in a transductive form.

So if we have labels of good quality but still not enough for supervised learning the good choice is transductive learning. One of the first approaches of transductive learning is T-SVM [7] by Vapnik. Also, the transductive approach was applied in LR by Rahangdale A. U., and Raut S. in [8]. The authors propose rank preserving clustering with PLocalSearch and get pseudo labels for unlabeled data. They present semi-supervised learning that adopts a clustering-based transductive method and combines it with a non-measure-specific listwise approach to learn the LTR model.

And another group of methods is proposed cost function for boosting algorithms with regularization on unlabeled data and. [9] presents a boosting-based algorithm for learning a bipartite ranking function (BRF) with partially labeled data. The proposed approach is a semi-supervised inductive ranking algorithm which, as opposed to transductive algorithms, can infer an ordering on new examples that were not used for its training. And in [10] the authors made the assumption that unlabeled data gives information about the data distribution $P(x)$ and the

structure of the unlabeled data tells us about the ranking function. The authors show it on non-convex clusters. So unlabeled data defines the extent and shape of clusters and labeled data determines the class/function value of each cluster. **Semi-supervised ranking function.** The proposed semi-supervised approach consists of two steps: pseudo labeling and learning of ranking function. The main focus of the paper is a second step.

During the pseudo labeling step we used the k -NN algorithm for label assignment. This algorithm is very common choice in semi-supervised learning. The k -NN algorithm uses the compactness hypothesis: nearest documents have the same labels. For each example in the labeled training set L , we assigned the same label to its nearest neighbors from U [11]. The k -NN algorithm with a small number of neighbors (<3) showed the best experimental result. It could be explained by complicated structure of cluster, overlapping between them and noise in the data.

As we described above the pairwise ranking approach based on the Bradley – Terry model, which could be expressed as sigmoid (fig. 2):

$$P(i \succ j) = \frac{o_i}{o_i + o_j} = \frac{e^{o_i - o_j}}{1 + e^{o_i - o_j}} = 1 - P(j \succ i). \quad (3)$$

So sigmoid usually optimized using cross-entropy loss:

$$\begin{aligned} C_{ij} &= -\bar{P}_{ij} \log P_{ij} - (1 - \bar{P}_{ij}) \log(1 - P_{ij}) = \\ &= -\bar{P}_{ij} [\log P_{ij} - \log(1 - P_{ij})] - \log(1 - P_{ij}) = \\ &= -\bar{P}_{ij} (o_i - o_j) + \log(1 + e^{o_j}), \end{aligned} \quad (4)$$

where \bar{P}_{ij}, P_{ij} – true and predicted probabilities respectively.

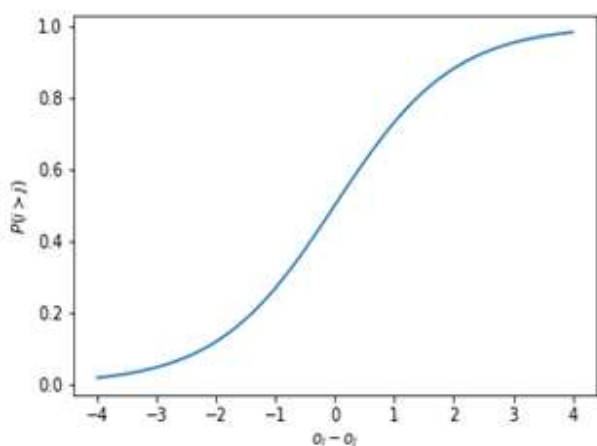


Fig. 2. Bradley – Terry model

Derivative of the loss function:

$$\frac{\delta C_{ij}}{\delta \mathbf{w}} = \frac{\delta C_{ij}}{\delta o_{ij}} \frac{\delta o_{ij}}{\delta \mathbf{w}} = \frac{\delta C_{ij}}{\delta o_{ij}} \left[\frac{\delta o_i}{\delta \mathbf{w}} - \frac{\delta o_j}{\delta \mathbf{w}} \right]. \quad (5)$$

Let's describe the derivative:

$$\frac{\delta C_{ij}}{\delta o_{ij}} = \frac{\delta(-\bar{P}_{ij} o_{ij})}{\delta o_{ij}} + \frac{\delta \log(1 + e^{o_j})}{\delta o_{ij}} = -\bar{P}_{ij} + P_{ij}. \quad (6)$$

Training process using mini-batch gradient descent, but the batch is a query. And the query consists of labeled and unlabeled documents, the LambdaMART learning looks like this:

$$\mathbf{w} \rightarrow \mathbf{w} - \eta \sum_L \frac{\delta C_L}{\delta \mathbf{w}}, \quad (7)$$

where η – learning rate.

Formula (7) should be modified for semi-supervised model:

$$\mathbf{w} \rightarrow \mathbf{w} - \eta \left[\sum_L \frac{\delta C_L}{\delta \mathbf{w}} + \alpha \sum_U \frac{\delta C_U}{\delta \mathbf{w}} \right]. \quad (8)$$

The equation above introduced a semi-supervised version of the LambdaMART model.

For clarity, we'll use an example: a query consisting of 5 documents: 1, 2, and 3 with labels and 4, 5 without labels. In this case, describe the equation (8):

$$\begin{aligned} \sum_L \frac{\delta C_L}{\delta \mathbf{w}} + \alpha \sum_U \frac{\delta C_U}{\delta \mathbf{w}} &= \\ &= \frac{\delta C_{12}}{\delta o_{12}} \left(\frac{\delta o_1}{\delta \mathbf{w}} - \frac{\delta o_2}{\delta \mathbf{w}} \right) + \frac{\delta C_{13}}{\delta o_{13}} \left(\frac{\delta o_1}{\delta \mathbf{w}} - \frac{\delta o_3}{\delta \mathbf{w}} \right) + \\ &+ \frac{\delta C_{23}}{\delta o_{23}} \left(\frac{\delta o_2}{\delta \mathbf{w}} - \frac{\delta o_3}{\delta \mathbf{w}} \right) + \alpha \frac{\delta C_{45}}{\delta o_{45}} \left(\frac{\delta o_4}{\delta \mathbf{w}} - \frac{\delta o_5}{\delta \mathbf{w}} \right) + \\ &+ \alpha \frac{\delta C_{14}}{\delta o_{14}} \left(\frac{\delta o_1}{\delta \mathbf{w}} - \frac{\delta o_4}{\delta \mathbf{w}} \right) + \alpha \frac{\delta C_{15}}{\delta o_{15}} \left(\frac{\delta o_1}{\delta \mathbf{w}} - \frac{\delta o_5}{\delta \mathbf{w}} \right) + \\ &+ \alpha \frac{\delta C_{24}}{\delta o_{24}} \left(\frac{\delta o_2}{\delta \mathbf{w}} - \frac{\delta o_4}{\delta \mathbf{w}} \right) + \alpha \frac{\delta C_{25}}{\delta o_{25}} \left(\frac{\delta o_2}{\delta \mathbf{w}} - \frac{\delta o_5}{\delta \mathbf{w}} \right) + \\ &+ \alpha \frac{\delta C_{34}}{\delta o_{34}} \left(\frac{\delta o_3}{\delta \mathbf{w}} - \frac{\delta o_4}{\delta \mathbf{w}} \right) + \alpha \frac{\delta C_{35}}{\delta o_{35}} \left(\frac{\delta o_3}{\delta \mathbf{w}} - \frac{\delta o_5}{\delta \mathbf{w}} \right). \end{aligned} \quad (9)$$

For optimization, we'll introduce factorization,

because $\frac{\delta C_{ij}}{\delta o_{ij}}$ computed faster than $\frac{\delta o_i}{\delta \mathbf{w}}$:

$$\begin{aligned} \frac{\delta o_1}{\delta \mathbf{w}} \left[\frac{\delta C_{12}}{\delta o_{12}} + \frac{\delta C_{13}}{\delta o_{13}} + \alpha \frac{\delta C_{14}}{\delta o_{14}} + \alpha \frac{\delta C_{15}}{\delta o_{15}} \right] + \\ + \frac{\delta o_2}{\delta \mathbf{w}} \left[\frac{\delta C_{23}}{\delta o_{23}} + \alpha \frac{\delta C_{24}}{\delta o_{24}} + \alpha \frac{\delta C_{25}}{\delta o_{25}} - \frac{\delta C_{12}}{\delta o_{12}} \right] + \\ + \frac{\delta o_3}{\delta \mathbf{w}} \left[\alpha \frac{\delta C_{34}}{\delta o_{34}} + \alpha \frac{\delta C_{35}}{\delta o_{35}} - \frac{\delta C_{13}}{\delta o_{13}} - \frac{\delta C_{23}}{\delta o_{23}} \right] + \\ + \frac{\delta o_4}{\delta \mathbf{w}} \left[\alpha \frac{\delta C_{45}}{\delta o_{45}} - \alpha \frac{\delta C_{14}}{\delta o_{14}} - \alpha \frac{\delta C_{24}}{\delta o_{24}} - \alpha \frac{\delta C_{34}}{\delta o_{34}} \right] + \\ + \frac{\delta o_5}{\delta \mathbf{w}} \left[-\alpha \frac{\delta C_{45}}{\delta o_{45}} - \alpha \frac{\delta C_{15}}{\delta o_{15}} - \alpha \frac{\delta C_{25}}{\delta o_{25}} - \alpha \frac{\delta C_{35}}{\delta o_{35}} \right]. \end{aligned} \quad (10)$$

Let's introduce lambda:

$$\lambda_{ij} = \frac{\delta C_{ij}}{\delta o_{ij}},$$

$$\lambda_i = \sum_{i:(i,j) \in I_L} \lambda_{ij} - \sum_{j:(i,j) \in I_L} \lambda_{ij} +$$

$$+ \alpha [\sum_{i:(i,j) \in I_U} \lambda_{ij} - \sum_{j:(i,j) \in I_U} \lambda_{ij}],$$
(11)

where I – set of pairs of indices $\{i, j\}$.

So in general:

$$\sum_{i,j \in I} \frac{\delta C_{ij}}{\delta o_{ij}} = \sum_i \frac{\delta o_i}{\delta \mathbf{w}} \lambda_i.$$
(12)

Only need to add an NDCG component to the gradient. The idea is quite straightforward: if the change in NDCG by swapping i and j is large, we expect the gradient to be large as well to make this happen in the hope of maximizing NDCG. And indeed, the paper showed that this works empirically:

$$\frac{\delta C_{\text{LambdaRank}, ij}}{\delta \mathbf{w}} = \frac{\delta C_{ij}}{\delta \mathbf{w}} |\Delta \text{NDCG}_{ij}|,$$
(13)

where $|\Delta \text{NDCG}_{ij}|$ – change in NDCG score if document i swapped with j .

$$|\Delta \text{NDCG}_{ij}| =$$

$$= \frac{1}{iDCG} | (\frac{2^{o_i} - 1}{\log(1+r_i)} + (\frac{2^{o_j} - 1}{\log(1+r_j)} + \dots) -$$

$$- (\frac{2^{o_j} - 1}{\log(1+r_j)} + (\frac{2^{o_i} - 1}{\log(1+r_i)} + \dots)) | =$$

$$= | \frac{1}{iDCG} (2^{o_i} - 2^{o_j}) (\frac{1}{\log(1+r_i)} - \frac{1}{\log(1+r_j)}) |,$$
(14)

where s_i – true label of document i ;

r_i – rank position of document i returned by the model [12].

Experiments. For the experiment was used LightGBM – an open-source gradient boosting framework that uses tree-based learning algorithm. It is based on decision tree algorithms and used for ranking, classification, and other machine learning tasks. The development focus is on performance and scalability. The

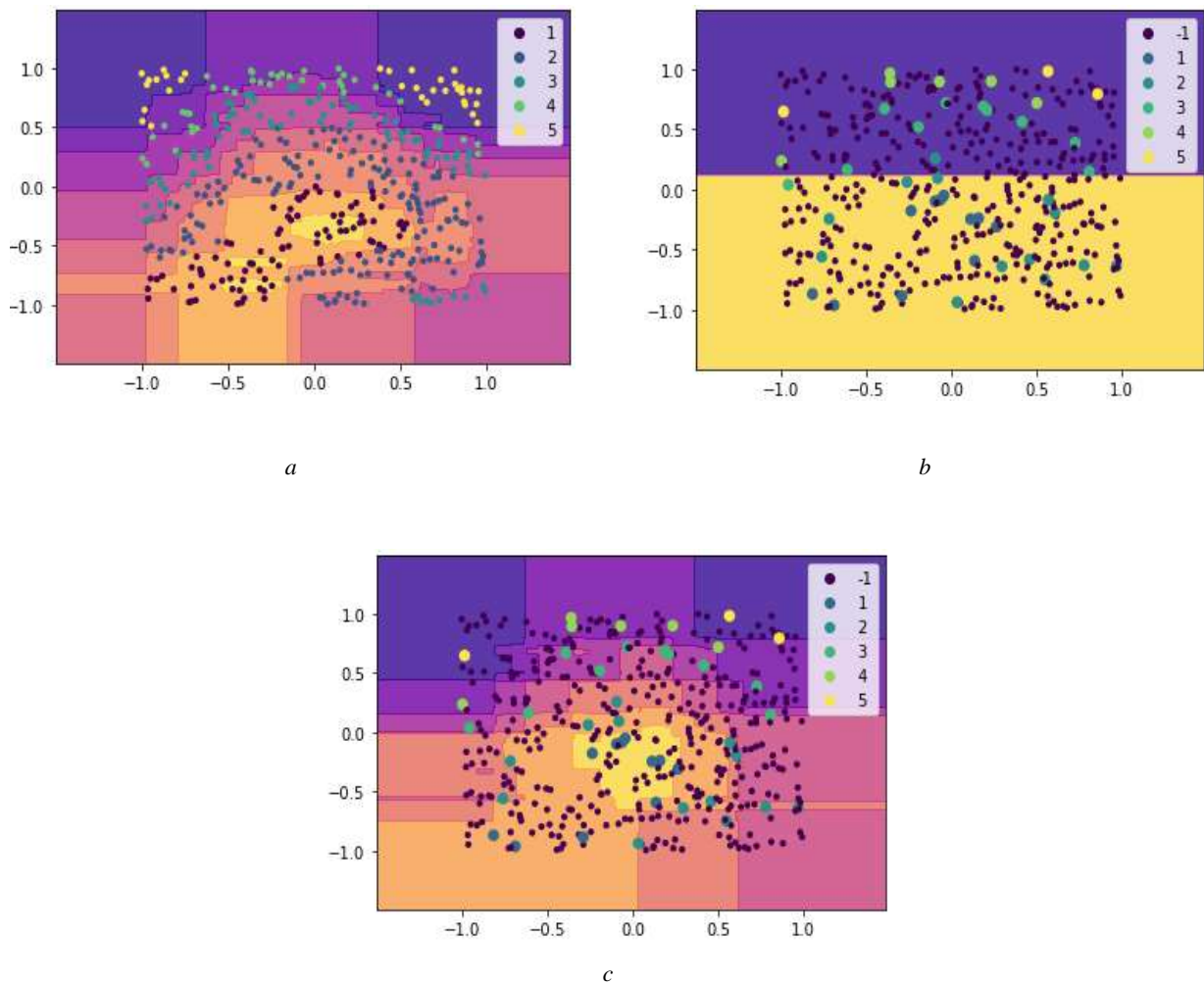


Fig. 3. Ranking methods comparison: *a* – supervised on full dataset, *b* – supervised using only available part of dataset, *c* – semi-supervised method. Here: -1 – no label, 1-5 – ranks

framework has an implementation of a pairwise ranking approach and allows the use of custom learning objections for ranking.

First of all, we've generated a dataset with 2D feature space and multiple queries. These allow us to visualize results and understand the models. The fig. 3 compares supervised and semi-supervised methods.

Fig. 3, *c* shows that a semi-supervised approach with a small number of labels can build decision function almost the same as the supervised with all labels. For pseudo labeling was used the k -NN method with the number of neighbors equal to 3. With synthetic data, the semi-supervised clustering method works also well, but for concordance with real-world, datasets were chosen exactly the k -NN approach.

A dataset was used Microsoft Learning to Rank Datasets (MSLR-WEB10K) and C14 – Yahoo! Learning to Rank Challenge (Yahoo LTRC). Both datasets consist of feature vectors extracted from query – URL pairs along with relevance judgment labels from 0 to 4.

As we can see in tab. 1 the datasets are imbalanced: the higher the rank, the fewer documents are found in the dataset.

So to make a semi-supervised dataset we removed labels of 90 % documents per query. Documents for label removal were chosen by uniform distribution and this aggravated the disbalance problem. Because in some queries there were only a couple of relevant labels and after labeling these queries lost their relevant document with labels at all. Let's say in this way we generate the worst case.

As a metric NDCG was chosen. The metric works with multiple relevance levels and always in range [0, 1]. For better measuring, we take NDCG on different levels 1, 3, and 10.

First, we measured baseline on full (upper) and only semi-supervised (lower) editions for both datasets (tab. 1). As we can see in the table below the additional 90 % of labeled data can increase metric only by 0.05.

Table 1 – Dataset description

	MSLR-WEB10K	Yahoo LTRC
Number of train docs	723412	473134
Number of train queries	6000	19944
Number of docs per query mean	120.57	23.72
Number of docs per query std	70.15	9.00
Number of docs per query mode	110.00	9.00
Ranks distribution	0 – 0.52 1 – 0.32 2 – 0.13 3 – 0.02 4 – 0.01	0 – 0.25 1 – 0.36 2 – 0.29 3 – 0.08 4 – 0.02

The tab. 2 and tab. 3 show that the proposed semi-supervised method can improve accuracy. Also, we can see that improvement grows with relevance level (an improvement on level 1 higher than on level 10).

For pseudo labeling was used the k -NN method with the number of neighbors equal to 1. This is caused by the

method's sensitivity to labeling quality. So datasets complex inner structure has a negative influence on pseudo labeling quality with a high number of neighbors or semi-supervised clustering approaches. And as a result, it's bringing the model accuracy decreasing.

Table 2 – MSLR-WEB10K results

	NDCG@1	NDCG@3	NDCG@10
Lower baseline	0.5169	0.5089	0.5163
Upper baseline	0.5621	0.5440	0.5416
Proposed method	0.5290	0.5118	0.5184

Table 3 – Yahoo LTRC results

	NDCG@1	NDCG@3	NDCG@10
Lower baseline	0.7187	0.7206	0.7738
Upper baseline	0.7583	0.7545	0.7965
Proposed method	0.7310	0.7286	0.7778

Conclusion. The given research described a semi-supervised adaptation of the LambdaMART algorithm. The proposed semi-supervised method is based on the regularization and pseudo labeling approach and improves the ranking accuracy on datasets with a small number of labeled documents. It especially improves accuracy on a high relevance level. But the proposed method is very sensitive to the quality of pseudo labeling. Also, the task is complicated by unbalancing the target datasets. Because of that, we should choose labeling algorithms very carefully.

References

- Burges C. J. C., Svore K. M., Wu Q., Gao J. *Ranking, boosting and model adaptation*. URL: <https://www.microsoft.com/en-us/research/publication/ranking-boosting-and-model-adaptation/> (accessed 07.04.2023).
- Chang Y., Chapelle O. *Yahoo! Learning to Rank Challenge Overview. JMLR: Workshop and Conference Proceedings 14*. 2011. P. 1–24.
- Xu H., Li H. *AdaRank: A Boosting Algorithm for Information Retrieval. Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2007. P. 391–398.
- Yilmaz E., Szummer M. *Semi-supervised Learning to Rank with Preference Regularization. Proceedings of the 20th ACM International Conference on Information and Knowledge Management*. 2011. P. 269–278.
- Burges C. J. C. *From RankNet to LambdaMART to LambdaMART: An Overview*. URL: <https://www.microsoft.com/en-us/research/publication/from-ranknet-to-lambdarank-to-lambdamart-an-overview/> (accessed 07.04.2023).
- Grira N., Crucianu M., Boujema N. *Unsupervised and Semi-supervised Clustering: a Brief Survey*. URL: <http://cedric.cnam.fr/~cruciann/src/BriefSurveyClustering.pdf> (accessed 07.04.2023)
- Vapnik V. N. *Statistical Learning Theory*. New York: Wiley, 1998. 768 p.
- Rahangdale A. U., Raut, S. *Clustering Based Transductive Semi-supervised Learning for Learning-to-Rank. International Journal of Pattern Recognition and Artificial Intelligence*. 2019. Vol. 33, no. 12. P. 1951007:1–1951007:27. DOI: 10.1142/s0218001419510078.
- Amini M., Truong T., Goutte C. *A Boosting Algorithm for Learning Bipartite Ranking Functions with Partially Labeled Data. Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2008*. 2008. P. 99–106.

10. Szummer M., Yilmaz E. Semi-supervised Learning to Rank with Preference Regularization. *Proceedings of the 20th ACM International Conference on Information and Knowledge Management*. 2011. P. 269–278.
11. Weston J., Leslie C., Ie E., Zhou D., Elisseeff A., Noble W. S. Semi-supervised protein classification using cluster kernels. *Bioinformatics*. 2005. Vol. 21, no. 15. P. 3241–3247.
12. Valizadegan H., Jin R., Zhang R., and Mao J. Learning to Rank by Optimizing NDCG Measure. *Advances in Neural Information Processing Systems 22: 23rd Annual Conference on Neural Information Processing Systems 2009*. 2009. P. 1883–1891.

References (transliterated)

1. Burges C. J. C., Svore K. M., Wu Q., Gao J. *Ranking, boosting and model adaptation*. Available at: <https://www.microsoft.com/en-us/research/publication/ranking-boosting-and-model-adaptation/> (accessed 07.04.2023).
2. Chang Y., Chapelle O. Yahoo! Learning to Rank Challenge Overview. *JMLR: Workshop and Conference Proceedings 14*. 2011, pp. 1–24.
3. Xu H., Li H. AdaRank: A Boosting Algorithm for Information Retrieval. *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2007, pp. 391–398.
4. Yilmaz E., Szummer M. Semi-supervised Learning to Rank with Preference Regularization. *Proceedings of the 20th ACM International Conference on Information and Knowledge Management*. 2011, pp. 269–278.
5. Burges C. J. C. *From RankNet to LambdaMART to LambdaMART: An Overview*. Available at: <https://www.microsoft.com/en-us/research/publication/from-ranknet-to-lambdamart-to-lambdamart-an-overview/> (accessed 07.04.2023).

6. Grira N., Crucianu M., Boujemaa N. *Unsupervised and Semi-supervised Clustering: a Brief Survey*. Available at: <http://cedric.cnam.fr/~crucianu/src/BriefSurveyClustering.pdf> (accessed 07.04.2023)
7. Vapnik V. N. *Statistical Learning Theory*. New York, Wiley, 1998. 768 p.
8. Rahangdale A. U., Raut, S. Clustering Based Transductive Semi-supervised Learning for Learning-to-Rank. *International Journal of Pattern Recognition and Artificial Intelligence*. 2019, vol. 33, no. 12, pp. 1951007:1–1951007:27. DOI: 10.1142/s0218001419510078.
9. Amini M., Truong T., Goutte C. A Boosting Algorithm for Learning Bipartite Ranking Functions with Partially Labeled Data. *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2008*. 2008, pp. 99–106.
10. Szummer M., Yilmaz E. Semi-supervised Learning to Rank with Preference Regularization. *Proceedings of the 20th ACM International Conference on Information and Knowledge Management*. 2011, pp. 269–278.
11. Weston J., Leslie C., Ie E., Zhou D., Elisseeff A., Noble W. S. Semi-supervised protein classification using cluster kernels. *Bioinformatics*. 2005, vol. 21, no. 15, pp. 3241–3247.
12. Valizadegan H., Jin R., Zhang R., and Mao J. Learning to Rank by Optimizing NDCG Measure. *Advances in Neural Information Processing Systems 22: 23rd Annual Conference on Neural Information Processing Systems 2009*. 2009, pp. 1883–1891.

Received 11.04.2023

УДК 519.2

К. С. ЯМКОВИЙ, Національний технічний університет «Харківський політехнічний інститут», асистент кафедри комп'ютерної математики і аналізу даних, м. Харків, Україна, e-mail: klym.yamkovyi@cs.khpi.edu.ua; ORCID: <https://orcid.org/0000-0001-9512-4150>

АДАПТАЦІЯ МОДЕЛІ LAMBDA MART ДО НАПІВКОНТРОЛЬОВАНОГО НАВЧАННЯ

Проблема пошуку інформації дуже поширена в епоху Інтернету та великих даних. Зазвичай існують величезні колекції документів, і лише кілька відсотків з них є актуальними. У цьому налаштуванні методи перебору неефективні. Пошукові системи допомагають вирішити цю проблему оптимальним способом. Більшість пошукових двигунів засновані на методах навчання ранжируванню, тобто спочатку алгоритм видає оцінки документам на основі їх ознак, а потім сортує їх відповідно до цих оцінок у відповідному порядку. Існує багато алгоритмів у цій галузі, але одним із найшвидших і надійних алгоритмів ранжирування є LambdaMART. Цей алгоритм заснований на бустингу та розроблений лише для навчання з вчителем, де кожен документ у колекції має ранг, оцінений експертом. Але зазвичай у цій сфері колекції містять масу документів, і їх анотація вимагає багато ресурсів, як-от часу, грошей, експертів тощо. У цьому випадку напіваавтоматичне навчання є потужним підходом. Напіваавтоматичне навчання – це підхід у машинному навчанні, який поєднує невелику кількість позначених даних із великою кількістю не позначених даних під час навчання. Дані без міток у поєднанні з невеликою кількістю мічених даних можуть значно підвищити точність навчання. Ця стаття присвячена адаптації LambdaMART до напіваавтоматичного навчання. Автор пропонує додавати різні ваги для розмічених і нерозмічених документів під час процедури навчання, щоб досягти більшої надійності і точності. Запропонований алгоритм було реалізовано з використанням мови програмування Python та фреймворку lightGBM, який уже має реалізацію LambdaMART для навчання з вчителем. Для цілей тестування було використано кілька наборів даних. Один синтетичний 2D-набір даних для візуального пояснення результатів і два реальних набори даних MSLR-WEB10K від Microsoft і Yahoo LTRC.

Ключові слова: навчання ранжируванню, пошук інформації, напіваавтоматичне навчання, парне ранжирування, LambdaMART, псевдомаркування, NDCG.

Повне ім'я автора / Author's full name

Ямковий Клим Сергійович, Yamkovyi Klym Serhiyovych

O. L. SOLOVEI, Candidate of Technical Sciences (Ph.D.), Kyiv National University of Construction and Architecture, Associate Professor at the Department of Information Technology for Design and Applied Mathematics, Kyiv, Ukraine, e-mail: solovey.ol@knuba.edu.ua; ORCID: <https://orcid.org/0000-0001-8774-7243>

D. A. GOBOV, Candidate of Technical Sciences (Ph.D.), National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Associate Professor at the Department of Computer Science and Software Engineering of the Faculty of Informatics and Computer Science, Kyiv, Ukraine, e-mail: d.gobov@kpi.ua; ORCID: <https://orcid.org/0000-0001-9964-0339>

PARSIMONIOUS MACHINE LEARNING MODELS IN REQUIREMENTS ELICITATION TECHNIQUES SELECTION

The subject of research in the article is machine learning algorithms used for requirement elicitation technique selection. The goal of the work is to build effective parsimonious machine learning models to predict the using particular elicitation techniques in IT projects that allow using as few predictor variables as possible without a significant deterioration in the prediction quality. The following tasks are solved in the article: design an algorithm to build parsimonious machine learning candidate models for requirement elicitation technique selection based on gathered information on practitioners' experience, assess parsimonious machine learning model accuracy, and design an algorithm for the best candidate model selection. The following methods are used: algorithm theory, statistics theory, sampling techniques, data modeling theory, and science experiments. The following results were obtained: 1) parsimonious machine learning candidate models were built for the requirement elicitation technique selection. They included less number of features that helps in the future to avoid overfitting problems associated with the best-fit models; 2) according to the proposed algorithm for best candidate selection – a single parsimonious model with satisfied performance was chosen. Conclusion: An algorithm is proposed to build parsimonious candidate models for requirement elicitation technique selection that avoids the overfitting problem. The algorithm for the best candidate model selection identifies when a parsimonious model's performance is degraded and decides on the suitable model's selection. Both proposed algorithms were successfully tested with four datasets and can be proposed for their extensions to others.

Keywords: requirements elicitation techniques, Bayesian Information Criterion, Bayes factor grades, log-likelihood, parsimonious model.

Introduction. Business analysis as an extension of requirements engineering is crucial to software development. The main business analysis deliverables are requirements and designs used as a basis for solution implementation, testing, and deployment. In turn, the critical input for the tasks of analysis, specification, and modeling of requirements and design for software is the information collected during the elicitation. Standard approaches to the requirements-gathering process have been systematized and described in the form of dozens of standard elicitation techniques. Industrial guidelines and empirical studies contain detailed descriptions of the techniques' elements and usage considerations but do not provide an elicitation selection process [1].

Consequently, one of the challenges for business analysts/requirement engineers, especially novice ones, is the selection of the appropriate requirements elicitation techniques that best fit their project. As a result, some of them are misused, others are never used, and only a few are constantly applied. To solve the problem, a machine learning model to predict/recommend using the following elicitation techniques as Interviews, Document Analysis, Process Analysis, and Interface Analysis depending on the project's context was proposed [1].

In the study [2], the model's prediction accuracy was increased by transforming the dataset from imbalanced to balanced, thus making a Random Forest Classifier learner unbiased to the majority class. Feature importance score was identified by mutual information criteria, i.e., independent from the machine learner classifier. It served as an assurance that the feature's score doesn't depend on the learning algorithm's bias. Ten features with the most

significant importance score were reported in tables 4–5 as predictors for choosing the elicitation technique.

However, in both [1] and [2], selecting the best model from the candidates remained based on the performance metrics such as Accuracy and AUC.

A model selected that way is also called a "best-fit" model. The "best fit" model is complex – it includes many parameters in order to better approximate training data. The more variables included in a model, the more dependent the model becomes on the observed data so that it can fail on the test data due to noisy, uninformative, and unrepresentative data being included in the model. i.e., a "best-fit" model is prone to overfit data [3].

Although the "best-fit" models included twenty features, we took ten features with the most significant importance score, which potentially may be incorrect if the model due to include less than ten features.

To eliminate the mentioned problems for the model proposed in works [1–2] in the current study, we will develop a parsimonious model that still accurately predicts/recommends using the techniques: Interviews, Document Analysis, Process Analysis, and Interface Analysis.

Analysis of last achievements and publications. The principle of parsimony suggests a model should be as simple as possible concerning the included variables, model structure, and several parameters. It is a desired characteristic of a model defined by a suitable trade-off between squared bias and variance of parameter estimators [4]. The construction of the parsimonious model happens in the following steps:

© Solovoei O. L., Gobov D. A., 2023



Research Article: This article was published by the publishing house of NTU "KhPI" in the collection "Bulletin of the National Technical University "KhPI" Series: System analysis, management and information technologies." This article is distributed under a Creative Common [Creative Common Attribution \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/). **Conflict of Interest:** The author/s declared no conflict of interest.



- build multiple candidate models from the same dataset but include a different number of features;
- compare and select the best candidate as a final parsimonious model;
- assess the fit of the selected candidate.

So to develop a parsimonious model, it is required to choose the methods:

- for selecting the variables for the model;
- for comparison of the candidate models;
- for assessing the fit.
- There are three commonly used methods for selecting the variables [5–7]:
- purposeful selection;
- stepwise selection;
- best subsets.

A purposeful selection – initially, a multivariable model is built with variables which are having a significant univariate test. Then, variables skipped on the 1st step are added one by one to the model to identify the variables that are not significant by themselves but contribute to the presence of other variables. The process of deleting, refitting, and verifying continues until it appears that all of the essential variables are included in the model, and those excluded are statistically unimportant. The major problem of purposeful selection is that a final model may "overfit".

A stepwise selection is based on a statistical algorithm that checks for the "importance" of variables and either includes or excludes them based on a fixed decision rule. The "importance" of a variable is defined in terms of a measure of the statistical significance of the coefficient(s) for the variable. The statistic used for linear regression is an F -test; for logistic regression – likelihood ratio, score, and Wald test.

A "best subsets" are the number of models containing one, two, three variables, and so on, which are fitted to determine the "best" according to specified criteria.

Due to meeting the current research's goals, only the "best subsets" approach from the listed able can be applied. The statistical measure that is commonly used to compare models with different numbers of parameters based on the parsimonious principle is the Akaike Information Criterion (AIC). It measures the distance between a candidate model and the accurate model – the closer the distance, the more similar the candidate is to the truth model. AIC calculates the distance between models as expected relative to Kullback – Leibler (K–L) divergence. Although AIC is a consistent estimator of K–L divergence, there is no statistical test to compare values of AIC [8–9].

Another criterion to compare candidate models is Bayesian Information Criterion (BIC), derived from Bayesian statistical analysis and estimates. BIC approximates a Bayes factor with desirable properties for hypothesis testing and model selection [10–13]. BIC is calculated for each candidate model by equation (1)

$$\text{BIC}_d = -2 \log L_d + p \log n, \quad (1)$$

where L_d is a maximized log-likelihood; p – the number of estimable parameters included in the model and n – the number of observations in the dataset; $d = \overline{1, D}$ where D

is the number of candidate models. To calculate BIC of binary classification models with predictors matrix $X \in R^{n \times p}$ and target $y \in R^n$ which takes values in the set $\{0, 1\}$ and which is built with learner algorithms: logistic regression, support vector machine (SVC), or decision tree classifiers (RandomForestClassifier), a maximized log-likelihood from (1) is calculated as a logistic loss function:

$$L_d(y, \hat{p}) = \sum_{i=1}^n -(y_i \log(\hat{p}_i) + (1 - y_i) \log(1 - \hat{p}_i)), \quad (2)$$

where \hat{p}_i is a probability with which a fit model predicts a positive class

$$\hat{p}_i = \frac{1}{1 + \exp(-X_{ij} \bar{\beta}_j - \beta_{j_0})}, \quad (3)$$

where $i = 1, \dots, n$, $j = 1, \dots, p$, $\bar{\beta}_j$, β_{j_0} – coefficients of the model.

The final look of equation (1) is specified in (4)

$$\text{BIC}_d = -2 \log L_d(y, \hat{p}) + p \log n \quad (4)$$

The model with the smallest BIC is preferable because the complex models are almost always likely to fit the data better, so the first term in definition (4) will have a low value; however, the second provides a way to penalize these extra parameters, therefore causes BIC is increasing. Candidate models' comparison by BIC_d is done by BIC weights w_d calculated by equation

$$w_d = \frac{\exp\left(-\frac{1}{2} \Delta_d\right)}{\sum_{r=1}^R \exp\left(-\frac{1}{2} \Delta_d\right)}, \quad (5)$$

where Δ_d is a difference between d -th candidate model with calculated BIC_d and the minimum value of BIC. BIC weights are the probability that d -th candidate model is the best among the candidate models. BIC weights use Bayes factor grades to evaluate a candidate model according to the rules:

- If $w_d \leq 0.5$, then this is a weak candidate model;
- If $0.5 < w_d < 0.75$, then this is a positive candidate model;
- If $0.75 \leq w_d < 0.95$, then this is a strong candidate model;
- If $0.95 \leq w_d < 1$, then this is very strong.

To assess the goodness of fit of the selected candidate models compared to etalon (or best-fit) models in works [14] is proposed to apply testing of the hypothesis based on a difference between sample means of the model's performance metric. When the mean accuracy of the selected parsimonious models is \bar{A}_1 and the mean accuracy of best-fit models is \bar{A}_2 then the parsimonious models fit if the null hypothesis is not rejected by the computed two-tailed p -value of the t -statistic (eq. 6).

$$t = \frac{\bar{A}_1 - \bar{A}_2}{\sqrt{\frac{1}{n} \sum_{i=1}^n (A_{1i} - A_{2i})^2 - \frac{\left(\sum_{i=1}^n (A_{1i} - A_{2i}) \right)^2}{n}}} \quad (6)$$

where n is the number of the parsimonious models included in the test; $ddof$ is the delta degree of freedom with a value equal to 1. Other classification metrics, such as AUC, f1 score, precision, recall, and Jaccard score, can be used to measure the goodness of the parsimonious model in the same manner as specified in equation 6 for the accuracy metric.

The problem statement. We aim to build parsimonious models for four datasets considered in works [1–2] to avoid overfitting problems associated with the best-fit models. To design an algorithm for assessing a parsimonious model's performance compared to the best-fit model and selecting the best candidate. To execute tests to prove that the proposed algorithms can be used with other datasets.

Experiment Methodology. Our experiment methodology for constructing and assessing the parsimonious model is specified per each phase of the supervised learner model's creation lifecycle [15].

Data preparation and acquisition. Original data was formed based on a survey conducted among business analysts and requirement engineers in Ukraine regarding their use of requirement elicitation techniques and their context. Three hundred twenty-eight practitioners completed the survey. Four respondents were disqualified due to incorrect data: non-filled industrial sector and non-filled team types. The features included in the dataset used in this study are two types:

- features to describe the project's context;
- features to list all elicitation techniques used in the project.

The following features belong to the first type:

- country;
- project size: small – till 15 people, mid-size – 15–30 people, big – 30–100 people, and very big – more than 100 people;
- industrial sector;
- company business model: IT company+outsourcing, IT company+outstaffing, IT company+product, Non-IT company;
- company size: small – till 80 employees, mid-size – 81–200 employees, upper-mid size – 201–800

employees, big – 801–1500; very big – over 1500 employees;

- Software type (current project): business, embedded, scientific, system, other;
- Team types: collocated, distributed;
- Experience in business analysis/requirement engineering: up to 3 years; 3–5 years; 5–10 years; over ten years;
- Methodology: Agile, Hybrid, plan-driven (e.g., Waterfall)
- Project Category: development from scratch, user interface engineering, solution reengineering (re-design and reimplementation), solution customization;
- business analysis activities in which the respondent is usually involved in: Business analysis planning & monitoring, Strategy analysis, Elicitation & Collaboration, Requirements analysis and design definition, Solution evaluation, and Requirements life cycle management;
- certification.

The following features belong to the second type:

- benchmarking and market analysis;
- brainstorming;
- business rules analysis;
- collaborative games;
- data mining;
- design thinking / lean startup;
- document analysis;
- interface analysis;
- interviews;
- observations;
- process analysis/process modeling;
- prototyping;
- reuse database and guidelines;
- stakeholders list, map, or personas;
- survey or questionnaire;
- workshops and focus groups.

The dataset contains information about the features, along with the names of target classes such as "Elicitation", "Document Analysis", "Interface Analysis", and "Process Analysis". However, a feature with the same name as a target class is not included in the list of features. Databases' characteristics and imbalanced ratios calculated as majority-to-minority samples are specified in table 1.

Data preprocessing. The imbalance predictors matrix X and a target vector y were transformed into balanced X^* , y^* by applying SMOTE method. This method allows us to

Table 1 – The characteristics of datasets

Target class name	Majority class	Minority class	Imbalance ratio	Machine learning task	Feature type	Missing values, Y/N?
Interviews	282	41	6.9	Binary classification	Discrete	N
Document Analysis	276	47	5.9			
Interface Analysis	232	91	2.5			
Process Analysis	213	110	1.9			

generate synthetic samples that diminish the class imbalance problem. SMOTE gets a new sample from two samples from the minority class (x^R and x) with $0 \leq k \leq 1$; x^R is randomly chosen among the 5-minority class nearest neighbors of x .

Modeling. The steps undertaken are described as per pseudocode (fig. 1).

```

Input:  $X^*$ ,  $y^*$ , Acc_min, AUC_min,  $F$ 
Output:  $\vec{S}$ 
 $\vec{S} = []$ 
1. FOR  $i$  in range ( $F$ )
2.   select  $i$  features with the biggest MI score
3.   get predictors matrix  $X^{**} \in R^{n \times i}$ 
4.   divide  $X^{**}$  and  $y^*$  on train/test subsets.
5.   create model object and fit with train data.
6.   calculate Acc, AUC on test subset
7.   IF (Acc  $\geq$  Acc_min) && (AUC  $\geq$  AUC_min)
8.     add model object to vector  $\vec{S}$ 
9.   ENDIF
10. ENDFOR

```

Fig. 1. Modeling algorithm

In lines 2–5, candidate models are fitted with increasing by one number of included features. The first candidate model includes one feature, and the last candidate includes F features, where F is the maximum number of features in our datasets. In line 2, i -features are sorted according to their mutual information (MI) score in descending order; the i -features are selected from the start of the sorted list with MI scores. In lines 3–4, the predictors' matrix is truncated to include only selected features, and a target variable and train and test subsets are formed from it. In lines 5–6, a model fits with the training subset, and performance metrics accuracy (Acc) and area under the ROC curve (AUC) are calculated on the test subset. In lines 7–8, if the model object's calculated performance satisfies the minimum performed level of accuracy (Acc_min) and AUC (AUC_min), then the model object is saved in the result vector \vec{S} .

A general guideline is used in supervised machine learning with the following intervals for accuracy and AUC metrics:

- if Accuracy/AUC = 0.5, then this is a guessing equal to flipping a coin;
- if $0.5 < \text{Accuracy/AUC} < 0.7$, then this is poor classification;
- if $0.7 < \text{Accuracy/AUC} < 0.8$, then this is an acceptable classification;
- if $0.8 < \text{Accuracy/AUC} < 0.9$, then this is an excellent classification;
- if Accuracy/AUC ≥ 0.9 , then this is outstanding discrimination.

The above rules are to be used to set minimum values of Accuracy and AUC for the algorithm (fig. 1). If in the result of the execution of the algorithm vector \vec{S} is empty, then we propose to lower the minimum values of the

performance metrics. If vector \vec{S} is not empty, then we can move on to grade candidate models by Bayes factor and grades the steps undertaken are described as per pseudocode (fig. 2).

For each model object from \vec{S} in line 2 is identified a BIC weight, denoted as w_m . Then in lines 3–6, each model is graded according to the Bays factor's rules. "Positive" models are saved to vector \vec{M}_1 . "Strong" – to vector \vec{M}_2 , "Very Strong" – to vector \vec{M}_3 . In current work, we ignored "weak" candidates.

```

Input:  $\vec{S}$ 
Output:  $\vec{M}_1, \vec{M}_2, \vec{M}_3$ 
1. FOR  $m$  in  $\vec{S}$ 
2.   calculate BIC  $w_m$ 
3.   IF  $w_m \geq 0.5$  &&  $w_m < 0.75$ 
4.     add the model object to  $\vec{M}_1$ 
5.   ELSEIF  $w_m \geq 0.75$  &&  $w_m < 0.95$ 
6.     add the model object to  $\vec{M}_2$ 
7.   ELSEIF  $w_m \geq 0.95$ 
8.     add the model object to  $\vec{M}_3$ 
9.   ENDIF
10. ENDFOR

```

Fig. 2. Steps to grade the candidate models

Model validation. The assessment of the goodness of fit of models from $\vec{M}_1, \vec{M}_2, \vec{M}_3$ compared to best-fit models \vec{B} was done through the steps as per pseudocode (fig. 3).

```

Input:  $\vec{M}_1, \vec{M}_2, \vec{M}_3, \vec{B}$ 
Output:  $\vec{R}, M$ 
 $\vec{R} = []$ 
1. calculate  $\bar{A}_2$  for models from  $\vec{B}$ .
2. calculate  $\bar{A}_1$  for models from  $\vec{M}_3$ .
3. calculate  $t$ -statistic.
4. compute the two-tailed  $p$ -value of the normal distribution.
5. IF  $p\text{-value} > 0.05$ 
6.   add the model object to  $\vec{R}$ 
7. ENDIF
8. IF  $\vec{R}$  is empty
9.   repeat steps 2–7 with models from  $\vec{M}_2$  and  $\vec{M}_1$ 
10. ENDIF
11. IF  $\vec{R}$  is empty
12.   select a model  $M$  with the best performance from  $\vec{M}_1, \vec{M}_2, \vec{M}_3$  or model from  $\vec{B}$ 
13. ENDIF

```

Fig. 3. Steps to assess goodness of fit of parsimonious models

In lines [1–4], mean values, the t -statistic, and the two-tailed p -value of the normal distribution for "very strong" models are computed. In lines 5–7, if the null hypothesis is not rejected, then a parsimonious model is added to the result vector \vec{R} . Lines 8–9 are executed if the goodness of fit test is failed for models from \vec{M}_3 . In this case, steps 2–7 are repeated with "strong" and "positive" models. Lines 11–14 are executed only if all models from $\vec{M}_1, \vec{M}_2, \vec{M}_3$ failed goodness of fit test. In that scenario, the model M with the best performance is selected from $\vec{M}_1, \vec{M}_2, \vec{M}_3$. The algorithm (fig. 3.) leaves experts to finally judge which model to use if all parsimonious candidate models failed the assessment. It could be either best-fit models from \vec{B} or the parsimonious model with the best performance metrics because their minimum values are set as an input parameter of the algorithm (fig. 2).

Study results and their discussion. Multiple candidate models are created according to designed algorithm (fig. 1). Applied Bayes factor grades as specified

in fig. 2 allowed to select: a "very strong" parsimonious model to recommend Interviews as an elicitation technique that included eight features and evaluated with performance Accuracy = 90 %; AUC = 98 % (fig. 4, a) which are 4 % and 1 % lower than Accuracy and AUC of best-fit model (table 2 – "Interviews"); a "very strong" parsimonious model to recommend Document analysis as an elicitation technique that included five features and evaluated with performance Accuracy = 90 %; AUC = 95 % (fig. 4, b) which are 1 % and 2 % lower than Accuracy and AUC of best-fit model (table 2 – "Document Analysis").

A "strong" parsimonious model to recommend Interface analysis as an elicitation technique that included nine features and evaluated with performance Accuracy = 81 %; AUC = 88 % (fig. 5, a) which are 3 % and 2 % lower than Accuracy and AUC of best-fit model (table 2 – "Interface Analysis"); a "strong" parsimonious model to recommend Process analysis as an elicitation technique that included fifteen features and evaluated with performance Accuracy = 81 %; AUC = 86 % (fig. 5, b) which are 1 %

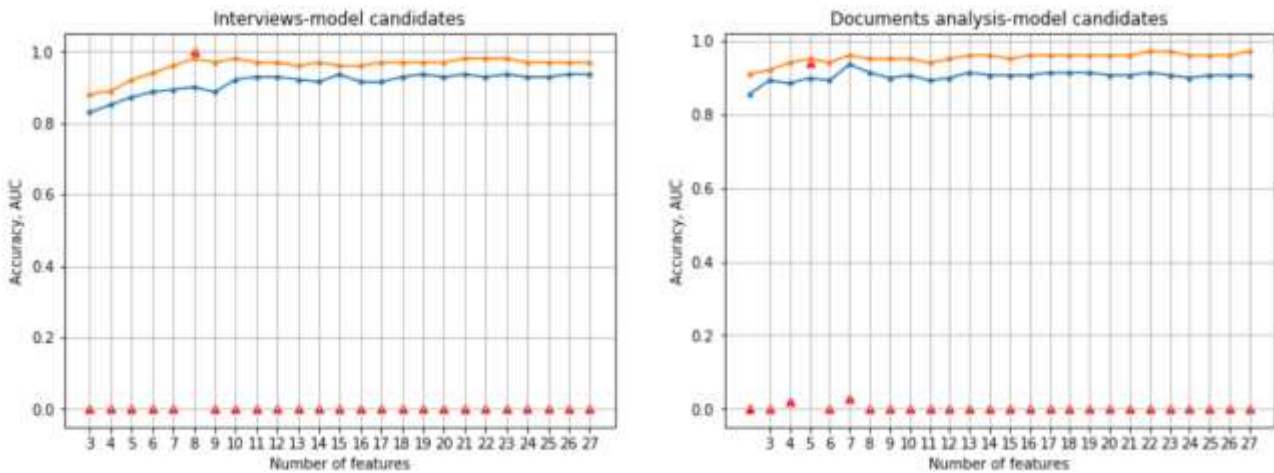


Fig. 4. Candidate model(s) BIC weight, Accuracy, AUC:
a – Interviews; b – Document analysis

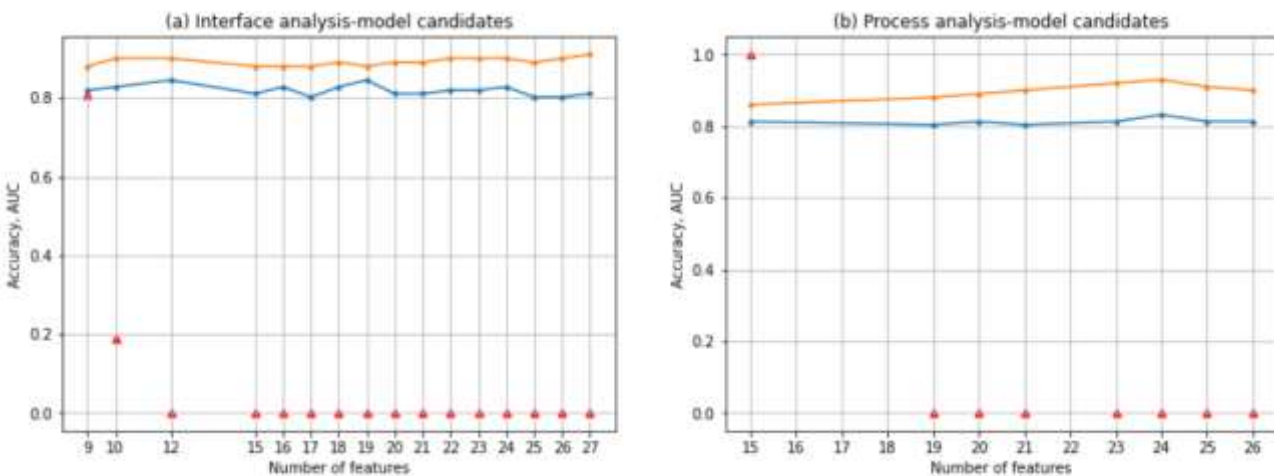


Fig. 5. Candidate model(s) BIC weight, Accuracy, AUC:
a – Interface analysis; b – Process analysis

Table 2 – Performance of parsimonious and best-fit models

Techniques	Accuracy			AUC		
	Parsimonious	Best-fit	Difference	Parsimonious	Best-fit	Difference
Interviews	0.90	0.94	–4 %	0.96	0.97	–1 %
Document Analysis	0.90	0.91	–1 %	0.95	0.97	–2 %
Interface Analysis	0.82	0.84	–3 %	0.88	0.90	–2 %
Process Analysis	0.81	0.82	–1 %	0.86	0.88	–2 %

and 2 % lower than Accuracy and AUC of best-fit model (table 2 – "Process Analysis").

As specified in fig. 3, the hypothesis test is applied with the models' performance metrics from table 2. A null hypothesis H_0 : the mean difference between the parsimonious and best-fit models' accuracies is 0. An H_1 hypothesis: the difference between the accuracies is different. T -statistic per equation 7 gives $t = -2.8$. The p -value with the degree of freedom equal to 3 is 0.066, which is greater than 0.05, so our H_0 is accepted, i.e., the parsimonious models are accepted, and best-fit can be ignored. Similarly, the hypothesis test with a null hypothesis H_0 : the mean difference between the values of AUC of the parsimonious model and the values of AUC of best-fit models is 0. H_1 hypothesis: the mean difference between AUC values is different. T -statistic per equation 7 gives $t = -7$. The p -value of $t = -7$ with the degree of freedom equal to 3 is 0.006, which is less than 0.05, so our H_0 is rejected, and the best-fit model is preferable due to the reduced parsimonious model's performance based on the mean value of AUC.

Thus, it can be concluded that applying the algorithm as per fig. 3 with each performance metric in sequence helps to identify when a parsimonious model's performance is degraded and decide on the suitable model's selection. We accepted the built parsimonious models in the current test experiment because the model's accuracy didn't deteriorate based on the goodness of fit test.

Conclusions and perspectives of further development. In the current study, the algorithms to build parsimonious candidate machine learning models and select the best candidate were designed and tested with four datasets collected for requirement elicitation technique selection. The results showed that the best candidate models graded as "very strong" and "strong" reduced the number of features: three times for Interviews and Interface analysis, five times for Document analysis, and 1.7 times for Process analysis. It helped to avoid the overfitting data problem.

The designed algorithm to assess the goodness of fit of the parsimonious models was applied with two performance metrics: accuracy and AUC in sequence. Based on the received results is concluded that by applying the proposed procedure, the gaps in the performance of the parsimonious model compared to the best-fit model can be detected, and a decision on the suitable model's selection can be made.

In summary, the obtained results allow us to recommend using a parsimonious model instead of the best-fit model to predict the using the particular elicitation

technique in IT projects and form recommendations based on that model.

Several directions for future research can be considered, such as creating machine learning models for other business analysis techniques, e.g., specification and modeling, prioritization, and structure of business analysis architecture.

References

- Gobov D., Huchenko, I. Influence of the Software Development Project Context on the Requirements Elicitation Techniques Selection. *Lecture Notes on Data Engineering and Communications Technologies*. Cham: Springer, 2021. Vol 83. P. 208–218. DOI: 10.1007/978-3-030-80472-5_18.
- Gobov D., Solovei O. Approaches to Improving the Accuracy of Machine Learning Models in Requirements Elicitation Techniques Selection. *arXiv:2303.14762*, 2023. DOI: 10.48550/arXiv.2303.14762.
- Harrell F. E. *Regression Modeling Strategies with Applications to Linear Models, Logistic Regression and Survival Analysis*. New York, USA: Springer, 2001. 582 p.
- Anderson D., Burnham K. *Model selection and multi-model inference*. Second ed. New York, USA: Springer-Verlag, 2004. 488 p. DOI: 10.1007/b97636.
- Bursac Z., Gauss H. C., Williams D. K., and Hosmer D. W. Purposeful selection of variables in logistic regression. *Source Code for Biology and Medicine*. 2008. Vol. 3 (17). P. 3–17. DOI: 10.1186/1751-0473-3-17.
- Zhang Z. Variable selection with stepwise and best subset approaches. *Annals of translational medicine*. 2016. Vol. 4(7). P. 136. DOI: <https://doi.org/10.21037/atm.2016.03.35>.
- Solovei O. New organization process of feature selection by filter with correlation-based features selection method. *Innovative Technologies and Scientific Solutions for Industries*. 2022. Vol. 3 (21). P. 39–50. DOI: 10.30837/ITSSI.2022.21.039.
- Vrieze S. Model selection and psychological theory: a discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). *Psychological methods*. 2012. Vol. 17, № 2. P. 228–243. DOI: 10.1037/a0027127.
- Arnold T. Uninformative parameters and model selection using Akaike's Information Criterion. *The Journal of Wildlife Management*. 2010. Vol. 74, № 6. P. 1175–1178. DOI: 10.1111/j.1937-2817.2010.tb01236.x.
- van de Schoot R., Depaoli S., King R., Kramer B., Märtens K., Tadesse M. G., Vannucci M., Gelman A., Veen D., Willemsen J., Yau C. Bayesian statistics and modelling. *Nature Reviews Methods Primers*. 2021. Vol. 1. DOI: 10.1038/s43586-020-00001-2.
- Lesaffre E., Lawson A. *Bayesian Biostatistics*. West Sussex, United Kingdom: John Wiley & Sons, 2012. 544 p. DOI: 10.1002/9781119942412.
- Wasserman L. Bayesian model selection and model averaging. *Journal of mathematical psychology*. 2000. Vol. 44. № 1. P. 92–107. DOI: 10.1006/jmps.1999.1278.
- Rouder J.N., Speckman P.L., Sun D., Morey R. D., Iverson G. Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic bulletin & review*. 2009. Vol.16. P. 225–237. DOI: 10.3758/PBR.16.2.225.
- Hosmer Jr. D., Lemeshov S., Sturdivant R. *Applied logistic regression*, West Sussex, United Kingdom: John Wiley & Sons, 2013. 510 p.

15. Ghoroghi A., Rezgui Y., Petri I., Beach T. Advances in application of machine learning to life cycle assessment: a literature review. *The International Journal of Life Cycle Assessment*. 2022. Vol. 27. P. 433–456. DOI: 10.1007/s11367-022-02030-3.

References (transliterated)

- Gobov D., Huchenko, I. Influence of the Software Development Project Context on the Requirements Elicitation Techniques Selection. *Lecture Notes on Data Engineering and Communications Technologies*. Springer, Cham, 2021, vol 83, pp. 208–218. DOI: 10.1007/978-3-030-80472-5_18.
- Gobov D., Solovei O. Approaches to Improving the Accuracy of Machine Learning Models in Requirements Elicitation Techniques Selection, *arXiv:2303.14762*, 2023. DOI: 10.48550/arXiv.2303.14762.
- Harrell F. E. *Regression Modeling Strategies with Applications to Linear Models, Logistic Regression and Survival Analysis*. New York, USA, Springer, 2001. 582 p.
- Anderson D., Burnham K. Model selection and multi-model inference. Second ed. New York, USA, Springer-Verlag, 2004. 488 p. DOI: 10.1007/b97636.
- Bursac Z., Gauss H. C., Williams D. K., and Hosmer D. W. Purposeful selection of variables in logistic regression. *Source Code for Biology and Medicine*. 2008, vol. 3 (17), pp. 3–17. DOI: 10.1186/1751-0473-3-17.
- Zhang Z. Variable selection with stepwise and best subset approaches. *Annals of translational medicine*. 2016, vol. 4 (7), p. 136. DOI: 10.21037/atm.2016.03.35.
- Solovei O. New organization process of feature selection by filter with correlation-based features selection method. *Innovative Technologies and Scientific Solutions for Industries*. 2022, vol. 3 (21), pp. 39–50. DOI: 10.30837/ITSSI.2022.21.039.
- Vrieze S. Model selection and psychological theory: a discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). *Psychological methods*. 2012, Vol. 17, no. 2, pp. 228–243. DOI:10.1037/a0027127.
- Arnold T. Uninformative parameters and model selection using Akaike's Information Criterion. *The Journal of Wildlife Management*. 2010, vol. 74, no. 6, pp. 1175–1178. DOI: 10.1111/j.1937-2817.2010.tb01236.x.
- van de Schoot R., Depaoli S., King R., Kramer B., Märtens K., Tadesse M. G., Vannucci M., Gelman A., Veen D., Willemsen J., Yau C. Bayesian statistics and modelling. *Nature Reviews Methods Primers*. 2021, 1, vol. 1. DOI: 10.1038/s43586-020-00001-2.
- Lesaffre E., Lawson A. *Bayesian Biostatistics*. West Sussex, United Kingdom, John Wiley & Sons, 2012. 544 p. DOI: 10.1002/9781119942412.
- Wasserman L. Bayesian model selection and model averaging. *Journal of mathematical psychology*. 2000, vol. 44, no. 1, pp. 92–107. DOI: 10.1006/jmps.1999.1278.
- Rouder J.N., Speckman P.L., Sun D., Morey R. D., Iverson G. Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic bulletin & review*. 2009, vol.16, pp. 225–237. DOI: 10.3758/PBR.16.2.225.
- Hosmer Jr. D., Lemeshov S., Sturdivant R. *Applied logistic regression*, West Sussex, United Kingdom, John Wiley & Sons, 2013, 510 p.
- Ghoroghi A., Rezgui Y., Petri I., Beach T. Advances in application of machine learning to life cycle assessment: a literature review. *The International Journal of Life Cycle Assessment*. 2022, vol. 27, pp. 433–456. DOI: 10.1007/s11367-022-02030-3.

Received 15.04.2023

УДК 004.415.25

О. Л. СОЛОВЕЙ, кандидат технічних наук (PhD), Київський Національний університет будівництва і архітектури, доцент кафедри інформаційних технологій проектування та прикладної математики, м. Київ, Україна; e-mail: solovey.ol@knuba.edu.ua; ORCID: <https://orcid.org/0000-0001-8774-7243>

Д. А. ГОБОВ, кандидат технічних наук, Національний технічний університет України "Київський політехнічний інститут імені Ігоря Сікорського", доцент кафедри інформатики та програмної інженерії факультету інформатики та обчислювальної техніки, Київ, Україна; e-mail: d.gobov@kpi.ua; ORCID: <https://orcid.org/0000-0001-9964-0339>

ЕКОНОМНІ МОДЕЛІ МАШИННОГО НАВЧАННЯ ДЛЯ ВИБОРУ ТЕХНІК ВІЯВЛЕННЯ ВИМОГ

Предметом дослідження в статті є алгоритми машинного навчання, що використовуються для вибору технік виявлення вимог. Метою роботи є побудова ефективних економних моделей машинного навчання для прогнозування використання методів виявлення вимог в IT-проектах, які дозволяють використовувати якомога менше незалежних змінних без значного погіршення якості прогнозу. У статті вирішуються наступні завдання: розробка алгоритму побудови економних моделей-кандидатів машинного навчання для вибору технік виявлення вимог на основі зібраної інформації про досвід практикуючих фахівців, оцінка точності моделі економного машинного навчання та розробка алгоритму вибору найкращої моделі-кандидата. Використовуються такі методи: теорія алгоритмів, теорія статистики, методи вибірки, теорія моделювання даних та наукові експерименти. Було отримано наступні результати: 1) для вибору технік виявлення вимог побудовано економні моделі-кандидати машинного навчання. Вони включали менше параметрів, що допомагає у майбутньому уникнути проблем із перенавчанням, пов'язаних із найкращими моделями; 2) відповідно до запропонованого алгоритму для відбору найкращого кандидата була обрана одна економна модель із задовільною продуктивністю. Висновок. Запропоновано алгоритм для побудови ошадливих моделей-кандидатів для вибору технік виявлення вимог, які дозволяють уникнути проблеми перенавчання. Алгоритм вибору найкращої моделі-кандидата визначає, коли продуктивність економної моделі погіршується, і приймає рішення щодо вибору відповідної моделі. Обидва запропоновані алгоритми були успішно протестовані з чотирма наборами даних і можуть бути запропоновані для їх розширення для інших.

Ключові слова: техніки виявлення вимог, байєсівський інформаційний критерій (BIC), фактор Байєса, довірчий інтервал, економна модель, точність, площа під кривою ROC.

Повні імена авторів / Author's full names

Автор 1 / Author 1: Соловей Ольга Леонідівна, Solovei Olga Leonidivna

Автор 2 / Author 2: Гобов Денис Андрійович, Gobov Denys Andriyovych

K. ONYSHCHENKO, Senior Lecturer at the Department of Software engineering, Kharkiv National university of Radio electronics, Kharkiv, Ukraine, e-mail: kostiantyn.onyshchenko@nure.ua, ORCID: <https://orcid.org/0000-0002-7746-4570>

Y. DANIEL, Assistant at the Department of Software engineering, Kharkiv National university of Radio electronics, Kharkiv, Ukraine, e-mail: yana.daniel@nure.ua, ORCID: <https://orcid.org/0000-0002-3895-0744>

USING LONG SHORT-TERM MEMORY NETWORKS FOR NATURAL LANGUAGE PROCESSING

The problem of emotion classification is a complex and non-trivial task of language interpretation due to the natural language structure and its dynamic nature. The significance of the study is in covering the important issue of automatic processing of client feedbacks, collecting opinions and trend-catching. In this work, a number of existing solutions for emotion classification problem were considered, having their shortcomings and advantages illustrated. The evaluation of performance of the considered models was conducted on emotion classification on four emotion classes, namely Happy, Sad, Angry and Others. The model for emotion classification in three-sentence conversations was proposed in this work. The model is based on smileys and word embeddings with domain specificity in state of art conversations on the Internet. The importance of taking into account the information extracted from smileys as an additional data source of emotional coloring is investigated. The model performance is evaluated and compared with language processing model BERT (Bidirectional Encoder Representations from Transformers). The proposed model achieved better performance at classifying emotions comparing to BERT (having F_1 score as 78 versus 75). It should be noted, that further study should be performed to enhance the processing by the model of mixed reviews represented by emotion class Others. However, modern performance of models for language representation and understanding did not achieve the human performance. There is a variety of factors to consider when choosing the word embeddings and training methods to design the model architecture.

Keywords: natural language processing, neural network, natural language, long short-term memory networks, text classification, emotional text analysis.

Introduction. Over the past few decades, the amount of text data produced by humanity has grown exceedingly. One of the reasons behind this fact is that we actively exchange information and publish our thoughts on websites and social media.

Such unstructured data is represented in arbitrary form and is often complemented by emojis, which makes it difficult for a computer to categorize and derive meaning from the source. On this basis, the challenge to teach computers to properly process the information arose.

Natural language processing (NLP) is widely used for text data analysis and classification [1].

The core aspects of language understanding include three parameters, which are morphology, semantics, and syntax. Morphology is the study of word or statement structure; semantics is the study of meaning or reference; syntax is the study of how words and morphemes combine to form larger units such as phrases and sentences [2].

Modern models consider all three parameters. The models are using data-driven approach through machine learning and deep learning.

The goal of this work is to consider existing solutions for text data processing in terms of emotional classification and propose the model that can solve such task.

Problem statement. The semantic evaluation problem of emotion classification will be considered in this work. The deep learning approach will be based on Keras and TensorFlow – the Python frameworks. These frameworks have ready to use functions for rapid optimization, model prediction, model tuning and many more. This will allow to achieve approximate state-of-art results with an original model architecture. BERT will also be implemented to obtain a state-of-art model in natural language

understanding. The received results will be evaluated and compared [3].

The theoretical fundamentals of emotional classification models will be discussed. The practical details of the proposed model's training and BERT on unseen test data will be presented. The obtained results will be evaluated and compared. The differences between the models will be illustrated.

It should be considered that the proposed model will be implemented using Keras and TensorFlow. The trained models will be applied to a specific pre-defined dataset to solve the semantic evaluation problem of emotion classification. The models will be trained on four CPU cores. These conditions apply specific limitations on the scope of the project.

Analysis of existing solutions. Machine learning is a modelling approach focused on finding underlying patterns in a dataset. An algorithm with learning function is applied to rich dataset. The parameters of model are modified to enhance the predictive performance of a model. The unseen data is used to evaluate the trained model.

Recurrent Neural Network (RNN) is a class of artificial neural networks, where connections between nodes form a directed sequential graph. By this, the sequential nature of input can be considered when making output predictions [1].

A set of feedback weights contained in a hidden state vector is computed at every step in the sequence that pass information from earlier states. The ability to predict to which class the sequence belongs to is provided by the model reformulation. This will allow to incorporate new time dependency by using the final recurrent hidden state vector to make softmax probability predictions:



$$\hat{y} = \varphi(Vh_p), \quad (1)$$

where

$$h_p = \delta(Ux_p + Wh_{(p-1)}). \quad (2)$$

The parametrization W of the model is comprised of the weight matrices U , V and W . The hidden layer is dependent on earlier states.

The fig. 1 provides more information about the recurrence mechanism applied.

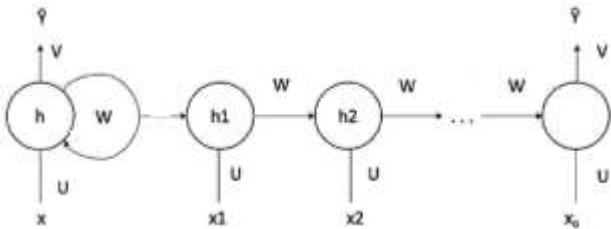


Fig. 1. The unfolding concept

New hidden state vector is computed at each step. The vector is trained to pass forward the most significant information for solving the problem through gradient descent with an appropriate loss function [3].

Long short-term memory (LSTM) networks have two enhancements comparing to previously considered RNNs [4]. The first enhancement is that each time step passes a hidden state vector and a local context vector to the next recurrent node. The second enhancement is that long short-term memory network contains a set of gating mechanisms (fig. 2). These mechanisms provide the ability to the model to decide which data to pass forward in recurrence. These enhancements allow the LSTM model to learn long-term dependencies in the sequence in a more stable way [5].

The gating mechanisms of LSTM contain an input gate, a context gate, a forgetting gate and an output gate, which are the activated matrix manipulations. The manipulations are based on gating weights optimally learned through training. The following relationships define the gates and their associated weights [1]:

$$f_p = \delta(x_p U^f + h_{p-1} W^f), \quad (3)$$

$$i_p = \delta(x_p U^i + h_{p-1} W^i), \quad (4)$$

$$o_p = \delta(x_p U^o + h_{p-1} W^o). \quad (5)$$

The functions are recurrent to the hidden state of the previous time step and the current input data.

A candidate \tilde{c}_p for the context state c_p is computed by the given formula:

$$c_p = \tanh(x_p U^c + h_{p-1} W^c). \quad (6)$$

The previous context information is filtered by the forgetting gate and the present information from input gate in an equation is filtered through the context gate form the context state.

The gating machinery provides an ability to determine which long-term and short-term data to filter and pass to the final output representation [5].

The hidden state representation of the sequence h_{p-1} is computed as a combination of the filtered output from the output gate and current context information:

$$h_p = o_p \tanh(c_p). \quad (7)$$

The received output state can be used for predictions. This output can be used in the same way as hidden states were used in RNN architecture equation (1) considered above [6].

On this basis, LSTM can represent complex sequences and make stable predictions.

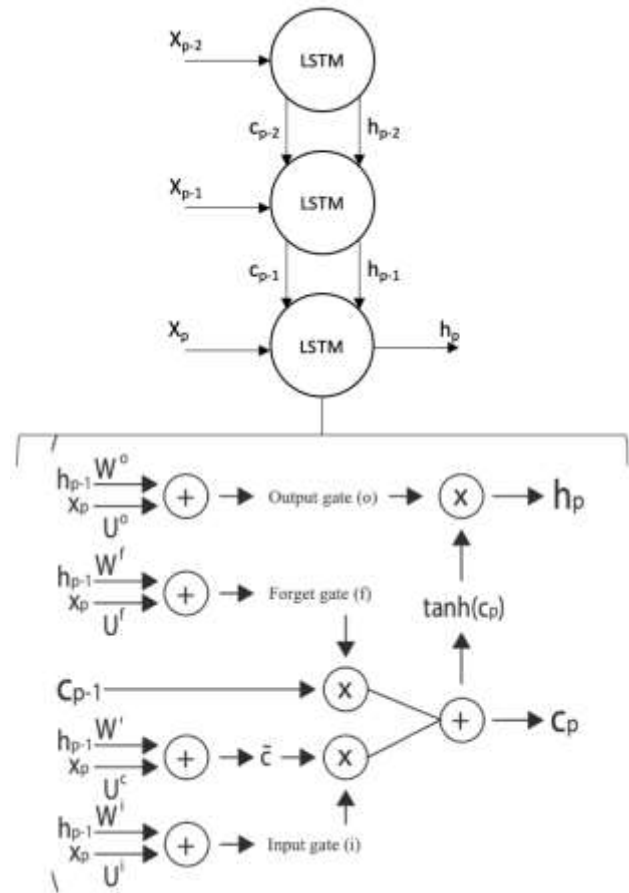


Fig. 2. The LSTM node

Bidirectionality. It should be considered that not all words can be predicted by the previous words in a sequence.

In the sequences like “bow tie” and “bow ring”, “bow” is contextually dependent both on the previous and following word sequences. This is a challenge for regular recurrent networks. The solution for this challenge is bidirectionality, which means reversing direction of the sequence and feeding it to network. Both resulting hidden states are concatenated, which is a standard practice for many language models [7].

In fig. 3, the simple RNN from fig. 1 is extended to work bidirectionally.

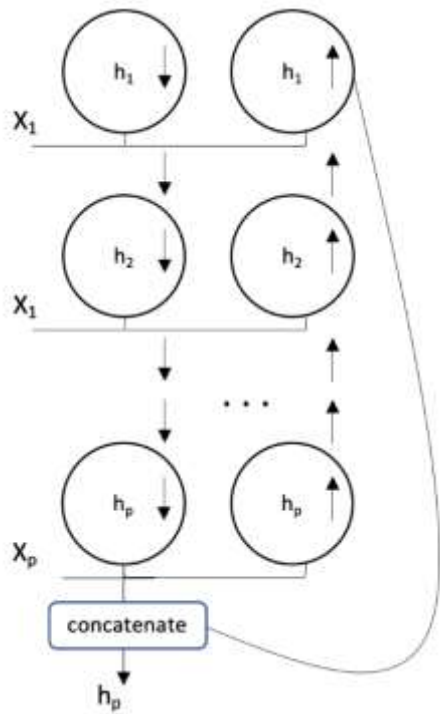


Fig. 3. Bi-directional RNN

The states h_p in fig. 3 can be extracted from regular nodes or LSTM. The network with the reverted states is a copy from the original network. The resulting hidden states are concatenated into form:

$$h_p = (\vec{h}_p, \overleftarrow{h}_p). \tag{8}$$

The final hidden state vector can be used similarly as for RNN through equation (8). The bidirectional models can be modelled by doubling the number of weights. This method is often applied to improve model representation and predictions due to its ability to add contextual information to language models [7].

Vector Representation of Language. Machine Learning models represented above have real valued vectors x_i . The language is represented as vectors which is an essential step in using neural networks as emotional classifiers. The words were gathered in a vocabulary to form the bag of words [8]:

$$V = \{w_j : j \in 1, \dots, N\}. \tag{9}$$

The phrase is divided into one-word hot vectors, which do not provide any extra information about the context. The vectors are also computationally hard to use as a single vector is a single word.

Word Embeddings. A more efficient solution is to pass a lower dimensional vector $x_j \in R_d, (d < N)$, which also contains language information about the word w_j . Less computational power is required to process such vector by the classificational model. This vector contains more language information than one-word hot vector. This type of lower dimensional representation is called word embedding [9].

Statistical language modelling is the other language feature. The words are used in conjunction with each other. This can extract more information about the semantic and syntactical use of a word through the context [10].

Principal Component Analysis (PCA) is a classic dimensional reduction technique, which is based on singular value decomposition of the co-occurrence matrix. This approach does not need linguistic rules to be fed to the system, which makes it unsupervised. It considers the text corpus in a whole, in different context, which also makes it computationally expensive.

This can be solved by using a feed forward neural network to predict the n -gram probabilities of the words in the vocabulary given by the context [10].

Trainable d -dimensional random initialized vector representation of each word and n previous words are used as a context. These vectors are concatenated and fed through several hidden layers. The output layer is a softmax probability over the vocabulary of all the probabilities to meet each word by the context words. The network is learning linguistically valuable information in the matrix of d -dimensional vectors while training to predict n -grams [11].

On this basis, words embedding can be used as the sole input to the other machine learning models for real NLP tasks, such as machine translation, and semantic sentence parsing. The fig.4 illustrates a scheme of such feed forward neural network.

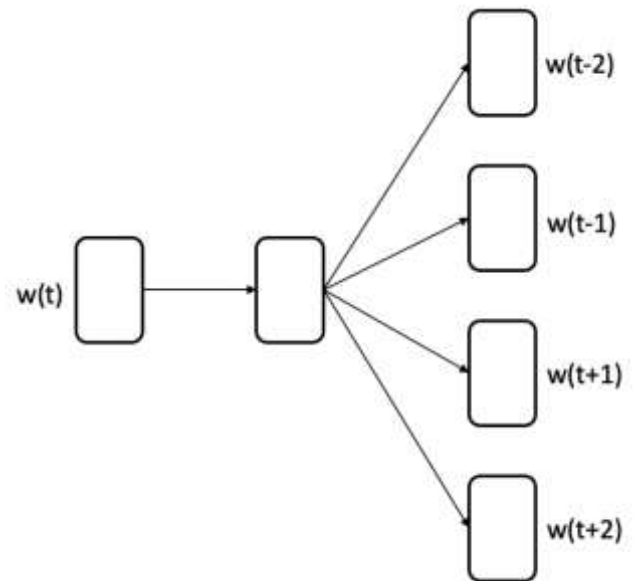


Fig. 4. Scheme of a feed forward neural network

It should be considered that this model is expensive to train. To increase the context window for the network to learn, there is a need to increase the number of input nodes, and the task of emotional classification requires a lot of context.

GloVe. The skip-gram model is another proposed RNN approach which is efficient on smaller datasets with rich context information. RNNs efficiently model rich context information due to the fact they have a sequence memory of the previously observed words. The model

starts with the opposite probability comparing to the n -gram prediction, because this is the probability of different context words. The basis of this probability is a target word [12].

GloVe is the other unsupervised learning algorithm for obtaining vector representations for words. The difference of GloVe from the skip-gram and feed forward word embedding is the fact that GloVe is trained on a loss function. The loss function considers both local co-occurrences from the n -length context windows and global count-based co-occurrence probabilities from the text corpus. This allows to encode more of the language features comparing to PCA. This ability is provided since including only local context data does not give enough information to features about the frequency the words occur in rare contexts [12]. The loss function has slightly richer context information embedded in the vectors comparing to skip-gram model.

BERT. Pre-trained word embeddings in language are easy to use and flexible. Their drawback is that one vector can represent only one word. This means homonyms or phrasal verbs, like 'key' or 'get' lose their language information [14]. This can be mitigated by using RNN or LSTM to carry information over sequences to create a context for a single word, but these networks cannot provide rich context data to predict context-heavy language constructions.

One of the proposed approaches to solve this issue is BERT (Bi-Directional Representations from Transformers). In its core, the transformer relies on attention mechanism for its representation to improve the contextual awareness [14].

The weighted representation of all hidden state vectors is trained at the same time with the recurrence relations. This allows the network to be aware of the previous hidden states and do not rely solely on the final hidden state representation when making predictions.

The transformers apply attention mechanism to underlying word vectors from the original sequence to extract relevant language features. This is called text encoding. Since transformers do not use any recurrent relations, this allows to conduct training in parallel [5].

On this basis, BERT model is applicable to solve NLP tasks. BERT pre-trained weights can be downloaded via the Internet and used for transfer learning. This will allow to compare BERT with the other models in the task of emotion classification.

BERT is a sequence-to-sequence language representation model, which is efficient for NLP tasks. A sequence of language information $X = (I_0, \dots, I_n)$ is used as inputs and outputs as contextualized vector representation $H = (h_0, \dots, h_n)$ of the elements of the input sequence.

During the pre-processing phase, the model splits words into word-parts. The placeholder tags are inserted before the sequence and after sentences, however this does not allow input vectors to directly correspond to the underlying words in the sequence [13].

Due to the design of BERT model, the output presentation h_0 becomes a distributed representation of the

underlying sequence. This sequence is similar to the final hidden state of RNN. A classification model can be obtained when adding an extra hidden layer to the BERT model and activating this model via a softmax function [9]:

$$\hat{y} = \varphi(Vh_0). \tag{10}$$

Encoders, which are stacking nodes, are the other part of this language representation framework. Encoders are used to create encoded text representation (fig. 5).

Every encoder layer abstracts language patterns from an input sequence. It allows to form more complicated patterns as the information flows to the further layers. The first encoder layer L receives language inputs and the last encoder layer outputs the final encoded language information [15]:

$$H_1 = \text{Encoder}(X) \tag{11}$$

$$\dots$$

$$\hat{y} = \varphi(Vh_0).$$

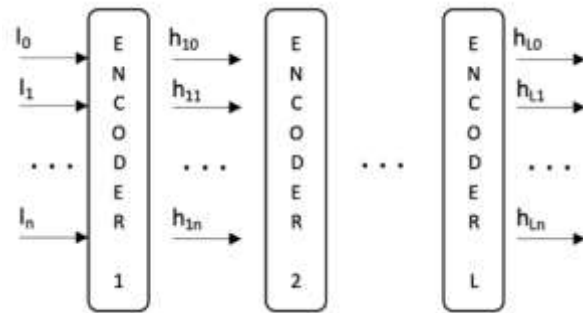


Fig. 5. Layers of the model

The first vector h of the final H representation is the basis for the classification task. Two sub-processes form each encoder layer. As the first step, the inputs to the encoder are passed through a multi-head self-attention layer. This layer uses a series of matrix manipulations to extract language information from the data inputs. The definition of this process is multi-head self-attention (fig. 6).

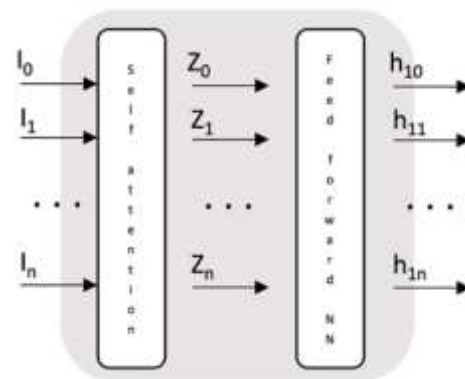


Fig. 6. The encoder dissection

The second sub-layer receives the outputs of the first layer after these outputs are residually connected and normalized. The second sub-layer retrieves the most valuable information from the attention layer. This data is

residually connected and normalized yet another time [15]. After this, the outputs are sent onwards to the following encoder layer.

Theoretical propositions. This section is dedicated to proposition of a model for classifying the last phrase of three-turn conversations into any of pre-defined emotions. These emotions are Happy, Angry, Sad and Others. The context is defined by the previous two sentences. The datasets are classified beforehand by the emotion they represent, since model training should be performed on high-quality data. The dataset examples (data samples) of three-turn conversations and their corresponding labels can be seen in table 1. These data samples may include emojis and smileys as separate complete response options with emotional coloring.

This task is solved by creation of neural network architecture. BERT is used as a state-of-art language representation model. The proposed model and BERT are compared in terms of their performance on the given problem.

The pre-processing is divided into two steps. The first step is data cleaning, which stands for defining and fixing data irregularities. The second step is tokenization, which is a mathematical mapping of text data into a suitable input for classification models.

Table 1 – Data sample

id	Turn 1	Turn 2	Turn 3	
112	It's cool meme	Ha-ha, yes	☺	Happy
154	I don't think it is a good idea	You better think twice	No, I won't!	Angry
186	In my hometown there are many chestnuts	Where is it?	In Kyiv	Others

Data cleaning provides an ability to the model to clarify the text content and boost the coverage. The coverage is measured by pre-trained embeddings as the number of words for which pre-trained embeddings exist divided by the total amount of unique words in the problem. Since our dictionary is problem-dependent and GloVe embedding matrix is generalized, it is not certain that the embedding matrix has a vector representation of a given word. Thus, data cleaning is recurrently applied to boost the coverage.

In the given problem, data cleaning is used to decompose abbreviations into their initial state and replace symbol emojis with Unicode emojis. This allows to improve sentence representation in terms of their meaning.

After the data cleaning, tokenization is applied. The words are vectorized and prepared for passing into classification models. The words are separated and mapped to unique numbers for further decoding. Keras, which is used to solve the problem, has this as a built-in feature.

As a result, tokenized texts are mapped with embedding matrices which represent the sequences in a vector space. The vector represented sequences are passed to LSTM models. The inputs have to be of equal length to

allow to train the model weights to recognize features at the same positions in different sequences. To solve this, the vectors are zero padded to the same length, which is constrained to 100 words per sentence.

The fig. 7 illustrates the word length distribution of the training data.

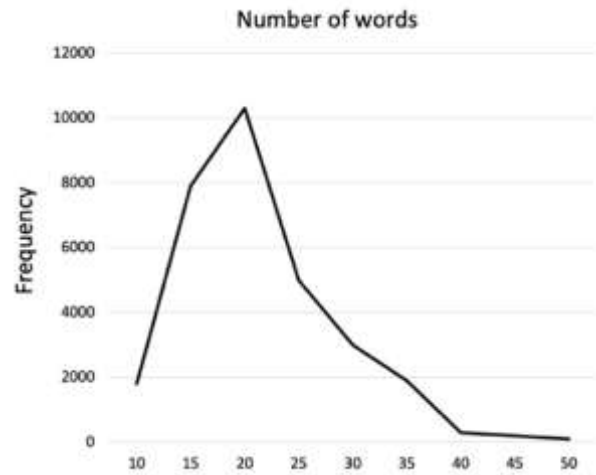


Fig. 7. Word lengths per sentence

Each sentence in a conversation is a matrix with elements of padded vectors representing the sentences. The corresponding labels are one-hot vectors. The index of the element 2 represents the emotion the training example is connected to. The indexes are as follows: 0 – Angry, 1 – Sad, 2 – Happy, 3 – Others. The given dataset is ready for further processing.

Proposed model. The proposed model (fig. 8) is built as follows. The standard LSTM outputs straight to a prediction layer and is filtered through the softmax function. The result is a probability vector. We use GloVe embeddings, trained on the text corpus represented by the data from Wikipedia. The 100-dimensional word embeddings and 128-dimensional LSTM layer with built in dropout are used by the model.

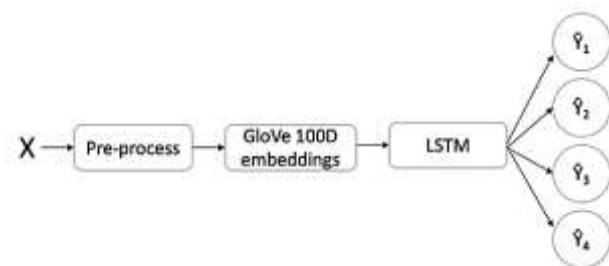


Fig. 8. The proposed model architecture

The model passes several improvements, which allow us to define features with high impact on model performance. This impact is quantitatively represented by F_1 score. Such features are isolated and used to create simple networks.

The first improvement is to separate the phrases of the conversation and pass them to LSTM nodes. Thus, the model is provided by a clearer vision of sentence differences.

The second improvement is to extract emojis and smileys from text for further separate processing. The emojis are represented with Emoji2Vec embeddings, which provides new language information to the model. Emoji2Vec is a 300-dimensional representation of Unicode smileys. Further concatenation of information contained in emojis with the data from the rest of the model is passed through the final dense layer. This allows the model to weight the relative importance of smileys for emotion classification. The rectified linear unit is the activation function in this dense layer.

BERT is used due to its applicability in wide range of problems. The model is hosted on TensorFlow hub. TensorFlow is a Google pre-trained machine learning repository.

It provides an ability to download an implementation which has to perform the following: the number of encoded layers, the number of heads in the multi-head attention, the number of dense layers in the feed forward network and the number of used training samples.

On the contrary to GloVe embeddings, BERT uses word-piece embeddings and follows different pre-processing and tokenizing which can be downloaded with the ready-to-use pre-trained model. This allows us not to perform the pre-processing step of the sentences but directly pass them to BERT.

Experimental results. The table 2 illustrates the comparative micro-averaged F_1 results for the models on emotion classification. The performance of the models was considered by evaluation metric on Happy, Angry and Sad emotions. The model performance on emotion class Others was ignored in this evaluation.

The LSTM model achieved overall micro-averaged F_1 as 0.616. This model required minimal data pre-processing and GloVe word embeddings. The word embeddings were trained on the Wikipedia data. It can be seen in the table, that the model achieved the worst results on Happy conversations and the best results on Angry conversations. The shortcoming of the model is the inability to distinguish the Happy, Angry and Sad emotion classes from the emotion class Others.

Table 2 – F_1 scores

Model	Emotion F_1			Micro F_1
	Happy	Sad	Angry	
LSTM	0.523	0.601	0.724	0.616
BERT	0.687	0.799	0.771	0.752
SS-LSTM	0.556	0.818	0.784	0.719
Proposed Model	0.784	0.767	0.811	0.787

BERT model achieved better results in macro F_1 comparing to LSTM due to the presence of extra context information. The Angry emotion class was predicted as 0.771 F_1 points. The Sad emotion class was also predicted better in F_1 points comparing to LSTM. The proposed model F_1 scores are shown in the table. The model has

shown the best results on Angry and Happy emotion classes.

The confusion matrix of the proposed model is shown in table 3. The matrix illustrates the distribution between the emotion labels in the test dataset. The most confusion comes from distinguishing the Others emotion class from Happy, Angry and Sad emotion classes. Angry and Sad emotion classes are never evaluated by the model as Happy and vice versa. The confusion between predicting Angry and Sad labels is present rarely. On this basis, the focus should be put on prediction improvement among Others emotion class and the other three emotion classes.

Table 3 – Confusion Matrix

TRUE	Predicted			
	Others	Happy	Sad	Angry
Others	4103	101	80	94
Happy	63	215	4	1
Sad	40	0	212	8
Angry	46	0	6	334

The idea of whether the early stopping effectiveness on overfitting prevention is got by looking at the loss function over the training steps. The decrease of training loss is present over the training steps. The validation loss is decreasing by reaching the minimum after three epochs and is rising afterwards (fig. 9).

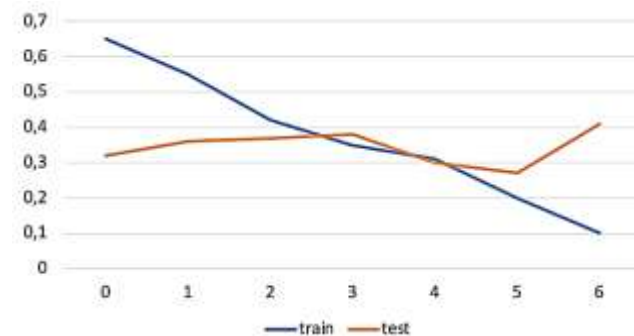


Fig. 9. Training and Validation Loss

The explanation behind the given tendency is the different distribution of emotions in the training dataset and validation dataset. This can be seen in table 4.

Table 4 – Emotion Class Distribution

	Others	Happy	Sad	Angry	Size
Train	0.441	0.980	0.211	0.194	29884
Validation	0.793	0.045	0.054	0.052	2467
Test	0.882	0.039	0.061	0.044	5804

The loss from the training set updates the gradients. The gradient updates change the validation loss in other ways. The amount of data points in the validation set is fewer comparing to the training set, which contributes to higher variance in the effects of updating the loss. On this basis, it is important to use early stopping after 3 epochs

since the improvements in training loss are not contributing into improvements in validation loss after this epoch.

Conclusions. The problem of emotion classification is a complex and non-trivial task of language interpretation. The existing generic solutions for text data processing were considered in this work, having their shortcomings and advantages illustrated.

The evaluation of performance of the considered models was conducted. New model for emotion classification was proposed. The model is based on 128-D LSTM neural network combined with 100-D GloVe embeddings and BERT as a state-of-art language representation model.

The proposed model shows better results comparing to generalized model with transfer learning (having F_1 score as 78 versus 75). It should be noted that the proposed model needs further improvements to better capture ambiguous or mixed emotional expression differences represented by Others emotion class. It also showed slightly lower results on Sad emotion class compared to SS-LSTM model (having Emotion F_1 score as 0.767 versus 0.818). However, the proposed model still overpowers all considered models in micro F_1 score.

There are many decisions applicable for solving the emotion classification problem. There is a variety of factors to consider when choosing the word embeddings and training methods to design the model architecture.

To date, machine learning models did not achieve the human performance in terms of language representation and emotion classification. The confusion is present in labeling complex sentence structures. Such nuances are explained by the language structure and its dynamic nature.

However, the area of capturing the language nuances is constantly evolving and new approaches are created every day.

References

- Graves A., Schmidhuber J. Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural networks*. 2005. Vol. 18(5). P. 602–610.
- Srivastava N., Hinton G., Krizhevsky A., Sutskever I., Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. *Journal of machine learning research*. 2014. Vol. 15. P. 1929–1958.
- Daniil Y., Onyshchenko K. Implementation of Recursive Deep Learning Algorithms for Natural Language Processing. *Information systems and technologies 2021*. Kharkiv-Odesa, 2021. P. 141–145.
- Afanasieva I., Golian N., Hnatenko O., Daniil Y., Onyshchenko K. Data exchange model in the internet of things concept. *Telecommunications and radio engineering*. 2019. Vol. 78(10). P. 869–878.
- Collobert R., Weston J., Bottou L., Karlen M., Kavukcuoglu K., Kuksa P. Natural language processing (almost) from scratch. *CoRR*. 2011. P. 201–244.
- Joulin A., Grave E., Bojanowski P., Mikolov T. Bag of tricks for efficient text classification. *CoRR*. 2016. P. 11–32.
- Otter D.W., Medina J.R., Kalita J.K. A survey of the usages of deep learning in natural language processing. *CoRR*. 2018. P. 112–123.
- Allen J. *Natural language understanding*. 2nd ed. Redwood City, CA, USA: Benjamin-Cummings Publishing Co., Inc., 1995. 512 p.
- Gupta U., Chatterjee A., Srikanth R., Agrawal P. A sentiment-and-semantic-based approach for emotion detection in textual conversations. *In neu-ir: the SIGIR 2017 workshop on neural information retrieval*. 2017. P. 21–28.
- Jurafsky D., Martin J.H. *Speech and language processing: an introduction to natural language processing, computational linguistics, and speech recognition*. Upper Saddle River, NJ, USA: Prentice Hall PTR, 2000. 1044 p.
- Christopher M., Schütze H. *Foundations of statistical natural language processing*. Cambridge, MA, USA: MIT Press, 1999. 718 p.
- Yao L., Guan Y. An improved LSTM structure for natural language processing. *2018 IEEE International Conference of Safety Produce Informatization (IICSPI)*. Chongqing, China, 10–12 December 2018. 2018. P. 565–569.
- Korti S. S., Kanakaraddi, S. G. Depression detection from Twitter posts using NLP and Machine learning techniques. *2022 Fourth International Conference on Emerging Research in Electronics, Computer Science and Technology (ICERECT)*. Mandya, India. 2022. P. 1–6.
- Sharma D., Dhiman C., Kumar D. Automated Image Caption Generation Framework using Adaptive Attention and Bi-LSTM. *IEEE Delhi Section Conference (DELCON)*. New Delhi, India. 2022. P. 1–5.
- Aziz A. A., Diamal E. C., Ilyas R. Paraphrase Detection Using Manhattan's Recurrent Neural Networks and Long Short-Term Memory. *6th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)*. Bandung, Indonesia, 2019. P. 432–437.
- Yang H., Feng Y. Authoritative Prediction of Website Based on Deep Learning. *IEEE Fourth International Conference on Big Data Computing Service and Applications (BigDataService)*. Bamberg, Germany, 2018. P. 208–212.

References (transliterated)

- Graves A., Schmidhuber J. Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural Networks*. 2005, vol.18(5), pp.602–610.
- Srivastava N., Hinton G., Krizhevsky A., Sutskever I., Salakhutdinov R. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*. 2014, vol. 15, pp. 1929–1958.
- Daniil Y., Onyshchenko K. Implementation of Recursive Deep Learning Algorithms for Natural Language Processing. *Information Systems and Technologies 2021*. Kharkiv-Odesa, 2021, pp. 141–145.
- Afanasieva I., Golian N., Hnatenko O., Daniil Y., Onyshchenko K. Data exchange model in the internet of things concept. *Telecommunications and Radio Engineering*. 2019, vol. 78(10), pp. 869–878.
- Collobert R., Weston J., Bottou L., Karlen M., Kavukcuoglu K., Kuksa P. Natural language processing (almost) from scratch. *CoRR*, 2011, pp. 201–244.
- Joulin A., Grave E., Bojanowski P., Mikolov T. Bag of tricks for efficient text classification. *CoRR*, 2016, pp. 11–32.
- Otter D.W., Medina J.R., Kalita J.K. A survey of the usages of deep learning in natural language processing. *CoRR*, 2018, pp. 112–123.
- Allen J. *Natural Language Understanding*. 2nd ed. Benjamin-Cummings Publishing Co., Inc. Redwood City, CA, USA, 1995. 512 p.
- Gupta U., Chatterjee A., Srikanth R., Agrawal P. A sentiment-and-semantic-based approach for emotion detection in textual conversations. *In neu-ir: the SIGIR 2017 workshop on neural information retrieval*. 2017, pp. 21–28.
- Jurafsky D., Martin J.H. *Speech and language processing: an introduction to natural language processing, computational linguistics, and speech recognition*. Upper Saddle River, NJ, USA, Prentice Hall PTR, 2000. 1044 p.
- Christopher M., Schütze H. *Foundations of Statistical Natural Language Processing*. MIT Press, Cambridge, MA, USA, 1999. 718 p.
- Yao L., Guan Y. An Improved LSTM Structure for Natural Language Processing, *2018 IEEE International Conference of Safety Produce Informatization (IICSPI)*. Chongqing, China. 2018, pp. 565–569.
- Korti S. S., Kanakaraddi, S. G. Depression detection from Twitter posts using NLP and Machine learning techniques, *2022 Fourth International Conference on Emerging Research in Electronics, Computer Science and Technology (ICERECT)*. Mandya, India. 2022, pp. 1–6.
- Sharma D., Dhiman C., Kumar D. Automated Image Caption Generation Framework using Adaptive Attention and Bi-LSTM, *IEEE Delhi Section Conference (DELCON)*. New Delhi, India, 2022, pp. 1–5.

14. Aziz A. A., Diamal E. C., Ilyas R. Paraphrase Detection Using Manhattan's Recurrent Neural Networks and Long Short-Term Memory, *6th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)*, Bandung, Indonesia, 2019, pp. 432–437.
15. Yang H., Feng Y. Authoritative Prediction of Website Based on Deep Learning. *IEEE Fourth International Conference on Big Data Computing Service and Applications (BigDataService)*. Bamberg, Germany, 2018, pp. 208–212.

Received 10.05.2023

УДК 004.8

К. Г. ОНИЩЕНКО, Харківський національний університет радіоелектроніки, старший викладач кафедри програмної інженерії, м. Харків, Україна, e-mail: kostiatyn.onyshchenko@nure.ua, ORCID: <https://orcid.org/0000-0002-7746-4570>
Я. Д. ДАНИЕЛЬ, Харківський національний університет радіоелектроніки, асистент кафедри програмної інженерії, м. Харків, Україна, e-mail: yana.daniiel@nure.ua, ORCID: <https://orcid.org/0000-0002-3895-0744>

ВИКОРИСТАННЯ МЕРЕЖ ДОВГОТРИВАЛОЇ ПАМ'ЯТІ ДЛЯ ОБРОБКИ ПРИРОДНОЇ МОВИ

Проблема класифікації емоцій є складним та нетривіальним завданням інтерпретації мови через структуру природної мови та її динамічний характер. Актуальність дослідження полягає в охопленні важливої проблеми автоматичної обробки відгуків клієнтів, збирання думок та виявлення тенденцій. У цій роботі розглянуто ряд існуючих рішень для проблеми класифікації емоцій, де продемонстровано їхні недоліки та переваги. Оцінка продуктивності розглянутих моделей була проведена на класифікації емоцій чотирьох класів: Happy, Sad, Angry та Other. У цій роботі запропоновано модель для класифікації емоцій в трирядкових розмовах. Модель базується на емодах та представленнях слів зі специфікою області сучасних розмов в Інтернеті. Досліджується важливість врахування інформації, отриманої зі емодах як додаткового джерела даних з емоційним забарвленням. Оцінено продуктивність моделі та порівняно її з мовною моделлю BERT (Bidirectional Encoder Representations from Transformers) для класифікації емоцій. Запропонована модель показала кращу продуктивність у класифікації емоцій порівняно з BERT (з F_1 -оцінкою 78 порівняно з 75). Слід зазначити, що потрібні додаткові дослідження для поліпшення обробки моделей змішаних відгуків, що представлені класом емоцій "Other". Однак, сучасна продуктивність моделей для представлення та розуміння природної мови не досягла рівня людини. Є різноманітні фактори, які необхідно враховувати при виборі представлень слів та методів навчання для проектування архітектури моделі.

Ключові слова: обробка природної мови, нейронна мережа, природна мова, мережі довготривалої пам'яті, текстова класифікація, емоційний аналіз тексту.

Повні імена авторів / Author's full names

Автор 1 / Author 1: Оніщенко Костянтин Георгійович, Onyshchenko Kostiantyn
Автор 2 / Author 2: Данієль Яна Дмитрівна, Daniiel Yana

O. M. RYBITSKYI, Postgraduate, Kharkiv National University of Radioelectronics, Kharkiv, Ukraine, e-mail: oleksandr.rybitskyi@nure.ua, ORCID: <https://orcid.org/0000-0002-8508-7269>

V. V. GOLIAN, Candidate of Technical Sciences (PhD), Docent, Kharkiv National University of Radioelectronics, Kharkiv, Ukraine, e-mail: vira.golian@nure.ua, ORCID: <https://orcid.org/0000-0001-5981-4760>

N. V. GOLIAN, Candidate of Technical Sciences (PhD), Docent, Kharkiv National University of Radioelectronics, Kharkiv, Ukraine, e-mail: natalia.golian@nure.ua, ORCID: <https://orcid.org/0000-0002-1390-3116>

Z. V. DUDAR, Candidate of Technical Sciences (PhD), Professor, Kharkiv National University of Radioelectronics, Kharkiv, Ukraine, e-mail: zoia.dudar@nure.ua, ORCID: <https://orcid.org/0000-0001-5728-9253>

O. V. KALYNYCHENKO, Candidate of Technical Sciences (PhD), Docent, Kharkiv National University of Radioelectronics, Kharkiv, Ukraine, e-mail: olga.kalynychenko@nure.ua, ORCID: <https://orcid.org/0000-0003-1466-3967>

D. M. NIKITIN, Postgraduate, Kharkiv National University of Radioelectronics, Kharkiv, Ukraine, e-mail: dmytro.nikitin1@nure.ua, ORCID: <https://orcid.org/0000-0003-4388-4996>

USING OBD-2 TECHNOLOGY FOR VEHICLE DIAGNOSTIC AND USING IT IN THE INFORMATION SYSTEM

This article considers the research of OBD-2 technology for interaction with on-board vehicle systems, the creation of a unified system that can work with different makes and models of cars. The history of OBD-2 technology, its development, existing standards and their implementation in modern cars is described. Basic diagnostic functions are described, with which you can get information about the car, as well as perform its settings and send various commands to it. In addition, the hardware part of the scanner, its technical features, nuances of interaction with it and the purpose of each of its pins are considered. Also reviewed couple models of scanners that can be used to develop this system. Conclusions are drawn on the practicability of their use, taking into account their capabilities. Existing software analogues, their main functions, advantages and disadvantages are considered. The general concept of interaction with the car via OBD-2, the interaction algorithm and step-by-step analysis of the interaction between the scanner and the car are also considered. The software part of the interaction between the car and the scanner, special codes used to send requests and responses, their further analysis and interpretation in a human-friendly form are considered. The possibilities for the unification of these codes for different car manufacturers and the common code base to create a universal system that will be suitable for different cars from different car manufacturers are highlighted. A model of the software system is proposed, which can embody a large set of useful functions for any motorist and will be compatible with a large number of modern cars equipped with universal diagnostic tools, while using an affordable scanner model without the use of expensive professional equipment.

Keywords: Car, on-board computer, computer diagnostics, scanner, OBD2, information system.

Introduction. Despite the rapid pace of development of the automotive industry, today a large number of cars are quite primitive and do not even have an on-board computer or have it, but it is endowed with a very narrow functionality. But at the same time, almost all modern cars are equipped with diagnostic ports that allow you to get a lot of useful information about the car.

OBD2 car computer diagnostics technology is a standard for diagnostics of various car systems performed by car control units. It was introduced in 1996 as mandatory for all vehicles sold in the United States of America, and subsequently for all other vehicles.

Now, more or less every car is equipped with this diagnostic port, and the number of those that are not equipped is so low that it can be written off as an error.

So we can say with great confidence that this technology is promising for research.

Today, there are a large number of similar programs that allow motorists to self diagnose their own car. But there is a number of certain disadvantages. All these programs can be divided into 2 types: professional software and those intended for general use.

Professional software usually has wide functionality and provides a large number of opportunities to the user. But this software, firstly, requires professional equipment, and secondly, it has a high cost. These factors negate any sense of mass use of this software.

Another category of programs are those that are intended for general use. It has a small cost, but the vast majority have narrow functionality, shortcomings in their work or simply an inconvenient and unintuitive interface.

Also, one of the most common problems faced by motorists is the problem of car theft. Standard alarms of all, without exception, car manufacturers are not reliable protection of the car. There are a huge number of illegal devices on the black market that allow thieves to overcome standard security measures and enter the car. Cars equipped with keyless access systems are especially prone to this threat. Thieves who have a special device for transmitting a signal from the key to the car can easily break into the car and steal it.

Today, a large number of motorists own not the most modern cars, which do not even have such a vital option as an on-board computer. This is a very functional tool that allows the driver to get rid of a very large number of different worries. But people don't only fill cars with fuel. Lubricant, antifreeze and other special fluids need to be replaced in it.

This led me to the idea of researching the technologies of computer diagnostics of the car to create a system that would allow any motorist to receive information about his car on his own, which was previously available only at specialized stations.

© Rybitskyi O. M., Golian V. V., Golian N. V., Dudar Z.V., Kalynychenko O.V., Nikitin D.M. 2023



Research Article: This article was published by the publishing house of *NTU "KhPI"* in the collection "Bulletin of the National Technical University "KhPI" Series: System analysis, management and information technologies." This article is distributed under a Creative Commons [Creative Commons Attribution \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/). **Conflict of Interest:** The author/s declared no conflict of interest.



The purpose of the research is to study the possibilities of OBD-2 technology and its application in a software system that is an analogue of a car's on-board computer, to study the possibilities of integration with services that provide the necessary information for car diagnostics, to design a motorist's assistant system that provides car diagnostic functions.

Proposed model. Analyzing similar programs, we can conclude that today there are a large number of software products on the market that provide real-time computer diagnostics of the car using universal OBD reading devices.

But most of these products are aimed primarily at professional auto mechanics who are engaged in car diagnostics, have professional equipment and extensive knowledge in the field of car diagnostics and its structure.

This means that such products are not suitable for use by the masses of motorists, since they have neither expensive equipment nor professional education. They generally do not use such professional tools.

In addition, such solutions often cost a lot of money and their purchase is simply unprofitable for an ordinary car enthusiast who needs to get some basic information about his car.

Research begins with the development of the program. A research program is a document that regulates all stages, stages of preparation, organization and conduct of a specific study. The research program contains theoretical substantiations of methodological approaches and methodical methods of studying a certain phenomenon or process.

In the research of the study, the main possibilities of OBD2 technology for computer diagnostics of the car were studied. The main goal was to study the possibility of creating a system that would allow diagnosing any modern car, while having a universal car scanner without of using of OEM equipment.

Results and discussion. A standard OBD-II diagnostic connector is used to connect to the car. Most of the serial cars produced after 1996 are already equipped with such a connector. Although such a diagnostic connector is standardized, it supports several protocols of different engine control systems at once, that is, different contacts on the connector are physically used, which the scanner communication module must know. Accordingly, different brands of cars may have different internal busses for obtaining diagnostic data from the engine management side.

One way or another, but all specialized solutions are more advanced industrial products, compared to the usual device for reading diagnostic codes based on the ELM327 microcircuit of the Canadian company Elm Electronics. ELM327 is a universal converter of protocols used in diagnostic tires of cars into a serial protocol of the RS-232 type.

Interaction with ELM327 is carried out by standard AT-commands supported by the chip. You just need to organize the exchange of text messages using the RS-232 protocol. Namely, a low-level physical connection via USB, Bluetooth or Wi-Fi is implemented using RS-232 serial protocol conversion chips. That is, to turn a car into

an IoT device, it is enough to connect the ELM327 chip to the OBD-II diagnostic connector and at the output of this chip, for example, put a serial interface converter in Bluetooth or Wi-Fi. Then, you can "read" car diagnostics from your smartphone or computer. However, there are more than enough such ready-made modules or blocks on the market. And their price on AliExpress ranges from a few dollars to thousands for professional devices.

The most common interface CAN (Controller Area Network) is used for interaction with the car. At one time, the CAN standard, proposed by Bosch, made significant progress in the development of systems for automotive electronics. If the car in the Internet appeared only recently, then the concept of the network inside the car has existed since the mid-80s. The idea is very simple, and just as Ethernet made a break-through in computer networks, CAN became the basis of reliable communications inside the car.

CAN bus is an automotive bus developed by Robert Bosch that is mostly adopted in the automotive and aerospace industries. CAN is a serial bus protocol, with the connection of individual systems and sensors as an alternative to the usual multiwire bundle. Allows you to connect automotive components in a single wire called a data bus network with a high speed of information transfer. Before the CAN bus was released, cars had a large amount of wiring that was necessary to connect various electronic components. That is, it is a bus that connects various blocks and components in the car and allows them to exchange information in real time.

Previously, in cars, tires of various modules and devices converged to the central engine control unit. The serial CAN bus made it possible to implement independent intelligent modules that communicate with each other actually using a network protocol. At the same time, the amount of wiring inside the car is significantly reduced.

Data transmission in the CAN bus resembles a publisher-subscriber model, where each device on the bus has a unique identifier, and when one device transmits data, all the others listen and make a decision based on this identifier - whether it need that data to receive and processing or not.

Moreover, there are two different CAN nets. One works with a higher speed (1 Mbit/s) and is used to monitor the engine and interconnect the ECU. The other one is used to communicate the rest of the parts of the vehicle such as doors, seats or lights and works with less speed (250 Kbit/s).[1]

It is this feature that allows the user to connect a diagnostic scanner through a special port and receive real-time information about his car, read errors, etc.

The main method of computer diagnostics is the connection of special scanners that are connected to car electronics and read data in digital form.

First, computer diagnostic tools are used to read not only the error codes, but also all the digital data that is directly related to the problem that occurred. Here you need to understand that the scanner can fully decipher detected malfunctions.

At the second stage, all these data need further confirmation. And above all, the car's electronic systems

need to be thoroughly checked to ensure that it is fully operational

Next, the scanner displays the data in real time. This function can be used to check signals from sensors and other controls to detect malfunctions.

Interaction of the scanner with the car. Modern cars are not only mechanical parts, it are also many electronic units. It monitor the operation of various systems and units, monitor their condition, and track their errors in operation. Reading these metrics makes computer diagnostics and troubleshooting possible.

The multifunctional adapter, which connects to a standard OBD-2 port, supports cars of various makes and models. Special software for them contains in the database both standard error codes that are common to most cars, and codes that are unique to each brand.

In addition to reading fault codes, the latest OBD-2 scanners offer several options for tuning the vehicle's electronics.

The OBD2 scanner can analyze both the entire on-board network and individual systems. The ability to monitor such parameters as engine speed, speed, oil and antifreeze temperature, oil and fuel pressure, intake manifold and exhaust manifold pressure in real time.

Within the OBDII diagnostic standard, there are 5 main communication protocols between the electronic control unit (ECU) and the diagnostic scanner. Physically, the auto scanner is connected to the ECU via the DLC (Diagnostic Link Connector), which complies with the SAE J1962 standard and has 16 pins (2x8). Below is the layout of the contacts in the DLC connector on the Figure 1, as well as the purpose of each of them.

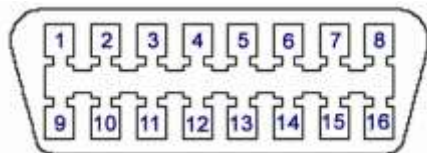


Fig. 1: Pin assignment in the DLC (Diagnostic Link Connector)

Pins list:

1. OEM (manufacturer's protocol);
2. Bus + (Bus positive Line). SAE-J1850 PWM, SAE-1850 VPW;
3. Defining by manufacturer;
4. Body grounding;
5. Signal ground;
6. CAN-High line of CAN Highspeed bus (ISO 15765-4, SAE-J2284);
7. K-Line (ISO 9141-2 and ISO 14230);
8. Defining by manufacturer;
9. CAN-Low line, CAN Low speed bus;
10. Bus - (Bus negative Line). SAE-J1850 PWM, SAE-1850 VPW;
11. Defining by manufacturer;
12. Defining by manufacturer;
13. Defining by manufacturer;
14. CAN-Low line of CAN Highspeed bus (ISO 15765-4, SAE-J2284);
15. L-Line (ISO 9141-2 and ISO 14230);
16. Power supply +12v from the battery.

The assignment of undefined contacts is at the discretion of the vehicle manufacturer. Next, we will consider in detail the format and physical layer of each communication protocol within the OBDII standard.

The OBD-II system uses 9 measuring modes, each one of them allows the access to the ECU data in the vehicle check Table 1. In order to request data, it is necessary to use PID [Parameter Identification] codes. Each PID is related with a specific measurement of the modes 1 and 2 of the OBD-II system. For instance, if the real-time datum of the vehicle speed is requested, the mode 1 should be chosen and the PID "OD" has to be used. [2]

Table 1 – OBD measuring modes

Mode	Features
01	Collecting updated data: it allows the real-time access to the ECU inputs and outputs.
02	Access to frozen data frames: the ECU takes a sample of the values related with the emissions at the exact moment when a failure arises.
03	Gathering of the failure codes: it allows to extract all the DTC [Data Trouble Codes] stored in the ECU memory.
04	Code erasing and failure in the stored values: it allows to delete all the stored codes in the ECU, including the DTC and the saved data frame.
05	Tests results in the oxygen transducers: it allows the access to the test results performed to the oxygen transducers.
06	Tests results of other transducers: results of the diagnostics in components not submitted to constant surveillance.
07	Pending failure codes sampling: it allows to read all the pending DTC from the ECU memory.
08	08 Components operating control: it permits the execution of tests in the actuators.
09	Vehicle information: it allows to request the VIN [Vehicle Identification Number].

Within the OBDII standard, there are 5 data exchange protocols between the ECU (electronic control unit) and the diagnostic scanner.

SAE J1850 PWM on the Figure 2.

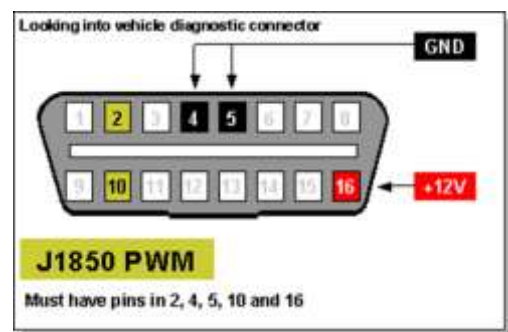


Fig. 2: SAE J1850 PWM protocol connection diagram

There are two types of J1850 protocol. The first of them is high-speed and provides a performance of 41.6 KB/s. This protocol is called PWM (Pulse Width Modulation). It is used by Ford, Jaguar and Mazda. For the first time this type of communication was used in Ford cars. In accordance with the PWM protocol, signals are

transmitted over two wires connected to pins 2 and 10 of the diagnostic connector.

SAE J1850 VPW (Variable Pulse Width) on the Figure 3. The VPW protocol supports data transfer at a rate of 10.4 KB/s and is used in General Motors (GM) and Chrysler vehicles. It is very similar to the protocol used in Ford vehicles, but is significantly slower. The VPW protocol provides for data transfer over a single wire connected to pin 2 of the diagnostic connector.

ISO 9141-2 on the Figure 4.

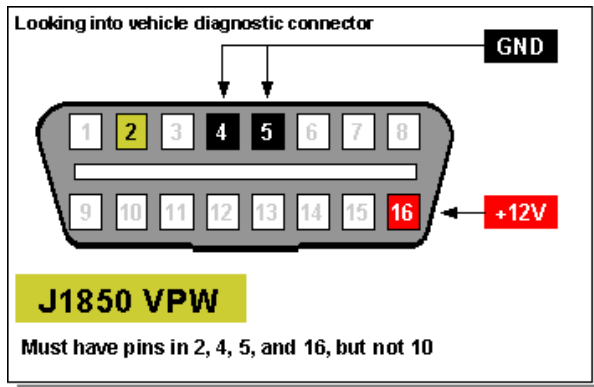


Fig. 3: SAE J1850 VPW protocol connection diagram

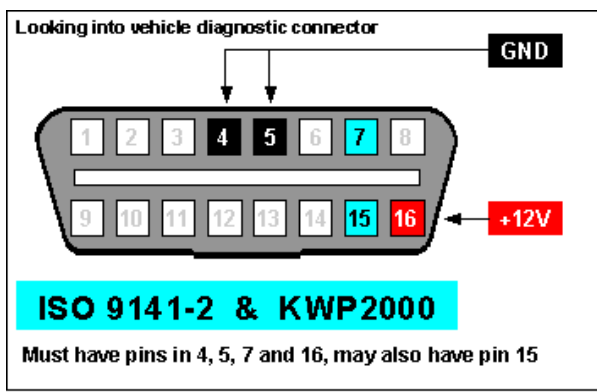


Fig. 4: ISO 9141-2 protocol connection diagram

Applies to most European and Asian vehicles and some Chrysler vehicles. The ISO9141 protocol is not as complex as the J1850 standards. While the latter require the use of special communication microprocessors, ISO9141 requires conventional serial communication microprocessors, which are found on store shelves. Uses pin 7 (K-line) and optionally pin 15 (L-line).

ISO 14230 KWP2000 (Keyword Protocol 2000).

Physically identical to ISO 9141. Also uses pin 7 (K-line) and optionally pin 15 (L-line).

ISO 15765 CAN (250 kBit / s or 500 kBit / s) on the Figure 5.

The CAN protocol was developed by Bosch for automotive and industrial applications. Unlike other OBD protocols, its variants are widely used outside of the automotive industry. CAN was not OBD-II compliant for vehicles in the US until 2003, but as of 2008, all vehicles sold in the US must support CAN. Within OBDII, it uses 2 pins: 6 and 14. It is the fastest and most modern.

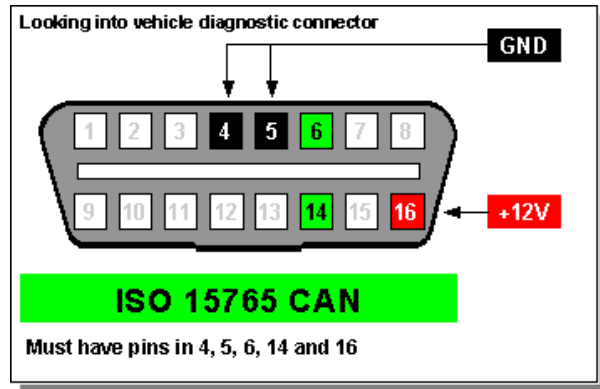


Fig. 5: ISO 15765 CAN protocol connection diagram

Existing software. TOAD is short for Total OBD & ECU Auto Diagnostics. It is an OBD diagnostic software that performs vehicle diagnostic systems and provides diagnostic reports. It monitors electronic sensors in the engine, transmission and emissions system [3].

Benefits:

- provides wide possibilities of diagnostics;
- error reading in real time;
- makes it possible to clean blocks of cars from errors;
- has the function of saving a snapshot of block data at a certain point in time;
- compatible with almost all makes and models of cars.

Disadvantages:

- very high price;
- only compatible with Windows OS;
- requires a deep understanding and skills in car diagnostics.

PCMSCAN is a full-featured generic OBD-II scanner and diagnostic tool that supports a wide range of OBD-II hardware interfaces [4]. It allows viewing, graphing, logging and playback of real-time diagnostic data through the vehicle's OBD-II diagnostic port. It also allows you to view vehicle diagnostic trouble codes, real-time auto data, and other vehicle information.

PCMSCAN supports almost all modern cars.

Benefits:

- provides wide possibilities of diagnostics;
- error reading in real time;
- error clearing;
- compatible with almost all makes and models of cars that were released after 1996.

Disadvantages:

- only compatible with Windows OS;
- requires a deep understanding and skills in car diagnostics.

Scanners. BAFX Products Bluetooth Scanner – this scanner has one of the best ratings on the Amazon platform. It is compatible with Android and Windows and allows you to read Check Engine codes, reset/clear the Check Engine Light, read data from various car sensors, check if the car is ready to pass the emissions test and much more.

Panlong Bluetooth OBD2 Car Diagnostic Scanner – compact scanner from the Panlong company has a good

rating on Amazon, plus it boasts an attractive price. It allows you to read errors and various car data that are not available on the instrument panel.

ELM327 – one of the most popular and widespread mobile scanners on the market, it provides the ability to connect using Bluetooth, supports Android and Windows. Also one of the cheapest. This scanner will be used for investigation.

All information from the car is sent using a special ELM327 adapter that connects via the OBD-2 port and transmits all data via Bluetooth. MCP-2551 is used for communication of ELM-327 with CAN. The communication with the ECU can be achieved with these two integrated tools. Wired or wireless communication can be used to transmit this communication to an external device. [5]

Before starting the computer diagnostics, you need to establish a connection with the scanner and execute the initialization commands.

Interaction with car. Access to ECU data with mobile devices of the drivers is performed by diagnostic device connected to OBD-II connector [6].

Once all the commands supported by the car have been defined, you can send requests and get the information you need. But to interpret it in a form acceptable to a person, you need to decipher the codes returned by the machine. That is, to have a certain base of all codes.

Each command has its own unique identifier, the response format, that is, the number of bytes it returns, the range of values, in which the response is returned.

As mentioned earlier, special codes are parameter identifiers that are used to request diagnostic information from cars. The SAE J1979 standard defines a standard list of such codes, but manufacturers are free to add their own specific codes. Also, manufacturers may not use all the codes regulated by the standard SAE J1979.

General schema of interaction depicted on Figure 6.

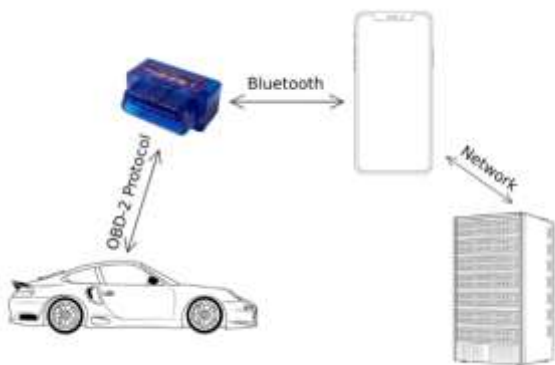


Fig. 6: Schema of interaction with car

First of all, it should be noted that all data from the car will be transmitted using a special adapter ELM327, which connects via the OBD-2 port and transmits all data via Bluetooth. [7]

To get started you need to perform initialization commands, here are the main ones you need to get started:

- ATZ - Resets the adapter to the factory settings;
- ATL0 - Disable line breaks;

- ATE1 - Echo on;
- ATH1 - Headers on;
- ATSTFF - Set the timeout to maximum;
- ATDP - [Describe the current Protocol] The scanner is able to independently determine the protocol of the vehicle to which it is connected;
- ATSP0 - [Set Protocol h] Command to select the protocol in automatic mode.

Based on the above commands, you can generate an initialization string. It will look like this:

Special PID's commands are used to read diagnostic data. PID (Parameter id's) - codes used to require data from the ECU, like RPM in idle speed [8].

The main pads support all cars that have an OBD-2 port. There are also sets of commands for certain makes and types of cars. In our case, the study focuses on basic car diagnostics, so we use a basic set of commands.

Each car supports a certain set of floors. That is, when working with a particular car, you need to operate with a certain set of basic in order not to send the car requests that it does not support. To do this, you need to fulfill a special request, to which the car will return all the floors that it supports. This should be done immediately after the connection is established.

All queries and answers to them are executed in hexadecimal.

To install all supported floors, run a special command with the following code - 0100.

Suppose the car returned the following value - BB1E3211. Next, we translate the result into the binary number system.

BB1E3211(16) is equal to 10111011000111100011001000010001(2).

Using the general floor table shown in Figure 7, determine which ones are supported.

1	0	1	1	1	0	1	1	0	0	0	1	1	1	1	0	0	0	1
+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
01	02	03	04	05	06	07	08	09	0A	0B	0C	0D	0E	0F	10	11	12	13
1	0	0	1	0	0	0	0	1	0	0	0	1						
+																		
14	15	16	17	18	19	1A	1B	1C	1D	1E	1F	20						

Fig. 7: Table of codes

Based on these data, we can determine that our car supports the following types: 01, 03, 04, 05, 07, 08, 0C, 0D, 0E, 0F, 13, 14, 17, 1C, 20.

Now we will use only supported commands to increase productivity.

This is usually not the whole list of floors, but only a small part of them, given as an example.

Car errors can also be different and there are separate commands for them. example:

- 03 - To display saved error codes;
- 0A - To display constant error codes.

As with other teams, car errors come in coded form, respectively, as in other teams, they need to be decoded to obtain the necessary information.

After the car responds to the request, you will need to decrypt its response and find the appropriate code in the database, and then display all the information about the error. [9].

The main advantage of OBD2 technology is the unification of ports, as well as standards for data transmission and interaction with them.

It was found that all modern cars, which are equipped with diagnostic ports of the OBD2 standard, have a universal standard of interaction with the car, as well as a certain unified set of commands, which is subject to the SAE J1979 standard, which allows creating a single interface for interaction with the car [10].

At the same time, there are nuances, such as the fact that automakers are free to deviate from this standard, to introduce their own innovations in their cars, but for my task of high-level diagnostics, this is not a problem. After all, based on open sources, the basic diagnostic commands are universal for most machines.

That is, if there is an open base of basic commands, as well as their interpretation, most modern cars can be diagnosed with a universal diagnostic device.

The end user needs to purchase an ELM327 type adapter, download a special application and use the system.

When developing a software product, it is advisable to use a flexible development methodology, because when writing a system, you need to be ready for changes in requirements and quick correction of shortcomings.

Since, according to the specifics of working with the ELM327 adapter, requests are sometimes executed for a long time, due to the fact that data transmission using the RS232 protocol occurs sequentially, it will be appropriate to use an asynchronous model of interaction with the adapter based on messages using broker messages.

The main problem is to check the correctness of working with this or that car, because the features of working with them may differ not even between brands or models, but even between complete sets of the same model, depending on the presence of certain electronic options [11].

So after developing a system with basic functions, a very important stage is testing on different makes and models of cars. The proposed system will greatly shorten the time to detect vehicle trouble condition [12].

Obviously, it is impossible to test the product on a significant percentage of all cars. Therefore, it will be advisable to organize open beta testing after closed testing, to encourage owners of different cars to cover a larger percentage of tested cars and collect data on defects related to specific models for their further elimination.

Conclusions. As a result of the article, a scheme of research was developed on the basis of which a software system will be developed, which will provide opportunities to use OBD-2 technology to diagnose the car. It will also contain the functions of the on-board computer, collect statistics. Also, was researched history of that technology development, as well as the features of its use in software systems were analyzed

Different models of scanners currently on the market were compared and the most affordable and suitable for this task of high-level car diagnostics was selected. The scheme of interaction with the car is considered using special commands that allow you to get any information about the vehicle in real time.

The practical significance of the obtained results is that the model of interaction of the car with the help of OBD-2 technology with the program proposed in the article is a theoretical basis for the development of a system analogous to the on-board computer of the car.

References

- Gallardo F. *Extraction and analysis of car driving data via OBD-II: Bachelor thesis: 09.18.* Ельче, 2018. 65 p.
- Rodríguez A., Alvarez J., Rodríguez R. Implementation of an OBD-II diagnostics tool over CAN-BUS with Arduino. *Sistemas & Telematic.* 2018. Vol. 16, no. 45. P. 45–53.
- TOAD Pro Homepage: Perfect OBD2 Programming Software For Laptop.* URL: <https://www.obdadvisor.com/toad-pro/> (дата звернення: 20.01.2023).
- PCMSCAN Features Homepage. Palmer performance engineering, inc.* URL: <https://www.palmerperformance.com/products/pcmscan/index.php> (дата звернення: 01.02.2023).
- Süzen A.A., Kayaalp K. Web based tracking of vehicle fault and performance data on OBD II. *Techno-Science.* 2018. P. 13–16.
- Türker G. F., Kutlu A. Survey of Smartphone Applications Based on OBD-II for Intelligent Transportation Systems. *Gül Fatma Türker Int. Journal of Engineering Research and Applications.* 2016. Vol. 6. P. 69–73.
- Golian V., Rybitskiy, O. Methods, models and means of interaction with the car using obd-2 diagnostic systems. *InterConf.* 2022. P. 467–472.
- Rimpasa D., Papadakis A., Samarakou M. OBD-II sensor diagnostics for monitoring vehicle operation and consumption. *Tmrees, EURACA.* 2019. Vol. 6. P. 55–63.
- OBD-II DTC List.* URL: <http://www.totalcardiagnostics.com/support/Knowledgebase/Article/View/21/0/complete-list-of-obd-codes-generic-obd2-obdii--manufacturer> (дата звернення: 10.02.2023).
- Moniaga J., Manalu S., Hadipurnawan, D. Sahidi F. Diagnostics vehicle's condition using obd-ii and raspberry pi technology. *IOP Conf. Series: Journal of Physics.* 2018. DOI: 10.1088/1742-6596/978/1/012011.
- Smith J., Johnson A. Integration of OBD-II Technology for Vehicle Diagnostic in Information Systems. *Proceedings of the International Conference on Information Systems (ICIS).* 2022. Vol. 1. P. 45–51.
- Jhou J., Chen S. The Implementation of OBD-II Vehicle Diagnosis System Integrated with Cloud Computation Technology. *Intelligent Data analysis and its Applications.* 2014. Vol. 1. P. 413–420.

References (transliterated)

- Gallardo F. *Extraction and analysis of car driving data via OBD-II: Bachelor thesis.* Elche, 2018. 65 p.
- Rodríguez A., Álvarez J., Rodríguez R. Implementation of an OBD-II diagnostics tool over CAN-BUS with Arduino. *Sistemas & Telematic.* 2018, vol. 16, no. 45, pp. 45–53.
- TOAD Pro Homepage: Perfect OBD2 Programming Software For Laptop.* Available at: <https://www.obdadvisor.com/toad-pro/> (access date: 20.01.2023).
- PCMSCAN Features Homepage. Palmer performance engineering, inc.* Available at: <https://www.palmerperformance.com/products/pcmscan/index.php> (access date: 01.02.2023).
- Süzen A.A., Kayaalp K. Web based tracking of vehicle fault and performance data on OBD II. *Techno-Science.* 2018, pp. 13–16.
- Türker G. F., Kutlu A. Survey of Smartphone Applications Based on OBD-II for Intelligent Transportation Systems. *Gül Fatma Türker Int. Journal of Engineering Research and Applications.* 2016, vol. 6, pp. 69–73.
- Golian V., Rybitskiy, O. Methods, models and means of interaction with the car using obd-2 diagnostic systems. *InterConf.* 2022, pp. 467–472.
- Rimpasa D., Papadakis A., Samarakou M. OBD-II sensor diagnostics for monitoring vehicle operation and consumption. *Tmrees, EURACA.* 2019, vol. 6, pp. 55–63.
- OBD-II DTC List.* Available at: <http://www.totalcardiagnostics.com/support/Knowledgebase/Article/>

- View/21/0/complete-list-of-obd-codes-generic-obd2-obdii--manufacturer (access date: 10.02.2023).
10. Moniaga J., Manalu S., Hadipurnawan, D. Sahidi F. Diagnostics vehicle's condition using obd-ii and raspberry pi technology. *IOP Conf. Series: Journal of Physics*. 2018, DOI:10.1088/1742-6596/978/1/012011.
 11. Smith J., Johnson A. Integration of OBD-II Technology for Vehicle Diagnostic in Information Systems. *Proceedings of the International Conference on Information Systems (ICIS)*. 2022, vol. 1, pp. 45–51.
 12. Zhou J., Chen S. The Implementation of OBD-II Vehicle Diagnosis System Integrated with Cloud Computation Technology. *Intelligent Data analysis and its Applications*. 2014, vol. 1, pp. 413–420.

Received 15.05.2023

УДК 004.42:004.45

О. М. РИБИЦЬКИЙ, аспірант кафедри програмної інженерії, Харківський національний університет радіоелектроніки, м. Харків, Україна, e-mail: oleksandr.rybitskyi@nure.ua, ORCID: <https://orcid.org/0000-0002-8508-7269>

В. В. ГОЛЯН, кандидат технічних наук (PhD), Харківський національний університет радіоелектроніки, доцент кафедри програмної інженерії, м. Харків, Україна, e-mail: vira.golan@nure.ua, ORCID: <https://orcid.org/0000-0001-5981-4760>

Н. В. ГОЛЯН, кандидат технічних наук (PhD), Харківський національний університет радіоелектроніки, доцент кафедри програмної інженерії, м. Харків, Україна, e-mail nataliia.golian@nure.ua, ORCID: <https://orcid.org/0000-0002-1390-3116>

З. В. ДУДАР, професор, кандидат технічних наук (PhD), Харківський національний університет радіоелектроніки, завідувач кафедри програмної інженерії, м. Харків, Україна, e-mail zoia.dudar@nure.ua, ORCID: <https://orcid.org/0000-0001-5728-9253>

О. В. КАЛИНИЧЕНКО, кандидат технічних наук (PhD), Харківський національний університет радіоелектроніки, доцент кафедри програмної інженерії, м. Харків, Україна, e-mail olga.kalynychenko@nure.ua, ORCID: <https://orcid.org/0000-0003-1466-3967>

Д. М. НІКІТІН, аспірант кафедри програмної інженерії, Харківський національний університет радіоелектроніки, м. Харків, Україна, e-mail: nikitin27959@gmail.com, ORCID: <https://orcid.org/0000-0003-4388-4996>

ВИКОРИСТАННЯ ТЕХНОЛОГІЇ OBD-2 ДЛЯ ДІАГНОСТИКИ АВТОМОБІЛІВ ТА ЇЇ ВИКОРИСТАННЯ В ІНФОРМАЦІЙНІЙ СИСТЕМІ

У статті розглядаються дослідження технології OBD-2 для взаємодії з бортовими системами автомобіля, створення уніфікованої системи, яка може працювати з різними марками і моделями автомобілів. Описана історія технології OBD-2, її розвитку, існуючі стандарти та їх втілення у сучасних автомобілях. Описано основні діагностичні функції, за допомогою яких можна отримувати інформацію про автомобіль, а також виконувати його налаштування та надсилання йому різних команд. До того ж розглянуто апаратну частину сканера, його технічні особливості, нюанси взаємодії з ним та призначення кожного з його пінів. Також розглянуто кілька моделей сканерів, які можна використовувати для розробки цієї системи. Зроблено висновки про доцільність їх використання з урахуванням їх можливостей. Розглянуто існуючі програмні аналоги, їх основні функції, переваги та недоліки. Також розглянуто загальну концепцію взаємодії з автомобілем через OBD-2, алгоритм взаємодії та покроково розібрано взаємодію сканера і автомобіля. Розглянуто програмну частину взаємодії між автомобілем і сканером, спеціальні коди, що використовуються для надсилання запитів та відповідей, їх подальший аналіз та інтерпретація у придатний для людини вигляд. Висвітлено можливості щодо уніфікації цих кодів для різних автовиробників та бази загальних кодів для створення універсальної системи, яка буде підходити для різних автомобілів різних автовиробників. Запропоновано модель програмної системи, що зможе втілити в собі великий набір корисних функцій для будь-якого автомобіліста та буде сумісною з великим числом сучасних машин обладнаних універсальними засобами діагностики у при цьому використовуючи доступну модель сканера без застосування дорогого професійного обладнання.

Ключові слова: Автомобіль, бортовий комп'ютер, комп'ютерна діагностика, сканер, OBD2. Інформаційна система.

Повні імена авторів / Author's full names

Автор 1 / Author 1: Рибіцький Олександр Максимович, Rybitskyi Oleksandr Maksymovych

Автор 2 / Author 2: Голян Віра Володимирівна, Golian Vira Volodymyrivna

Автор 3 / Author 3: Голян Наталія Вікторівна, Golian Nataliia Viktorivna

Автор 4 / Author 4: Дудар Зоя Володимирівна, Dudar Zoia Volodymyrivna

Автор 5 / Author 5: Калиниченко Ольга Вікторівна, Kalynychenko Olga Viktorivna

Автор 6 / Author 6: Нікітін Дмитро Михайлович, Nikitin Dmytro Mykhailovych

D. M. NIKITIN, Postgraduate Student of the Department of Software Engineering, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine, e-mail: nikitin27959@gmail.com, ORCID: <https://orcid.org/0000-0003-4388-4996>

SPECIFICATION FORMALIZATION OF STATE CHARTS FOR COMPLEX SYSTEM MANAGEMENT

This article presents a formalization approach for the requirements of object-oriented programs with state machines, using a spacecraft control system as a case study. It proposes a state pattern implementation, where each state is represented as a class with clearly defined responsibilities, and the transitions between states are controlled by the state objects themselves. Additionally, the application of model checking, theorem proving, and code generation techniques are discussed. The effectiveness of the proposed approach in ensuring compliance with the specified requirements is demonstrated, while also identifying potential drawbacks and limitations of the approach. The implementation is validated using a range of formal verification techniques, including model checking and theorem proving. The article also discusses how the approach can be extended and applied to other complex systems. Overall, the valuable insights into the formalization of requirements for object-oriented programs with state machines are provided, offering a practical and effective approach for verifying the correctness and completeness of such implementations. The results of this work have important implications for the development of safety-critical systems and can potentially improve the quality and reliability of software systems in various domains. By using mathematical models and rigorous formal methods, it is possible to detect and eliminate errors early in the development process, leading to higher confidence in the correctness of the final product. Future research in this area could explore the use of more advanced techniques, such as model-driven development and automatic code synthesis, to further streamline the software development process. Additionally, the development of more efficient and user-friendly tools could make these techniques more accessible to a wider range of developers and organizations. Altogether, the combination of formal methods and software engineering has the potential to revolutionize the way software systems are designed, developed, and verified, leading to safer and more reliable software for critical applications.

Keywords: formal methods, automated programming, state machines, model checking, theorem proving, code generation, object-oriented programming, spacecraft control, requirements formalization, verification and validation.

Introduction. Formalization of requirements for automated object-oriented programs involves the process of translating natural language requirements into a precise and unambiguous specification that can be used to guide the development of software systems. Object-oriented programming (OOP) is a popular approach to software development that emphasizes modular design, code reuse, and encapsulation of data and functionality within objects.

To formalize requirements for OOP programs, developers use a combination of textual descriptions, graphical models, and formal languages such as Unified Modeling Language (UML) or Object Constraint Language (OCL). UML provides a standard notation for modeling software systems, including class diagrams, sequence diagrams, and state machine diagrams, which can be used to visualize the structure and behavior of software components. OCL is a formal language for specifying constraints and operations on objects in an OOP system.

One key benefit of formalizing requirements for OOP programs is that it helps to minimize ambiguity and inconsistency in the software development process. By using a formal language to express requirements, developers can identify potential issues or conflicts early in the development cycle and ensure that the resulting software system meets the desired functional and non-functional requirements. Formalization of requirements also facilitates collaboration among developers, stakeholders, and end-users by providing a common language for discussing and refining requirements.

However, formalization of requirements for OOP programs can also be a challenging and time-consuming process. It requires a deep understanding of both the application domain and the OOP paradigm, as well as expertise in modeling and formal languages. Additionally, there is a risk of over-specifying requirements, which can

lead to inflexibility and difficulties in adapting to changing user needs or system requirements. Thus, it is important to strike a balance between formalization and flexibility, and to involve all relevant stakeholders in the requirements engineering process.

Specification Formalization of State Charts.

Specification formalization of state charts involves defining the behavior of a system using a graphical notation that represents states, transitions, and actions in a structured and systematic way [1]. State Charts can be used to model complex systems and provide a clear and concise way to specify the behavior of a system [2]. Such formalization involves creating a precise and unambiguous specification that can be used to verify the correctness of the system.

There are several formal methods that can be used to specify and analyze state charts, including model checking, theorem proving, and code generation. These methods can help to detect errors in the system design and ensure that the system meets its requirements.

Formalization of state charts is particularly important for safety-critical systems, where errors in the system design could have serious consequences. In these systems, formal methods can be used to verify that the system meets safety requirements and that it behaves correctly under all possible conditions.

An automated state machine can be used to control the behavior of a vehicle during its mission. The state machine can be defined using a set of mathematical formulas that describe the transition between states and the actions to be taken in each state [3].

For example, let us consider a spacecraft that is designed to perform a series of maneuvers, including attitude control, trajectory correction, and payload deployment. The state machine for this spacecraft could be

© Nikitin D. M. 2023



Research Article: This article was published by the publishing house of *NTU "KhPI"* in the collection "Bulletin of the National Technical University "KhPI" Series: System analysis, management and information technologies." This article is distributed under a Creative Common [Creative Common Attribution \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/). **Conflict of Interest:** The author/s declared no conflict of interest.



defined using a set of differential equations that govern the spacecraft's motion and the forces acting on it.

The state variables in this case could include the spacecraft's position, velocity, and attitude, as well as the state of its propulsion and control systems. The state machine could be defined using a set of mathematical equations that specify the transition between states and the actions to be taken in each state.

For instance, the state machine could transition from the attitude control state to the payload deployment state when certain conditions are met, such as reaching a certain altitude or orientation. The mathematical formula for this transition could involve calculating the spacecraft's position and velocity relative to the payload and adjusting the attitude and thruster firing accordingly.

Let's present a spacecraft that needs to perform an attitude control maneuver to align its sensors with a target object in space. The spacecraft's attitude can be described by its orientation relative to a reference frame, such as the Earth-centered inertial (ECI) frame.

The spacecraft's attitude is described by a quaternion

$$q = [q_0, \mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3], \quad (1)$$

where q_0 represents the scalar component and $\mathbf{q}_1, \mathbf{q}_2$ and \mathbf{q}_3 represent the vector component of the quaternion. The goal of the attitude control maneuver is to adjust the quaternion to a desired value that corresponds to the desired orientation of the spacecraft. The equations could be represented using matrix algebra, as follows:

$$\mathbf{I}\dot{\mathbf{w}} + \mathbf{W}\mathbf{I}\mathbf{w} = \mathbf{B}\mathbf{U}, \quad (2)$$

where \mathbf{I} – spacecraft's inertia matrix of the spacecraft, which is a 3x3 matrix representing the distribution of mass in the spacecraft about its center of mass;

\mathbf{w} – spacecraft's angular velocity vector;

\mathbf{W} – skew-symmetric matrix of \mathbf{w} ;

\mathbf{B} – spacecraft's control torque matrix, which represents the external torque applied to the spacecraft by the control system. It is a 3x3 matrix that is determined by the control law used to adjust the spacecraft's attitude;

\mathbf{U} – control input vector, which represents the control commands issued by the spacecraft's control system. It is a 3-dimensional vector that is determined by the control law.

The dot notation denotes the time derivative of a variable, representing its rate of change over time.

\mathbf{W} is a matrix representation of the cross product of \mathbf{w} with itself, defined as follows:

$$\mathbf{W} = \begin{pmatrix} 0 & -w_3 & w_2 \\ w_3 & 0 & -w_1 \\ -w_2 & w_1 & 0 \end{pmatrix}. \quad (3)$$

The equation (1) states that the rate of change of the spacecraft's angular momentum ($\mathbf{I}\dot{\mathbf{w}}$) is equal to the external torque applied to it ($\mathbf{B}\mathbf{U}$), with the skew-symmetric matrix of \mathbf{w} (i.e., $\mathbf{W}\mathbf{I}\mathbf{w}$) representing the Coriolis and centrifugal forces acting on the spacecraft. The equation is a second-order ordinary differential equation and can be solved

numerically to obtain the angular velocity vector \mathbf{w} as a function of time.

The external torque is generated by the spacecraft's control system, which adjusts the angular velocity of the spacecraft in response to the error between the desired and actual quaternion values.

The control torque matrix \mathbf{B} and the control input vector \mathbf{U} can be derived using a control law that minimizes the error between the desired and actual quaternion values. This control law could be represented using a formula such as:

$$\begin{cases} \mathbf{B} = -\mathbf{k}_p\mathbf{Q} - \mathbf{k}_d\mathbf{W}, \\ \mathbf{U} = \mathbf{k}_p\mathbf{q} + \mathbf{k}_d\dot{\mathbf{q}}, \end{cases} \quad (4)$$

where \mathbf{q} is the error quaternion between the desired and actual orientations, \mathbf{Q} is the skew-symmetric matrix of \mathbf{q} , and \mathbf{k}_p and \mathbf{k}_d are the proportional and derivative gain matrices, respectively.

The control law uses the error quaternion and the angular velocity of the spacecraft to calculate the control torque and input vectors that adjust the spacecraft's attitude. The proportional and derivative gains, \mathbf{k}_p and \mathbf{k}_d are tuning parameters that determine the response of the control system to changes in the error and velocity.

These equations form the basis of a closed-loop control system that adjusts the spacecraft's attitude to the desired orientation. The state machine can transition to the next state once the desired orientation is achieved, such as when the spacecraft's sensors are aligned with the target object in space.

Here is an example Python code listing that implements the formulas described earlier:

```
import numpy as np
def spacecraft_dynamics(I, w, B, U):
    """
    Computes the derivative of angular
    velocity vector w
    Args:
        I: 3x3 inertia matrix
        w: 3-dim angular velocity vector
        B: 3x3 control torque matrix
        U: 3-dim control input vector
    Returns:
        3-dimensional array representing the
        time derivative of the angular velocity
        vector w
    """
    w_dot = np.linalg.inv(I).dot(-
np.cross(w, I.dot(w), axisa=0, axisb=0) +
B.dot(U))
    return w_dot
```

This code defines a function called `spacecraft_dynamics` that takes as input the spacecraft's inertia matrix \mathbf{I} , angular velocity vector \mathbf{w} , control torque matrix \mathbf{B} , and control input vector \mathbf{U} , and computes the time derivative of \mathbf{w} using the spacecraft dynamics equation. The `numpy` module is used to perform the necessary matrix operations, such as matrix inversion and cross products.

State Charts in OOP Domain. In the "State" object-oriented design pattern, each component of the model would have its own state object that encapsulates the behavior and data specific to that state [4]. Here's how the state objects could be defined and described for each component of the spacecraft dynamics model:

1. Angular Velocity State: represents the current angular velocity of the spacecraft. It has a single variable, w , that stores the current angular velocity vector. The state object provides methods to update the angular velocity vector and to compute its time derivative using the spacecraft dynamics equation.

2. Inertia State: represents the current inertia matrix of the spacecraft. It has a single variable, I , that stores the current inertia matrix. The state object provides methods to update the inertia matrix.

3. Control Torque State: represents the current control torque applied to the spacecraft. It has a single variable, B , which stores the current control torque matrix. The state object provides methods to update the control torque matrix.

4. Control Input State: represents the current control input commands issued by the spacecraft's control system. It has a single variable, U , which stores the current control input vector. The state object provides methods to update the control input vector.

Each state object has its own set of methods that allow it to interact with other states and components in the spacecraft dynamics model. For example, the Angular Velocity State might have a method that computes the Coriolis and centrifugal forces acting on the spacecraft, given the current inertia matrix, and control torque and input states. Similarly, the Control Input State might have a method that generates control input commands based on the current spacecraft state.

Using the "State" design pattern can help to modularize the spacecraft dynamics model, making it easier to modify and extend in the future. By encapsulating the behavior and data specific to each state in its own object, the overall complexity of the model can be reduced, and its overall structure made more maintainable [5].

In the provided spacecraft example, the states can be replaced with each other through a process of state transitions. This means that as the spacecraft system runs, each state object can update its own internal state and then transition to a new state object, which will take over control of the system.

The process of state transition can be controlled by the spacecraft control software, which can determine when a state object should transition to a new state based on certain conditions. For example, the control software may trigger a state transition when a certain time has elapsed, when a certain event occurs, or when certain sensor readings meet certain thresholds [6].

To implement state transitions in the spacecraft control system, each state class should implement an `update()` method that updates its own internal state, and then returns a new instance of a state class that represents the next state of the system. The control software can then update the current state object with the new state object, allowing the system to transition to the new state.

For example, the `AngularVelocityState` class might implement an `update()` method that reads sensor data to calculate the current angular velocity of the spacecraft, and then returns a new instance of a state class that represents the next state of the system based on that velocity. This new state object might be an instance of the `InertiaState` class, which would update the system's internal state based on the current inertia of the spacecraft.

Overall, by implementing state transitions in this way, the spacecraft control software can dynamically switch between different state objects to control the spacecraft system in a safe and reliable manner.

Formal Methods of State Chart Analysis. To use model checking, theorem proving, and code generation with the spacecraft example, we can first start by creating a formal specification of the system using a modeling language such as Statecharts or Mermaid. This formal specification will represent the desired behavior of the system, including the states and transitions between them, as well as any constraints or requirements that must be satisfied [7].

Once the formal specification is created, we can use model checking and theorem proving techniques to verify that the specification is correct and satisfies the desired requirements. Model checking involves automatically verifying that a model of a system satisfies a given set of properties [8]. Theorem proving, on the other hand, involves manually proving that a model satisfies a set of logical properties using formal logic and mathematical reasoning [9].

Finally, once we have verified that the formal specification is correct, we can use code generation techniques to automatically generate code that implements the desired behavior of the system. This code can be written in a programming language such as C or Python and can be used to control the behavior of the spacecraft in accordance with the formal specification as illustrated in fig. 1.

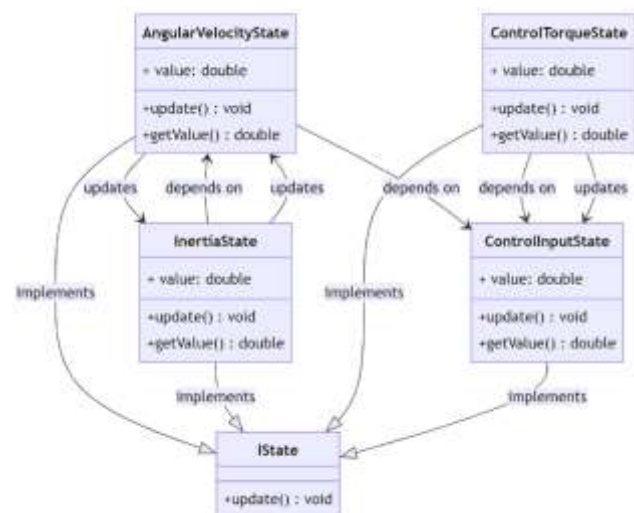


Fig. 1. State chart implementation for the spacecraft

In fig. 1, each state class (`AngularVelocityState`, `InertiaState`, `ControlTorqueState`, and `ControlInputState`) contains a `value` property, `update()` and `getValue()` methods. The `update()` method changes the state of the

object, while the `getValue()` method retrieves the current value of the object's state.

The `IState` interface contains only an `update()` method, which is implemented by each of the state classes.

This diagram represents the relationships between the four state classes and the `IState` interface that they all implement. It also shows the dependencies between the states (for example, `AngularVelocityState` depends on `ControlInputState`) and the state updates (for example, `InertiaState` updates `AngularVelocityState`).

Overall, the use of model checking, theorem proving, and code generation can help ensure the correctness and reliability of the spacecraft's control system and can help mitigate the risks associated with space missions [10].

First, model checking can be used to verify that the state transitions in the spacecraft control software are correct and satisfy the system requirements. A model of the spacecraft control software can be created, and the model checker can explore all system states to check that the software behaves as intended under all possible conditions.

Second, theorem proving can be used to formally prove that the state transitions in the spacecraft control software are correct and satisfy the system requirements. A formal specification of the spacecraft control software can be created, and the theorem prover can use mathematical logic to prove that the software satisfies the specified requirements.

Finally, code generation can be used to automatically generate executable code from the formal specification of the spacecraft control software. The formal specification can serve as a precise and unambiguous description of the software behavior, and the code generator can automatically generate code that faithfully implements the specified behavior.

By applying these formal methods to the spacecraft control software, we can ensure that the software is correct and reliable, and that it satisfies the system requirements. This can help to reduce the risk of errors and malfunctions in the software.

Model Checking. To verify that the spacecraft control software behaves correctly under all conditions, we can model the software using a state machine and use a model checker to explore all possible system states and check that the software behaves as intended [11]. Here is an example of a state machine model in Promela, a modeling language for the Spin model checker:

```
mtype      =      {IDLE,      ACCELERATING,
DECELERATING};
mtype state = IDLE;
active proctype spacecraft() {
do
:: state == IDLE ->
state = ACCELERATING;
:: state == ACCELERATING ->
state = DECELERATING;
:: state == DECELERATING ->
state = IDLE;
od
}
```

In this model, the spacecraft can be in one of three states: `IDLE`, `ACCELERATING`, or `DECELERATING`. The spacecraft transitions between states based on certain conditions, and the model checker can explore all system states to check that the software behaves correctly under all conditions.

Theorem Proving. If we want to prove that the spacecraft operates as expected we can use a theorem prover to create a formal specification of the software, and then use mathematical logic to prove that the software satisfies the requirements [12]. Here is an example of a specification of the spacecraft software in Z notation:

```
state ::= IDLE | ACCELERATING |
DECELERATING
SPC ::= [state: state]
InitSPC == state = IDLE
AccelerateSPC == state = ACCELERATING
DecelerateSPC == state = DECELERATING
NextSPC      ==      AccelerateSPC      \/
DecelerateSPC \/ (state = IDLE /\ state' /=
IDLE)
```

This specification defines the initial state, the state transitions, and the constraints on the possible state transitions. We can use a theorem prover to prove that the specification is correct, and that the software satisfies the specified requirements.

Code Generation. When there is a formal specification of the spacecraft control software, and we need to automatically generate executable code that implements the specified behavior, we can use a code generator to automatically generate code from the formal specification [13]. Here is an example of how we could generate C# code from the Z specification above using the Zing code generator:

```
public class SPC {
private enum state { IDLE,
ACCELERATING, DECELERATING };
private state _state = state.IDLE;
public void AccelerateSPC() {
_state = state.ACCELERATING;
}
public void DecelerateSPC() {
_state = state.DECELERATING;
}
public void NextSPC() {
if (_state == state.IDLE) {
_state = state.ACCELERATING;
} else if (_state ==
state.ACCELERATING) {
_state = state.DECELERATING;
} else if (_state ==
state.DECELERATING) {
_state = state.IDLE;
}
}
}
```

Contract Compliance. To ensure compliance with the contract of the states in the spacecraft example, we can use a combination of static code analysis tools and automated testing.

Static code analysis tools can help detect violations of coding standards and best practices, as well as potential errors and vulnerabilities in the code. These tools can analyze the code and flag any violations of the interface contract, such as missing or incorrect method signatures or non-compliant access modifiers.

Automated testing can help ensure that the code adheres to the interface contract by verifying that each state class behaves as expected [14]. Unit tests can be written to verify that the `update()` and `getValue()` methods of each state class perform their expected functions, and that the state transitions between the classes are correct. Integration tests can also be written to test the system as a whole, and to verify that the interactions between the state classes are correct.

By using both static code analysis tools and automated testing approach, we can ensure that the code adheres to the contract of the states and prevent bugs and errors in the system.

State Chart Implementation Concerns. The state pattern provides a way to encapsulate state-specific behavior in separate classes and allows for the object's behavior to change dynamically as its state changes [15]. While the state pattern can be a useful tool for designing complex software systems, there are potential drawbacks to its implementation:

1. Increased memory usage: as each state object maintains its own internal state and any associated data, this can lead to increased memory usage, which may be a concern in systems with limited memory resources;

2. Complex object interactions: as the state objects interact with each other to transition between states, the code can become more complex and difficult to follow. This may make it harder to debug and maintain the code over time;

3. Potential for errors in state transition logic: the transition logic between states is implemented in each state object's update method, which may lead to errors if not implemented correctly. For example, if the state transition conditions are not properly defined, the system may get stuck in a certain state or transition to the wrong state.

4. Increased development time: implementing the state pattern can be more time-consuming than other approaches to managing state, such as using a switch statement or if-else blocks, which may lead to longer development times and increased costs.

5. Potential for performance issues: as the state objects update and transition between states at run time, this may introduce performance overhead and impact the system's overall performance.

Conclusions. The use of formalization techniques such as model checking, theorem proving, and code generation can greatly enhance the reliability and safety of automated systems such as spacecraft control systems. By formalizing the system requirements and specifications, potential errors and bugs can be caught early in the development process, reducing the risk of catastrophic failures.

In the context of the discussed spacecraft example, it was demonstrated how the use of state machines and the

state design pattern can provide a structured approach to modeling and implementing complex control systems.

The use of formal methods for developing reliable and correct software systems was explored, specifically in the context of state machines.

The results present a detailed example of using the state pattern to model the behavior of a spacecraft and how formal verification techniques can help ensure the correctness and completeness of the system design. The article demonstrates how model checking can be used to detect potential errors and violations of the system requirements and how theorem proving can be used to formally verify the accuracy of the system's behavior.

Despite the benefits of using formal methods for developing state machine-based systems, there are also some limitations and challenges that need to be addressed. Some of the drawbacks of using this approach, such as the complexity of the mathematical models and the high computational costs of verification techniques, have been discussed.

Overall, the use of formal methods is a promising approach for developing reliable and correct software systems based on state machines. By ensuring the correctness and robustness of state machine-based systems, we can increase their security and overall quality, which is especially important for safety-critical systems such as those used in aerospace and transportation.

References

1. Lodi S., Mesiti M., Orsi G. A state machine for relational databases. *In Proceedings of the «34th IEEE/ACM International Conference on Automated Software Engineering»*. 2019. P. 114–125.
2. Liggesmeyer P., Seib E., Prehofer C. A state machine approach for modeling and testing autonomous driving functions. *In Proceedings of the «2021 IEEE Intelligent Vehicles Symposium»*. 2021. P. 2854–2859.
3. Giannakopoulou D., Pasareanu C., Rungta N. An integrated approach to analyzing and testing stateful systems. *In Proceedings of the «32nd IEEE/ACM International Conference on Automated Software Engineering»*. 2017. P. 943–948.
4. Saenz J. C., Perez-Palacin D., d'Amorim M. Behavior-driven development of stateful systems: a case study of a medical information system. *In Proceedings of the «2019 IEEE/ACM International Conference on Automated Software Engineering»*. 2019. P. 1011–1016.
5. Azevedo G., Ribeiro M., Medeiros F. Model-based test generation for stateful systems using an FSM language. *In Proceedings of the «2018 IEEE International Conference on Software Testing, Verification and Validation»*. 2018. P. 237–247.
6. Daumke P., Laroche L., Graubner S. A state machine-based approach for the dynamic adaptation of software systems. *In Proceedings of the «14th International Conference on Software Technologies»*. 2019. P. 466–475.
7. Lo D., Liu Y., Xie X., Wong S. Symbolic execution of stateful programs with abstract state machines. *In Proceedings of the «28th ACM SIGSOFT International Symposium on Software Testing and Analysis»*. 2019. P. 285–296.
8. Li Y., Li Y., Dong J. S. Formally verifying the state machine-based software through the UPPAAL model checker. *Journal of Intelligent & Fuzzy Systems* 36(4). 2019. P. 3657–3668.
9. Chen Q., Liu S., Wang S., Sun J. Efficient generation of state machine models from Java source code for vulnerability detection. *Journal of Systems and Software*, 177. 2021. 237 p.
10. Jovanovic J., Rackovic M., Milicic M. Analysis of the role of state machine diagrams in software development: An exploratory study. *Information and Software Technology*, 103. 2019. 132–146 p.
11. Chen X., Chen J., Wang J. Research on Formalization Method of State Transition Rules for Automated Vehicle Systems. *IEEE Access*, 8. 2020. P. 134116–134127.

12. Eltoweissy M., Alnemr R., Seliem M., Ali M. Formal specification and verification of state-based software systems: A systematic literature review. *Journal of Systems and Software*, 163. 2021. P. 1105–1128.
13. Gaber A., Elragal A. Formalizing requirements for automated driving systems: A systematic literature review. *Safety Science*, 142. 2021. P. 538–549.
14. Gu C., Li X., Liu Y. A Formal Method for Analyzing and Validating the Functionality of Statecharts. *IEEE Access*, 7. 2021. P. 135–154.
15. Taha I., Ahmed E., Al-Mamory S., Karama S. Formalizing Requirements for State Machine Models of Safety-Critical Systems: A Review. *IEEE Access*, 9. 2021. P. 315–333.
6. Daumke P., Laroche L., Graubner S. A state machine-based approach for the dynamic adaptation of software systems. In *Proceedings of the 14th International Conference on Software Technologies*. 2019. P. 466–475.
7. Lo D., Liu Y., Xie X., Wong S. Symbolic execution of stateful programs with abstract state machines. In *Proceedings of the 28th ACM SIGSOFT International Symposium on Software Testing and Analysis*. 2019. P. 285–296.
8. Li Y., Li Y., Dong J. S. Formally verifying the state machine-based software through the UPPAAL model checker. *Journal of Intelligent & Fuzzy Systems* 36(4). 2019. P. 3657–3668.
9. Chen Q., Liu S., Wang S., Sun J. Efficient generation of state machine models from Java source code for vulnerability detection. *Journal of Systems and Software*, 177. 2021. 237 p.
10. Jovanovic J., Rackovic M., Milicic M. Analysis of the role of state machine diagrams in software development: An exploratory study. *Information and Software Technology*, 103. 2019. 132–146 p.
11. Chen X., Chen J., Wang J. Research on Formalization Method of State Transition Rules for Automated Vehicle Systems. *IEEE Access*, 8. 2020. P. 134116–134127.
12. Eltoweissy M., Alnemr R., Seliem M., Ali M. Formal specification and verification of state-based software systems: A systematic literature review. *Journal of Systems and Software*, 163. 2021. P. 1105–1128.
13. Gaber A., Elragal A. Formalizing requirements for automated driving systems: A systematic literature review. *Safety Science*, 142. 2021. P. 538–549.
14. Gu C., Li X., Liu Y. A Formal Method for Analyzing and Validating the Functionality of Statecharts. *IEEE Access*, 7. 2021. P. 135–154.
15. Taha I., Ahmed E., Al-Mamory S., Karama S. Formalizing Requirements for State Machine Models of Safety-Critical Systems: A Review. *IEEE Access*, 9. 2021. P. 315–333.

References (transliterated)

1. Lodi S., Mesiti M., Orsi G. A state machine for relational databases. In *Proceedings of the 34th IEEE/ACM International Conference on Automated Software Engineering*. 2019. P. 114–125.
2. Liggemeyer P., Seib E., Prehofer C. A state machine approach for modeling and testing autonomous driving functions. In *Proceedings of the 2021 IEEE Intelligent Vehicles Symposium*. 2021. P. 2854–2859.
3. Giannakopoulou D., Pasareanu C., Rungta N. An integrated approach to analyzing and testing stateful systems. In *Proceedings of the 32nd IEEE/ACM International Conference on Automated Software Engineering*. 2017. P. 943–948.
4. Saenz J. C., Perez-Palacin D., d'Amorim M. Behavior-driven development of stateful systems: a case study of a medical information system. In *Proceedings of the 2019 IEEE/ACM International Conference on Automated Software Engineering*. 2019. P. 1011–1016.
5. Azevedo G., Ribeiro M., Medeiros F. Model-based test generation for stateful systems using an FSM language. In *Proceedings of the 2018 IEEE International Conference on Software Testing, Verification and Validation*. 2018. P. 237–247.

Received 12.05.2023

УДК 004.053:004.428.2:519.688

Д. М. НИКИТИН, аспірант кафедри програмної інженерії, Харківський національний університет радіоелектроніки
м. Харків, Україна, e-mail: nikitin27959@gmail.com, ORCID: <https://orcid.org/0000-0003-4388-4996>

ФОРМАЛІЗАЦІЯ СПЕЦИФІКАЦІЇ СХЕМ СТАНУ ДЛЯ УПРАВЛІННЯ СКЛАДНИМИ СИСТЕМАМИ

У статті представлено підхід формалізації для вимог об'єктно-орієнтованих програм із кінцевими автоматами з використанням як прикладу системи керування космічним апаратом. Запропоновано реалізацію шаблону стану, де кожен стан представлено як клас із чітко визначеними обов'язками, а переходи між станами контролюються самими об'єктами стану. Крім того, обговорюється застосування методів перевірки моделі, доведення теорем і генерації коду. Продемонстровано ефективність запропонованого підходу щодо забезпечення відповідності зазначеним вимогам, а також виявлено потенційні недоліки та обмеження підходу. Реалізація перевіряється за допомогою низки формальних методів перевірки, включаючи перевірку моделі та доведення теорем. У статті також обговорюється, як цей підхід можна розширити та застосувати до інших складних систем. Загалом, надано детальну інформацію щодо формалізації вимог до об'єктно-орієнтованих програм із кінцевими автоматами, що пропонує практичний та ефективний підхід для перевірки правильності та повноти таких реалізацій. Результати цієї роботи мають важливе значення для розробки критично важливих для безпеки систем і потенційно можуть підвищити якість і надійність програмних систем у різних областях. За допомогою математичних моделей і строгих формальних методів можна виявити й усунути помилки на ранніх стадіях процесу розробки, що веде до більшої впевненості в правильності кінцевого продукту. Майбутні дослідження в цій галузі можуть вивчити використання більш передових методів, таких як розробка на основі моделі та автоматичний синтез коду, для подальшої оптимізації процесу розробки програмного забезпечення. Крім того, розробка більш ефективних і зручних інструментів може зробити ці методи більш доступними для широкого кола розробників і організацій. Загалом, поєднання формальних методів і розробки програмного забезпечення має потенціал революціонізувати спосіб проектування, розробки та перевірки систем програмного забезпечення, створюючи більш надійні та надійніші програмне забезпечення для критичних програм.

Ключові слова: формальні методи, автоматизоване програмування, кінцеві автомати, перевірка моделі, доведення теорем, генерація коду, об'єктно-орієнтоване програмування, управління космічним кораблем, формалізація вимог, верифікація та валідація.

Повне ім'я автора / Author's full name

Нікітін Дмитро Михайлович, Nikitin Dmytro Mykhailovych

DOI: 10.20998/2079-0023.2023.01.17
УДК 519.2

О. М. НИКУЛІНА, д-р техн. наук, професор, завідувачка кафедри інформаційних систем та технологій Національного технічного університету «Харківський політехнічний інститут», Харків, Україна; e-mail: elniknik02@gmail.com; ORCID: <https://orcid.org/0000-0003-2938-4215>

В. П. СЕВЕРИН, д-р техн. наук, професор, професор кафедри системного аналізу та інформаційно-аналітичних технологій Національного технічного університету «Харківський політехнічний інститут», Харків, Україна; e-mail: valerii.severyn@khpri.edu.ua; ORCID: <https://orcid.org/0000-0002-2969-6780>

О. М. КОНДРАТОВ, аспірант кафедри інформаційних систем та технологій Національного технічного університету «Харківський політехнічний інститут», Харків, Україна; e-mail: kondratovolexiy@gmail.com; ORCID: <https://orcid.org/0000-0001-6367-9944>

Н. Ю. РЕКОВА, д-р економічних наук, професор, професор кафедри цифрових технологій та проектно-аналітичних рішень ТОВ «Технічний університет «МЕТІНВЕСТ ПОЛІТЕХНІКА», Запоріжжя, Україна, e-mail: natarekova@gmail.com; ORCID: <https://orcid.org/0000-0003-0956-6564>

АНАЛІЗ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ ДЛЯ ДИСТАНЦІЙНОЇ ІДЕНТИФІКАЦІЇ ДИНАМІЧНИХ ОБ'ЄКТІВ

Розглянуто проблему ідентифікації динамічних об'єктів з використанням інформаційних технологій дистанційної ідентифікації. Зазначено, що ідентифікація рухомих об'єктів має важливе значення в різних сферах, включаючи автономні транспортні засоби, медичну діагностику та робототехніку. Мета статті полягає в аналізі різних інформаційних технологій виявлення об'єктів, які можуть бути використані в майбутніх дослідженнях з дистанційної ідентифікації. Проведено аналіз методів визначення швидкості як динамічного параметру, аналіз двокрокових та однокрокових методів віддаленої ідентифікації об'єктів, аналіз ранніх методів ідентифікації, а також аналіз методів покращення віддаленої ідентифікації об'єктів. Розглянуто кілька засобів визначення руху об'єктів, зокрема, пропорційно-інтегрально-диференціальний регулятор, метод блоку вирівнювання, фазова кореляція, алгоритми піксельної рекурсії та методи оптичного потоку Лукаса – Канаді, Хорна – Шунка, Фарнбека, густого оптичного потоку. Ці засоби можуть бути використані для ефективного визначення руху об'єктів та ідентифікації їхньої швидкості незалежно від розміру та положення об'єктів. Розглянуті двокрокові та однокрокові методи виявлення об'єктів: метод регіонів зі згортковими нейронними мережами, його покращення, мережі пулінгу просторової піраміди, метод "Ти дивишся лише один раз", однокроковий багаторамковий метод, мережі сітківки, кутова мережа, центральна мережа та трансформер виявлення, які використовують різні підходи для покращення продуктивності та точності виявлення об'єктів. Підкреслено необхідність використання методів згорткових нейронних мереж та мереж пулінгу просторової піраміди для ефективної ідентифікації об'єктів незалежно від їхнього розміру та положення. Пропонуються нові підходи, які дозволяють створювати представлення фіксованої довжини для обробки зображень та регіонів інтересу, а також методи Віоли – Джонса, гістограми орієнтованих градієнтів, моделі деформованих частин. Дослідження в області виявлення об'єктів сприяють розвитку інформаційних технологій та покращенню ефективності систем ідентифікації динамічних об'єктів. Шляхом огляду та аналізу різних методів надані рекомендації для дослідників і практиків, що працюють у галузі дистанційної ідентифікації динамічних об'єктів.

Ключові слова: Дистанційна ідентифікація динамічних об'єктів, виявлення об'єктів, оптичний потік, ідентифікація швидкості, глибоке навчання, згорткові нейронні мережі.

Вступ. Дистанційна ідентифікація параметрів динамічних об'єктів (ДІПДО) спрямована на ідентифікацію параметрів об'єктів на відстані з використанням різних датчиків та зображень і є підгалуззю виявлення об'єктів (ВО). ВО є фундаментальним завданням в області комп'ютерного зору з різноманітними застосуваннями, такими як нагляд, автономні транспортні засоби, робототехніка та розумне виробництво. Основною метою ВО є ідентифікація параметрів виявлених об'єктів, включаючи положення, розмір, форму та орієнтацію. Надано короткий огляд підходів до віддаленого визначення параметрів об'єкта, включаючи методи визначення швидкості як динамічного параметру, методи на основі моделей, методи на основі машинного навчання, методи на основі ознак. Підхід на основі машинного навчання включає використання різних алгоритмів навчання, таких як глибоке навчання для визначення параметрів об'єкта з великого набору даних з мітками зображень. Це дослідження підкреслює важливість віддаленого визначення параметрів об'єкта і надає цінні посилання на ресурси для

дослідників та практиків, що працюють в галузі комп'ютерного зору.

Мета та задачі дослідження. Мета статті полягає в аналізі методів ДІПДО для використання в майбутніх дослідженнях. Для досягнення мети поставлено задачі:

- 1) проаналізувати методи визначення швидкості як динамічного параметру.
- 2) проаналізувати методи ВО на основі згорткових нейронних мереж (ЗНМ);
- 3) проаналізувати ранні методи ВО;
- 4) проаналізувати методи, які покращують ВО.

Методи визначення руху об'єктів. Пропорційно-інтегрально-диференціальний (ПІД) регулятор використовується для визначення руху або швидкості об'єкта. Цей регулятор включає три компоненти: пропорційний, інтегральний та диференціальний. Пропорційна складова враховує поточну помилку між заданою і фактичною величиною, і пропорційно збільшує або зменшує вихідний сигнал для наближення до заданого значення. Це дозволяє швидко реагувати на зміни і відповідати на них пропорційно. Інтегральна складова накопичує помилки в часі та враховує їх у

© Нікуліна О. М., Северин В. П., Кондратов О. М., Рекова Н. Ю., 2023



Дослідницька стаття: Цю статтю опубліковано видавництвом *НТУ «ХПІ»* у збірнику «Вісник Національного технічного університету "ХПІ" Серія: Системний аналіз, управління та інформаційні технології». Ця стаття поширюється за міжнародною ліцензією [Creative Common Attribution \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/). **Конфлікт інтересів:** Автор/и заявив/или про відсутність конфлікту.



вихідному сигналі. Це допомагає усунути сталий стан помилки і забезпечує точніше регулювання. Диференціальна складова оцінює швидкість зміни помилки і змінює вихідний сигнал для зменшення осциляцій та збільшення стабільності системи. Завдяки комбінації цих трьох складових ПДД-регулятор дозволяє ефективно визначати рух або швидкість об'єкта і намагається зменшити помилку між заданим значенням і фактичним станом об'єкта. Метод блоку вирівнювання (БВ) використовується для визначення руху об'єктів на відеозаписах. Зображення розбивається на блоки, і для кожного блоку виконується пошук відповідного блоку на наступному кадрі. Метод блоку вирівнювання може бути застосований на різних рівнях, наприклад, на піксельному рівні або на рівні піраміди зображень. Він є простим та ефективним методом для виявлення руху на відеозаписах, однак його точність може бути недостатньою. Фазова кореляція є одним з методів, який може використовуватися для визначення руху або швидкості об'єкта в комп'ютерному зорі та обробці зображень. Цей метод ґрунтується на використанні фазової інформації в сигналах зображень для виявлення переміщення між кадрами або шаблоном та вхідним зображенням. Алгоритми піксельної рекурсії є методами, які використовують аналіз пікселів в зображенні. Кожен піксель у зображенні аналізується окремо і порівнюється зі своїми сусідніми пікселями або попередніми кадрами зображення. За допомогою алгоритмів та методів обробки, таких як порівняння інтенсивності, градієнтів, кластеризація або використання шаблонів, можна виявити зміну позиції або значення пікселя і визначити швидкість руху об'єкта. Також розроблені методи оптичного потоку. Метод оптичного потоку Лукаса-Канади використовує інтенсивність пікселів для визначення оптичного потоку і базується на апроксимації локального рівняння різниці інтенсивності з використанням методу найменших квадратів. Метод оптичного потоку Хорна – Шунка також використовує інтенсивність пікселів і гладкість оптичного потоку, формулює задачу оптимізації для знаходження глобального оптичного потоку з мінімізацією функціоналу енергії. Метод оптичного потоку Фарнбека використовує метод апроксимації поліномами для визначення оптичного потоку, алгоритм побудови піраміди для обробки зображень різних розмірів і отримання точних результатів. Метод густого оптичного потоку використовує густу сітку точок для визначення оптичного потоку на зображенні та надає інформацію про рух в областях зображення [13].

Методи на основі ЗНМ. У 2012 р. відбулося відродження ЗНМ. Основу яких ще покладено у 1998 р. Я. Лекуном та використані для розпізнавання рукописних цифр. Оскільки ЗНМ можуть навчатися стійким і високорівневим ознакам зображення, у 2014 р. для ВО запропоновано метод регіонів зі згортковими нейронними мережами (МРЗНМ). З того часу ВО почало швидко еволюціонувати. У глибинному навчанні виділяють дві групи методів: "методи з двокроковою архітектурою" і "методи з однокроковою архітектурою", де перші розглядають виявлення як "грубе до точного", а другі як "завершення за один

крок". У МРЗНМ спочатку здійснюється виділення набору пропозицій об'єктів, тобто кандидатів у ВО. Потім кожен пропозицію масштабують до зображення фіксованого розміру та вводять у попередньо навчену модель ЗНМ, щоб отримати ознаки. Нарешті, лінійні опорно-векторні машини (ОВМ), які називаються ОВМ-класифікаторами, використовуються для прогнозу наявності об'єкта в кожній області та для визначення класу об'єктів. МРЗНМ значно покращує результати на вибірці візуальних об'єктів класів. Недоліком МРЗНМ є надмірні обчислення ознак на великій кількості пропозицій. З одного зображення може бути понад 2000 прямокутників, що призводить до надзвичайно повільної швидкості виявлення, понад 14 секунд на одне зображення з використанням графічної карти. У 2014 р. були запропоновані мережі пулінгу просторової піраміди (МППП), які дозволяють генерувати представлення фіксованої довжини незалежно від розміру зображення або регіону інтересу. Використання МППП дозволяє обчислити ознаки лише один раз для всього зображення і потім згенерувати представлення фіксованої довжини для регіонів інтересу, що прискорює процес виявлення без втрати точності. Недоліками МППП є необхідність багатокрокового навчання та формування повнозв'язних шарів, ігноруючи попередні шари. У 2015 р. Р. Гіршик запропонував швидкий МРЗНМ (ШМРЗНМ), який є удосконаленням МРЗНМ та МППП. ШМРЗНМ дозволяє одночасно навчати метод та регресор обведення прямокутників в рамках однакової конфігурації мережі. Хоча ШМРЗНМ успішно поєднує переваги МРЗНМ та МППП, його швидкість виявлення все ще обмежена виявленням пропозицій. У 2015 р. запропоновано надшвидкий МРЗНМ (НМРЗНМ), який працює майже в реальному часі. У НМРЗНМ впроваджені мережі регіонів пропозицій, що дозволяє практично безкоштовно генерувати пропозиції регіонів. При переході від МРЗНМ до НМРЗНМ більшість окремих блоків системи ВО, такі як виділення пропозицій, вилучення ознак, регресія обведення прямокутників та інші, були поступово інтегровані в єдину навчальну рамку з кінця до кінця. Хоча все ще збереглася велика кількість обчислень на наступному кроці ВО. У 2017 р. запропоновані пірамідальні мережі ознак (ПМО). Щоб не виконувати ВО лише на картах ознак верхнього шару мережі, розроблена архітектура зверху вниз з бічними зв'язками для побудови високорівневої семантики на всіх шкалах. Оскільки ЗНМ природно формує піраміду ознак під час прямого поширення, ПМО удосконалили ВО з різноманітними масштабами. Використання ПМО в базовій системі НМРЗНМ досягло найкращих результатів ВО з використанням однієї моделі. ПМО стали основним будівельним блоком багатьох майбутніх методів. Таким чином, відродження згорткових нейромереж у 2012 р. спричинило революцію в виявленні об'єктів. Двокрокові методи, починаючи з МРЗНМ, розвивалися з небаченою швидкістю, що призвело до появи НМРЗНМ – першого метода глибинного навчання майже в реальному часі. ПМО покращили ВО за допомогою архітектури згори донизу і бічних зв'язків для побудови високорівневої

семантики на всіх шкалах. Методи з одним кроком на основі згорткових нейронних мереж можуть виявляти всі об'єкти за один крок. Вони популярні серед мобільних пристроїв з функціями, які використовуються в реальному часі та простоті розгортання, але їх продуктивність помітно знижується при виявленні густо розташованих та малих об'єктів. У 2015 р. запропоновано метод «Ти дивишся лише один раз» (МТДЛОР). МТДЛОР працює надзвичайно швидко і використовує одну нейронну мережу для обробки повного зображення. Для МТДЛОР характерне зниження точності локалізації порівняно з двокроковими методами, особливо для деяких малих об'єктів. Також у 2015 р. був запропонований одноходовий багаторамковий метод (ОХБМ), який значно покращує точність однокрокових методів ВО, особливо для певних малих об'єктів. ОХБМ виявляє об'єкти різних масштабів на різних шарах мережі. У 2017 р. досліджені причини відставання за точністю від двокрокових методів ВО і запропоновані мережі сітківки (МС). Вони виявили, що основною причиною є незбалансованість класів переднього та заднього плану. З цією метою в МС була введена нова функція втрат під назвою "фокусована втрата", щоб метод більше уваги звертав на складні невірно класифіковані приклади під час навчання та досягнення такої ж точності в порівнянні з двокроковими методами. У 2018 р. введено кутову мережу (КМ). Для вирішення проблем з нерівномірністю класів, великою кількістю ручних гіперпараметрів і тривалим часом збіжності була відкинута попередня парадигма виявлення та розглянута задача передбачення ключових точок (КТ). Після отримання КТ кутові мережі розгрупувають та знову групують КТ. У 2019 р. К. Жоу та ін. запропонували Центральну мережу (ЦМ), яка також використовує КТ, але виключає витратні післяпроцеси, такі як групове призначення ключових точок (як у КМ, екстремальній мережі тощо) та не максимальне придушення (НМП), що призводить до повноцінної мережі виявлення кінець до кінця. ЦМ вважає об'єкт однією точкою, тобто центром об'єкта, і регресує всі його атрибути, такі як розмір, орієнтація, розташування, на основі посилення на центральну точку. Модель є простою та елегантною, і може використовуватись для трьохвимірних ВО, оцінки пози людини, вивчення оптичного потоку, визначення глибини та інших завдань. У 2020 р. Н. Каріон та ін. запропонували трансформер виявлення (ТРВИ), де вони розглядали ВО як задачу передбачення множини та запропонували мережу виявлення з використанням трансформерів. Пізніше К. Жу та інші запропонували деформований ТРВИ для вирішення проблеми великого часу збіжності ТРВИ та обмеженої продуктивності виявлення малих об'єктів, що дало найкращі результати. ВО є критичним завданням у комп'ютерному зорі. Однокрокові та двокрокові методи ВО мають свої переваги та недоліки. Однокрокові методи ВО, такі як МТДЛОР, ОХБМ, МС, КМ та ЦМ мають переваги у мобільних застосунках і сценаріях реального часу завдяки своїй простоті та швидкості, тоді як двокрокові методи є більш точними, але повільними. ТРВИ відображає нову

еру у виявленні об'єктів, де використовуються трансформери, а деформований ТРВИ досяг найкращої продуктивності.

Ранні методи ВО. У 1990 х роках ранні алгоритми ВО були побудовані на основі штучно створених ознак. Це методи Віоли – Джонса, гістограми орієнтованих градієнтів, моделі на основі деформованих частин. У 2001 р. П. Віола та М. Джонс вперше досягли виявлення людських обличчя в реальному часі. Працюючи на процесорі Pentium III з тактовою частотою 700 МГц, метод Віоли–Джонса (ВД) був в десятки або навіть в сотні разів швидшим за інші алгоритми свого часу при порівнянні точності ВО. Метод ВД використовує вікна пошуку для проходження через всі можливі місця та масштаби на зображенні, щоб перевірити, чи містить вікно людське обличчя. Метод ВД значно покращив швидкість ВО, поєднавши три важливі техніки: інтегральне зображення, вибір ознак та каскадні класифікатори. У 2005 р. Н. Далал та Б. Тріггс запропонували метод гістограми орієнтованих градієнтів (ГОГ). Метод ГОГ є вдосконаленням методу інваріантного до масштабу перетворення ознак та методу контекстів форми. Для забезпечення балансу інваріантності ознак, що включають трансляцію, масштабування, освітлення і т. д., та нелінійності метод ГОГ розраховується на щільній сітці рівномірно розташованих комірок і використовує локальну нормалізацію контрасту на "блоках". Для ВО різного розміру метод ГОГ змінює масштаб вхідного зображення декілька разів, при цьому розмір вікна виявлення залишається незмінним. Метод ГОГ став важливим фундаментом для багатьох методів ВО і широкого спектру застосувань у галузі комп'ютерного зору протягом багатьох років. Метод моделі деформованих частин (МДЧ) запропонований П. Фельценшвалбом у 2008 р. як розширення методу ГОГ. Він слідує філософії "розділяй і пануй", де навчання розглядається як вивчення способу розкладання об'єкта. Наприклад, задачу ВО "автомобіль" можна розкласти на ВО його вікон, кузова та коліс. Метод МДЧ Р. Гіршик розширив для роботи з об'єктами у реальному часі з більшими варіаціями. Ці методи мали потенціал розвитку, сучасні техніки ВО, засновані на глибокому навчанні, ефективніші завдяки автоматичному вивченню розпізнавальних ознак з великих наборів даних [4–20].

Методи, які покращують ВО на основі ЗНМ. У 2017 р. Ху та інші запропонували мережі відношень (МВ) для ВО: адаптований модуль уваги та модуль відношень об'єкта, який враховує взаємодію між різними цілями на зображенні, включаючи їх характерні ознаки та геометричну інформацію. Модуль відношень заміняє крок після обробки НМП, що дозволяє досягти більшої точності порівняно з НМП. У 2018 р. Дай та інші запропонували деформовані згорткові мережі (ДСМ). КНМ можуть фокусуватися тільки на ознаках фіксованого квадратного розміру. ДСМ можуть створювати деформоване ядро, а зміщення фіксованого розміру від початкового конволюційного ядра знаходиться шляхом навчання мереж. Деформований пулінг також може адаптуватися до розташування частин для об'єктів з різними формами.

ДСМв1 досяг значного покращення точності. ДСМв2 використовує більше деформованих конволюційних шарів, ніж ДСМв1. Авторі з Гугл Мозку використали пошук нейронних архітектур для знаходження нової архітектури піраміди ознак, яку назвали пошуком архітектури нейронної мережі (ПАНМ) ПМО. Вона складається з підключень зверху вниз і знизу вгору для об'єднання ознак різних масштабів. Шляхом повторення архітектури ПМО одного і більше разів та їх конкатенації створюється велика архітектура під час пошуку, високорівневі шари ознак вибирають ознаки різних рівнів. Додавання додаткових мереж пірамід, збільшення розмірності ознак і використання великомістких архітектур значно підвищують точність ВО [21–23].

Ху та співавтори запропонували модуль об'єктових відношень для ВО, який враховує взаємодію між цілями на зображенні, замінюючи крок післяобробки НМП і досягаючи високої точності. ДСМв2 використовує більше деформованих шарів та функцію мімікування ознак для покращення деформованого ефекту та точності ВО. ПАН ПМО – це пошук нейроархітектури для нових архітектур піраміди ознак, що складаються з підключень зверху вниз і знизу вгору, які можуть генерувати особливості високої роздільної здатності для виявлення малих об'єктів, а стекинг додаткових пірамідних мереж і додавання розмірності ознак значно підвищують точність виявлення. Швидкість виявлення може бути покращена завдяки вивченню трьохвимірної форми об'єкта [24].

Складність відстеження та ідентифікації об'єктів камерами спостереження в реальних умовах полягає в нестабільності результатів спостереження, наприклад, збуреннях, викликаних зміною видимості об'єкта спостереження [25].

Перспективи ДПДО. Розвиток ДПДО є важливим для реалізації безпілотних автомобілів та інших автономних транспортних засобів. Ідентифікація швидкості, розміру, положення та орієнтації об'єктів на дорозі може допомогти в уникненні зіткнень та забезпеченні безпеки. ДПДО є важливим елементом в робототехніці, де роботи повинні взаємодіяти з рухомими об'єктами в навколишньому середовищі. ДПДО є важливим компонентом систем відеоспостереження та забезпечення безпеки. ДПДО може бути застосована для моніторингу навколишнього середовища, такого як виявлення рухомих об'єктів на вулицях, в аеропортах або в промислових майданчиках. Це може забезпечити безпеку, виявлення незвичайної активності та вчасне реагування на випадки аварій чи порушень. ДПДО може бути застосована в медичній діагностиці для визначення характеристик руху органів та тканин. Наприклад, визначення швидкості кровотоку або руху серця може допомогти у виявленні аномалій та допомогти в ранній діагностиці хвороб. Вона може бути використана для виявлення підозрілих дій, відстеження руху та ідентифікації осіб або об'єктів у режимі реального часу. ДПДО може бути використана для покращення віртуальної реальності та розширеної реальності. Це дозволяє взаємодіяти з рухомими об'єктами у віртуальному середовищі або

розширювати реальний світ з додатковою інформацією про рух та параметри об'єктів. В ігрових і спортивних областях ДПДО може бути використана для аналізу руху гравців, тренування та покращення стратегій. Вона дозволяє вимірювати швидкість, траєкторію та інші характеристики руху об'єктів у режимі реального часу.

Висновки. Під час періоду ВО на основі глибокого навчання відбулося зростання використання сгорткових нейронних мереж, які самостійно вивчають ознаки та здійснюють ВО. Метод ВД використовував ознаки Хаара і алгоритм адаптивного підсилення для виявлення обличчя у реальному часі. ДМЧ базувався на моделі пікторіальної структури і використовував суміш частин для ВО. В НМРЗНМ вперше використані МРП і отримані сучасні результати на різних контрольних наборах даних для ВО. У період виявлення на основі глибокого навчання відзначаються кілька важливих методів, які базувалися на глибоких нейронних мережах, зокрема ОХМ та МТДЛР у 2016 р., МС у 2017 р. ОХМ використовував одну мережу як для локалізації об'єктів, так і для класифікації. МТДЛР використовував одну мережу для прогнозування охоплюючих прямокутників та ймовірності класів безпосередньо з повних зображень, а МС ввів функцію фокусованої втрати, яка вирішувала проблему незбалансованості класів у ВО. ДПДО відноситься до процесу ідентифікації та відстеження руху та поведінки об'єкта у реальному часі. Цей процес залежить від ВО, яке передбачає визначення наявності та місцезнаходження об'єкта на зображенні або відеокадрі. ДПДО є важливою та актуальною проблемою, а застосування різних методів, включаючи ті, що базуються на згорткових нейронних мережах та оптичному потоці, дозволяє досягти високої точності та продуктивності в процесі виявлення та ідентифікації об'єктів.

Список використаної літератури

1. Severin V. P., Nikulina E. N., Buriakovskiy V. S. Development of the controller for the quadcopter finkenin simulation enviroment vrep. *Вісник Національного технічного університету «Харківський політехнічний інститут»: сб. наук. пр. Темат. вип.: Системний аналіз, управління та інформаційні технології.* Харків: НТУ «ХПІ», 2016. № 37 (1209). С. 9–12.
2. Ammar A., Chebbah A., Fredj H., Souani C. *Comparative Study of latest CNN based Optical Flow Estimation.* URL: <https://ieeexplore.ieee.org/document/9806070/references#references>. (дата звернення: 27.04.2023).
3. Нікуліна О. М., Северин В. П., Коцюба Н. В. Розробка інформаційної технології оптимізації управління складними динамічними системами. *Вісник Національного технічного університету «Харківський політехнічний інститут»: сб. наук. пр. Темат. вип.: Системний аналіз, управління та інформаційні технології.* Харків: НТУ «ХПІ», 2020. № 2 (4). С. 63–69.
4. Zou Z., Chen K., Shi Z., Shi Z., Guo Y., Ye J. *Object Detection in 20 Years: A Survey.* URL: https://arxiv.org/pdf/1905.05055.pdf?fbclid=IwAR0ILGAWTwU-9-iH6lZyPFXYXA5JRWArM_XoSJ78QEhmnn-txrv_iGEzCio (дата звернення: 27.04.2023).
5. Krizhevsky A., Sutskever I., Hinton G. E. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems.* 2012. P. 1097–1105.
6. Girshick R., Donahue J., Darrell T., Malik J. *Rich feature hierarchies for accurate object detection and semantic segmentation.* URL: <https://arxiv.org/abs/1311.2524> (дата звернення: 27.04.2023).

7. Girshick R., Donahue J., Darrell T., and Malik J. Region-based convolutional networks for accurate object detection and segmentation. *IEEE transactions on pattern analysis and machine intelligence*. 2016. Vol. 38, no. 1. P. 142–158.
8. Uijlings J. R., Van De Sande K. E., Gevers T., Smeulders A. W. Selective search for object recognition. *International journal of computer vision*. 2013. Vol. 104, no. 2. P. 154–171.
9. Viola P., Jones M. *Rapid object detection using a boosted cascade of simple features*. URL: <https://www.cs.cmu.edu/~efros/courses/LBMV07/Papers/viola-cvpr-01.pdf> (дата звернення: 27.04.2023).
10. Viola P., Jones M. J. Robust real-time face detection. *International journal of computer vision*. 2004. Vol. 57, no. 2. P. 137–154.
11. Dalal N., Triggs B. Histograms of oriented gradients for human detection. URL: <https://ieeexplore.ieee.org/document/1467360> (дата звернення: 27.04.2023).
12. Lowe D. G. *Object recognition from local scale-invariant features*. URL: <https://ieeexplore.ieee.org/document/790410> (дата звернення: 27.04.2023).
13. Distinctive image features from scale-invariant keypoints. *International Journal of computer vision*. 2004. Vol. 60, no. 2. P. 91–110.
14. Belongie S., Malik J., Puzicha J. Shape matching and object recognition using shape contexts. *California univ san diego la jolla dept of computer science and engineering. Tech. Rep.* 2002. Vol. 24, no. 24. P. 509–522.
15. Felzenszwalb P., McAllester D., Ramanan D. *A discriminatively trained, multiscale, deformable part model*. URL: <https://cs.brown.edu/people/pfelzens/papers/latent.pdf> (дата звернення: 27.04.2023).
16. Felzenszwalb P. F., Girshick R. B., McAllester D. *Cascade object detection with deformable part models*. URL: <https://cs.brown.edu/people/pfelzens/papers/cascade.pdf> (дата звернення: 27.04.2023).
17. Malisiewicz T., Gupta A., Efros A. A. *Ensemble of exemplar-SVMs for object detection and beyond*. URL: <https://ieeexplore.ieee.org/document/6126229> (дата звернення: 27.04.2023).
18. Felzenszwalb P. F., Girshick R. B., McAllester D., Ramanan D. Object detection with discriminatively trained part-based models. *IEEE transactions on pattern analysis and machine intelligence*. 2010. Vol. 32, no. 9, P. 1627–1645.
19. Girshick R. B., Felzenszwalb P. F., McAllester D. A. Object detection with grammar models. *Advances in Neural Information Processing Systems*, 2011. P. 442–450.
20. Girshick R. B. *From rigid templates to grammars: Object detection with structured models*. URL: <https://dl.dropboxusercontent.com/s/o9m1nq0jdax2on9/rbg-phd-dissertation.pdf?dl=0> (дата звернення: 27.04.2023).
21. Jiao L., Zhang F., Liu F., Yang S., Li L., Feng Z., Qu R. *A Survey of Deep Learning-based Object Detection*. URL: <https://ieeexplore.ieee.org/document/8825470> (дата звернення: 27.04.2023).
22. Hu H., Gu J., Zhang Z., Dai J., Wei Y. Relation networks for object detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018. P. 3588–3597.
23. Zhu X., Hu H., Lin S., Dai J. *Deformable ConvNets v2: More Deformable, Better Results*. URL: <https://arxiv.org/abs/1811.11168> (дата звернення: 27.04.2023).
24. Kitakaze H., Yoshihara R., Okabea S., Matsumura R. Development of Harmful Bird Recognition System using Object Detection YOLO. *The Japanese Journal of the Institute of Industrial Applications Engineers Online edition*. URL: https://www.jstage.jst.go.jp/article/jjiaie/8/1/8_10/_pdf/-char/ja (дата звернення: 27.04.2023).
25. Inomata T., Kimura K., Hagiwara M. *Object Tracking and Classification System Using Agent Search*. URL: https://www.jstage.jst.go.jp/article/ieejieiss/129/11/129_11_2065/_pdf/-char/ja (дата звернення: 27.04.2023).
2. Ammar A., Chebbah A., Fredj H., Souani C. *Comparative Study of latest CNN based Optical Flow Estimation*. URL: <https://ieeexplore.ieee.org/document/9806070/references#references>. (accessed: 27.04.2023).
3. Nikulina E. N., Severyn V. P., Kotsiuba N. V. Rozrobka informatsiynoyi tekhnologii optimizatsii upravlinnya skladnymy dynamichnymy systemamy [Development of information technology for optimizing the control of complex dynamic systems]. *Vestnik Nats. tekhn. un-ta "KhPI": sb. nauch. tr. Temat. vyp.: Sistemnyy analiz, upravlenie i informatsionnye tekhnologii* [Bulletin of the National Technical University "KhPI": a collection of scientific papers. Thematic issue: System analysis, management and information technology]. Kharkov, NTU "KhPI" Publ., 2020, no. 2 (4), pp. 63–69.
4. Zou Z., Chen K., Shi Z., Shi Z., Guo Y., Ye J. *Object Detection in 20 Years: A Survey*. URL: https://arxiv.org/pdf/1905.05055.pdf?fbclid=IwAR0ILGAWTuwU-9-iH6lZyPFXYXA5JRWarM_XoSJ78QEhmn-txvr_iGEzCio (accessed: 27.04.2023).
5. Krizhevsky A., Sutskever I., Hinton G. E. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*. 2012, pp. 1097–1105.
6. Girshick R., Donahue J., Darrell T., Malik J. *Rich feature hierarchies for accurate object detection and semantic segmentation*. URL: <https://arxiv.org/abs/1311.2524> (accessed: 27.04.2023).
7. Girshick R., Donahue J., Darrell T., and Malik J. Region-based convolutional networks for accurate object detection and segmentation. *IEEE transactions on pattern analysis and machine intelligence*. 2016, vol. 38, no. 1, pp. 142–158.
8. Uijlings J. R., Van De Sande K. E., Gevers T., Smeulders A. W. Selective search for object recognition. *International journal of computer vision*. 2013, vol. 104, no. 2, pp. 154–171.
9. Viola P., Jones M. *Rapid object detection using a boosted cascade of simple features*. URL: <https://www.cs.cmu.edu/~efros/courses/LBMV07/Papers/viola-cvpr-01.pdf> (дата звернення: 27.04.2023).
10. Viola P., Jones M. J. Robust real-time face detection. *International journal of computer vision*. 2004, vol. 57, no. 2, pp. 137–154.
11. Dalal N., Triggs B. *Histograms of oriented gradients for human detection*. URL: <https://ieeexplore.ieee.org/document/1467360> (accessed: 27.04.2023).
12. Lowe D. G. *Object recognition from local scale-invariant features*. URL: <https://ieeexplore.ieee.org/document/790410> (accessed: 27.04.2023).
13. Distinctive image features from scale-invariant keypoints. *International Journal of computer vision*. 2004, vol. 60, no. 2, pp. 91–110.
14. Belongie S., Malik J., Puzicha J. Shape matching and object recognition using shape contexts. *California univ san diego la jolla dept of computer science and engineering. Tech. Rep.* 2002, vol. 24, no. 24, pp. 509–522.
15. Felzenszwalb P., McAllester D., Ramanan D. *A discriminatively trained, multiscale, deformable part model*. URL: <https://cs.brown.edu/people/pfelzens/papers/latent.pdf> (accessed: 27.04.2023).
16. Felzenszwalb P. F., Girshick R. B., McAllester D. *Cascade object detection with deformable part models*. URL: <https://cs.brown.edu/people/pfelzens/papers/cascade.pdf> (accessed: 27.04.2023).
17. Malisiewicz T., Gupta A., Efros A. A. *Ensemble of exemplar-SVMs for object detection and beyond*. URL: <https://ieeexplore.ieee.org/document/6126229> (accessed: 27.04.2023).
18. Felzenszwalb P. F., Girshick R. B., McAllester D., Ramanan D. Object detection with discriminatively trained part-based models. *IEEE transactions on pattern analysis and machine intelligence*. 2010, vol. 32, no. 9, pp. 1627–1645.
19. Girshick R. B., Felzenszwalb P. F., McAllester D. A. Object detection with grammar models. *Advances in Neural Information Processing Systems*, 2011, pp. 442–450.

References (transliterated)

1. Severin V. P., Nikulina E. N., Buriakovskiy V. S. Development of the controller for the quadcopter finkenin simulation environment vrep. *Vestnik Nats. tekhn. un-ta "KhPI": sb. nauch. tr. Temat. vyp.:*

20. Girshick R. B. *From rigid templates to grammars: Object detection with structured models*. URL: <https://dl.dropboxusercontent.com/s/o9m1nq0jdax2on9/rbg-phd-dissertation.pdf?dl=0> (accessed: 27.04.2023).
21. Jiao L., Zhang F., Liu F., Yang S., Li L., Feng Z., Qu R. *A Survey of Deep Learning-based Object Detection*. URL: <https://ieeexplore.ieee.org/document/8825470> (accessed: 27.04.2023).
22. Hu H., Gu J., Zhang Z., Dai J., Wei Y. Relation networks for object detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018, pp. 3588–3597.
23. Zhu X., Hu H., Lin S., Dai J. *Deformable ConvNets v2: More Deformable, Better Results*. URL: <https://arxiv.org/abs/1811.11168> (accessed: 27.04.2023).
24. Kitakaze H., Yoshihara R., Okabe S., Matsumura R. Development of Harmful Bird Recognition System using Object Detection YOLO. *The Japanese Journal of the Institute of Industrial Applications Engineers Online edition*. URL: https://www.jstage.jst.go.jp/article/jjiaae/8/1/8_10/_pdf/-char/ja (accessed: 27.04.2023).
25. Inomata T., Kimura K., Hagiwara M. *Object Tracking and Classification System Using Agent Search*. URL: https://www.jstage.jst.go.jp/article/ieejieiss/129/11/129_11_2065/_pdf/-char/ja (accessed: 27.04.2023).

Надійшла (received) 10.05.2023

UDC 519.2

O. M. NIKULINA, Doctor of Technical Sciences, Associate Professor, Head of Department Information Systems and Technologies National Technical University «Kharkiv Polytechnic Institute», Kharkiv, Ukraine; ORCID: <https://orcid.org/0000-0003-2938-4215>; e-mail: elniknik02@gmail.com

V. P. SEVERYN, Doctor of Technical Sciences, Professor, Professor of Department System Analysis and Information-Analytical Technologies National Technical University «Kharkiv Polytechnic Institute», Kharkiv, Ukraine; ORCID: <https://orcid.org/0000-0002-2969-6780>; e-mail: valerii.severyn@kpi.edu.ua

O. M. KONDRATOV, Postgraduate of Department Information Systems and Technologies National Technical University «Kharkiv Polytechnic Institute», Kharkiv, Ukraine; ORCID: <https://orcid.org/0000-0001-6367-9944>; e-mail: kondratovolexiy@gmail.com

N. Y. REKOVA, Doctor of Economics, Professor, Professor of Department IT, Analysis and Project Decisions, Ltd «Technical University «METINVEST POLYTECHNICS», Zaporizhzhia, Ukraine; ORCID: <https://orcid.org/0000-0003-0956-6564>; e-mail: natarekova@gmail.com

ANALYSIS OF INFORMATION TECHNOLOGIES FOR REMOTE IDENTIFICATION OF DYNAMIC OBJECTS

The problem of identification of dynamic objects using remote identification information technologies is considered. It is noted that the identification of moving objects is important in various fields, including autonomous vehicles, medical diagnostics and robotics. The purpose of the article is to analyze various information technologies for detecting objects that can be used in future research on remote identification. Analysis of methods for determining speed as a dynamic parameter, analysis of two-step and one-step methods of remote identification of objects, analysis of early identification methods, as well as analysis of methods for improving remote identification of objects was carried out. Several means of determining the motion of objects are considered, in particular, the proportional-integral-differential controller, the leveling block method, phase correlation, pixel recursion algorithms, and the optical flow methods of Lucas – Kanady, Horn – Shunk, Farnbeck, dense optical flow. These tools can be used to effectively determine the movement of objects and identify their speed regardless of the size and position of the objects. Two-step and one-step object detection methods are considered: region method with convolutional neural networks, its improvements, spatial pyramid pooling networks, "You only look once" method, one-step multi-frame method, retinal networks, corner network, central network and detection transformer, which use different approaches to improve the performance and accuracy of object detection. The necessity of using methods of convolutional neural networks and spatial pyramid pooling networks for effective identification of objects regardless of their size and position is emphasized. New approaches are proposed that allow creating fixed-length representations for image processing and regions of interest, as well as Viola – Jones methods, oriented gradient histograms, and deformed part models. Research in the field of object detection contributes to the development of information technologies and the improvement of the efficiency of dynamic object identification systems. Through the review and analysis of various methods, recommendations for researchers and practitioners working in the field of remote identification of dynamic objects are provided.

Keywords: Remote identification of dynamic objects, object detection, optical flow, velocity identification, deep learning, convolutional neural networks.

Повні імена авторів / Author's full names

Автор 1 / Author 1: Нікуліна Олена Миколаївна, Nikulina Olena Mykolaivna

Автор 2 / Author 2: Северин Валерій Петрович, Severyn Valerii Petrovych

Автор 3 / Author 3: Кондратов Олексій Михайлович, Kondratov Oleksii Mikhailovich

Автор 4 / Author 4: Рекова Наталія Юріївна, Reкова Nataliia Yuriiivna

ЗМІСТ

СИСТЕМНИЙ АНАЛІЗ І ТЕОРІЯ ПРИЙНЯТТЯ РІШЕНЬ.....	3
<i>Pavlov A. A., Holovchenko M. N., Drozd V. V.</i> Efficiency substantiation for a synthetical method of constructing a multivariate polynomial regression given by a redundant representation	3
<i>Левтеров А. І., Плехова Г. А., Костікова М. В., Бережна Н. Г., Окунь А. О.</i> Дослідження методів безпечної маршрутизації у програмно-конфігурованих мережах.....	10
УПРАВЛІННЯ В ТЕХНІЧНИХ СИСТЕМАХ.....	19
<i>Жученко О. А., Коротинський А. П., Абрамова А. О. Чепов Д. Є.</i> Розробка системи керування випарним апаратом у процесі виробництва оцтової кислоти на базі МРС-регулятора	19
<i>Мазманішвілі О. С., Решетняк М. Г., Сидоренко Г. Ю.</i> Електронні пучки у градієнтному магнітному полі: управління перетворенням поздовжнього руху у поперечний	27
УПРАВЛІННЯ В ОРГАНІЗАЦІЙНИХ СИСТЕМАХ.....	33
<i>Кунанець Н. Е., Федорка П. П., Кут В. І.</i> Формування рекомендаційної системи для «Розумного регіону» з метою обрання інформаційних технологій та їх реалізацій при створенні застосунків.....	33
<i>Годлевський М. Д., Булаков Г. О.</i> Вербальний опис технології планування підвищення якості підмножини процесів еталонної моделі зрілості SPICE.....	41
МАТЕМАТИЧНЕ І КОМП'ЮТЕРНЕ МОДЕЛЮВАННЯ	49
<i>Мельников О. Ю., Закабула О. Ю.</i> Моделювання рівня незадоволення потреб мешканців малих міст у системі підтримки прийняття рішень для водопостачання в екстремальних випадках.....	49
<i>Корп А. М., Orlovskiy D. L.</i> An algorithm for NLP-based similarity measurement of activity labels in a database of business process models.....	54
<i>Левтеров А. І., Плехова Г. А., Костікова М. В., Окунь А. О.</i> Геометричне моделювання трас і потоків	60
<i>Нікуліна О. М., Северин В. П., Шаров В. О.</i> Моделювання та аналіз кодерів завадостійких каскадних кодів для динамічних систем	64
<i>Чалий С. Ф., Лециньська І. О.</i> Концептуальна ментальна модель пояснення в системі штучного інтелекту	70
ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ	76
<i>Yatkovyi K. S.</i> Adaptation of LambdaMART model to semi-supervised learning	76
<i>Solovei O. L., Gobov D. A.</i> Parsimonious machine learning models in requirements elicitation techniques selection	82
<i>Onyshchenko K., Daniil Y.</i> Using long short-term memory networks for natural language processing	89
<i>Rybitskiy O. M., Golian V. V., Golian N. V., Dudar Z. V., Kalynychenko O. V., Nikitin D. M.</i> Using OBD-2 technology for vehicle diagnostic and using IT in the information system.....	97
<i>Nikitin D. M.</i> Specification formalization of state charts for complex system management.....	104
<i>Нікуліна О. М., Северин В. П., Кондратов О. М., Рекова Н. Ю.</i> Аналіз інформаційних технологій для дистанційної ідентифікації динамічних об'єктів	110

CONTENT

SYSTEM ANALYSIS AND DECISION-MAKING THEORY.....	3
<i>Pavlov A. A., Holovchenko M. N., Drozd V. V.</i> Efficiency substantiation for a synthetical method of constructing a multivariate polynomial regression given by a redundant representation	3
<i>Levterov A. I., Pliekhova H. A., Kostikova M. V., Berezhna N. G., Okun A. O.</i> Enhancing security in software-defined networking through routing techniques exploration	10
CONTROL IN TECHNICAL SYSTEMS.....	19
<i>Zhuchenko O. A., Korotynskiy A. P., Abramova A. O., Chepov D. E.</i> Development of the evaporator control system in the acetic acid production process based on the MPC-regulator	19
<i>Mazmanishvili O. S., Reshetnyak M. G., Sydorenko G. Yu.</i> Electron beams in the gradient magnetic field: control for converting longitudinal motion into transversal	27
MANAGEMENT IN ORGANIZATIONAL SYSTEMS.....	33
<i>Kunanets N. E., Fedorka P. P., Kut V. I.</i> Formation of a recommendation system for the "Smart Region" for the purpose of choosing information technologies and their implementation when creating applications	33
<i>Godlevskiy M. D., Burlakov G. O.</i> Verbal description of the technology for planning the quality improvement of a subset of the processes of the spice maturity reference model	41
MATHEMATICAL AND COMPUTER MODELING	49
<i>Melnykov O. Yu., Zakabula O. Ju.</i> Modeling the level of dissatisfaction of the needs of residents of small towns in the decision support system for water supply in extreme cases.....	49
<i>Kopp A. M., Orlovskiy D. L.</i> An algorithm for NLP-based similarity measurement of activity labels in a database of business process models	54
<i>Levterov A. I., Pliekhova H. A., Kostikova M. V., Okun A. O.</i> Geometric modeling: tracks and flows.....	60
<i>Nikulina O. M., Severyn V. P., Sharov V. O.</i> Simulation and analysis of encoders of interference-resistant cascade codes for dynamic systems.....	64
<i>Chalyi S. F., Leshchynska I. O.</i> The conceptual mental model of explanation in an artificial intelligence system.....	70
INFORMATION TECHNOLOGY	76
<i>Yamkovyi K. S.</i> Adaptation of LambdaMART model to semi-supervised learning.....	76
<i>Solovei O. L., Gobov D. A.</i> Parsimonious machine learning models in requirements elicitation techniques selection	82
<i>Onyshchenko K., Daniil Y.</i> Using long short-term memory networks for natural language processing	89
<i>Rybitskiy O. M., Golian V. V., Golian N. V., Dudar Z. V., Kalynychenko O. V., Nikitin D. M.</i> Using OBD-2 technology for vehicle diagnostic and using IT in the information system	97
<i>Nikitin D. M.</i> Specification formalization of state charts for complex system management	104
<i>Nikulina O. M., Severyn V. P., Kondratov O. M., Rekova N. Y.</i> Analysis of information technologies for remote identification of dynamic objects	110

НАУКОВЕ ВИДАННЯ

**ВІСНИК НАЦІОНАЛЬНОГО ТЕХНІЧНОГО УНІВЕРСИТЕТУ «ХПІ».
СЕРІЯ: СИСТЕМНИЙ АНАЛІЗ, УПРАВЛІННЯ ТА ІНФОРМАЦІЙНІ
ТЕХНОЛОГІЇ**

Збірник наукових праць

№ 1 (9) 2023

Наукові редактори: М. Д. Годлевський, д-р техн. наук, професор, НТУ «ХПІ», Україна
О. С. Куценко, д-р техн. наук, професор, НТУ «ХПІ», Україна
Технічний редактор: М. І. Безменов, канд. техн. наук, професор, НТУ «ХПІ», Україна

Відповідальний за випуск М. І. Безменов, канд. техн. наук, професор

АДРЕСА РЕДКОЛЕГІЇ ТА ВИДАВЦЯ: 61002, Харків, вул. Кирпичова, 2, НТУ «ХПІ».
Кафедра системного аналізу та інформаційно-аналітичних технологій
Тел.: (057) 707-61-03, (057) 707-66-54; e-mail: Mykola.Bezmenov@khpі.edu.ua

Підп. до друку 07.07.2023 р. Формат 60×84 1/8. Папір офсетний.
Друк цифровий. Гарнітура Times New Roman. Умов. друк. арк. 9,8. Облік.-вид. арк. 10.
Тираж 100 пр. Зам. № 284. Ціна договірна.

Видавництво «Стильна типографія»
Свідоцтво суб'єкта видавничої справи: ДК № 5493 від 22.08.2017 р.
61002, м. Харків, вул. Чернишевська, 28А, тел. (057) 754-49-42; e-mail: zebraprint.zakaz@gmail.com