

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**

№ 2 (12) 2024

Збірник наукових праць

Видання засноване у 1961 р.

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

No. 2 (12) 2024

Collection of Scientific papers

The edition was founded in 1961

Харків
НТУ «ХПІ», 2024

Kharkiv
NTU "KhPI", 2024

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», 2024. — № 2 (12) 2024. — 114 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.

Для науковців, викладачів вищої школи, аспірантів, студентів та спеціалістів у галузі системного аналізу, управління та комп'ютерних технологій.

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.

Ідентифікатор медіа R30-01544, згідно з рішенням Національної ради України з питань телебачення і радіомовлення від 16.10.2023 № 1075.

Мова статей – українська, 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

Бентаєб 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

Копп А. М., доц., НТУ «ХПІ», Україна

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

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

Павлов О. А., проф., НТУ «ХПІ», 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

Рекомендовано до друку Вченою радою НТУ «ХПІ».

Протокол № 10 від 26 грудня 2024 р.

Editorial

Editor-in-chief

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

Deputy editor-in-chief

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

Editorial staff members

Akhiiezer 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

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

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

Kersten Wolfgang, prof., Hamburg University of Technology,

Germany

Kopp A. V., associate professor, NTU "KhPI", Ukraine

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

СИСТЕМНИЙ АНАЛІЗ І ТЕОРІЯ ПРИЙНЯТТЯ РІШЕНЬ

SYSTEM ANALYSIS AND DECISION-MAKING THEORY

DOI: 10.20998/2079-0023.2024.02.01

UDC 004:519.24

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>

MODIFICATION OF THE DECOMPOSITION METHOD OF CONSTRUCTING MULTIVARIATE POLYNOMIAL REGRESSION WHICH IS LINEAR WITH RESPECT TO UNKNOWN COEFFICIENTS

The authors created a universal method of constructing multivariate polynomial regression given by a redundant representation. The method is synthetic, it organically combines a decomposition method and the modified group method of data handling. First, the decomposition method is implemented, it consists in the decomposition of the multivariate problem into a sequence of subproblems of constructing univariate polynomial regressions and the corresponding systems of linear equations, the variables of which are estimates for the nonlinear terms of the multivariate polynomial regression. Partial cases that guarantee the finding of estimates with a predetermined value of their variances were considered. The formal algorithm for constructing coefficient estimates for nonlinear terms of the multivariate polynomial regression stops working on the first coefficient whose estimation with a predetermined accuracy is not achieved under the specified limitations on the number of tests. The estimation of all coefficients that were not found by the decomposition method is done by a heuristic method, which is an efficient modification of the group method of data handling. The increase in the efficiency of the synthetic method is achieved primarily by finding such new theoretically substantiated algorithmic procedures (aggregated operators) of the decomposition method, which significantly, in comparison with its previous version, increases the number of coefficients for nonlinear terms of a multivariate polynomial regression that can be found in advance given accuracy. The authors showed that this effect is achieved due to new theoretical provisions used in the visual analysis of the structure of the multivariate polynomial regression given by the redundant representation by a professional user. The given illustrative example facilitates the use of the presented results when solving practical problems.

Keywords: regression analysis, multivariate polynomial regression, redundant representation, decomposition method, individual algorithm, least squares method.

1. Introduction. Construction of a multivariate regression based on the results of an active experiment is still theoretically and practically relevant as shown, in particular, in [1–15]. As noted in the abstract, the authors proposed a universal synthetic method for constructing a multivariate polynomial regression (MPR) given by a redundant representation. The originality and efficiency of the method were protected by relevant publications. In this paper, we propose to qualitatively increase its efficiency by expanding the theoretical and practical capabilities of the decomposition method, which is a part of the synthetic method.

1.1. The problem statement. An MPR is given by the following redundant representation [14]:

$$Y(\bar{x}) = \sum_{\forall (i_1, \dots, i_j) \in K, \forall (j_1, \dots, j_l) \in K(i_1, \dots, i_j)} b_{i_1 \dots i_j}^{j_1 \dots j_l} (x_{i_1})^{j_1} \dots (x_{i_j})^{j_l} + E, \quad (1)$$

where $\bar{x} = (x_1, \dots, x_m)^T$ is a deterministic vector of input variables, $x_i \in [c_i, d_i]$, $0 \leq c_i < d_i$ (this is the strictest con-

dition for finding estimates with a given accuracy), E is a random variable, its mathematical expectation is $ME = 0$, its variance $\text{Var}(E) = \sigma^2 < \infty$. σ^2 is known or its efficient upper estimate is known. The values of the coefficients $b_{i_1 \dots i_j}^{j_1 \dots j_l}$ are unknown (b_0^0 is an unknown constant).

We need to estimate the unknown coefficients at nonlinear terms of the MPR according to the results of an active experiment $(x_i \rightarrow y_i, i = \overline{1, n})$.

1.2. Basic provisions of the decomposition method [1]. The decomposition method [14] implements a methodology of reducing the finding of estimates for nonlinear terms of the MPR (1) to the sequential construction of univariate polynomial regressions (UPR) and the solution of the corresponding systems of linear equations. Variables in these equations are estimates for nonlinear terms of the MPR (1). The general formal algorithmic procedure for obtaining estimates of all coefficients for nonlinear terms of the MPR

© Pavlov A. A., Holovchenko M. N., Drozd V. V., 2024



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.



(1) consists in the sequential execution of the first and then the second subalgorithm. The first subalgorithm is implemented step-by-step for all members of the MPR, each of which contains at least one scalar variable of power greater than or equal to two.

Remark 1. The procedure for choosing nonlinear terms and choosing for each of them a single coefficient of the form $(x_{i_p})^{j_p}$, $j_p \geq 2$, is substantiated in [1].

At each step of the first subalgorithm, the scalar variable x_{i_p} is replaced by a virtual scalar variable z :

$$x_{i_p} = a_{i_p} z + b_{i_p}. \quad (2)$$

The coefficients a_{i_p} , b_{i_p} are found according to [1].

The values $x_{i_p,i}$, $i = \overline{1, n}$, of the real input variable x_{i_p} are found according to (2) by the given in advance values of the virtual variable z . Other input variables take fixed values. The results of the active experiment on the obtained real univariate regression are transformed as the result of the virtual active experiment on the virtual UPR relative to the virtual scalar input variable z .

Remark 2. It is substantiated in [1] how many virtual UPRs are built at each step of the first subalgorithm. Their number determines the dimension of the systems of linear equations, which variables are the estimates of the coefficients at the nonlinear terms of the MPR (1).

Remark 3. It is shown in [1] that the reduction of real UPRs to virtual ones in the case when a system of linear equations has only one variable allows one to efficiently find the variance of the estimate for the nonlinear term of the MPR (1) before conducting a real experiment and to find the necessary number of repetitions of the main experiment to obtain the variance of the given value. If this quantity is excessive, the corresponding coefficient cannot be found with the specified accuracy by the decomposition method.

The steps of the second subalgorithm [14] relate to finding nonlinear terms of the MPR of the form $b_{i_1 \dots i_q}^{j_1 \dots j_q}$. For this case it is proposed in [1] to replace all input variables included in this term by the virtual variable z , other input variables take fixed values in the experiments.

It was proved in [1] that the implementation of the first and second subalgorithms formally allow obtaining estimates of the coefficients for all nonlinear members of the MPR (1).

1.3. The main drawback of the formal algorithmic procedure of the decomposition method. The theoretical substantiation of the decomposition method guarantees that the successive use of the first and second subalgorithms at the preliminary stage of the problem analysis leads to the guaranteed obtaining of estimates of all coefficients at nonlinear terms of the MPR. But with the limited number of experiments given in advance, the specified accuracy (by the variance value) is not guaranteed. When finding the estimates of the following coefficients, the estimates of previously found coefficients at nonlinear terms of the MPR are used. Thus, the result obtained may be unsatisfactory in general. Therefore, when solving practical problems, the algorithm of the decomposition method stops working on the

first nonlinear term of the MPR that has the estimate of the coefficient significantly different from the given one by variance at the stage of the preliminary analysis.

2. Methodology of qualitative improvement of the efficiency of the decomposition method. The methodology of qualitative improvement of the efficiency of the decomposition method is that a qualified user, analyzing the structure of the MPR, the coefficients of the nonlinear terms of which have to be estimated (by the variance value) with a given accuracy, using extended theoretical possibilities of the decomposition method (see Section 3), finds by himself the step by step sequence of the algorithm at the stage of the preliminary analysis of the problem solution (before conducting real experiments). That is, in what sequence will the active experiment be implemented to construct the found number of UPRs for a user-defined sequence of nonlinear terms of the MPR, which guarantees at each step finding the coefficient estimates with a given accuracy. Conducting an active experiment for each UPR in the found order after the preliminary analysis stage is used to build the estimates of the coefficients at the highest degree of each UPR. This sequence of steps can be obtained by a formal procedure only by enumerating all possible options, which is obviously unconstructive.

3. Expanding the theoretical possibilities of the decomposition method for the efficient construction of an individual algorithm for the preliminary analysis of the problem solution. Section 3 describes four different theoretically substantiated methods (aggregated operators) of constructing an UPR for estimating the coefficients at nonlinear terms of the MPR at the stage of preliminary analysis of the individual algorithm of the decomposition method. The sequence of their use is set by the user as a result of their analysis of the MPR structure. As a result of the implementation of the previous stage, the individual algorithm (the second stage) becomes fully formalized, namely, the sequence of active experiments to be conducted (with known values of the input variables) is specified, the result of each of the experiments is the estimate of the coefficient at the highest degree of a specially designed UPR, which allows to estimate with a specified accuracy the coefficient at the corresponding nonlinear term of the MPR.

3.1. The first human-computer algorithmic procedure (the first aggregated operator (AO)). The user selects the nonlinear term of the MPR

$$b_{i_1 \dots i_q}^{j_1 \dots j_q} (x_{i_1})^{j_1} \dots (x_{i_k})^{j_k} \dots (x_{i_q})^{j_q}, \quad (3)$$

that contains his chosen input variable x_{i_k} in the power of $j_k \geq 2$. Algorithmically, the verification of the following condition is carried out: there should not be any nonlinear term of the MPR containing $(x_{i_k})^{j_k} \forall j_k \geq 2$. If such terms are present, the coefficients for them with the specified accuracy will be found as a result of the implementation of an active experiment in the previous steps of the second stage of the individual algorithm. In the main active experiment, the variable x_{i_k} will take the value $x_{i_k,i} = a_{i_k} z_i + b_{i_k}$, $i = \overline{1, n}$, where the coefficients a_{i_k} , b_{i_k} , and the value z_i ,

$i = \overline{1, n}$, of the input virtual scalar variable z of the specially constructed virtual UPR [1] are given by formulas (7), (8) [1]. Other input variables $x_i, i = \overline{1, m}, i \neq i_k$, take fixed values in the active experiment, and the input variables included in (3) take maximally large modulo values (all others take minimally small modulo values). At the same time, according to [1], the statistical data of the results of the virtual active experiment are given in the form

$$\left(z_i \rightarrow \bar{y}_i, \bar{y}_i = y_i - \sum b_{i_1 \dots i_{j_i}}^{j_{i_1} \dots j_{i_{j_i}}} (x_{i_1, i})^{j_{i_1}} \dots (x_{i_{j_i}, i})^{j_{i_{j_i}}} \right), \quad (4)$$

where the coefficients $b_{i_1 \dots i_{j_i}}^{j_{i_1} \dots j_{i_{j_i}}}$ will be found with the specified accuracy in the previous steps of the second stage of the individual algorithm. In this case, according to [1], the variance of the estimate of the coefficient $b_{i_1 \dots i_{j_i}}^{j_{i_1} \dots j_{i_{j_i}}}$ (see (3)) has the form

$$D\hat{b}_{i_1 \dots i_{j_i}}^{j_{i_1} \dots j_{i_{j_i}}} = \frac{1}{(a_{i_k})^{2j_k}} \cdot \frac{1}{\prod_{l=1, l \neq i_k}^t (x_{i_l, \phi})^{2j_l}} \cdot D\hat{\gamma}_{j_k}, \quad (5)$$

where, according to [1], $\hat{\gamma}_{j_k}$ is an estimate of the coefficient at z of the maximum power j_k of the virtual UPR (for simplicity, both the random variable and its realization are denoted by $\hat{\gamma}_{j_k}$).

As shown in [1], for the virtual variable z , which in the virtual active experiment takes the value $z_1 < z_2 < \dots < z_n$, where $\forall i z_i - z_{i-1} = \text{const}, n = 10, z_1 = -50, z_{10} = 50$, the variances of the estimates of the coefficients of the virtual UPR are the following:

$$D\hat{\gamma}_2 = 1.7 \cdot 10^{-6} \sigma^2, \quad D\hat{\gamma}_3 = 4.7 \cdot 10^{-9} \sigma^2,$$

$$D\hat{\gamma}_4 = 2.6 \cdot 10^{-13} \sigma^2, \quad D\hat{\gamma}_5 = 4.5 \cdot 10^{-16} \sigma^2,$$

that is, the variance of each subsequent term decreases by three orders of magnitude. The analysis of expression (5) allows at the previous stage of the individual algorithm to set the values of the input variables in the active experiment (considering possible repetitions of the main active experiment [1]). This allows us to obtain an estimate of the coefficient (3) with the given accuracy or conclude that the coefficient (3) cannot be obtained with the given accuracy by the individual algorithm of the decomposition method.

3.2. The second human-computer algorithmic procedure (the second AO). The user selects the nonlinear term of the MPR

$$b_{i_1 \dots i_t}^{j_{i_1} \dots j_{i_t}} (x_{i_1})^{j_{i_1}} \dots (x_{i_k})^{j_k} \dots (x_{i_t})^{j_t}, \quad (6)$$

sets the input variables $x_{i_1}, x_{i_2}, \dots, x_{i_t}$, which in the main experiment will take values

$$\forall l x_{i_l, i} = a_{i_l} z_i + b_{i_l}, i = \overline{1, n}, \quad (7)$$

where, as in Subsection 3.1, coefficients a_{i_l}, b_{i_l} are set according to formulas (7), (8) [1]. Other input variables in the active experiment take fixed values, as in Subsection 3.1. Algorithmically, the following condition is checked: there should not be any nonlinear term of the MPR containing input variables x_{i_1}, \dots, x_{i_t} or their subset, the total degree

of which is greater than or equal to $\sum_{i=1}^t j_{i_l}$. If such terms exist, then their coefficients must be found with the specified accuracy in the previous steps of the second stage of the individual algorithm of the decomposition method.

If this condition is fulfilled, the description of the sequence of steps repeats Subsection 3.1 completely. In this case, according to [1], we obtain the variance of the coefficient estimate $b_{i_1 \dots i_t}^{j_{i_1} \dots j_{i_t}}$ at the stage of preliminary analysis:

$$D\hat{b}_{i_1 \dots i_t}^{j_{i_1} \dots j_{i_t}} = \frac{1}{\prod_{i=1}^t (a_{i_l})^{2j_{i_l}}} \cdot \frac{1}{\prod_{\substack{m=1 \\ i_m \neq i_1, \dots, i_t}}^t (x_{i_m, \phi})^{2j_m}} \cdot D\hat{\gamma}_{\sum_{i=1}^t j_{i_l}}. \quad (8)$$

As well as in Subsection 3.1, the analysis of expression (8) allows at the preliminary stage of the individual algorithm of the decomposition method to set the input values of the active experiment (considering possible repetitions of the main experiment [1]). This allows us to obtain an estimate of (8) with the given accuracy or to conclude about the impossibility of implementing such an active experiment.

3.3. The third human-computer algorithmic procedure (the third AO). The user selects the nonlinear term of the MPR

$$b_{i_1 \dots i_t}^{l_1 \dots l_t} x_{i_1}^{l_1} \cdot x_{i_2}^{l_2} \dots x_{i_t}^{l_t}. \quad (9)$$

In the main active experiment, the input variables will take the values $x_{i_l, i} = a_{i_l} z_i + b_{i_l}, l = \overline{1, t}, i = \overline{1, n}$; a_{i_l}, b_{i_l}, z_i are set as in Subsection 3.1. Other input variables in the active experiment take the minimum possible values. Algorithmically, the verification of the following condition is carried out: there should not be any nonlinear term of the MPR containing input variables $x_{i_l}, l = \overline{1, t}$, or their subset, the total degree of which is greater than or equal to t . If such terms exist, then their coefficients with the specified accuracy must be found in the previous steps of the second stage of the individual algorithm.

If this condition is fulfilled, similarly to Subsections 3.1, 3.2, we obtain the variance of the coefficient estimate of $b_{i_1 \dots i_t}^{l_1 \dots l_t}$ (8) at the preliminary stage of the individual algorithm:

$$D\hat{b}_{i_1 \dots i_t}^{l_1 \dots l_t} = \frac{1}{\prod_{l=1}^t (a_{i_l})^2} \cdot D\hat{\gamma}_t. \quad (10)$$

Remark 4. Implementation of Subsection 3.3 is real if the values of all positive coefficients a_{i_l} [1] are close to or exceed one. For example, if in [1] $z_{10} - z_1 = 100$, then the

values of $d_i - c_i, i = \overline{1, 10}$, should be close to or greater than 100.

3.4. The fourth human-computer algorithmic procedure (the fourth AO). The user selects the nonlinear term of the MPR

$$b_{i_1 \dots i_t}^{1 \dots 1} x_{i_1} \cdot x_{i_2} \cdots x_{i_t}. \quad (11)$$

A subset of its input variables $i_1, \dots, i_m, m < t, m \geq 2$, in the main active experiment will take the value $x_{i_l, i} = a_{i_l} z_i + b_{i_l}, l = \overline{1, m}, m < t, i = \overline{1, n}$, as in the previous Subsections 3.1–3.3, according to [1]. In the main active experiment, the other input terms of (11) take maximally large values, and the rest take minimally possible values. The fulfillment of the following condition is checked: there is no nonlinear term of the MPR containing input variables x_{i_1}, \dots, x_{i_m} or their subset, the total degree of which is greater than or equal to m . If there are such terms, then their estimates will be found with the given accuracy in the previous steps of the second stage of the individual algorithm.

When this condition is met, the variance of the estimate of the coefficient (11) is equal to

$$D\hat{b}_{i_1 \dots i_t}^{1 \dots 1} = \frac{1}{\prod_{l=1}^m (a_{i_l})^2} \cdot \frac{1}{\prod_{l=1}^t (x_{j_l, \phi})^2} \cdot D\hat{\gamma}_m. \quad (12)$$

Remark 5. According to [1], the estimate of the coefficient at the corresponding term of the virtual UPR according to the results of the main virtual active experiment is given by the expression

$$\hat{\gamma}_j = \sum_{i=1}^n \bar{y}_i Q_j(z_i), \quad (13)$$

where $Q_j(z)$ are normalized orthogonal polynomials of Forsythe (NOPF) found by the values $z_i, i = \overline{1, n}$, of the virtual input variable z . Thus, the estimates of coefficients, based on the results of real active experiments, are found using only a single set of NOPFs, the coefficients of which are found in advance with a given accuracy.

Remark 6. Estimates of the coefficients for nonlinear terms of the MPR are found by the formulas:

a) for the first human-computer algorithmic procedure (Subsection 3.1):

$$\hat{b}_{i_1 \dots i_t}^{j_1 \dots j_t} = \frac{1}{(a_{i_k})^{j_k} \cdot \prod_{l=1, l \neq k}^t (x_{i_l, \phi})^{j_l}} \cdot \hat{\gamma}_{j_k}; \quad (14)$$

b) for the second human-computer algorithmic procedure (Subsection 3.2):

$$\hat{b}_{i_1 \dots i_t}^{j_1 \dots j_t} = \frac{1}{\prod_{i=1}^l (a_{i_i})^{j_i} \cdot \prod_{m=1}^t (x_{i_m, \phi})^{j_m}} \cdot \hat{\gamma}_{\sum_{i=1}^l j_i}; \quad (15)$$

c) for the third human-computer algorithmic procedure (Subsection 3.3):

$$\hat{b}_{i_1 \dots i_t}^{1 \dots 1} = \frac{1}{\prod_{l=1}^t a_{i_l}} \cdot \hat{\gamma}_t; \quad (16)$$

d) for the fourth human-computer algorithmic procedure (Subsection 3.4):

$$\hat{b}_{i_1 \dots i_t}^{1 \dots 1} = \frac{1}{\prod_{l=1}^m a_{i_l} \cdot \prod_{\substack{l=1 \\ \forall j_l \notin \{i_1, \dots, i_m\}}}^t x_{j_l, \phi}} \cdot \hat{\gamma}_m. \quad (17)$$

Formulas (14)–(17) are a consequence of the equation expressing the coefficient at the maximum term of the virtual UPR through its corresponding coefficient $b_{i_1 \dots i_t}^{j_1 \dots j_t}$. The corollaries of these equations are also the formulas for the variances of the coefficients (5), (8), (10), (12).

Remark 7. If there are input variables, the range of admissible values of which contains zero, then using the possibility to fix a zero value for an input variable in an active experiment can increase the number of coefficients for nonlinear terms of the MPR, the estimate of which is the solution of a linear equation with a single unknown (see the illustrative example in Section 4).

Remark 8. The redundant representation of the MPR may contain input variables of a power greater than the maximum degree of the NOPFs, the coefficients of which are found in advance with a given accuracy. Then, the values of such variables should be fixed in all active experiments.

4. Illustrative example. Let us set a redundant representation of the MPR in the form:

$$Y(x_1, x_2, x_3, x_4, x_5, x_6) = b_0 + b_1 x_1^2 x_3^3 x_2 + b_2 x_1^2 x_2^2 x_4 + b_3 x_2^2 x_1 x_4 + b_4 x_2^2 x_4^2 x_1 + b_5 x_3^3 x_4^2 x_5 + b_6 x_3^3 x_1 x_2 x_6 + b_7 x_5^6 x_4^3 x_1^2 + b_8 x_4^3 x_6^2 x_1 + b_9 x_1 x_2 x_3 x_4 x_6 + E, \quad (18)$$

where $ME = 0, DE = 4, E$ distributed according to the normal law.

The true values of the coefficients are as follows: $b_0 = 1, b_1 = 2, b_2 = 0, b_3 = 1, b_4 = 0, b_5 = 2, b_6 = 1, b_7 = 2, b_8 = 1, b_9 = 2$.

Thus, the true representation of the MPR lacks the terms with coefficients b_2, b_4 .

Areas in which input variables can take values are as follows: $x_1 \in [1, 10], x_2 \in [1, 10], x_3 \in [1, 10], x_4 \in [1, 10], x_5 \in [0, 5], x_6 \in [1, 5]$.

Remark 9. The values of the input variables, which take different values in the simulation of a real active experiment at each step, are set according to formulas (7), (8) [1]:

$$a_j = \frac{d_j - c_j}{z_{10} - z_1}; b_j = c_j - \frac{d_j - c_j}{z_{10} - z_1} z_1;$$

$$x_{ji} = a_j z_i + b_j, i = \overline{1, 10}.$$

Remark 10. The MPR (18) does not contain linear components because, according to the theoretical provisions of the decomposition method, their estimate is inefficient and is found using the modified group method of data handling [1], and the presence of a linear part in an MPR does not affect the accuracy of coefficient estimates for nonlinear terms of an MPR.

Remark 11. The maximum degree of a NOPF is equal to five.

The user visually sets the sequence of steps of the individual algorithm at the preliminary stage of the analysis:

the first step for the coefficient b_1 (the second AO, Subsection 3.2), the second step for b_2 (the second AO), the third step for b_4 (the second AO), the fourth step for b_3 (the first AO, Subsection 3.1), the fifth step for b_6 (the first AO), the sixth step for b_5 (the first AO), the seventh step for b_7 (the second AO), the eighth step for b_8 (the second AO), the ninth step for b_9 (the fourth AO, Subsection 3.4). The results of the calculations are given in tables 1–4.

Remark 12. Due to the limited size of this article, the values of the input and output variables in the simulation of

Table 1 – Realizations of the random variable E used according to steps 1–9 in the simulation of the main active experiment

Step 1	Step 2	Step 3
-1.0659694104455328	-2.195369811977885	0.807950141301234
3.7710681456850645	-0.12063419142686603	2.6127415396270233
-0.7377368040485024	-1.7907751802465064	1.9923563564870557
-0.0769108594611502	-2.1026649535817645	-0.005506315074662242
1.659240003527758	-1.5447658191004847	1.2290579479034371
0.6385687090711256	-0.2002561262551154	0.1419919936945139
1.393884665512836	2.6798399955112573	0.7288220117347707
-0.9162780984774989	-2.7977613703718434	-0.5038414467160576
-4.381571522845034	2.4754704037663915	-0.24207886186565378
-5.414570636039612	1.056266088850859	1.4330913734546347
Step 4	Step 5	Step 6
-0.6893135072283657	-0.6893135072283657	1.540866648948249
-1.1931899114947138	-1.1931899114947138	1.869030207418431
3.0249102665371534	3.0249102665371534	-2.4896342020289994
0.5270108534316472	0.5270108534316472	-1.3621035826330083
1.7601025580636376	1.7601025580636376	-1.3059043929109533
2.666542829364775	2.666542829364775	2.3330867000197006
3.0458456722762604	3.0458456722762604	0.12292969415601586
2.130621053328643	2.130621053328643	1.6969924752545953
-2.6190016516637336	-2.6190016516637336	0.5081198977551179
-0.2229956051703742	-0.2229956051703742	1.0298668767026709
Step 7	Step 8	Step 9
-1.0904479093952715	-1.789867368121218	-1.789867368121218
0.9699664610876225	1.6833116410256515	1.6833116410256515
3.2737368583418776	-0.8595534529753378	-0.8595534529753378
0.6159984517967693	-0.7063717798945163	-0.7063717798945163
-1.4712753850010287	-0.5767283796110184	-0.5767283796110184
1.7289268018974024	-0.8432436883496297	-0.8432436883496297
-1.2971084475562134	-1.2037914210764402	-1.2037914210764402
-0.7872578872282519	1.1739094496264404	1.1739094496264404
0.7516094848925081	-3.719826795861956	-3.719826795861956
0.3787300161358095	2.5730133007139497	2.5730133007139497

Table 2 – Found coefficients of the NOPFs

	Q_0	Q_1	Q_2	Q_3	Q_4	Q_5
q_{j0}	0.31622776601683 794	0	-0.3590304652533 0308508	0	0.37601029391503 510256	0
q_{j1}		0.00990875675817 27680483	0	-0.0237237389744 04719567	0	0.04073337742261 610065
q_{j2}			0.00035250853500 849863344	0	-0.0012936966199 861974775	0
q_{j3}				0.00001311694894 4919936589	0	-0.0000674335625 32976568096
q_{j4}					5.111689344494366 79076·10 ⁻⁷	0
q_{j5}						2.11434860415668 52428·10 ⁻⁸

Table 3 – Calculation results for the illustrative example

Step number	The maximum degree of virtual UPR	The coefficient variance at the maximum degree of virtual UPR	Estimates of the corresponding coefficients of the virtual UPR
1	5	$D\hat{\gamma}_5 = 4 \cdot 4.5 \cdot 10^{-16}$	$\hat{\gamma}_5 = 0.000118172133464$
2	4	$D\hat{\gamma}_4 = 4 \cdot 2.6 \cdot 10^{-13}$	$\hat{\gamma}_4 = -0.00000022924202$
3	4	$D\hat{\gamma}_4 = 4 \cdot 2.6 \cdot 10^{-13}$	$\hat{\gamma}_4 = -0.0000000981927$
4	2	$D\hat{\gamma}_2 = 4 \cdot 1.7 \cdot 10^{-6}$	$\hat{\gamma}_2 = 0.80686953376847$
5	3	$D\hat{\gamma}_3 = 4 \cdot 4.7 \cdot 10^{-9}$	$\hat{\gamma}_3 = 0.36436409237099$
6	3	$D\hat{\gamma}_3 = 4 \cdot 4.7 \cdot 10^{-9}$	$\hat{\gamma}_3 = 0.72882533637497$
7	5	$D\hat{\gamma}_5 = 4 \cdot 4.5 \cdot 10^{-16}$	$\hat{\gamma}_5 = 0.1845281295144$
8	5	$D\hat{\gamma}_5 = 4 \cdot 4.5 \cdot 10^{-16}$	$\hat{\gamma}_5 = 0.0000117372678$
9	2	$D\hat{\gamma}_2 = 4 \cdot 1.7 \cdot 10^{-6}$	$\hat{\gamma}_2 = 8.09982106030674$

Table 4 – Calculation results for the illustrative example (continued)

Step number, coefficient, input variables which do not take a fixed value in the active experiment	Equation for estimating the coefficient at the nonlinear term of the MPR	Estimate of the coefficient at the nonlinear term	The estimate variance
1, b_1, x_1, x_3	$\hat{b}_1 \cdot a_1^2 \cdot a_3^3 \cdot 10 = \hat{\gamma}_5$	$\hat{b}_1 = 2.001$	$5.128473 \cdot 10^{-7}$
2, b_2, x_1, x_2	$\hat{b}_2 \cdot a_1^2 \cdot a_2^2 \cdot 10 = \hat{\gamma}_4$	$\hat{b}_2 = -0.0003$	$2.428001 \cdot 10^{-6}$
3, b_4, x_2, x_4	$\hat{b}_4 \cdot a_2^2 \cdot a_4^2 \cdot 10 = \hat{\gamma}_4$	$\hat{b}_4 = -0.0001$	$2.428001 \cdot 10^{-6}$
4, b_3, x_2	$\hat{b}_3 \cdot a_2^2 \cdot 10 \cdot 10 = \hat{\gamma}_2$	$\hat{b}_3 = 0.9961$	$1.096121 \cdot 10^{-5}$
5, $b_6, x_3, x_{5,\phi} = 0$	$\hat{b}_6 \cdot a_3^3 \cdot 10 \cdot 10 \cdot 5 = \hat{\gamma}_3$	$\hat{b}_6 = 0.9996$	$1.420843 \cdot 10^{-7}$
6, b_5, x_3	$\hat{b}_5 \cdot a_3^3 \cdot 10^2 \cdot 5 = \hat{\gamma}_3$	$\hat{b}_5 = 1.9995$	$1.420843 \cdot 10^{-7}$
7, b_7, x_1, x_4	$\hat{b}_7 \cdot a_1^2 \cdot a_4^3 \cdot 5^6 = \hat{\gamma}_5$	$\hat{b}_7 = 2.0000$	$2.100622 \cdot 10^{-13}$
8, b_8, x_4, x_6	$\hat{b}_8 \cdot a_6^2 \cdot a_4^3 \cdot 10 = \hat{\gamma}_5$	$\hat{b}_8 = 1.0062$	$1.314371 \cdot 10^{-5}$
9, b_9, x_2, x_4	$\hat{b}_9 \cdot a_2 \cdot a_4 \cdot 10 \cdot 10 \cdot 5 = \hat{\gamma}_2$	$\hat{b}_9 = 1.9999$	$4.384485 \cdot 10^{-7}$

a real active experiment at each step are not presented. Below are the tables for NOPFs and intermediate results of calculations necessary to obtain estimates of coefficients b_1, \dots, b_9 .

The NOPFs were built for $z_i, i = \overline{1, 10}, z_1 = -50, z_{10} = 50$, and their coefficients were found with an accuracy of up to 17 decimal places (Table 2).

Formulas for finding coefficient estimates of the virtual UPR:

$$\hat{\gamma}_j = \sum_{i=1}^{10} \bar{y}_i Q_j(z_i), j = \overline{0, 5};$$

$$\hat{y}_j = \hat{w}_5 q_{5j} + \dots + \hat{w}_j q_{jj}, j = \overline{0, 5},$$

where \bar{y}_i is the value of the output variable of the virtual active experiment (given by formula (4)), y_i is the value of the output variable of the corresponding real main experiment.

Remark 13. In all experiments, the minimum values of the fixed input variables are equal to one.

Remark 14. If the formal procedure of the first subalgorithm [1] were used to find coefficient estimates, then their

estimates would be the solution of the corresponding linear systems of algebraic equations of the dimension 2×2 . The coefficients at the variables of these equations depend on the choice of fixed values of the corresponding input variables from the area of their definitions. But, in contrast to equations with a single variable, there is currently no strategy for the efficient selection of their values to minimize the variance of estimates of unknown coefficients.

Conclusions.

1. The possibility of using the theoretical provisions of the decomposition method to estimate unknown coefficients at nonlinear terms of an MPR is being investigated in order to build an individual algorithm created by a user based on the analysis of the structure of the MPR given by a redundant representation.

2. Four aggregated operators are proposed, which allow solving the formulated problem. The first two generalized operators are used to estimate the coefficients at nonlinear members of the MPR which contain at least one variable to the power greater than or equal to two; the last two ones are for the case when all variables in the nonlinear term of the MPR are to the power of one.

3. An example is given that illustrates the efficiency of building an individual algorithm for estimating coefficients

at nonlinear terms of an MPR in comparison with the formal algorithmic procedure of the decomposition method.

References

- Pavlov A. A., Holovchenko M. N., Drozd V. V. Efficiency substantiation for a synthetical method of constructing a multivariate polynomial regression given by a redundant representation. *Вісник Нац. техн. ун-ту «ХПІ»: зб. наук. пр. Темат. вип.: Системний аналіз, управління та інформаційні технології*. Харків: HTU «ХПІ», 2023. № 1 (9). С. 3–9. DOI: 10.20998/2079-0023.2023.01.01.
- 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.
- 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.
- 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.
- 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.
- 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
- 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.
- 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.
- 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.
- Настенко Е., Павлов В., Бойко Г., Носовец О. Многокритеріальний алгоритм шагової регресії. *Біомедична інженерія і технологія*, 2020. № 3. С. 48–53. DOI: 10.20535/2617-8974.2020.3.195661.
- 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.
- Худсон Д. Статистика для физиков: Лекции по теории вероятностей и элементарной статистике. Москва: Мир, 1970. 296 с.
- 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: HTU «ХПІ», 2022. № 1 (7). С. 3–8. DOI: 10.20998/2079-0023.2022.01.01.
- 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. *Lecture Notes on Data Engineering and Communications Technologies*. 2023. Vol. 181. P. 207–222. DOI: 10.1007/978-3-031-36118-0_19
- 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: HTU «ХПІ», 2022. № 2 (8). С. 3–8. DOI: 10.20998/2079-0023.2022.02.01.

References (transliterated)

- Pavlov A. A., Holovchenko M. N., Drozd V. V. Efficiency substantiation for a synthetical method of constructing a multivariate polynomial regression given by a redundant representation. *Visnyk Nats. tekhn. un-tu "KhPI": zb. nauk. pr. Temat. vyp.: Systemnyy analiz, upravlinnya ta informatsiyni tekhnologiyi* [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 (9), P. 3–9. DOI: 10.20998/2079-0023.2023.01.01.
- 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.
- 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.
- 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.
- 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.
- 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
- 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.
- 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.
- 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, P. 3006–3013. DOI: 10.1039/C9AY00890J.
- Nastenko E., Pavlov V., Boyko G., Nosovets O. Mnogokriterial'nyj algoritm shagovoj regressii. *Biomedychna inzheneriya i tekhnolohiya* [Biomedical engineering and technology]. 2020, no. 3, P. 48–53. DOI: 10.20535/2617-8974.2020.3.195661.
- 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.
- 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 jelementarnoj statistike*. Moscow, Mir Publ., 1970. 296 p.). DOI: 10.5170/CERN-1964-018.
- 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 tekhnologiyi* [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), P. 3–8. DOI: 10.20998/2079-0023.2022.01.01.
- 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. *Lecture Notes on Data Engineering and*

Communications Technologies. 2023, vol. 181, P. 207–222. DOI: 10.1007/978-3-031-36118-0_19

15. 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 tekhnologiyi* [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), P. 3–8. DOI: 10.20998/2079-0023.2022.02.01.

Received 01.11.2024

УДК 004:519.24

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>

МОДИФІКАЦІЯ ДЕКОМПОЗИЦІЙНОГО МЕТОДУ ПОБУДОВИ БАГАТОВИМІРНОЇ ПОЛІНОМІАЛЬНОЇ РЕГРЕСІЇ, ЛІНІЙНОЇ ВІДНОСНО НЕВІДОМИХ КОЕФІЦІЄНТІВ

Авторами був створений універсальний метод побудови багатовимірної поліноміальної регресії, заданої надлишковим описом. Метод є синтетичним, що органічно поєднує декомпозиційний метод та модифікований метод групового урахування аргументів. Спочатку реалізується декомпозиційний метод, що полягає в декомпозиції багатовимірної задачі на послідовність підзадач побудови одновимірних поліноміальних регресій та відповідних систем лінійних рівнянь, змінними яких є оцінки при нелінійних членах багатовимірної поліноміальної регресії. Розглянуті часткові випадки, що гарантують знаходження оцінок з наперед заданою величиною їх дисперсій. Формальний алгоритм побудови оцінок коефіцієнтів при нелінійних членах багатовимірної поліноміальної регресії припиняє роботу на першому коефіцієнті, оцінка якого з наперед заданою точністю не досягається при заданих обмеженнях на кількість випробувань. Оцінка всіх коефіцієнтів, що не були знайдені декомпозиційним методом, знаходиться евристичним методом, що є ефективною модифікацією метода групового урахування аргументів. Підвищення ефективності синтетичного методу досягається в першу чергу за рахунок знаходження таких нових теоретично обґрунтованих алгоритмічних процедур (агрегованих операторів) декомпозиційного методу, що суттєво, в порівнянні з його попередньою версією, збільшує кількість коефіцієнтів при нелінійних членах багатовимірної поліноміальної регресії, що можуть бути знайдені з наперед заданою точністю. Автори показали, що цей ефект досягається за рахунок нових теоретичних положень, що використовуються при візуальному аналізі професійним користувачем структури багатовимірної поліноміальної регресії, заданої надлишковим описом. Наведений ілюстративний приклад полегшує використання приведених результатів при розв'язанні практичних задач.

Ключові слова: регресійний аналіз, багатовимірна поліноміальна регресія, надлишковий опис, декомпозиційний метод, індивідуальний алгоритм, метод найменших квадратів.

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

R. R. LAVSHCHENKO, Master of Computer Science, postgraduate student, National Technical University "Kharkiv Polytechnic Institute", Kharkiv, Ukraine; e-mail: ruslan.lavshchenko@infiz.khpi.edu.ua; ORCID: <https://orcid.org/0009-0001-3649-1118>

G. I. LVOV, Doctor of Technical Sciences, Professor at the Department of mathematical modeling and intelligent computing in engineering, National Technical University "Kharkiv Polytechnic Institute", Kharkiv, Ukraine; e-mail: lvovdpm@ukr.net; ORCID: <https://orcid.org/0000-0003-0297-9227>

ANALYSIS OF THE APPLICATIONS OF THE DATA-DRIVEN APPROACH IN EVALUATING THE THERMAL-PHYSICAL PROPERTIES OF COMPOSITES

This research analyzes the potential and prospects of a data-driven methodology for examining the thermo-physical properties of composite materials. The research is to provide an analysis of the potential and benefits of employing data-driven procedures, especially in contrast to conventional methods. The analysis examines fundamental principles and advanced machine learning approaches utilized in materials science, highlighting their ability to improve the knowledge, optimization, and overall quality of composite materials. This study thoroughly examines the application of neural networks in forecasting thermal characteristics, highlighting its predictive skills and potential to transform the analysis of thermal properties in composite materials. Additionally, the research underscores the growing reliance on big data analytics in addressing complex challenges in material behavior, particularly under variable environmental conditions. A comparison assessment is performed between the data-driven methodology and traditional analytical methodologies, emphasizing the distinct advantages and drawbacks of each. This comparison elucidates how data-driven methodologies can enhance and refine the precision of thermo-physical analysis. The convergence of machine learning and material science is shown to not only facilitate more accurate predictions but also reduce experimentation time and costs. The report also delineates contemporary techniques for measuring and forecasting the thermo-physical properties of composites, emphasizing the advancements in new technologies in recent years. The function of computational tools and computer technology is elaborated upon, especially with the modeling of thermo-physical properties and the simulation of production processes for composite materials. This paper highlights the growing significance of these technologies in enhancing both theoretical and practical dimensions of material science. The research provides novel insights into composite manufacture, thereby advancing the future of materials science and the practical applications of composite materials. The results have significant implications for enhancing production processes, fostering innovation, and progressing the research of composite materials across diverse industries.

Keywords: data-driven approach, composites, thermo-physical properties, data analysis, mathematical modeling, machine learning, process optimization, simulations.

Introduction. The thermophysical properties of composites govern their ability to transmit, dissipate, and conduct heat, consequently impacting their efficacy in diverse technical and engineering applications. These parameters include thermal conductivity, heat capacity, thermal expansion coefficient, and others that are important for heat transfer processes and stress analysis in structural elements. Composites, which are made up of two or more components with different physical properties, provide a challenging field for investigating thermophysical processes. In contrast to homogeneous materials, which have thermal conductivity and thermal expansion specified by scalar parameters, composites have these qualities described by second-rank tensors. Performing actual physical investigations to ascertain all the constituents of these tensors is not always viable or pragmatic. Predicting the physical properties of composites using theory is typically more effective.

The study of thermophysical characteristics has gained significance in materials science and engineering due to the progress in novel materials and technologies. Both the precision of measurements and the effectiveness of data analysis methods are vital in this setting. This underscores the need for enhancing information processing methods, which becomes particularly relevant in the context of employing a data-driven approach [1].

There has been an increasing interest in using data-driven methods in scientific and engineering research, such as materials analysis, in recent decades. The data-driven approach involves the use of machine learning algorithms

and other data processing techniques to identify patterns, trends, and predict outcomes based on large datasets [2]. This opens new avenues for improving the accuracy and efficiency of analyzing the thermophysical properties of composites.

The data-driven approach offers advantages over traditional research approaches in terms of efficiency and cost-effectiveness. It enables the automation of analysis processes, facilitates the rapid identification of intricate linkages, and allows for prediction based on current data. This holds great importance in the current scientific landscape, since the growing amount of data requires novel methods for its analysis and interpretation.

Examining the thermophysical characteristics of composites using a data-driven methodology can result in the identification of novel patterns and the enhancement of forecasting techniques. This approach facilitates a more precise comprehension of thermal processes in intricate materials and promotes the advancement of cutting-edge technologies in materials science and engineering. This literature review will analyze the fundamental elements of adopting a data-driven approach to analyze the thermo-physical properties of composites. The focus will be on current research and important discoveries that could influence the future direction of this area.

Fundamental principles of data-driven analysis in materials science. Data-driven analysis in materials science is a novel method for studying and optimizing material properties. It is crucial to have a thorough comprehension of the fundamental ideas of this technique in order to

© Lavshchenko R. R., Lvov G. I., 2024



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 licensed under an international license [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.



comprehend its principles and possibilities in scientific and engineering research. This notion entails the utilization of analytical and statistical techniques to derive insights from vast amounts of data. Data-driven research allows for the discovery of intricate linkages and patterns that may go unnoticed when employing traditional methods [1].

Big data processing methods are essential in materials research, where there is a large volume of experimental and computational data. These methods aid in the examination and identification of patterns within extensive datasets [3]. Clustering algorithms are utilized in data-driven analysis in materials science to group materials that have similarities, while classification methods are used to assign these materials to specific groups. This methodology facilitates the comprehension of the disparities in structure and characteristics between different materials [2]. The use of visualization becomes an effective means of representing complex data, with data-driven analysis incorporating graphical tools to make the results of the analysis more accessible [4]. Regression analysis is used in data-driven approaches to identify the relationships between distinct material parameters. This helps in predicting the qualities and behavior of these parameters under different scenarios.

Machine learning in materials science involves the utilization of algorithms and models that have the ability to adapt and enhance their performance by analyzing and learning from data. Machine learning approaches are essential for data-driven analysis in materials science. Machine learning in this field [5] includes both supervised learning techniques, where models are trained based on input-output pairs, and unsupervised learning methods, where algorithms analyze data without a clearly defined training set. Machine learning allows the automated processing and analysis of vast amounts of data. Algorithms have the ability to rapidly identify intricate connections and create models that can anticipate outcomes [3]. Machine learning approaches provide accurate prediction of thermophysical properties of materials by leveraging existing data and considering several aspects [2]. Machine learning enables the efficient investigation of thermophysical qualities as well as other material aspects, including structure and composition. This enhances comprehension of the interconnectedness within the field of materials science.

This article provides an analysis of the basic principles of data-driven analysis and the use of machine learning methods in materials science. It emphasizes the promise of these approaches in studying and improving composite properties. Continuing study in this direction is of utmost importance in order to attain novel scientific breakthroughs and technological progress in the field of material development.

Thermophysical properties of composites and their measurement and prediction. The thermophysical properties of composites play a crucial role in determining their behavior and usefulness in different operating circumstances. An examination of the primary thermophysical characteristics is an essential requirement for comprehending and managing thermal processes in composite materials.

Thermal conductivity is a fundamental property that governs a material's capacity to transmit heat. The property

of composites is contingent upon the specific type and amount of components, their microstructure, and the interactions that occur between them. Thermal conductivity in composites exhibits distinct characteristics that can result in unforeseen consequences and fluctuations when compared to uniform materials.

Heat capacity is a measure of how much heat a material can absorb as its temperature changes. The property of composites is governed by the mass content and heat capacity of each component. Heat capacity can be altered by phase transitions or other thermodynamic interactions in the system.

The coefficient of thermal expansion quantifies the extent to which the dimensions of a material alter in response to changes in temperature. The variation of this feature in composites is contingent upon the orientation and relative positioning of the components.

Aside from the primary thermophysical parameters, it is crucial to take into account the thermal stability and durability of composites under severe temperatures. Alterations in the microstructure and chemical composition can have an impact on the durability and effectiveness of composite materials when exposed to high temperatures.

Composites might get dampness, chemicals, harsh environments, and other elements in actual operating settings. Considering these factors is essential for a comprehensive comprehension of the thermophysical characteristics of composites.

In addition to experimental approaches used to determine these characteristics, theoretical prediction plays a crucial role in research. The utilization of contemporary computational technology to simulate and analyze the thermophysical characteristics of composites is of utmost significance [6]. The abundance of experimental data enables the creation of novel theoretical models that can accurately forecast the characteristics of composites under different situations and mixtures. The focus on employing advanced computational approaches allows for accurate forecasting and enhancement of the thermophysical characteristics of composites across many applications.

In addition, the study in reference [7] employed a continuous material property model to analyze hardened composites. Thermocouples were used to assess temperatures inside the composite, and controlling them was an important part of the production process to assure the correct thermophysical properties.

The advancement of materials science and computational technology enables the utilization of efficient techniques to forecast the thermophysical characteristics of composites, hence creating new opportunities for enhancing and extensively utilizing them. Accurate prediction techniques for thermal conductivity and thermal expansion are crucial in the development of materials tailored for specific technological applications and operating environments.

An advanced field is the application of molecular dynamics modeling techniques to forecast the thermophysical characteristics of composites on a molecular scale. These techniques enable the examination of atomic and molecular motion, which is crucial for comprehending heat transport and material effectiveness.

By utilizing computational methods and materials science databases, it is possible to make predictions about thermal conductivity and thermal expansion by considering the properties of the constituent components of a composite. Utilizing this data can enhance forecasts for intricate composite systems [8].

The utilization of artificial intelligence (AI) and machine learning techniques for examining the thermophysical characteristics of composites is gaining popularity. AI models possess the capability to scrutinize vast quantities of data and discern intricate connections among many elements, hence playing a crucial role in making precise predictions [2].

Computer modeling of composite structures enables the consideration of multiple aspects, including fiber arrangement, porosity, and material architecture. This allows for the estimation of thermal conductivity and thermal expansion based on the geometric and chemical properties [4].

In [9], a method for quantitatively determining nonlinear uncertainty and propagation based on data for studying the stochastic characteristics of unidirectional (UD) carbon fiber-reinforced polymer (CFRP) composites was presented. This method allows for the determination of nonlinear uncertainty and propagation using data. The suggested methodology integrates the study of microscopic images, the reconstruction of stochastic microstructures, and the implementation of efficient multiscale finite element modeling. This is further enhanced by the utilization of self-consistent clustering analysis (SCA). The UQ approaches suggested aim to represent intricate variations in microstructure by considering sources of uncertainty that are not Gaussian. This is achieved by a sampling methodology that does not rely on specific distribution assumptions, but instead utilizes non-parametric and asymptotic statistical tools. A hierarchical conditional sampling technique enables the simultaneous selection from various sources of uncertainty. This approach allows for the comprehension of the influence of microstructural alterations on the nonlinear reactions of unidirectional carbon fiber components when subjected to gradual compression loads, ultimately impacting failure rates over time. Accurately identifying the duration of failure is crucial for making reliable predictions about the performance of Carbon Fiber Reinforced Polymer (CFRP) designs.

The study in [10] investigated the practicality of using thermal conductivity and thermal expansion prediction techniques to analyze an automotive crash box. The analysis was based on data collected during the assessment of the building blocks. An automotive crash box is a vehicle component specifically engineered to dissipate energy in the event of a collision. An assessment was conducted to evaluate the overall efficacy of the crash box across various thicknesses and testing velocities. It can be concluded that when the thickness and testing speed are the same, the specific energy absorption (SEA) and peak load values are similar for both the construction block and the crash box. Hence, it is feasible to employ thermal conductivity and thermal expansion prediction techniques for assessing the design of intricate structures, hence streamlining the testing process. Nevertheless, it is crucial to prioritize the evalu-

ation of testing speed, since the proportion of damaged fibers can fluctuate based on it when subjected to quasi-static conditions.

Combining experimental data with computational methods is increasingly proving to be a powerful tool for making predictions. This methodology enables the improvement of models by utilizing actual data acquired from experiments, hence facilitating the development of precise forecasts. A viable option to just relying on experimental techniques for determining the thermophysical properties of composites is to use theoretical calculations based on the known properties of the structural components. A thorough numerical method was introduced in [6] to determine the components of the thermal conductivity and thermal expansion tensors of fibrous composites. This method offers a unified approach and utilizes specialized software. A technique was introduced in [11] to ascertain the thermal conductivity properties of composites with a periodic structure. This method is applicable to composites with any number of phases and inclusion shapes. The method relies on employing Fourier series to depict temperature fields and heat fluxes that exhibit periodic variations. A study conducted in [12] utilized multilevel modeling to analyze the thermal expansion of silicon carbide matrix composites reinforced with carbon fibers, while also considering the presence of ceramic porosity.

The study in [13] explored an experimental technique using a multiscale methodology to identify connections between global stress and local material reaction at the fiber scale. The technique of high-resolution optical digital image correlation was employed to detect regions of low and high strain in woven composite samples that were subjected to uniaxial tension. The measured full-field strains were utilized to develop a relationship between the overall and localized deformation characteristics based on the orientation of the fibers. In addition, the stress-strain curves of samples with varying fiber orientation angles were examined. The correlation between overall stress and strain created at a specific location, along with the incorporation of a plasticity model, enables the creation of a phenomenological model that considers both the overall and local elements of deformation in composite samples under off-axis tensile loading.

A three-dimensional multiscale finite element model was created in [14] to forecast the effective thermal conductivity of CNT/Al composites. A comprehensive investigation was conducted to examine the impact of carbon nanotube (CNT) arrangement, direction of heat transmission, interfacial thermal resistance, and volume fraction. Five topologies of carbon nanotubes (CNTs) were generated, which included randomly organized, uniformly orientated, multilayered, bundled, and networked arrangements. The findings validate that the arrangement of carbon nanotubes (CNT) plays a vital role in determining the thermal conductivity of CNT/aluminum composites. Additionally, the interface between the CNT and aluminum, as well as the direction of heat load, have a substantial impact on the thermal conductivity of these composites. Ultimately, the projected effective thermal conductivity was juxtaposed with empirical findings.

Ultimately, the thermophysical characteristics of composites, in conjunction with sophisticated computational techniques, are crucial in the manufacturing and utilization of novel materials across many industries. By employing these techniques in conjunction, it becomes possible to conduct a thorough examination and make precise forecasts regarding the response of materials under various circumstances. An integrative approach, encompassing both empirical and theoretical methodologies, has the potential to offer novel insights in the field of materials science and engineering. Comprehending and precisely forecasting thermophysical qualities are essential for the advancement of new materials, enhancement of current ones, and guaranteeing their efficacy in practical applications.

Analysis of the applications of the data-driven approach for analyzing the thermophysical properties of composites. In today's world, where the amount of data is rapidly expanding and scientific research is getting more advanced, employing a data-driven methodology in materials science is acknowledged as a viable strategy for investigating the thermophysical characteristics of composites. This section will provide a comprehensive examination of the key elements involved in utilizing data-driven techniques for analyzing the thermophysical properties of composite materials.

Neural networks have emerged as a potent method in current materials research for accurately predicting the thermal properties of composites [5]. Neural networks, which imitate the functioning of the human brain system, are becoming a crucial element of the data-driven method. They allow for the analysis and prediction of thermal properties of materials without explicitly establishing physical equations [1, 6].

Neural networks are computational models designed to mimic the intricate connections and interactions among neurons in the human brain [3, 5]. Neural networks are composed of stacked layers of neurons, with each neuron being interconnected to others by weighted connections. The primary concept is that the network has the ability to acquire knowledge about intricate connections between input and output data.

Predicting the thermal conductivity is a crucial component in studying the thermophysical properties of composites. Neural networks possess the ability to adjust to intricate nonlinear relationships and have demonstrated their efficacy in simulating heat transfer in composite materials [8].

When developing neural networks, selecting the correct architecture is of utmost importance. In the field of materials science, many structural designs are employed, such as conventional layered networks, recurrent networks for analyzing sequences, and deep networks for handling more intricate jobs [2].

The collection of large volumes of experimental data is critical for the successful training of neural networks. The accuracy and efficiency of a neural network model directly depend on the quality and representativeness of the training set [15].

Neural networks possess both benefits and constraints. The benefits encompass the capacity to comprehend

intricate interconnections and handle substantial volumes of data. Nevertheless, there are certain constraints associated with this approach, such as the requirement for a substantial volume of data for training and difficulties in elucidating the outcomes of the model [3].

In general, the utilization of neural networks in forecasting the thermal properties of composites introduces new possibilities in the field of material research [15]. This methodology is especially efficient in situations when conventional analytic approaches fail to offer the required precision or when handling substantial volumes of data.

Examining the thermophysical characteristics of composites using data-driven methodologies and comparing them to traditional approaches is an important undertaking in contemporary materials science. Traditional methodologies, relying on physical models and empirical measurements, and contemporary data-driven techniques, employing machine learning and big data analysis, are distinct approaches for investigating and enhancing the thermophysical characteristics of composite materials.

The conventional approach involves employing experimental techniques to quantify thermophysical parameters, such as thermal conductivity and thermal expansion [8]. It is usual to use theoretical models, such as compatibility models and macroscopic model theory, to anticipate the behavior of composites. Conventional techniques offer significant precision in numerous scenarios, but their effectiveness can be hindered by the intricacy of measurements and the exorbitant expense of research.

The data-driven technique is employed to reveal intricate relationships between thermophysical properties [2]. Machine learning, neural networks, and big data analysis are utilized to discover trends and forecast composite features without relying on explicit physical models. They can be particularly advantageous in circumstances when conventional approaches may be unattainable or costly.

Conventional techniques can offer exceptional precision for particular operational circumstances and substances. Conversely, data-driven techniques, because of their capacity to adjust to intricate connections in extensive data sets, may demonstrate to be more adaptable yet may necessitate a substantial quantity of training data to attain ideal accuracy [3].

Data-driven approaches exhibit a reduced susceptibility to fluctuations in input data and operating conditions as compared to conventional methods. This is due to their ability to incorporate unforeseen or intricate patterns that may not be accounted for in traditional models.

Utilizing data-driven methods can enhance the flexibility of forecasting thermophysical properties in novel settings and with different materials. These algorithms have the ability to adjust to new patterns if there is a enough amount of data available [16].

Conventional approaches can incur significant expenses and consume a considerable amount of time, particularly when dealing with a substantial number of tests. Conversely, the data-driven method can be more easily understood and effective, particularly when there is a substantial amount of preexisting data that can be analyzed [17].

The article [18] presents a macro-micro model that accurately predicts the thermophysical properties of fiber-reinforced composites, including their nonlinear plastic behavior. The current micromechanical model, which is based on two-stage orientation averaging, is utilized at the tiny level. This model is capable of including several microstructural factors, including constitutive parameters of the matrix and fibers, fiber volume fraction, and fiber aspect ratio. The micromechanical model takes into account several interactions, such as Voigt, Reuss, and self-consistent assumptions. Subsequently, this micromechanical model is incorporated into a finite element model to enable macroscopic simulation of actual composite structures and samples.

This method of studying the thermophysical characteristics of composites focuses on creating a comprehensive model that combines both micro and macro elements. When comparing it to classic methods like experiments and physical models, as well as the latest data-driven approaches like machine learning and big data analysis, we can observe the distinct advantages and characteristics of each approach. This paper discusses the accuracy, adaptability, sensitivity, and predictive capacities of different methods used to optimize the thermophysical properties of composites. It also considers how these methods are affected by input data and operating conditions.

The study [19] examines how microstructural factors affect the macromechanical properties of composites reinforced with short fibers. The suggested modeling methodology takes into consideration the variety of these characteristics, including the volume percentage of fibers, the aspect ratios of fibers, and the distributions of fiber orientations. The micromechanical model described here, which was built using finite element analysis and orientation averaging, can be used to forecast the thermophysical properties of composites. When analyzing and optimizing composite qualities, it is crucial to give weight to microstructural features, rather than relying solely on old methodologies or new data-driven approaches.

The study [20] introduces a novel micromechanical model that accurately predicts the nonlinear elastic-plastic characteristics of composites reinforced with short fibers. The model is constructed using a two-stage orientation averaging technique and is capable of incorporating a broad spectrum of microstructural characteristics. During the initial phase, a finite element analysis is conducted on a basic unit known as an elementary cell, which consists of a single fiber surrounded by a matrix. Next, the reaction of the basic unit is adjusted by calibrating its response to an elastic-plastic surrogate constitutive model. In the subsequent phase of homogenization, a self-consistent interaction system is put forward. The predictive capacity of the resulting two-stage homogenization scheme, as well as versions employing more conventional averaging schemes (Voigt and Reuss, which give upper and lower limits, respectively), is subsequently evaluated by comparing it with experiments and direct numerical modeling of realistic representative volume elements.

When comparing traditional and data-driven methodologies, it is crucial to take into account their strengths and

weaknesses, as well as the specific circumstances and objectives of the study. Both methodologies are valuable in the scientific investigation of the thermophysical characteristics of composites, and their synergy can result in novel findings and advancements in materials research.

Conclusions. This article examined the application of the data-driven approach in assessing the thermophysical characteristics of composite materials. The research findings emphasize the substantial impact of this methodology on the progress of contemporary materials science and its utility in forecasting the thermal properties of composites.

Neural networks have demonstrated high efficacy in forecasting the thermal conductivity of composites. Machine learning has facilitated the development of models capable of adjusting to intricate thermal processes and considering non-linear connections between various material characteristics.

The research demonstrated that selecting the suitable neural network design is essential for achieving high accuracy and versatility in the model. Different designs, such as recurrent and deep networks, are employed based on the individual characteristics of the materials and the task being performed.

Analyzing extensive datasets was a critical process in researching thermophysical characteristics. The utilization of statistical methodologies, machine learning, and computational techniques has facilitated the attainment of more precise outcomes and a more profound comprehension of patterns in heat transfer and other characteristics.

The data-driven approach offers more adaptable and pliable models in comparison to traditional approaches. The efficacy of such models relies on the abundance of extensive data, although their prediction prowess is notably enhanced under specific circumstances. Data-driven solutions mitigate the impact of fluctuations in input data and operational conditions. This attribute renders them highly efficient instruments for forecasting material performance in diverse settings.

The research findings demonstrate promising potential for employing the data-driven approach in future investigations of the thermophysical characteristics of composites. Additionally, they aid in identifying avenues for future research and development in this field.

Exploring the field of multimaterial systems and composites composed of diverse materials has the potential to become a significant focus of research. Utilizing data-driven approaches offers the capacity to accurately model thermophysical parameters for a wide range of material combinations. Future advancements in data-driven methods could prioritize the optimization of thermophysical characteristics in composites. Employing machine learning algorithms to automatically seek for optimal material configurations helps streamline and expedite the design process. Enhancing the synergy between data-driven models and experimental research is crucial for achieving precision and verifying outcomes. This will enable the validation and augmentation of models with empirical data.

The data-driven method can effectively be utilized in various domains, including energy, medicine, and the aviation industry. Modifying and expanding models to

various domains can potentially result in novel applications and findings.

Further investigation into the influence of micro-structure on the thermophysical characteristics of composites could be a significant avenue to explore. Utilizing data-driven methodologies can assist in the identification and modeling of intricate linkages.

The establishment of open databases and the implementation of standardized data processing methods have the potential to promote community development and improve the accuracy and dependability of outcomes. This stage is crucial for the development of collective expertise in the topic.

The primary objective of future research is to further advance the data-driven technique as a highly efficient tool for assessing the thermophysical properties of composites and integrating it into other sectors and scientific domains.

References

- Wu L., Zhang P., Xu B., Liu J., Yin H., Zhang L., Jiang X., Zhang C., Zhang R., Wang Y., Qu X. Data-driven design of brake pad composites for high-speed trains. *Journal of Materials Research and Technology*. 2023. Vol. 27. P. 1058–1071. DOI: <https://doi.org/10.1016/j.jmrt.2023.09.280>.
- He Z. C., Huo S. L., Li E., Cheng H. T., Zhang L. M. Data-driven approach to characterize and optimize properties of carbon fiber non-woven composite materials. *Composite Structures*. 2022. Vol. 297. Article 115961. DOI: <https://doi.org/10.1016/j.compstruct.2022.115961>.
- Fathidoost M., Yang Y., Oechsner M., Xu B.-X. Data-driven thermal and percolation analyses of 3D composite structures with interface resistance. *Materials & Design*. 2023. Vol. 227. Article 111746. DOI: <https://doi.org/10.1016/j.matdes.2023.111746>.
- Ciampaglia A. Data-driven statistical method for the multiscale characterization and modelling of fiber reinforced composites. *Composite Structures*. 2023. Vol. 320. Article 117215. DOI: <https://doi.org/10.1016/j.compstruct.2023.117215>.
- Malley S., Reina C., Nacy S., Gilles J., Koohbor B., Youssef G. Predictability of mechanical behavior of additively manufactured particulate composites using machine learning and data-driven approaches. *Computers in Industry*. 2022. Vol. 142. Article 103739. DOI: <https://doi.org/10.1016/j.compind.2022.103739>.
- Lvov G. I. Numerical Homogenization of the Thermophysical Properties of Fibrous Composites. *Mechanics of Composite Materials*. 2022. Vol. 58, no. 5. P. 613–628. DOI: <https://doi.org/10.1007/s11029-022-10054-x>.
- Barkanov E., Akishin P., Namsone E., Auzins J., Morozovs A. Optimization of Pultrusion Processes for an Industrial Application. *Mechanics of Composite Materials*. 2021. Vol. 6. P. 697–712.
- Eleftheroglou N., Zarouchas D., Benedictus R. An adaptive probabilistic data-driven methodology for prognosis of the fatigue life of composite structures. *Composite Structures*. 2020. Vol. 245. Article 112386. DOI: <https://doi.org/10.1016/j.compstruct.2020.112386>.
- Huang T., Gao J., Sun Q., Zeng D., Su X., Liu W. K., Chen W. Stochastic nonlinear analysis of unidirectional fiber composites using image-based microstructural uncertainty quantification. *Composite Structures*. 2021. Vol. 260. Article 113470. DOI: <https://doi.org/10.1016/j.compstruct.2020.113470>.
- Saenz-Dominguez I., Tena I., Esnaola A., Sarrionandia M., Torre J., Aurrekoetxea J. Design and characterisation of cellular composite structures for automotive crash-boxes manufactured by out of die ultraviolet cured pultrusion. *Composites Part B: Engineering*. 2019. Vol. 160. P. 217–224. DOI: <https://doi.org/10.1016/j.compositesb.2018.10.046>.
- Tian W., Chao X., Fu M. W., Qi L., Ju L. New numerical algorithm for the periodic boundary condition for predicting the coefficients of thermal expansion of composites. *Mechanics of Materials*. 2021. Vol. 154. Article 103737.
- Sun Z., Shan Z., Shao T., Li J., Wu X. A multiscale modeling for predicting the thermal expansion behaviors of 3D C/SiC composites considering porosity and fiber volume fraction. *Ceramics International*. 2021. Vol. 47, no. 6. P. 7925–7936.
- Koohbor B., Ravindran S., Kidane A. A multiscale experimental approach for correlating global and local deformation response in woven composites. *Composite Structures*. 2018. Vol. 194. P. 328–334. DOI: <https://doi.org/10.1016/j.compstruct.2018.04.016>.
- Zhou L., Yuan T. B., Yang X. S., Liu Z. Y., Wang Q. Z., Xiao B. L., Ma Z. Y. Microscale prediction of effective thermal conductivity of CNT/Al composites by finite element method. *International Journal of Thermal Sciences*. 2022. Vol. 171. Article 107206.
- Wu L., Adam L., Noels L. Micro-mechanics and data-driven based reduced order models for multi-scale analyses of woven composites. *Composite Structures*. 2021. Vol. 270. Article 114058. DOI: <https://doi.org/10.1016/j.compstruct.2021.114058>.
- Veenstra S. W. P., Wijskamp S., Rosic B., Akkerman R. Bending behaviour of thermoplastic composites in melt: A data-driven approach. *Composites Science and Technology*. 2022. Vol. 219. Article 109220. DOI: <https://doi.org/10.1016/j.compscitech.2021.109220>.
- Cheung H. L., Mirkhalaf M. A multi-fidelity data-driven model for highly accurate and computationally efficient modeling of short fiber composites. *Composites Science and Technology*. 2024. Vol. 246. Article 110359. DOI: <https://doi.org/10.1016/j.compscitech.2023.110359>.
- Castricum B. A., Fagerström M., Ekh M., Larsson F., Mirkhalaf S. M. A computationally efficient coupled multi-scale model for short fiber reinforced composites. *Composites Part A: Applied Science and Manufacturing*. 2022. Vol. 163. Article 107233. DOI: <https://doi.org/10.1016/j.compositesa.2022.107233>.
- Mirkhalaf S. M., Eggels E. H., van Beurden T. J. H., Larsson F., Fagerström M. A finite element based orientation averaging method for predicting elastic properties of short fiber reinforced composites. *Composites Part B: Engineering*. 2020. Vol. 202. Article 108388. DOI: <https://doi.org/10.1016/j.compositesb.2020.108388>.
- Mirkhalaf S. M., van Beurden T. J. H., Ekh M., Larsson F., Fagerström M. An FE-based orientation averaging model for elastoplastic behavior of short fiber composites. *International Journal of Mechanical Sciences*. 2022. Vol. 219. Article 107097. DOI: <https://doi.org/10.1016/j.ijmecsci.2022.107097>.

References (transliterated)

- Wu L., Zhang P., Xu B., Liu J., Yin H., Zhang L., Jiang X., Zhang C., Zhang R., Wang Y., Qu X. Data-driven design of brake pad composites for high-speed trains. *Journal of Materials Research and Technology*. 2023, vol. 27, pp. 1058–1071. DOI: <https://doi.org/10.1016/j.jmrt.2023.09.280>.
- He Z. C., Huo S. L., Li E., Cheng H. T., Zhang L. M. Data-driven approach to characterize and optimize properties of carbon fiber non-woven composite materials. *Composite Structures*. 2022, vol. 297, article 115961. DOI: <https://doi.org/10.1016/j.compstruct.2022.115961>.
- Fathidoost M., Yang Y., Oechsner M., Xu B.-X. Data-driven thermal and percolation analyses of 3D composite structures with interface resistance. *Materials & Design*. 2023, vol. 227, article 111746. DOI: <https://doi.org/10.1016/j.matdes.2023.111746>.
- Ciampaglia A. Data-driven statistical method for the multiscale characterization and modelling of fiber reinforced composites. *Composite Structures*. 2023, vol. 320, article 117215. DOI: <https://doi.org/10.1016/j.compstruct.2023.117215>.
- Malley S., Reina C., Nacy S., Gilles J., Koohbor B., Youssef G. Predictability of mechanical behavior of additively manufactured particulate composites using machine learning and data-driven approaches. *Computers in Industry*. 2022, vol. 142, article 103739. DOI: <https://doi.org/10.1016/j.compind.2022.103739>.
- Lvov G. I. Numerical Homogenization of the Thermophysical Properties of Fibrous Composites. *Mechanics of Composite Materials*. 2022, vol. 58, no. 5, pp. 613–628. DOI: <https://doi.org/10.1007/s11029-022-10054-x>.
- Barkanov E., Akishin P., Namsone E., Auzins J., Morozovs A. Optimization of Pultrusion Processes for an Industrial Application. *Mechanics of Composite Materials*. 2021, vol. 6, pp. 697–712.
- Eleftheroglou N., Zarouchas D., Benedictus R. An adaptive probabilistic data-driven methodology for prognosis of the fatigue life of composite structures. *Composite Structures*. 2020, vol. 245, article 112386. DOI: <https://doi.org/10.1016/j.compstruct.2020.112386>.

9. Huang T., Gao J., Sun Q., Zeng D., Su X., Liu W. K., Chen W. Stochastic nonlinear analysis of unidirectional fiber composites using image-based microstructural uncertainty quantification. *Composite Structures*. 2021, vol. 260, article 113470. DOI: <https://doi.org/10.1016/j.compstruct.2020.113470>.
10. Saenz-Dominguez I., Tena I., Esnaola A., Sarrionandia M., Torre J., Aurrekoetxea J. Design and characterisation of cellular composite structures for automotive crash-boxes manufactured by out of die ultraviolet cured pultrusion. *Composites Part B: Engineering*. 2019, vol. 160, pp. 217–224. DOI: <https://doi.org/10.1016/j.compositesb.2018.10.046>.
11. Tian W., Chao X., Fu M. W., Qi L., Ju L. New numerical algorithm for the periodic boundary condition for predicting the coefficients of thermal expansion of composites. *Mechanics of Materials*. 2021, vol. 154, article 103737.
12. Sun Z., Shan Z., Shao T., Li J., Wu X. A multiscale modeling for predicting the thermal expansion behaviors of 3D C/SiC composites considering porosity and fiber volume fraction. *Ceramics International*. 2021, vol. 47, no. 6, pp. 7925–7936.
13. Koohbor B., Ravindran S., Kidane A. A multiscale experimental approach for correlating global and local deformation response in woven composites. *Composite Structures*. 2018, vol. 194, pp. 328–334. DOI: <https://doi.org/10.1016/j.compstruct.2018.04.016>.
14. Zhou L., Yuan T. B., Yang X. S., Liu Z. Y., Wang Q. Z., Xiao B. L., Ma Z. Y. Microscale prediction of effective thermal conductivity of CNT/Al composites by finite element method. *International Journal of Thermal Sciences*. 2022, vol. 171, article 107206.
15. Wu L., Adam L., Noels L. Micro-mechanics and data-driven based reduced order models for multi-scale analyses of woven composites. *Composite Structures*. 2021, vol. 270, article 114058. DOI: <https://doi.org/10.1016/j.compstruct.2021.114058>.
16. Veenstra S. W. P., Wijskamp S., Rosić B., Akkerman R. Bending behaviour of thermoplastic composites in melt: A data-driven approach. *Composites Science and Technology*. 2022, vol. 219, article 109220. DOI: <https://doi.org/10.1016/j.compscitech.2021.109220>.
17. Cheung H. L., Mirkhalaf M. A multi-fidelity data-driven model for highly accurate and computationally efficient modeling of short fiber composites. *Composites Science and Technology*. 2024, vol. 246, article 110359. DOI: <https://doi.org/10.1016/j.compscitech.2023.110359>.
18. Castricum B. A., Fagerström M., Ekh M., Larsson F., Mirkhalaf S. M. A computationally efficient coupled multi-scale model for short fiber reinforced composites. *Composites Part A: Applied Science and Manufacturing*. 2022, vol. 163, article 107233. DOI: <https://doi.org/10.1016/j.compositesa.2022.107233>.
19. Mirkhalaf S. M., Eggels E. H., van Beurden T. J. H., Larsson F., Fagerström M. A finite element based orientation averaging method for predicting elastic properties of short fiber reinforced composites. *Composites Part B: Engineering*. 2020, vol. 202, article 108388. DOI: <https://doi.org/10.1016/j.compositesb.2020.108388>.
20. Mirkhalaf S. M., van Beurden T. J. H., Ekh M., Larsson F., Fagerström M. An FE-based orientation averaging model for elastoplastic behavior of short fiber composites. *International Journal of Mechanical Sciences*. 2022, vol. 219, article 107097. DOI: <https://doi.org/10.1016/j.ijmecsci.2022.107097>.

Received 17.09.2024

УДК 004

Р. Р. ЛАВЩЕНКО, магістр комп'ютерних наук, аспірант, Національний технічний університет «Харківський політехнічний інститут», м. Харків, Україна; e-mail: ruslan.lavshchenko@infiz.khpi.edu.ua; ORCID: <https://orcid.org/0009-0001-3649-1118>

Г. І. ЛЬВОВ, доктор технічних наук, професор кафедри математичного моделювання та інтелектуальних обчислень в інженерії (ММІ), Національний технічний університет «Харківський політехнічний інститут», м. Харків, Україна; e-mail: lvovdpm@ukr.net; ORCID: <https://orcid.org/0000-0003-0297-9227>

АНАЛІЗ ЗАСТОСУВАНЬ DATA-DRIVEN ПІДХОДУ ДЛЯ ОЦІНЮВАННЯ ТЕРМОФІЗИЧНИХ ВЛАСТИВОСТЕЙ КОМПОЗИТІВ

Це дослідження аналізує потенціал і перспективи використання data-driven підходу для вивчення теплофізичних властивостей композитних матеріалів. Дослідження має на меті надати аналіз потенціалу і переваг застосування data-driven методологій, особливо в порівнянні з традиційними методами. Аналіз охоплює основні принципи та сучасні підходи машинного навчання, які використовуються в матеріалознавстві, підкреслюючи їх здатність покращувати розуміння, оптимізацію та загальну якість композитних матеріалів. У дослідженні детально розглядається застосування нейронних мереж для прогнозування теплових характеристик, підкреслюючи їх прогностичні можливості та потенціал для трансформації аналізу теплових властивостей композитних матеріалів. Крім того, дослідження підкреслює зростаючу залежність від big-data аналітики у вирішенні складних проблем у поведінці матеріалів, особливо в змінних умовах навколишнього середовища. Проводиться порівняльна оцінка між data-driven підходом та традиційними аналітичними методами, підкреслюючи чіткі переваги та недоліки кожного. Це порівняння висвітлює, як data-driven методології можуть покращити точність теплофізичного аналізу. Зближення машинного навчання та матеріалознавства сприяє не лише точнішим прогнозам, але й зменшенню часу та вартості експериментів. Звіт також окреслює сучасні методи вимірювання та прогнозування теплофізичних властивостей композитів, підкреслюючи досягнення в нових технологіях останніх років. Особливо розглядається роль обчислювальних інструментів і комп'ютерних технологій, особливо в моделюванні теплофізичних властивостей і симуляції виробничих процесів для композитних матеріалів. У статті підкреслюється зростаюче значення цих технологій для покращення як теоретичних, так і практичних аспектів матеріалознавства. Дослідження надає нові погляди на виробництво композитів, сприяючи розвитку матеріалознавства та практичних застосувань композитних матеріалів у майбутньому. Результати мають значний вплив на покращення виробничих процесів, стимулювання інновацій та прогрес у дослідженні композитних матеріалів у різних галузях.

Ключові слова: data-driven підхід, композити, теплофізичні властивості, аналіз даних, математичне моделювання, машинне навчання, оптимізація процесів, симуляції.

Повні імена авторів / Author's full names

Автор 1 / Author 1: Лавщенко Руслан Ровшан огли / Lavshchenko Ruslan Rovshan ohly

Автор 2 / Author 2: Львов Геннадій Іванович / Lvov Gennadiy Ivanovych

УПРАВЛІННЯ В ТЕХНІЧНИХ СИСТЕМАХ

CONTROL IN TECHNICAL SYSTEMS

DOI: 10.20998/2079-0023.2024.02.03

УДК 004.93

О. В. ПОЛЯРУС, доктор технічних наук, професор, Харківський національний автомобільно-дорожній університет, м. Харків, Україна; e-mail: poliarus.kharkov@ukr.net; Researcher ID: AАН-1197-2020; SCOPUS Author ID: 14024556800; ORCID: <https://orcid.org/0000-0002-8023-5189>

Ю. С. ХОМЕНКО, аспірант, Харківський національний автомобільно-дорожній університет, м. Харків, Україна; e-mail: khomenko.yura40@gmail.com; Researcher ID: GSD-3718-2022; ORCID: <https://orcid.org/0000-0002-9242-1202>;

МЕТОД ВИЯВЛЕННЯ НАЗЕМНИХ ОРІЄНТИРІВ ДЛЯ НАВІГАЦІЇ АВТОНОМНИХ МОБІЛЬНИХ РОБОТІВ З ВИКОРИСТАННЯМ ОСОБЛИВОСТЕЙ РОЗПОДІЛУ СЕРЕДНЬОЇ ІНТЕНСИВНОСТІ КОЛЬОРІВ

Застосування відеокамер при навігації автономних мобільних роботів є один з можливих шляхів реалізації пасивних дистанційних методів виявлення наземних орієнтирів. Було запропоновано метод виявлення наземних орієнтирів при навігації автономних мобільних роботів, що ґрунтується на використанні особливостей розподілу середньої інтенсивності кольорів в стовпцях матриці відеокамери автономного мобільного робота. Головна особливість розподілу проявляється в тому, що при появі в зоні огляду відеокамери стовпоподібного об'єкту як можливого орієнтиру в ньому з'являється стрибок або провал, амплітуда яких може служити в якості критерію виявлення орієнтиру. В роботі показано, що цю операцію можна ефективно виконати на основі аналізу матриці зображення, якщо колір орієнтиру істотно відрізняється від кольору фонового зображення. В інших випадках запропоновано використовувати усереднення інтенсивності червоного, зеленого та синього кольорів по стовпцях матриці відеокамери. Вказаний метод для підвищення ймовірності виявлення орієнтиру в широких умовах застосування відеокамери автономного мобільного робота запропоновано в якості критерію виявлення застосовувати добуток модуля похідної від розподілу середніх кольорів в стовпцях матриці на модуль різниці зазначеного розподілу і його середнього значення по всім стовпцям. Було встановлено, що добуток модуля зазначеної похідної на модуль різниці між розподілом середніх кольорів і середнім значенням цього розподілу, який названо визначальним добутком, може служити в якості критерію для виявлення наземного орієнтиру. Показано, що перевищення максимального значення визначального добутку над пороговим значенням, яке визначають на основі аналізу статистичних даних, в будь-якому з каналів червоного, зеленого і синього кольорів свідчить про виявлення наземного орієнтиру. Дані дослідження показують, що визначальний добуток за своїм впливом на ймовірнісні характеристики виявлення схожий з відношенням сигнал-шум в радіолокації.

Ключові слова: автономний мобільний робот, наземний орієнтир, кольоровість, стрибки параметрів, розподіл інтенсивності кольору.

Вступ. При роботі по ліквідації наслідків катастроф і аварій на місцевостях, коли існує небезпека для життя і здоров'я людей, підвищується корисна роль автономних мобільних роботів (АМР), які можуть виконати частину робіт на зазначених місцевостях. Навігація роботів на місцевості здійснюється, як правило, з допомогою GPS. Однак, ефективність цієї навігаційної системи може значно знижуватися внаслідок перекриття області, істотної для поширення електромагнітних хвиль, зруйнованими елементами будівель, рельєфом місцевості тощо. Крім того, можлива постановка завдань приймальним елементам GPS, що розміщені на АМР. Отже, нагальною необхідністю стає використання автономних мобільних роботів, які здатні здійснювати навігацію без використання GPS. Автономна навігація роботів без GPS можлива по цифровій карті конкретної місцевості, яка, однак, при значних руйнуваннях об'єктів на місцевості може виявитися малоприматною. Оскільки ситуація на місцевості може піддаватися змінам внаслідок пожеж, вибухів та інших факторів, створення нової карти місцевості в умовах дефіциту

часу є недоцільним. Таким чином, АМР самостійно повинен знаходити нерухомі об'єкти на місцевості, до яких він може «прив'язуватися».

Постановка проблеми. «Прив'язка» АМР повинна здійснюватися до об'єктів, які будемо називати наземними орієнтирами (НО). Як правило, наземний орієнтир повинен бути високим об'єктом з малими лінійними розмірами у горизонтальній площині. Він забезпечує при належних умовах високу кутову точність «прив'язки» робота в горизонтальній площині. Як правило, потрібно мати декілька НО для створення АМР своєї системи координат на місцевості, де робот виконує поставлене завдання. Дальність до НО може визначатися активними засобами випромінювання. При використанні роботом активних засобів (радар, ультразвуковий далекомір, лазерна система тощо) наземний орієнтир екранується ненавмисними завадами, що обумовлені відбиттям хвиль від об'єктів, що оточують НО [1]. Отже, при наявності фонового відбиття від елементів місцевості доцільно використовувати для автономної навігації АМР невідпромінюючі системи,

© Полярус О. В., Хоменко Ю. С., 2024



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.



до яких належать, зокрема, відеокамери та камери нічного бачення. В статті аналізується ефективність використання в АМР відеокамер для світлого часу доби при наявності складних фонових зображень. Якщо ж за орієнтиром є фонові зображення, що близькі за кольором з НО, то завдання виявлення відеокамерою наземного орієнтиру не завжди може бути вирішено успішно. Отже, на цей час проблема виявлення і розпізнавання наземного орієнтиру відеокамерами робота при наявності щільних фонових зображень залишається вирішеною тільки частково і тому потребує розв'язання.

Аналіз останніх досліджень і публікацій. Багато методів обробки кольору зображень [2] є ефективними в прикладних задачах, коли проведена попередня обробка зображення при належному виборі характеристик відео камери [3]. Для розпізнавання НО часто використовуються шаблони [4], а належність реальних зображень до шаблонів оцінюється з допомогою нечітких множин. Для підвищення показників розпізнавання застосовуються декілька відеокамер в одному полі огляду [5]. Істотно може покращити характеристики виявлення НО апріорна інформація про орієнтир [6]. Всі сучасні алгоритми оброблення зображень працюють, як правило, в реальному часі, причому навіть при обробці зображень від рухомих об'єктів [7]. У роботі [8] синтезована система диференціальних рівнянь для виявлення, розпізнавання орієнтирів і для оптимального визначення цих параметрів та їх дисперсій. Для задач виявлення і розпізнавання НО починають використовуватися інтелектуальні системи на основі нейронних мереж [9] та глибокого машинного навчання в хмарах точок [10]. У [11] запропоновано два методи виявлення стовпоподібних НО. Один з них ґрунтується на накопиченні зсунутих в часі сигналів кольоровості при скануванні відеокамерою в азимутальній площині, а другий використовує моди Гільберта – Хуанга, які містять інформацію про стрибки кольоровості.

Виділення недосліджених частин загальної проблеми. Для розпізнавання об'єктів на матриці зображення в [11] використовується двовірсна картина розподілу інтенсивності кольорів, в якій здійснюється усереднення інтенсивностей (або координат) кольорів по стовпцям матриці. Це означає, що здійснюється відображення двовірсного розподілу кольорів матриці зображення в одномірний розподіл складових червоного, зеленого та синього кольорів $R(x), G(x), B(x)$, тобто залежностей цих параметрів від горизонтальної координати x матриці відеокамери. Прийнятний наземний орієнтир, як правило, є витягнутим у вертикальній площині і має приблизно однаковий колір, інтенсивність якого майже співпадає з середнім по тим стовпцям, що охоплюють цей орієнтир. Усереднені значення кольорів R, G, B в інших стовпцях кадру з неоднорідним фоном у більшості випадків відрізняються від параметрів R, G, B орієнтиру. Ці усереднені значення інтенсивності або координат кольорів будемо називати параметрами кольоровості зображення.

При скануванні навколишнього простору відеокамерою в зону огляду, де є наземний орієнтир, для кожного моменту часу НО буде розміщений в різних частинах кадру [11]. При введенні відповідної затримки в розподіл кольоровості, який є функцією азимутальної координати та часу сканування, стрибки або провали кольоровості, що обумовлені наявністю орієнтира, будуть складатися в фазі, а в інших напрямках, де немає НО, – довільно. Така фізична картина спостерігається, якщо радіус кореляції флуктуацій кольоровості фонових зображень $R(x), G(x), B(x)$ є невеликим [11]. В інших випадках накопичення стрибків чи провалів кольоровості також відбувається, але істотного перевищення стрибка чи провалу кольоровості над фоновією кольоровістю не відбувається. Складання параметрів кольоровості в кутових напрямках, де орієнтир відсутній є хаотичним і тому дає деяку усереднену амплітуду параметру кольоровості. Сума залежностей $R(x)$ для різних кутових напрямків є функцією радіусу кореляції флуктуацій параметрів кольоровості. Звідси випливає необхідність розробки методу виявлення НО, що не залежить від зазначеного радіусу кореляції.

Метою статті є розробка методу виявлення наземних орієнтирів відеокамерою автономного мобільного робота, що слабо залежить від радіусу кореляції флуктуацій кольоровості в широкому діапазоні умов фонових освітлення та кольорів, впливу заважаючих факторів тощо.

Виклад основного матеріалу. Аналіз можливостей з виявлення наземного орієнтиру проведемо з використанням середнього значення кольорів в стовпцях матриці зображення, тобто кольоровості. Для зображення, що приведене на рис. 1, розподіл складових кольоровості R, G, B вздовж горизонтальної координати матриці, показано на рис. 2.



Рис. 1. Приклад зображення

Як впливає з рис. 2, складові кольору R, G, B мають близький за характером розподіл і тому далі будемо в багатьох випадках обмежуватися тільки розподілом одного з кольорів, найчастіше, червоного. В усіх трьох кривих є один великий провал розподілу і декілька менших. Порівняння рис. 2 з рис. 1 свідчить, що ці провали обумовлені наявністю окремих дерев, які в якості орієнтирів є не дуже зручними, оскільки не зосереджені в горизонтальній площині. Для створення моделей вузьких в цій площині НО розроблений код програми встановлення моделей стовпоподібних орієн-

тирів будь-який розмірів на будь-яке зображення. Щоб встановити фізичні закономірності процесів виявлення НО, моделі вибиралися різних розмірів і кольорів.



Рис. 2. Розподіл середніх червоного, зеленого та синього кольорів вздовж горизонтальної координати матриці відеокамери

Приклад орієнтури червоного кольору, що накладений на зображення (рис. 1), приведено на рис. 3.



Рис. 3. Зображення з накладеною моделлю наземного орієнтури

Розподіл кольоровості для складової червоного кольору показано на рис. 4.



Рис. 4. Розподіл складової червоного кольору вздовж горизонтальної координати матриці відеокамери

Порівняно з провалами кольоровості від дерев провал від моделі стовпа не спостерігається, а навпаки, замість провалу реєструється великий стрибок. Слід зазначити, що передбачити стрибок чи провал від НО важко, але це не потрібно, оскільки критерієм виявлення може бути як стрибок, так і провал, причому в розподілі трьох кольорів R, G, B деякі можуть мати стрибок, а інші кольори – провали, чи навпаки. Стрибок кольоровості на рис. 4, по-перше, майже повторює форму моделі стовпа, а, по-друге, він має велику амплітуду порівняно з кольоровістю фону. Для складових зеленого та синього кольорів спостерігаються аналогічні стрибки, хоча їх відносна амплітуда менша, ніж для червоного кольору. Причина цього зрозуміла: модель стовпа є червоною. Звідси випливає висновок про можливість використання стрибків або провалів кольоровості для виявлення НО, однак, на практиці перевищення кольоровості в місці розміщення НО над кольоровістю фону не завжди є великим. Ось чому потрібні інші підходи, які дозволяють досягти прийнятної ймовірності виявлення НО в умовах, коли стрибки або провали кольоровості мало відрізняються від фонові кольоровості.

Логічно припустити, що похідні в тих точках, в яких відбуваються різкі зміни кольоровості, мають бути великими, що підтверджує рис. 5.

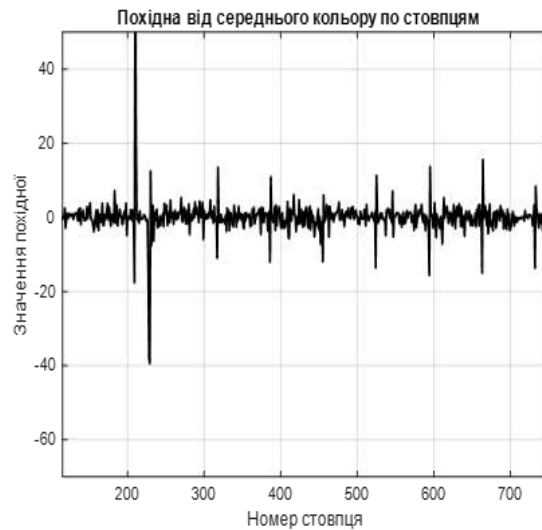


Рис. 5. Похідна від розподілу складової червоного кольору, яка показана на рис. 4

Як випливає з рис. 5, у місці розміщення НО похідна є великою для фронту і спаду розподілу кольоровості, але має різні знаки. Ця ознака може використовуватися для реєстрації початку і кінця стрибка чи провалу кольоровості. Похідні для кольоровості, що описує фонове зображення, є меншими, ніж в місці розміщення орієнтури. Це обумовлено усередненням кольорів в стовпцях, що привело до згладжування флуктуацій кольорів. В даному прикладі похідна від стрибка кольоровості від НО є великою, а в окремих ситуаціях вона може бути набагато меншою. Отже, логічно використати для виявлення НО як амплітуди стрибків (провалів) кольоровості, так і значення похідних від розподілу кольоровості. У

зв'язку з цим в якості критерію виявлення орієнтиру доцільно використати добуток модуля похідної від розподілу кольоровості на модуль різниці розподілу кольоровості і середнього значення цієї кольоровості по всім стовпцям. Для скорочення запису будемо називати цей добуток визначальним, оскільки при порівнянні його максимального значення з пороговим значенням визначається наявність НО в зоні огляду відеокамери. Для розглянутого прикладу залежність визначального добутку від номеру стовпця матриці зображення показано на рис. 6.

Тут створені умови для явного виявлення і виділення наземного орієнтиру на фоні зображення (рис. 1).



Рис. 6. Залежність визначального добутку від номеру стовпця матриці відеокамери

Якщо колір стовпа змінити на зелений, то визначальний добуток зміниться лише незначно (рис. 7), проте умови фіксації стрибка як по фронту, так і по спаду, помітно покращуються, що сприяє більшій стабільності системи.

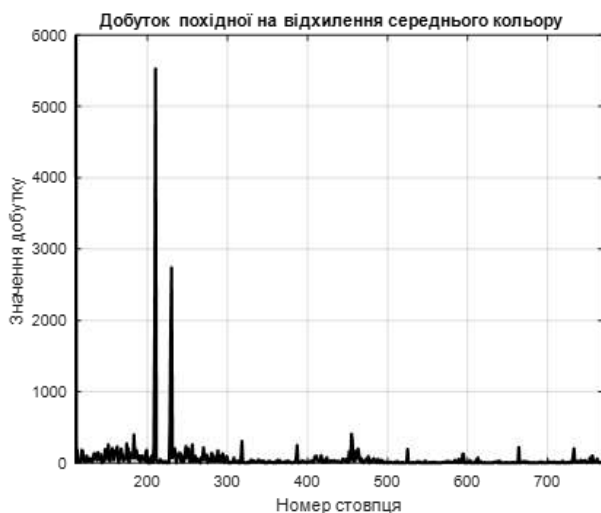


Рис. 7. Залежність визначального добутку від номеру стовпця матриці зображення для моделі стовпа зеленого кольору

Розглянемо тепер реальний стовп замість модельного (рис. 8).



Рис. 8. Зображення реального стовпа на фоні гірської місцевості

Розподіл визначального добутку для цього зображення має вигляд (рис. 9).

З рис. 9 випливають можливості виявлення наземного орієнтиру за критерієм перевищення максимального значення визначального добутку над пороговим значенням.

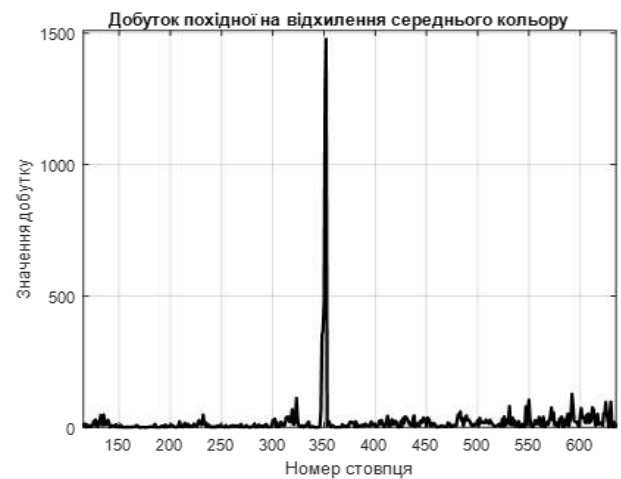


Рис. 9. Розподіл визначального добутку вздовж горизонтальної координати матриці відеокамери для зображення (рис. 8)

Нарешті, розглянемо можливість виявлення багатьох дерев, які можуть слугувати надійними наземними орієнтирами завдяки своїй природній розташованості та візуальній помітності у різних умовах (рис. 10).



Рис. 10. Зображення дерев як можливих наземних орієнтирів

Розподіл кольоровості для цього зображення характеризується стрибками для зеленого і синього кольорів та провалами для червоного кольору (рис. 11).

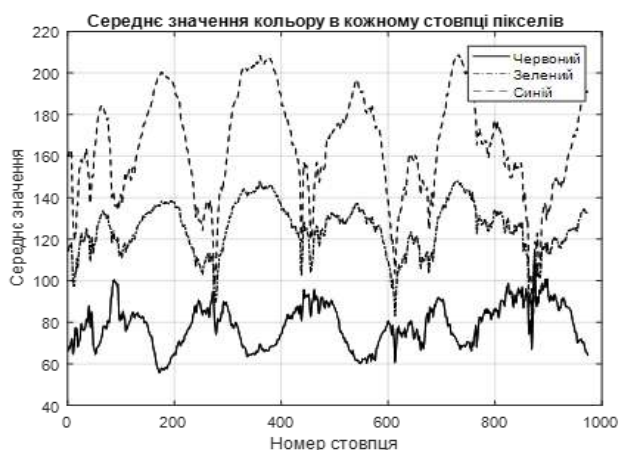


Рис. 11. Розподіл середніх по стовпцях складових червоного, зеленого і синього кольорів вздовж горизонтальної координати матриці відеокамери

Визначальний добуток для зображення (рис. 10) має вигляд, наведений на рис. 12. Він характеризується піковим розподілом, а кількість локальних максимумів розподілу приблизно співпадає з кількістю дерев.

Оскільки один з піків набагато перевищує інші (глобальний максимум), то дерево, що йому відповідає, можна прийняти в якості наземного орієнтиру, якщо амплітуда цього максимуму перевищує поріг виявлення. Зображення з багатьма орієнтирами показано на рис.13.

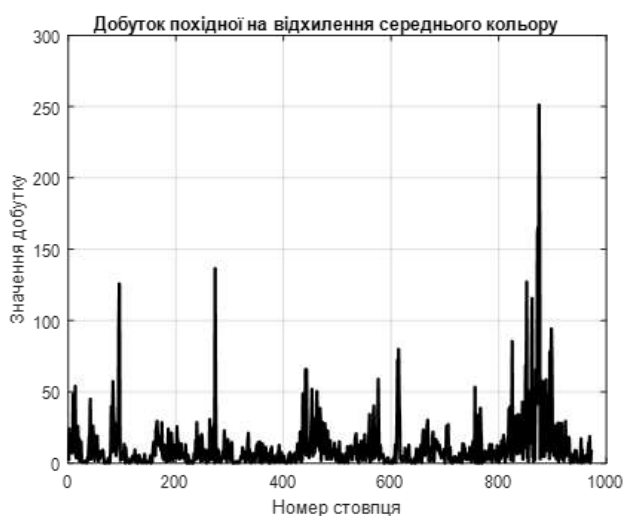


Рис. 12. Розподіл визначального добутку вздовж горизонтальної координати матриці відеокамери для зображення (рис. 10)



Рис. 13. Зображення з багатьма можливими наземними орієнтирами

Для зображення (рис. 13) кількість можливих наземних орієнтирів стає великою і визначальний добуток (рис. 14) характеризується збільшеною кількістю піків, що в свою чергу породжує складність у виборі необхідного наземного орієнтиру.

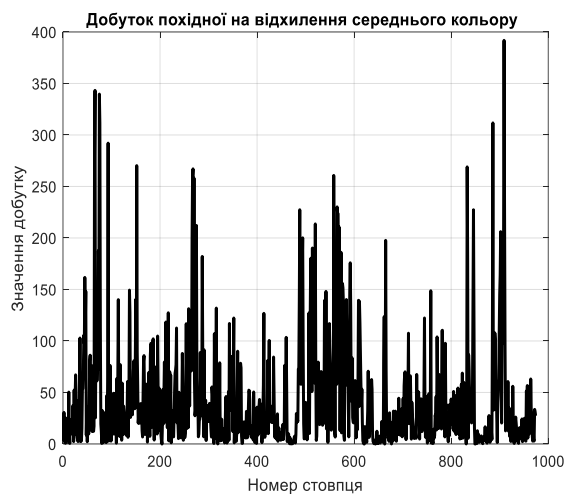


Рис. 14. Розподіл визначального добутку вздовж горизонтальної координати матриці відеокамери для зображення (рис. 13)

Зрозуміло, що для підвищення надійності запропонованого методу можуть застосовуватися додатково інші засоби виявлення, в тому числі інтелектуальні.

Обґрунтування порогу виявлення являє собою окреме завдання, яке використовує статистичні дані щодо визначального добутку. На цей час таких даних не існує. Їх отримання не є складним завданням, однак воно вимагає витрат часу на збір та обробку результатів. Статистичні дані можуть також використовуватися для машинного навчання з використанням штучних нейронних мереж, що підвищить в майбутньому ефективність рішень щодо виявлення та розпізнавання наземних орієнтирів для АР.

З проведених досліджень випливає, що визначальний добуток за своїм впливом на ймовірнісні характеристики виявлення орієнтиру схожий з відношенням сигнал-шум в радіолокації. У [11, 12] аналогічним показником є відношення значення кольоровості в області стрибка чи провалу до фонового значення і розраховані ймовірності виявлення НО при фіксованому значенні хибної тривоги. В даній статті розрахунки є аналогічними з єдиною різницею, що аналогом відношення сигнал-фон використовується визначальний добуток. Він використовує добуток модуля похідної на параметр, що схожий з енергетичним показником, тобто модуль різниці між розподілом середніх кольорів і середнім значенням цього розподілу по всім стовпцям. Отже, результати розрахунків ймовірності виявлення орієнтирів можна вважати такими ж, як в [11, 12], однак за рахунок використання визначального добутку вдається розширити область застосування методу, зокрема в умовах, коли радіус кореляції кольорів фону в пікселях є сумірним з горизонтальними розмірами матриці відеокамери. Крім того, запропонований метод є простішим

для технічної реалізації, ніж метод, що приведений в [11, 12]. Отже, метод не залежить від радіусу кореляції флуктуацій кольоровості і є універсальним в широкому діапазоні умов застосування АМР.

Висновки. Застосування відеокамер при навігації автономних мобільних роботів є один з можливих шляхів реалізації пасивних дистанційних методів виявлення наземних орієнтирів. В статті показано, що цю операцію можна ефективно виконати на основі аналізу матриці зображення, якщо колір орієнтиру істотно відрізняється від кольору фонового зображення. Для інших випадків запропоновано використовувати усереднення інтенсивності червоного, зеленого та синього кольорів по стовпцям матриці відеокамери. Встановлено, що чим ближче об'єкт схожий з циліндром або стовпом за формою, тим краще він виділяється на розподілі середніх кольорів вздовж горизонтальної координати матриці зображення. При цьому в місці розміщення орієнтиру на розподілах середніх кольорів виникає стрибок або провал і кожний з них може використовуватися в якості ознаки для виявлення орієнтиру. Для фіксації положення фронту та спаду стрибка або провалу рекомендовано застосовувати похідну від розподілу середніх кольорів. Встановлено, що добуток модуля зазначеної похідної на модуль різниці між розподілом середніх кольорів і середнім значенням цього розподілу, який названо визначальним добутком, може служити в якості критерію для виявлення наземного орієнтиру. Показано, що перевищення максимального значення визначального добутку над пороговим значенням, яке визначають на основі аналізу статистичних даних, в будь-якому з каналів червоного, зеленого і синього кольорів свідчить про виявлення наземного орієнтиру. Проведені дослідження показують, що визначальний добуток за своїм впливом на ймовірнісні характеристики виявлення схожий з відношенням сигнал-шум в радіо-локації.

Список використаних джерел

1. Poliarus O., Poliakov Y. The Methods of Radar Detection of Landmarks by Mobile Autonomous Robots. *Machine Vision and Navigation*. 2020. P. 171–196. DOI: doi.org/10.1007/978-3-030-22587-2_6.
2. Yang J., Liu L., Li S. Z. Separating specular and diffuse reflection components in the HSI color space. *IEEE International Conference on Computer Vision Workshops*. 2013. P. 891–898. DOI: doi.org/10.1007/978-3-030-22587-2_6.
3. Anuncia S. M., Joseph J. G. Approaches for Automated Object Recognition and Extraction from Images a Study. *Journal of Computing and Information Technology*. 2009. Vol. 17, no. 4. P. 359–370. DOI: doi.org/10.2498/cit.1001363.
4. Fukuda T., Ito S., Arai F., Yokoyama Y., Abe Y., Tanaka K., Tanaka Y. Navigation system based on ceiling landmark recognition for autonomous mobile robot-landmark detection based on fuzzy template matching (FTM). *1995 IEEE/RSJ International Conference on Intelligent Robots and Systems. Human Robot Interaction and Cooperative Robots*. 1995. Vol. 2. P. 150–155. DOI: doi.org/10.1109/iros.1995.526153.
5. Novak G., Bais A., Mahlknecht S. Simple stereo vision system for real-time object recognition for an autonomous mobile robot. *Second IEEE International Conference on Computational Cybernetics*. 2004. P. 213–216. DOI: doi.org/10.1109/ICCCYB.2004.1437710.
6. Zhengt J. Y., Barth M., Tsuji S. Autonomous Landmark Selection for Route Recognition by A Mobile Robot. *Proceedings of the 1991 IEEE International Conference on Robotics and Automation*. 1991, vol. 3, pp. 2004–2009. DOI: doi.org/10.1109/robot.1991.131922.
7. Abdullah Q., Shah N., Mohamad M., Ali M., Farah N., Salh A., Aboali M., Mohamad M., Saif A. Real-time Autonomous Robot for Object Tracking using Vision System. *Solid State Technology*. 2020. Vol. 63, pp. 16278–16294. DOI: doi.org/10.48550/arXiv.2105.00852.
8. Poliarus O. V., Poliakov Y. O. Detection of Landmarks by Mobile Autonomous Robots Based on Estimating the Color Parameters of the Surrounding Area. *Examining Optoelectronics in Machine Vision and Applications in Industry 4.0*. 2021. P. 224–257. DOI: doi.org/10.4018/978-1-7998-6522-3.ch008.
9. Thin K., Zaw N. Object Detection System of an Autonomous Mobile Robot by using Artificial Neural Network. *International Journal of Scientific and Research Publications*. 2019, vol. 9, pp. 374–380. DOI: doi.org/10.29322/ijrsp.9.03.2019.p8758.
10. Zhou Z., Li L., Fürsterling A., Durocher H., Mouridsen J., Zhang X. Learning-based object detection and localization for a mobile robot manipulator in SME production. *Robotics and Computer-Integrated Manufacturing*. 2022. Vol. 73, no. 102229. P. 1–12. DOI: doi.org/10.1016/j.rcim.2021.102229.
11. Poliarus O., Poliakov Ye., Lebedynskiy A. Detection of landmarks by autonomous mobile robots using camera-based sensors in outdoor environments. *IEEE Sensors Journal*. 2021. Vol. 21. P. 11443–11450. DOI: doi.org/10.1109/JSEN.2020.3010883.
12. Kriegl H., Kroger P., Schubert E., Zimek A. Outlier Detection in Arbitrarily Oriented Subspaces. *IEEE 12th International Conference on Data Mining*. 2012. P. 379–388. DOI: https://doi.org/10.1109/icdm.2012.21.

References (transliterated)

1. Poliarus O., Poliakov Y. The Methods of Radar Detection of Landmarks by Mobile Autonomous Robots. *Machine Vision and Navigation*. 2020, pp. 171–196. DOI: doi.org/10.1007/978-3-030-22587-2_6.
2. Yang J., Liu L., Li S. Z. Separating specular and diffuse reflection components in the HSI color space. *IEEE International Conference on Computer Vision Workshops*. 2013, pp. 891–898. DOI: doi.org/10.1007/978-3-030-22587-2_6.
3. Anuncia S. M., Joseph J. G. Approaches for Automated Object Recognition and Extraction from Images a Study. *Journal of Computing and Information Technology*. 2009, vol. 17, no. 4, pp. 359–370. DOI: doi.org/10.2498/cit.1001363.
4. Fukuda T., Ito S., Arai F., Yokoyama Y., Abe Y., Tanaka K., Tanaka Y. Navigation system based on ceiling landmark recognition for autonomous mobile robot-landmark detection based on fuzzy template matching (FTM). *1995 IEEE/RSJ International Conference on Intelligent Robots and Systems. Human Robot Interaction and Cooperative Robots*. 1995, vol. 2, pp. 150–155. DOI: doi.org/10.1109/iros.1995.526153.
5. Novak G., Bais A., Mahlknecht S. Simple stereo vision system for real-time object recognition for an autonomous mobile robot. *Second IEEE International Conference on Computational Cybernetics*. 2004, pp. 213–216. DOI: doi.org/10.1109/ICCCYB.2004.1437710.
6. Zhengt J. Y., Barth M., Tsuji S. Autonomous Landmark Selection for Route Recognition by A Mobile Robot. *Proceedings of the 1991 IEEE International Conference on Robotics and Automation*. 1991, vol. 3, pp. 2004–2009. DOI: doi.org/10.1109/robot.1991.131922.
7. Abdullah Q., Shah N., Mohamad M., Ali M., Farah N., Salh A., Aboali M., Mohamad M., Saif A. Real-time Autonomous Robot for Object Tracking using Vision System. *Solid State Technology*. 2020, vol. 63, pp. 16278–16294. DOI: doi.org/10.48550/arXiv.2105.00852.
8. Poliarus O. V., Poliakov Y. O. Detection of Landmarks by Mobile Autonomous Robots Based on Estimating the Color Parameters of the Surrounding Area. *Examining Optoelectronics in Machine Vision and Applications in Industry 4.0*. 2021, pp. 224–257. DOI: doi.org/10.4018/978-1-7998-6522-3.ch008.
9. Thin K., Zaw N. Object Detection System of an Autonomous Mobile Robot by using Artificial Neural Network. *International Journal of Scientific and Research Publications*. 2019, vol. 9, pp. 374–380. DOI: doi.org/10.29322/ijrsp.9.03.2019.p8758.
10. Zhou Z., Li L., Fürsterling A., Durocher H., Mouridsen J., Zhang X. Learning-based object detection and localization for a mobile robot manipulator in SME production. *Robotics and Computer-Integrated Manufacturing*. 2022, vol. 73, no. 102229, pp. 1–12. DOI: doi.org/10.1016/j.rcim.2021.102229.

- Manufacturing*, 2022, vol. 73, no. 102229, pp. 1–12. DOI: doi.org/10.1016/j.rcim.2021.102229.
11. Poliarus O., Poliakov Ye., Lebedynskiy A. Detection of landmarks by autonomous mobile robots using camera-based sensors in outdoor environments. *IEEE Sensors Journal*. 2021, vol. 21, pp. 11443–11450. DOI: doi.org/10.1109/JSEN.2020.3010883.
12. Kriegel H., Kroger P., Schubert E., Zimek A. Outlier Detection in Arbitrarily Oriented Subspaces. *IEEE 12th International Conference on Data Mining*. 2012, pp. 379–388. DOI: https://doi.org/10.1109/icdm.2012.21.

Надійшла (received) 17.07.2024

UDC 004.93

O. V. POLIARUS, doctor of engineering science, professor, Kharkiv national automobile and highway university, Kharkiv, Ukraine; e-mail: poliarus.kharkov@ukr.net; Researcher ID: AAH-1197-2020; SCOPUS Author ID: 14024556800; ORCID: https://orcid.org/0000-0002-8023-5189

Y. S. KHOMENKO, graduate student, Kharkiv national automobile and highway university; Kharkiv, Ukraine; email: khomenko.yura40@gmail.com; Researcher ID: GSD-3718-2022; ORCID: https://orcid.org/0000-0002-9242-1202

METHOD OF DETECTING LANDMARKS FOR NAVIGATION OF AUTONOMOUS MOBILE ROBOTS USING FEATURES OF AVERAGE COLOR INTENSITY DISTRIBUTION

The use of video cameras in the navigation of autonomous mobile robots is one of the possible ways of implementing passive remote methods of detecting ground landmarks. A method for detecting ground landmarks during the navigation of autonomous mobile robots was proposed, which is based on the features of the distribution of average color intensity in the columns of the video camera matrix of the autonomous mobile robot. The main feature of the distribution is manifested in the fact that when a pillar-like object appears in the field of view of the video camera as a possible landmark, a jump or dip appears in it, the amplitude of which can serve as a criterion for landmark detection. The work shows that this operation can be effectively performed on the basis of the image matrix analysis, if the color of the landmark is significantly different from the color of the background image. In other cases, it is proposed to use the averaging of the intensity of red, green, and blue colors along the columns of the video camera matrix. The specified method to increase the probability of landmark detection in the broad conditions of application of a video camera of an autonomous mobile robot is proposed to use as a detection criterion the product of the modulus of the derivative of the distribution of average colors in the columns of the matrix by the modulus of the difference of the specified distribution and its average value across all columns. It was established that the product of the module of the specified derivative by the module of the difference between the distribution of average colors and the average value of this distribution, which is called the determining product, can serve as a criterion for identifying a landmark. It is shown that exceeding the maximum value of the determining product above the threshold value, which is determined based on the analysis of statistical data, in any of the channels of red, green, and blue colors indicates the detection of a ground landmark. Research data show that the determining product in its influence on the probabilistic characteristics of detection is similar to the signal-to-noise ratio in radar.

Keywords: autonomous mobile robot, landmark, parameter jumps, color intensity distribution, color intensity distribution.

Повні імена авторів / Author's full names

Автор 1 / Author 1: Полярус Олександр Васильович / Poliarus Oleksandr Vasyliovych

Автор 2 / Author 2: Хоменко Юрій Сергійович / Khomenko Yurii Serhiiovych

O. S. KUTSENKO, Doctor of Technical Sciences, Professor, National Technical University "Kharkiv Polytechnic Institute", professor of the Department of System Analysis and Information-Analytical Technologies, Kharkiv, Ukraine;

e-mail: Oleksandr.Kutsenko@khp.edu.ua; ORCID: <https://orcid.org/0000-0001-6059-3694>;

M. E. ALFOROV, PhD Student of National Technical University "Kharkiv Polytechnic Institute", Kharkiv, Ukraine;

e-mail: Mykhailo.Alforov@cs.khpi.edu.ua; ORCID: <https://orcid.org/0009-0005-4920-5626>

A. E. ALFOROV, PhD Student of National Technical University "Kharkiv Polytechnic Institute", Kharkiv, Ukraine;

e-mail: Andrii.Alforov@cs.khpi.edu.ua; ORCID: <https://orcid.org/0009-0007-9506-7468>

SYNTHESIS OF DESIGN PARAMETERS OF MULTI-PURPOSE DYNAMIC SYSTEMS

Two problems related to the optimization of linear stationary dynamic systems are considered. A general formulation of the multi-purpose problem of optimal control with the choice of design parameters is given. As a special case, the problem of multi-objective optimization of a linear system according to an integral quadratic criterion with a given random distribution of initial deviations is considered. The solution is based on the method of simultaneously reducing two positive-definite quadratic forms to diagonal form. Analytical results have been obtained that make it possible to calculate the mathematical expectation of the criterion under the normal multidimensional distribution law of the vector of random initial perturbations. The inverse problem of stability theory is formulated: to find a vector of structural parameters that ensure the stability of the system and a given average value of the quadratic integral quality criterion on a set of initial perturbations. The solution of the problem is proposed to be carried out in two stages. The first stage involves deriving a general solution to the Lyapunov matrix equation in terms of the elements of the system matrix. To achieve this, the state space is mapped onto the eigen-subspace of the positive-definite matrix corresponding to the integral quadratic performance criterion. It has been established that this solution is determined by an arbitrary skew-symmetric matrix or by the corresponding set of arbitrary constants. In contrast, when the system matrix depends linearly on the vector of design parameters, a linear system of equations can be formulated with respect to the unknown parameters and arbitrary constants present in the general solution of the inverse stability problem. In general, such a system is consistent and admits an infinite number of solutions that satisfy the initial requirements for the elements of the symmetric matrices in the Lyapunov.

Keywords: stability, integral quadratic functional, Lyapunov matrix equation, inverse stability problem, multi-purpose dynamic systems, linear stationary systems, parametric optimization.

1. Introduction. The problem of optimizing the parameters of stable dynamic systems based on integral quadratic functional (IQF) of transient processes has been studied and solved in various theoretical and applied contexts. Fundamental results in this area were obtained in [1, 2, 3]. These results have been regarded as classical for a long time.

They differ somewhat in the formulation of the problem of optimal control for dynamic systems combined with the selection of design parameters. The mathematical formulation of the optimal control problem with parameters based on the maximum principle is addressed in [4, 5]. The issue of parametric optimization of the IQF with a known statistical distribution of initial perturbations is discussed in [3]. This work also formulates and proposes a solution to the problem of selecting the elements of the matrix of a linear stationary dynamic system that optimize the mathematical expectation of the IQF. The general solution to this problem is expressed in terms of matrix algebraic equations involving several parameters and a set of auxiliary variables. Analyzing the entire spectrum of research in the field of theory and practice of ensuring the stability of dynamic systems, the field can be divided into two primary classes: problems of analysis and problems of synthesis of stable systems, corresponding to the direct and inverse problems of stability theory. Direct problems focus on determining whether a system with specified parameters is stable. In contrast, inverse problems aim to identify a vector of parameters that ensures the stability of the system to be constructed.

Among the methods for solving inverse problems of stability theory, two fundamental directions can be distinguished. The first direction is related to the control of the distributed roots of the characteristic equation of the system of differential equations. These methods are indirectly related to the control of the coefficients of the characteristic equation by selecting the appropriate design parameters [6]. The theory of modal control is closely related to this direction [7]. Information on the distribution of roots also makes it possible to find indirect characteristics of transient processes, such as the degree of stability and the degree of oscillation of the synthesized system but does not make it possible to calculate direct indicators: the time of the transient process and the degree of oscillation.

The second direction gained significant momentum in the second half of the last century and is associated with the works [8-10], particularly in the context of the Lyapunov matrix equation, as well as related studies [11,12] on IQF. The direct stability problem, based on Lyapunov equations, involves evaluating the solutions for given parameters of the dynamic system matrix. The inverse problem can be formulated as the task of finding the parameters of the dynamic system matrix that satisfy the Lyapunov matrix equation.

2. Formulation of a parametric multi-criteria optimal control problem. Among the variety of formulations of optimal control problems for continuous dynamic systems, there exists an important problem that has been formulated and solved in general form. Let us briefly outline the mathe-

© Kutsenko O.S., Alforov M. E., Alforov A.E., 2024



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.



mathematical formulation. A stationary dynamic system of the following form is given:

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \mathbf{u}, \boldsymbol{\alpha}), \quad (1)$$

where: $\mathbf{x} \in X$ – is the state vector of the system, $\dim X = n$; $\mathbf{u} \in U$ – vector of control parameters, $\dim U = m$; $\boldsymbol{\alpha} \in A$ – vector of design parameters, $\dim A = s$. Vector function \mathbf{f} and its partial derivatives $\frac{\partial \mathbf{f}}{\partial \mathbf{x}'}, \frac{\partial \mathbf{f}}{\partial \boldsymbol{\alpha}}$ are assumed to be continuous at $\mathbf{u} \in U$ and any $\mathbf{x}, \boldsymbol{\alpha}$. The control objective is defined by a target vector

$$\mathbf{c} = (\mathbf{x}_0, \mathbf{x}_1, \tau), \quad (2)$$

where $\mathbf{x}_0 \in X$ – is the vector of the initial state, $\mathbf{x}_1 \in X$ – the vector of the final state, τ – is the time of transition of the system from \mathbf{x}_0 to \mathbf{x}_1 . The quality criterion of the controlled process is expressed as

$$J(\mathbf{c}, \mathbf{u}, \boldsymbol{\alpha}) = \int_0^\tau f_0(\mathbf{x}, \mathbf{u}, \boldsymbol{\alpha}) dt. \quad (3)$$

The problem of optimal control of a system (1) with variable parameters is formulated as follows: for a given system objective, find the constant vector $\boldsymbol{\alpha}$ and the function $\mathbf{u}(t)$ that minimizes the quality functional (3). In works [4, 5], necessary conditions for solving the formulated problem were obtained based on the maximum principle. It follows from these conditions that the optimal functions $\boldsymbol{\alpha}$ and $\mathbf{u}(t)$ are functions of the vector \mathbf{c} , and the explicit form of these functions in general is unknown.

Let the system being constructed be multi-criteria, i.e., a subset is given:

$$C \subseteq X \times X \times T, \quad T \in (0, \infty) \quad (4)$$

defines the set of objectives of the form (2) that can be a goal for the control system and let a probability measure $p(\mathbf{c})$ be defined on the subset C that defines the probability of achieving the given objective. It would be natural to pose the following parametric optimization problem for the multi-objective system (1) and a set of objectives (4), find a parameter $\boldsymbol{\alpha}$ vector that minimizes criterion (3) at each point of the set C . In general, such a formulation of the problem is not correct, since the vector of parameters that is optimal for one of the goals of the system will not be optimal for another. In this regard, we will optimize a certain average criterion for all permissible goals of the system. Let's introduce the function:

$$\Psi(\mathbf{c}, \boldsymbol{\alpha}) = \min_{\mathbf{u} \in U} J(\mathbf{c}, \mathbf{u}, \boldsymbol{\alpha}) \quad (5)$$

which is the minimum value of criterion (3) with a fixed vector of constructive parameters $\boldsymbol{\alpha}$ and a fixed target vector \mathbf{c} . Let's find the mathematical expectation of function (5):

$$I(\boldsymbol{\alpha}) = \int_C p(\mathbf{c}) \Psi(\mathbf{c}, \boldsymbol{\alpha}) dC, \quad (6)$$

determining the average quality of the system with a fixed parameter vector.

Now the formulation of the parametric optimization problem can be proposed as follows: for the system (1), the set of goals (4) and criterion (3), choose a vector of design parameters $\boldsymbol{\alpha}$ that minimizes the average quality (6). The vector of design parameters determined in this way provides optimization of criterion (3) on average for a set of system goals. In many important practical cases, the system is designed for a single use. In these cases, the criterion of the form (6) loses its physical meaning, since it presupposes the repeated use of the system. In this case, it is proposed to optimize the worst conditions of the system's functioning, i.e., instead of functional (6), to optimize the functional

$$I(\boldsymbol{\alpha}) = \max_{\mathbf{c} \in C} p(\mathbf{c}) \Psi(\mathbf{c}, \boldsymbol{\alpha}).$$

The formulation of the optimization problem in this case is similar to the previous one. In the implementation of the above-mentioned formulation of the problem, the main difficulty is to determine the function $\Psi(\mathbf{c}, \boldsymbol{\alpha})$ for the finding of which at each point $(\mathbf{c}, \boldsymbol{\alpha})$ it is necessary to solve the problem of optimal control, which in most cases does not have an analytical solution. However, in many cases, and particularly the problem of optimal stabilization, it is possible to obtain an analytical dependence $\Psi(\mathbf{c}, \boldsymbol{\alpha})$

3. Multi-purpose stabilization of a linear dynamic system. Let us consider a dynamic system of the form

$$\dot{\mathbf{x}} = \mathbf{F}(\boldsymbol{\alpha})\mathbf{x}, \quad (7)$$

where $\mathbf{F}(\boldsymbol{\alpha}) - n \times n$ is the matrix-function of the s -dimensional vector $\boldsymbol{\alpha}$ of the constructive parameters. Let the control objective be defined as

$$\mathbf{c} = (\mathbf{x}_0 \in X, 0, \tau = \infty). \quad (8)$$

In other words, a set of stabilization problems under arbitrary initial conditions and infinite time is considered. The perturbed motion of the system caused by the deviation \mathbf{x}_0 from the zero-equilibrium point will be evaluated using the IQF.

$$I = \int_0^\infty \mathbf{x}^T(t) \mathbf{Q} \mathbf{x}(t) dt, \quad (9)$$

where \mathbf{Q} – given positive symmetric matrix. In this case [8] the value of the criterion I is in the form of $I = \mathbf{x}_0^T \mathbf{S} \mathbf{x}_0$, where \mathbf{S} – positive solution of the Lyapunov matrix equation

$$\mathbf{F}^T(\boldsymbol{\alpha})\mathbf{S} + \mathbf{S}\mathbf{F}(\boldsymbol{\alpha}) + \mathbf{Q} = \mathbf{0}. \quad (10)$$

It should be noted [13], that the positivity of the matrix \mathbf{S} for arbitrarily positive matrix \mathbf{Q} is a necessary and sufficient condition for the matrix $\mathbf{F}(\boldsymbol{\alpha})$ in the context of Hurwitzness criteria.

Let the initial perturbations \mathbf{x}_0 be distributed according to a normal law with zero mathematical expectation and a covariance matrix \mathbf{P} :

$$p(\mathbf{x}_0) = \frac{1}{(2\pi)^{n/2} \sqrt{|\mathbf{P}|}} e^{-\frac{1}{2}(\mathbf{x}_0)^T \mathbf{P}^{-1} \mathbf{x}_0}. \quad (11)$$

It is known [13], that there exists a coordinate transformation $\mathbf{z} = \mathbf{H}\mathbf{x}$, such that the quadratic forms in $I = \mathbf{x}_0^T \mathbf{S} \mathbf{x}_0$ and $V = \mathbf{x}_0^T \mathbf{P}^{-1} \mathbf{x}_0$ the coordinate system z_1, z_2, \dots, z_n will have a diagonal form $I = \mathbf{z}_0^T \mathbf{A} \mathbf{z}_0$, $V = \mathbf{z}_0^T \mathbf{z}_0$, where $\mathbf{A} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_k)$; λ_k – solutions of the characteristic equation of a bundle of quadratic forms (I, V) [13]:

$$|\mathbf{S} - \lambda \mathbf{P}^{-1}| = 0. \quad (12)$$

In coordinates z_1, z_2, \dots, z_n the mathematical expectation of the IQF on the set of initial perturbations with distribution (11) will take the form

$$\bar{J} = k \int_{-\infty}^{\infty} e^{-\frac{1}{2} \sum_{k=1}^n (\lambda_k)^2 z_{0,k}^2} \sum_{j=1}^n \lambda_j (z_{0,j})^2 dz_{0,1} dz_{0,2} \dots dz_{0,n}, \quad (13)$$

where $k = \frac{1}{(2\pi)^{n/2} |\mathbf{P}|^{n/2}}$.

The ratio (13) is equivalent to the following:

$$\bar{J} = k \sum_{j=1}^n \lambda_j \int_{-\infty}^{\infty} z_{0,j}^2 e^{-\frac{1}{2} \sum_{k=1}^n (\lambda_k)^2 z_{0,k}^2} dz_{0,1} dz_{0,2}, \dots, dz_{0,n}. \quad (14)$$

It is not difficult to show that the integral in (14) can be calculated by the formula $(2\pi)^{\frac{n}{2}}$, which after substituting in (14), will give the following result:

$$\bar{J} = |\mathbf{P}|^{-n/2} \sum_{j=1}^n \lambda_j. \quad (15)$$

Ratio (12) can be written in the equivalent form

$$|\mathbf{P}\mathbf{S} - \lambda \mathbf{E}| = 0, \quad (16)$$

from which it follows that the numbers $\lambda_1, \lambda_2, \dots, \lambda_n$ coincide with the eigenvalues of the matrix $\mathbf{P}\mathbf{S}$. Therefore, the average values of the IQF on the set of random initial perturbations (15) will take the form

$$\bar{J} = |\mathbf{P}|^{-n/2} \text{tr}(\mathbf{P}\mathbf{S}). \quad (17)$$

Thus, the problem of optimal stabilization of a linear dynamical system over the mean value of the IQF on a set of initial states distributed according to a normal law can be formulated as follows: minimize the criterion (17) at \mathbf{S} , satisfying equation (10) and the given dependencies for $\mathbf{F}(\mathbf{a})$ where $\mathbf{a} \in A$. It is not difficult to see that the solution of such an optimization problem in a general form is impossible and quite difficult in numerical formulation. In this regard, let us consider a slightly different approach to solving the problem of stabilization based on the inverse problem of the theory of stability.

4. Inverse problem of stability theory. Let two symmetric positive-definite matrices \mathbf{Q} and \mathbf{S} be given, which, as before, represent the matrices of the IQF and the solution to the Lyapunov equations

$$\mathbf{F}^T \mathbf{S} + \mathbf{S} \mathbf{F} + \mathbf{Q} = \mathbf{0}. \quad (18)$$

It is necessary to find a solution to equation (18) with in terms of the matrix \mathbf{F} . Since the matrix equation (18) is equivalent to a system $\frac{n(n+1)}{2}$ of linear equations for n^2 unknowns, such a system will generally be compatible and have an infinite number of solutions. To find these solutions, we will proceed as follows. Find the coordinate transformation

$$\mathbf{y} = \mathbf{R} \mathbf{x} \quad (19)$$

such that the matrices \mathbf{S} and \mathbf{Q} simultaneously take a diagonal form \mathbf{E} and \mathbf{A} , where $\mathbf{A} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$, $\lambda_1, \lambda_2, \dots, \lambda_n$ are the solutions of the equation $|\mathbf{Q} - \lambda \mathbf{S}| = 0$.

In [13] the algorithm for finding the transformation (19) is given. Since the matrix \mathbf{R} satisfies the ratio:

$$\mathbf{R}^T \mathbf{S} \mathbf{R} = \mathbf{E}, \quad (20)$$

then from (20) it follows

$$\mathbf{S} = (\mathbf{R}^T)^{-1} \mathbf{R}^{-1}. \quad (21)$$

Substituting (21) in (18), we get

$$\mathbf{F}^T (\mathbf{R}^T)^{-1} \mathbf{R}^{-1} + (\mathbf{R}^T)^{-1} \mathbf{R}^{-1} \mathbf{F} + \mathbf{Q} = \mathbf{0}. \quad (22)$$

Multiplying (22) on the left by \mathbf{R}^T and right by \mathbf{R} , we get

$$\bar{\mathbf{F}}^T + \bar{\mathbf{F}} = -\mathbf{A}, \quad (23)$$

where $\bar{\mathbf{F}} = \mathbf{R}^{-1} \mathbf{F} \mathbf{R}$. According to the general theory of solving systems of linear equations, the solution (23) can be represented as $\bar{\mathbf{F}} = \bar{\mathbf{F}}_1 + \bar{\mathbf{F}}_2$, where $\bar{\mathbf{F}}_1$ – partial solution of the inhomogeneous system (23), and $\bar{\mathbf{F}}_2$ – general solution of the corresponding homogeneous system

$$\bar{\mathbf{F}}^T + \bar{\mathbf{F}} = \mathbf{0}. \quad (24)$$

Let's present an arbitrary skew-symmetric matrix \mathbf{K} , as a linear combination of basic skew-symmetric matrices \mathbf{K}_{ij} , consisting of zeros, except for the elements $k_{ij} = -k_{ji} = 1$. Then the matrix \mathbf{K} will take the form $\mathbf{K} = \sum_{i,j=1, i < j}^n c_{ij} \mathbf{K}_{ij}$, where $c_{ij} = \frac{n(n-1)}{2}$, are arbitrary constants.

From (23) it follows directly that $\bar{\mathbf{F}}_1 = -\frac{1}{2} \mathbf{A}$, and from (24) it follows that $\bar{\mathbf{F}}_2 = \mathbf{K}$. Thus

$$\bar{\mathbf{F}} = -\frac{1}{2} \mathbf{A} + \mathbf{K}. \quad (25)$$

Let's move on to the initial basis by converting the inverse to (24)

$$\mathbf{F} = \mathbf{R}\bar{\mathbf{F}}\mathbf{R}^{-1}. \quad (26)$$

After substituting (26) in (25), we get the final form of the set of solutions of Lyapunov's equations in terms of the matrix \mathbf{F} :

$$\mathbf{F} = -\frac{1}{2}\mathbf{R}\mathbf{A}\mathbf{R}^{-1} + \mathbf{R}\mathbf{K}\mathbf{R}^{-1}. \quad (27)$$

As a partial solution of equation (18), we can also take the matrix

$$\mathbf{F} = -\frac{1}{2}\mathbf{S}^{-1}\mathbf{Q}, \quad (28)$$

which is not difficult to verify by the direct substitution of (28) in (18). It is also evident that the general solution corresponding to (18) of a homogeneous system has the form $\mathbf{F}_2 = \mathbf{S}^{-1}\mathbf{K}$.

Thus, the general solution of the Lyapunov equation in terms of the matrix \mathbf{F} can be written as

$$\mathbf{F} = \mathbf{S}^{-1}\left(-\frac{1}{2}\mathbf{Q} + \mathbf{K}\right). \quad (29)$$

Finally, (29) let us present it in the form of

$$\mathbf{F} = -\frac{1}{2}\mathbf{S}^{-1}\mathbf{Q} + \sum_{(i,j)} c_{ij}\mathbf{F}_{ij}, \text{ where } \mathbf{F}_{ij} = \mathbf{S}^{-1}\mathbf{K}_{ij}.$$

Now the problem of choosing a vector of parameters α of a dynamical system that provides the specified dynamic characteristics based on the IQF is reduced to solving the system of equation (30) in terms of the α and c_{ij} :

$$\mathbf{F}(\alpha) = \mathbf{F}_1 + \sum_{(i,j)} c_{ij}\mathbf{F}_{ij}. \quad (30)$$

Confine ourselves to the case of a linear matrix $\mathbf{F}(\alpha)$

in terms of parameters α : $\mathbf{F}(\alpha) = \mathbf{F}^0 + \sum_{i=1}^m \alpha_i \mathbf{F}^i$, where

$\mathbf{F}^0, \mathbf{F}^1, \dots, \mathbf{F}^m$ are the given fixed matrices. In this case, the problem of synthesis of a stable system under consideration is reduced to the solution of a linear system of equations:

$$\sum_{i=1}^m \alpha_i \mathbf{F}^i - \sum_{(i,j)} c_{ij} \mathbf{F}_{ij} = \mathbf{F}_1 - \mathbf{F}^0. \quad (31)$$

System (31) can be represented as a system n^2 of linear equations in context of $m + \frac{n(n-1)}{2}$ unknown a_i and c_{ij} .

The condition for the compatibility of such a system is the condition $m + \frac{n(n-1)}{2} \geq n^2$. The number of unknowns must be not less than the number of equations. The last inequality can be rewritten as

$$m \geq \frac{n(n-1)}{2}. \quad (32)$$

In the general case, let us assume that instead of (32) there is a strict inequality. Let us also assume that the

system of matrices \mathbf{F}_{ij} and \mathbf{F}^i , included in (31), has a maximum rank. Let's vectorize the matrices in (31). As a result, we get a linear system of the equation

$$\mathbf{A}\mathbf{z} = \mathbf{b}, \quad (33)$$

where $\mathbf{z}^T = (\alpha_{1m}, \alpha_{m1}, c_{12}, \dots, c_{(n-1)n})$ – vector of unknown constants, and the matrix \mathbf{A} and vectors are constructed from the elements of the matrices \mathbf{F}^i , \mathbf{F}_{ij} and \mathbf{F}^0 in accordance with the matrix equation (31). Since the matrix \mathbf{A} is rectangular, of $n^2 \times \left(m + \frac{n(n-1)}{2}\right)$ dimensions, then,

in accordance with the general theory of systems of linear equations, the general solution of the system of equations (33) can be found in the form:

$$\mathbf{z} = \mathbf{z}_0 + \sum_{i=1}^r \xi_i \mathbf{z}_i, \quad r = \frac{n(n+1)}{2} - m, \quad (34)$$

where \mathbf{z}_0 – partial solution of the system (33). \mathbf{z}_i – linearly independent solutions corresponding to a homogeneous system, ξ_i – arbitrary constants.

The ratio (34) sets a set of parameters of the system α that ensure stability (7). To select specific values $\alpha_1, \alpha_2, \dots, \alpha_m$ it is necessary to consider the system of restrictions on their values, but this issue is not the subject of investigation in this work.

Conclusions. The paper presents a general formulation of the multi-objective optimal control problem with the selection of design parameters for the controlled system. As a special case, the problem of selecting the average value of quality is formulated based on the integral quadratic functional over a set of initial deviations from the zero equilibrium position, distributed according to a normal distribution.

The inverse stability problem for linear dynamic systems is considered, which involves finding a set of system matrices that satisfy the Lyapunov equation, given the matrices of intensive quadratic forms that appear in the equation. A general expression for the set of the sought stable matrices of the dynamic system is derived.

References

1. Фельдбаум А. А. Интегральные критерии качества регулирования. Автоматика и телемеханика. 1948. Т. IX, № 1. С. 3–19.
2. Красовский А. А. *Интегральные критерии оценки качества процессов регулирования*. Москва: Машгиз, 1949. 24 с.
3. Красовский А. А. *Статистическая теория переходных процессов в системах управления*. Москва: Наука, 1965. 228 с.
4. Понтрягин Л. С., Болтянский В. Г., Гамкрелидзе Р. В. и др. *Математическая теория оптимальных процессов*. Москва: Физматгиз, 1961. 391 с.
5. Болтянский В. Г. *Математические методы оптимального управления*. Москва: Физматгиз, 1966. 307 с.
6. Андреев Ю. Н. *Управление конечномерными линейными объектами*. Москва: Физматгиз, 1976. 424 с.
7. Кузовков Н. Т. *Модальное управление и наблюдающие устройства*. М.: Машиностроение, 1976. 184 с.
8. Ляпунов А. М. *Общая задача об устойчивости движения*. Ленинград–Москва: ГИТТЛ, 1950. 471 с.
9. Лурье А. И. *Некоторые нелинейные задачи теории автоматического управления*. Москва: Госэнергоиздат, 1951.

10. Якубович В. А. Решение некоторых матричных неравенств, встречающихся в теории автоматического регулирования. *ДАН СССР*. 1962. Т. 143, № 6. С. 1304–1307.
11. Kutsenko A. S., Kovalenko S. V. Some Approaches to a Quantitative Measure of the Stability of Dynamical Systems. *Eurasian Journal of Mathematical and Computer Applications*. 2019, vol. 7, pp. 276–281.
12. Куценко А. С., Свиридова И. А. Выбор параметров линейной динамической системы по квадратичному критерию качества. *Вестник Харьковского государственного политехнического университета*. Харьков: ХГПУ, 1999, вып. 57. С. 17–22.
13. Гантмахер Ф. Р. Теория матриц. Москва: Наука. 1967. 575 с.
6. Andreev Yu. N. *Upravlenie konechnomernymi lineynymi ob'ektami*. [Management of finite-dimensional linear objects]. Moscow, Fizmatgiz Publ., 1976. 424 p. (In Russ.).
7. Kuzovkov N. T. *Modal'noye upravleniye i nablyudayushchiye ustroystva* [Modal Control and Observer Devices]. Moscow, Mashinostroyeniye Publ., 1976. 184 p. (In Russ.).
8. Lyapunov A. M. *General Problems of the Stability of Motion* [General problem of motion stability]. Leningrad–Moscow, GITTL Publ., 1950. 471 p. (In Russ.).
9. Lur'e A. I. *Nekotorye nelineinye zadachi teorii avtomaticheskogo upravleniya* [Some nonlinear problems of automatic control theory]. Moscow, Gosenergoizdat Publ., 1951. (In Russ.).
10. Yakubovich V. A. Resheniye nekotorykh matrichnykh neravenstv, vstrechayushchikhsya v teorii avtomaticheskogo regulirovaniya [Solution of some matrix inequalities encountered in the theory of automatic control]. *DAN SSSR [RAS USSR]*. 1962, vol. 143, no. 6, pp. 1304–1307. (In Russ.).

References (transliterated)

1. Feldbaum A. A. Integral'nye kriterii kachestva regulirovaniya [Integral criteria for quality of regulation]. *Avtomatika i telemekhanika* [Automation and telemechanics]. 1948, vol. IX, no. 1. pp. 3–19. (In Russ.).
2. Krasovskiy A. A. *Integral'nye kriterii otsenki kachestva protsessov regulirovaniya* [Integral criteria for assessing the quality of regulatory processes]. Moscow, Mashgiz Publ., 1949. 24 p. (In Russ.).
3. Krasovskiy A. A. *Statisticheskaya teoriya perekhodnykh protsessov v sistemakh upravleniya* [Statistical theory of transient processes in control systems]. Moscow, Nauka Publ., 1965. 228 p. (In Russ.).
4. Pontryagin L. S., Boltyansky V. G., Gamkrelidze R.V. et al. *Matematicheskaya teoriya optimal'nykh protsessov* [Mathematical theory of optimal processes]. Moscow, Fizmatgiz Publ., 1961. 391 p. (In Russ.).
5. Boltyanskiy V. G. *Matematicheskie metody optimal'nogo upravleniya*. [Mathematical methods of optimal control]. Moscow, Fizmatgiz Publ., 1966. 307 p. (In Russ.).
11. Kutsenko A. S., Kovalenko S.V. Some Approaches to a Quantitative Measure of the Stability of Dynamical Systems. *Eurasian Journal of Mathematical and Computer Applications*. 2019, vol. 7, pp. 276–281.
12. Kutsenko A. S., Sviridova I.A. Vybor parametrov lineynoy dinamicheskoy sistemy po kvadratichnomu kriteriyu kachestva [Selection of parameters of a linear dynamic system using a quadratic quality criterion]. *Vestnik Kharkovskogo gosudarstvennogo politekhnicheskogo universiteta* [Bulletin of the Kharkov State Polytechnic University]. Kharkov, KhGPU Publ., 1999, issue. 57, pp. 17–22. (In Russ.).
13. Gantmakher F. R. *Teoriya matrits* [Matrix theory]. Moscow, Nauka Publ., 1967. 575 p. (In Russ.).

Received 01.11.2024

УДК 681.5

О. С. КУЦЕНКО доктор технічних наук, професор, Національний технічний університет «Харківський політехнічний інститут», професор кафедри системного аналізу та інформаційно-аналітичних технологій, м. Харків, Україна; e-mail: Oleksandr.Kutsenko@khpri.edu.ua; ORCID: <https://orcid.org/0000-0002-7964-1286>

М. Є. АЛФЬОРОВ, аспірант, Національний технічний університет «Харківський політехнічний інститут», кафедра системного аналізу та інформаційно-аналітичних технологій, м. Харків, Україна; e-mail: Mykhailo.Alforov@cs.khpri.edu.ua; ORCID: <https://orcid.org/0009-0005-4920-5626>

А. Є. АЛФЬОРОВ, аспірант, Національний технічний університет «Харківський політехнічний інститут», кафедра системного аналізу та інформаційно-аналітичних технологій, м. Харків, Україна; e-mail: Andrii.Alforov@cs.khpri.edu.ua; ORCID: <https://orcid.org/0009-0007-9506-7468>

СИНТЕЗ КОНСТРУКТИВНИХ ПАРАМЕТРІВ БАГАТОЦІЛЬОВИХ ДИНАМІЧНИХ СИСТЕМ

Розглядаються дві задачі, пов'язані з оптимізацією лінійних стаціонарних динамічних систем. Запропоновано загальну постановку багатоцільової задачі оптимального керування з вибором конструктивних параметрів. Як окремий випадок, аналізується задача багатокритеріальної оптимізації лінійної системи за інтегральним квадратичним критерієм за умови заданого випадкового розподілу початкових збурень. Розв'язок ґрунтується на методі одночасного зведення двох виразно позитивних квадратичних форм до діагонального вигляду. Отримано аналітичні вирази, що дають змогу обчислити математичне очікування критерію для випадку багатовимірного нормального розподілу вектора початкових випадкових збурень. Сформульовано обернену задачу теорії стійкості: необхідно віднайти вектор структурних параметрів, що забезпечує стійкість системи та задане середнє значення інтегрального квадратичного критерію якості на множині початкових збурень. Розв'язання пропонується виконувати у два етапи. Спершу виводиться загальний розв'язок матричного рівняння Ляпунова в термінах елементів матриці системи. Для цього простір станів відображається на власний підпростір додатноозначеної матриці, що відповідає інтегральному квадратичному критерію. Встановлено, що загальний розв'язок зумовлений довільною кососиметричною матрицею або відповідним набором довільних сталих. З іншого боку, коли матриця системи лінійно залежить від вектора конструктивних параметрів, можливо сформулювати лінійну систему рівнянь відносно цих невідомих параметрів та довільних сталих, що фігурують у загальному розв'язку оберненої задачі стійкості. Загалом, така система є сумісною та допускає нескінченну кількість розв'язків, які задовольняють початкові вимоги до елементів симетричних матриць, присутніх у рівнянні Ляпунова.

Ключові слова: стійкість, інтегральний квадратичний функціонал, матричне рівняння Ляпунова, обернена задача стійкості, багатоцільові динамічні системи, лінійні стаціонарні системи, параметрична оптимізація.

Повні імена авторів / Author's full names

Автор 1 / Author 1: Куценко Олександр Сергійович, Kutsenko Oleksandr Serhiyovych

Автор 2 / Author 2: Алфьоров Михайло Євгенович, Alforov Mykhailo Evgenovich

Автор 3 / Author 3: Алфьоров Андрій Євгенійович, Alforov Andrii Evgeniovich

УПРАВЛІННЯ В ОРГАНІЗАЦІЙНИХ СИСТЕМАХ

MANAGEMENT IN ORGANIZATIONAL SYSTEMS

DOI: 10.20998/2079-0023.2024.02.05
UDC 004.9

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>
I. V. LIUTENKO, 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: cherliv68@gmail.com; ORCID: <https://orcid.org/0000-0003-4357-1825>
V. V. YAMBURENKO, National Technical University "Kharkiv Polytechnic Institute", PhD Student, Kharkiv, Ukraine; e-mail: yamburenko@gmail.com; ORCID: <https://orcid.org/0009-0005-4797-8684>
A. A. PASHNIEV, Candidate of Technical Sciences (PhD), Senior Researcher, National Technical University "Kharkiv Polytechnic Institute", Associate Professor of the Department of Information Systems and Technologies, Kharkiv, Ukraine; e-mail: pashniev@email.ua; ORCID: <https://orcid.org/0000-0002-9150-6108>

A SOFTWARE SOLUTION FOR REAL-TIME COLLECTION AND PROCESSING OF MEDICAL DATA FOR EPILEPSY PATIENTS

The rapid development of computer technologies has significantly impacted various sectors, including healthcare. The ability to collect, process, and visualize medical data in real time is becoming increasingly important, especially for managing chronic conditions such as epilepsy. This paper presents a web-based application designed for real-time monitoring of health indicators, enabling healthcare professionals to track patient data efficiently. The system automates the process of collecting data from fitness trackers, transmitting it via a mobile device to a server, and visualizing it in a web application. Its architecture employs a thin-client model with Node.js for backend logic and React.js for the user interface, ensuring scalability and responsiveness. Key features include real-time data visualization, historical trend analysis, and the ability to export health metrics for further examination. The system architecture follows a modular approach, with a clear separation of concerns between the client-side, server-side, and database components. MongoDB is used as the database provider, offering flexibility in handling large volumes of health data. The system underwent extensive testing in two stages. During the first stage, real-world data collection demonstrated an average data transmission time of less than 112 ms, ensuring compliance with real-time requirements. In the second stage, stress testing with up to 100 simultaneous users showed an average server response time of 145.8 ms and a 95th percentile response time of 167.1 ms. These results confirm the system's robustness and suitability for deployment in medical facilities. Future work aims to enhance the system by incorporating advanced real-time alert mechanisms and additional health metrics, such as oxygen saturation and activity levels, to provide comprehensive monitoring. The presented solution showcases the potential of integrating modern web technologies into healthcare, contributing to improved patient outcomes and more efficient workflows for medical professionals.

Keywords: real-time monitoring, medical data processing, automation, health indicators, software, system architecture, development technologies.

Introduction. The development of contemporary society is characterized by the substantial impact of computer technologies, which have found application across all areas of human activity. Information technologies facilitate the dissemination of information within society, thereby forming a global information space. Consequently, they have become a crucial driving force in the development of the global economy and other spheres of human activity. It is challenging to identify sectors where information technologies are not currently utilized. The primary domains of their implementation include construction, mechanical engineering, education, banking, and, notably, medicine. In numerous medical studies, the use of computers and specialized software is indispensable. In today's context, proficiency in computer use is becoming one of the most vital professional skills for medical

workers. Data collection and analysis methods, particularly in the realm of intellectual analysis, although relatively young among computer sciences, are gaining increasing relevance. This trend is driven by the growing volume of information and the enhancement of methods for its collection and processing. The relevance of this work lies in the ongoing development of the field of medical data collection and processing, especially in real-time scenarios.

Epilepsy is a highly significant issue within modern neurology and psychiatry. According to the World Health Organization (WHO), the lack of comprehensive epidemiological data on epilepsy across various countries results in notable deficiencies in the organization of medical care. Seizures associated with epilepsy may be attributed to brain trauma or hereditary factors; however, in many cases, the etiology remains unknown. Individuals with epilepsy

© Kopp A. M., Liutenko I. V., Yamburenko V. V., Pashniev A. A., 2024



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.



frequently exhibit multiple seizure types and may also suffer from other neurological disorders. Recurrent epileptic seizures are often exacerbated by specific external and internal factors. For instance, it is well-documented that seizures can be triggered by sensory stimuli such as exposure to bright, flashing lights or cardiovascular stress. The management of these seizures presents significant challenges, and preventing their consequences without the employment of specialized (auxiliary, invasive) monitoring devices is virtually unfeasible.

Fitness trackers represent a cornerstone of trends within the field of healthcare web applications. The synchronization of these devices with mobile platforms and their integration into personalized dashboards significantly aids in the collection of patient data. These trackers facilitate the monitoring of variables such as heart rate, blood glucose levels, among others. Notably, 49% of the North American population utilizes smart devices, highlighting the remarkable proliferation of such trends within healthcare.

The primary objective of this work is to enhance the existing approach to data collection from fitness trackers, focusing on the analysis and presentation of final data to the physician of a specific patient.

Analysis of recent research and publications. The creation and maintenance of modern computer databases concerning patients' health status, disease progression, laboratory tests, and treatments no longer pose significant challenges for information technology specialists. However, a more complex issue is the lack of effective information technology for the processing and analysis of medical data, which would enable medical analysts to identify patterns and correlations between various medical indicators that are not immediately apparent. This, in turn, would enhance the efficacy of treatment by selecting therapy intensity that is adequate to the patient's condition, assessed based on a combination of identified risk factors [1]. The determination of specific features in the processing and analysis of statistical data from medical research, as well as the improvement of information technology that accounts for these features, remains critical. The application of traditional statistical methods in medical data analysis has a long history [2, 3]. Despite their significant utility and importance, it is essential to acknowledge their limitations concerning medical data. The specific nature of medical data used in causal relationship analysis does not allow for the correct application of many existing mathematical methods currently used for data processing and analysis. The primary reason for this is that classical statistical methods work well for testing pre-formulated hypotheses, which is rarely achievable in medical practice. Additionally, various limitations of statistical methods related to specific distribution laws of samples, data types, sample sizes, and other factors must be noted.

Methods of data mining developed in the late 1990s, such as neural networks of various types, fuzzy logic, genetic algorithms, etc., can address the limitations of statistical methods. Examples of versatile software systems based on these methods include Intelligent Miner (USA), Knowledge Studio (Canada), SAS (USA), MineSet (USA), Clementine (UK), ST: Neural Networks, among others.

However, this versatility becomes a drawback in medical applications, as it does not account for all the specific characteristics of medical data. These characteristics include extremely large data volumes, the selection of the most informative features, the development of decision rules considering risks, and the evaluation of the validity of information.

Medical information systems (MIS) are employed to automate tasks at each stage of the diagnostic and therapeutic process. The creation of a medical information system serves several purposes:

- improving the quality of work of healthcare workers and institutions by organizing the comprehensive processing of medical information that aligns with the level of utilized technical means, including the enhancement of management and planning processes;
- facilitating the labor of healthcare workers by eliminating labor-intensive and inefficient manual processes of medical data processing and analysis;
- ensuring effective information exchange with other information systems.

Based on their purpose, MIS are classified into:

- systems whose primary function is data accumulation (automated data and/or information processing systems, automated information and reference systems);
- diagnostic and consultation systems;
- systems that provide medical services.

In Ukraine, public healthcare institutions can choose any medical information system provided that it has successfully passed testing, is connected to the central database of the Electronic Health System (EHS), has the necessary functionality, and meets the technical specifications defined by the Ministry of Health. Nowadays, medical information systems are a crucial means of establishing interaction between healthcare facilities and patients. Recently, the demand for acquiring information systems in hospitals has increased more than ever, leading to heightened competition among developers of medical information systems. Therefore, to compete with other MIS, developers must improve their systems or create them with a level of functionality that allows healthcare institutions to provide a better level of service to their patients [4].

Let us consider the most popular medical information systems operating in Ukraine.

“HELSI” is a modern, convenient, and reliable electronic medical system designed for both public and private medical institutions, doctors, and patients. The system features a data center certified by the State Service for Special Communication and Information Protection of Ukraine (SSSCIP) for comprehensive information security (CSIS).

The system automates the majority of processes within healthcare institutions, including the work of doctors and registration departments. It maintains electronic medical records for patients and manages doctors' schedules.

Additionally, the system ensures the accounting of payments and medications, the generation of statistics and reports, and includes a form constructor [5].

Among the advantages, it can be noted that the system automates numerous processes, records user actions, and sets access rights while maintaining data confidentiality.

However, the system has a very low rating in online surveys, with most negative reviews indicating that the system does not effectively interact with healthcare institutions. Appointments made in the system do not correspond to those in the hospital. The MIS also faces issues with patient registration, user interface convenience, and overall design.

The company “eLife” has been operating in the field of process automation for nine years. This company has developed a medical information system called “Medeir”. Among its advantages, the system is noted for planning interactions with patients, optimizing the operations of healthcare institutions, and easily scaling data management. The MIS promises high reliability, optimization of staff working time, and cost optimization [6]. Reviews of the system are available only on its own website, making it difficult to evaluate its overall rating. By examining the system’s images, the main disadvantages identified include inconvenient, complex, and outdated interface, making it challenging to assess the functionality in detail at this stage.

“Medstar Solutions” is a medical information system that automates medical processes, ensuring comfortable patient care. It is the first cloud-based MIS in Ukraine with a state security certificate for the protection of personal and medical data. The system has been operating in the market for nearly eight years, improving the patient care process [7]. The MIS has a decent rating, with approximately half of the reviews on social media being positive. Most negative reviews focus on the slow processing of requests and the creation of episodes, diagnoses, etc.

“Doctor Eleks” is the medical information system has been providing services for 17 years and has extensive experience working with over 1,400 healthcare facilities. The system automates all business processes within institutions and ensures the creation of reports based on the data entered into the system and their analysis. Among the advantages, it is noted that the MIS, due to its extensive experience in this field, has a better and more detailed understanding of all healthcare facility processes. Additionally, the MIS can store data both in cloud storage and on servers, ensuring data security in accordance with the comprehensive information security system (CSIS) [8]. Among the disadvantages are slow request processing, slow creation of medical leave documents, the presence of errors, a non-user-friendly interface, and unsatisfactory customer support service.

“Health 24” allows any patient to quickly make an appointment or consultation with any doctor without spending much time. To receive quality medical services, there is no need to stand in long queues at hospitals or clinics. The comprehensive service provides online doctor searches, selection of the best clinics, online registration, access to electronic medical records, and price comparison for services [9]. The majority of reviews about this product are positive, making this MIS the highest-rated among those listed. Customers highlight the convenience and clarity of the program, well-organized document flow, and the implementation of standardization in record-keeping. However, it faces challenges in working with multidisciplinary healthcare institutions.

Let us consider the most popular medical information

systems operating abroad. England has developed systems such as NHS Director, which allows for advice and consultation directly from the National Health Service. The National Center for Emergency Medical Services performs numerous tasks related to patient treatment at home. However, the issue is that the work of these services is not integrated with the activities of other health services and the health management system throughout the United Kingdom.

The Danish automation of the healthcare system serves as an exemplary model. In 1994, Denmark established a healthcare network system, which was later portalized in 2004. Currently, Health Net is an essential resource for citizens, professionals, and general practitioners, facilitating home treatment. The system operates on three levels: national (National Health Board), primary (nursing homes, pharmacies), and secondary (hospitals). All levels are integrated into a unified portal infrastructure, enabling participants at any level to access all types of information.

The implementation of a hospital automation information system in Austria is exemplified by the project carried out at the Vienna General Hospital. One of the operational principles of this institution is the integration of therapeutic, educational, and research activities, necessitating the implementation of an information system that supports clinical processes as well as research and teaching activities, serving approximately 8,000 users. Consequently, the hospital introduced electronic medical records, a patient administration module, a content management system for the general clinical repository (including medical device data, multimedia data, scanned documents, videos), a planning and documentation module for operations, a data migration system, integration of existing subsystems, a clinical drug trial system, and a research integration platform (to combine clinical data and research data, such as phenotype and genotype) [10].

Electronic Medical System in Canada. The Ministry of Health of the Canadian province of Alberta decided to distribute medical information through a secure network to ensure informed decision-making by healthcare professionals regarding the provision of medical care. To implement this project, an integrated medical system (Wellnet) was developed. This system supports decision-making and ensures the availability of accurate and up-to-date information on population health and current medication information (including evidence-based medicine).

It is evident that the primary task of information and communication technologies in healthcare is to ensure patient safety. Based on this premise, healthcare informatization in European Commonwealth countries follows the path of creating decision support systems for physicians, establishing a new informational environment for their activities. This environment allows access to professional information resources (information and reference systems and thematic forums) and navigation between professional associations and recognized experts. It creates a continuous education environment for physicians through regular monitoring of the latest achievements in medicine and pharmacy. Computerized

physician order entry systems can reduce medication errors by nearly 80% and adverse side effects by 55%. In some countries, physicians cannot obtain a license to practice without appropriate knowledge in information technology (e.g., in the Netherlands) [10].

Experts, based on many years of experience in creating and using medical information systems for working with visual images obtained from radiological, tomographic, mammographic, ultrasound, and other studies, assert that the benefits of automating instrumental diagnostics departments are evident. This automation significantly expands the professional capabilities of physicians across various specialties.

The automation of clinical laboratory studies is of great importance in improving the quality of medical care. It significantly enhances laboratory productivity and creates a unified database of research results, which, when integrated with the medical information system of healthcare facilities, becomes immediately accessible to all necessary medical professionals. According to most experts, the most adequate indicator of the system's organizational efficiency is the reduction of medical personnel's working time spent on preparing reporting documentation. The preparation of a large number of various documents and reports has become a bottleneck in the operation of healthcare facilities, which inevitably impacts the quality of treatment. Therefore, the benefit of automation is currently linked to solving this issue. As the MIS continues to develop, this aspect of work can be completely removed from the responsibilities of medical personnel. The system should automatically retrieve all necessary information from medical documentation and transmit it to specialists responsible for report preparation. The implementation of an MIS can be considered

successful if, after a period of time – which may span months or years—practically all medical personnel at the healthcare facility are using the system, at which point the question of quantitatively evaluating efficiency arises [10].

Proposed solutions. The article examines a developed web application that allows physicians to monitor patient health indicators in real-time, thereby automating the data monitoring process, as well as the overall data control process through the use of a dashboard. The software product can be utilized in various medical institutions. The web application includes patient search, patient information, and infographics. The primary user of the software is the physician.

We have a device in the form of a fitness tracker that transmits user data via Bluetooth protocol to an Android operating system device, storing this data in local storage. There is also a server entity that receives real-time data from the mobile device and transmits this data to the web application through a WebSocket connection. The data flow diagram is shown in fig. 1.

Functional requirements for the software:

- the doctor has the ability to log into the system;
- has the ability to log out of the system;
- the doctor has the ability to create a new patient profile;
- the doctor has the ability to interact with patient data;
- has the ability to edit patient data;
- has the ability to delete patient data from the system;
- the doctor has the ability to view patient health indicators;
- display of indicators for a specific time;

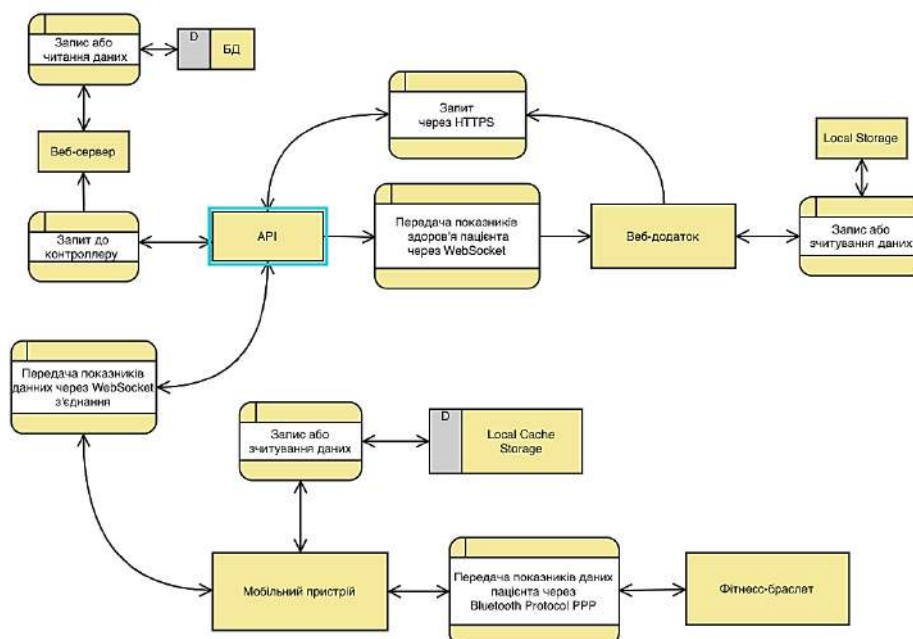


Fig. 1. Data Flow Diagram for proposed information system

- real-time display of indicators;
- real-time display of charts;
- display of indicators for the entire monitoring period;
- export of health indicator data for the entire period;
- export of health indicator data for a specific period.

The chosen architecture is a “thin client,” with information processing tasks performed on the server [11]. The deployment diagram is shown in fig. 2. Development technologies include NodeJS as the web application framework for user interaction with the website. The programming language is JavaScript (JS), with the React.JS library used for creating user interfaces. The platform utilized is Node.js. HTML5 is used for the external interface, and CSS3 is used for web page design.

The main components in the diagram are:

- “UserClient” – a computer or laptop used by a doctor or administrator.
 - “Web Browser” – a browser installed on the computer or laptop used by the doctor or administrator.
 - “Engine V8” – the browser engine for compiling and translating JavaScript code.
 - “React.JS” – a component library for project initialization.
 - “Client-Side-App” – a component for server-side rendering of application pages.
 - “web-socket-client.js” – a component for initializing WebSocket connections.
- “webpack.js” – a component library for optimizing the project bundle during compilation.
- “WebServer/LocalServer” – a device or machine leased for deployment.
 - “NodeJS” – a framework component for creating the server-side application.
 - “ExpressJS Application” – a component library for additional functionality and handling application services.

For data management, we have chosen to utilize MongoDB, an open-source NoSQL database that offers flexible data modeling without the need for predefined table schemas. Its ability to efficiently handle large volumes of diverse medical data makes it particularly suitable for healthcare applications [12,13]. The data model is illustrated in fig. 3.

As a result of developing the software system, all necessary functionality was implemented, including the user interface and data retrieval from the mobile device.

The system was tested in two stages. In the first stage, data collection and analysis were conducted under real conditions.

A sample of the collected data, shown in fig. 4, includes key parameters such as pulse rate, number of steps, calories burned at specific moments, and an indicator of the presence of an episode based on sharp deviations in pulse rate from the patient’s historical data. System logs were also analyzed, revealing that the average data transmission time from the fitness tracker to the web application was less than 112 ms, meeting the best practice ranges of patient

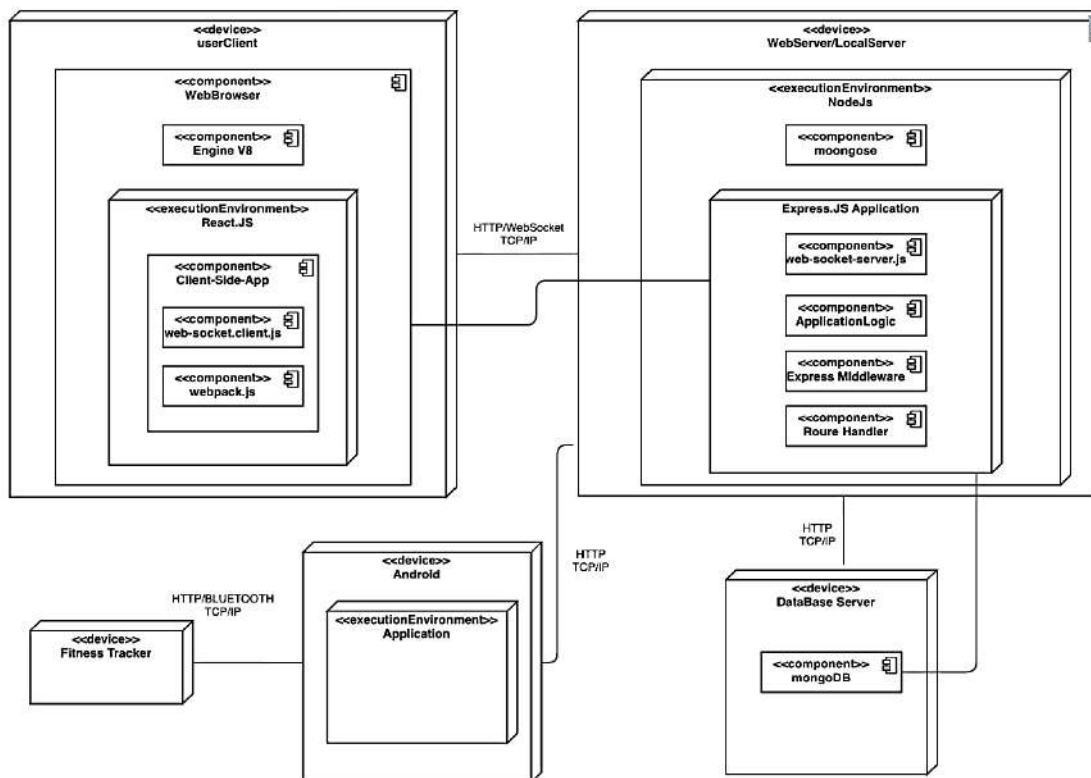


Fig. 2. Deployment Diagram for proposed system

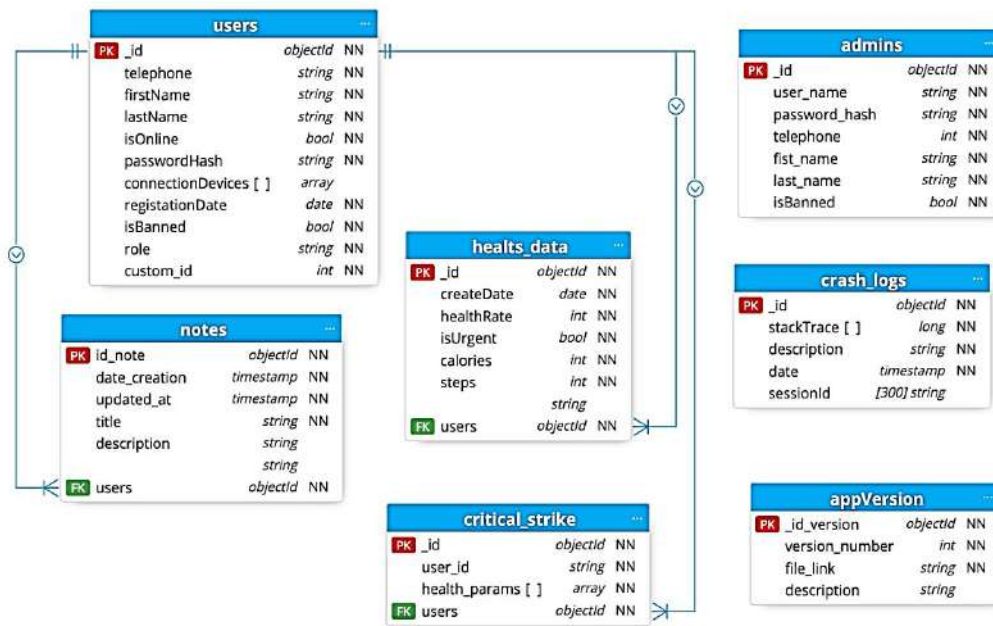


Fig. 3. Logical Data Model for proposed system

Идентификатор	Пользователь	Прístup	Дата	Пульс	Кроки	Калорії
609c8b133b474100173f5784	834124	FALSE	13.05.2021, 5:12:35 GMT+3	72	860	1344
609c8b0e3b474100173f5783	834124	FALSE	13.05.2021, 5:12:30 GMT+3	75	860	1344
609c8b093b474100173f5782	834124	FALSE	13.05.2021, 5:12:25 GMT+3	76	860	1344
609c8b043b474100173f5781	834124	FALSE	13.05.2021, 5:12:20 GMT+3	78	860	1344
609c8aff3b474100173f5780	834124	FALSE	13.05.2021, 5:12:15 GMT+3	79	860	1344
609c8afa3b474100173f577f	834124	FALSE	13.05.2021, 5:12:10 GMT+3	79	860	1344
609c8af53b474100173f577e	834124	FALSE	13.05.2021, 5:12:05 GMT+3	80	860	1344
609c8af03b474100173f577d	834124	FALSE	13.05.2021, 5:12:00 GMT+3	80	860	1344
609c8ae63b474100173f577c	834124	FALSE	13.05.2021, 5:11:55 GMT+3	79	860	1344
609c8ae3b474100173f577b	834124	FALSE	13.05.2021, 5:11:50 GMT+3	79	860	1344
609c8ae13b474100173f577a	834124	FALSE	13.05.2021, 5:11:45 GMT+3	80	860	1344
609c8ad3b474100173f5779	834124	FALSE	13.05.2021, 5:11:40 GMT+3	80	860	1344
609c8ad73b474100173f5778	834124	FALSE	13.05.2021, 5:11:35 GMT+3	79	860	1344
609c8ad23b474100173f5777	834124	FALSE	13.05.2021, 5:11:30 GMT+3	76	860	1344
609c8acd3b474100173f5776	834124	FALSE	13.05.2021, 5:11:25 GMT+3	78	860	1344
609c8ac83b474100173f5775	834124	FALSE	13.05.2021, 5:11:20 GMT+3	78	860	1344
609c8ac33b474100173f5774	834124	FALSE	13.05.2021, 5:11:15 GMT+3	79	860	1344
609c8abe3b474100173f5773	834124	FALSE	13.05.2021, 5:11:10 GMT+3	72	851	1342
609c8ab93b474100173f5772	834124	FALSE	13.05.2021, 5:11:05 GMT+3	75	850	1341
609c8ab43b474100173f5771	834124	FALSE	13.05.2021, 5:11:00 GMT+3	78	841	1340
609c8aa13b474100173f5770	834124	FALSE	13.05.2021, 5:10:55 GMT+3	65	840	1337
609c8aaa3b474100173f576f	834124	FALSE	13.05.2021, 5:10:50 GMT+3	95	840	1337
609c8aa53b474100173f576e	834124	FALSE	13.05.2021, 5:10:45 GMT+3	90	840	1337
609c8aa03b474100173f576d	834124	FALSE	13.05.2021, 5:10:40 GMT+3	78	837	1333
609c8a9b3b474100173f576c	834124	FALSE	13.05.2021, 5:10:35 GMT+3	75	835	1329
609c8a963b474100173f576b	834124	FALSE	13.05.2021, 5:10:30 GMT+3	74	834	1327

Fig. 4. Generated report with collected data

monitoring systems [14]. The average server request processing time was less than 71 ms.

As a result, the system visualized heartbeat rate changes as an interactive chart, enabling the doctor to assess the patient’s condition in real-time or over a specific period and identify potentially critical situations. fig. 5 presents the chart.

In the second stage of testing, load testing was conducted using the K6 Load Testing Tool, simulating up to 100 concurrent patient devices sending medical

indicators. Under load, the average server request time was 145.8 ms, and the 95th percentile time was 167.1 ms. These metrics confirm the system’s ability to operate stably even under high levels of concurrent requests, ensuring reliable performance in real-world medical settings.

Conclusions. The developed system provides real-time monitoring of health indicators for patients with epilepsy, enabling early detection of critical conditions and timely medical response. The system utilizes historical patient data to identify deviations in heart rate trends, which



Fig. 5. Interactive chart of heartbeat rate changes for Real-Time patient monitoring

can serve as predictors of epileptic seizures. This functionality enhances the accuracy and efficiency of medical decision-making.

In the future, the system can be improved by implementing the following features:

- integration of advanced algorithms for predicting seizures based on heart rate variability and other parameters;
- support for real-time notifications to alert medical staff about potential critical conditions.
- expansion of data collection to include additional health metrics, such as oxygen saturation and activity levels, for comprehensive monitoring.

References

1. Brown D. E. *Introduction to data mining for medical informatics*. URL: <https://pubmed.ncbi.nlm.nih.gov/18194716/> (access date: 12.09.2023).
2. Comak E., Polat K., Güneş S., Arslan A. *A new medical decision making system: least square support vector machine (LSSVM) with fuzzy weighting pre-processing*. URL: https://journals.scholarsportal.info/details/09574174/v32i0002/409_anmdmsvmwfw.xml&sub=all (access date: 18.09.2023).
3. *7 Ways Data Science Is Reshaping Healthcare*. URL: <https://www.altexsoft.com/blog/datascience/7-ways-data-science-is-reshapinghealthcare/> (access date: 20.09.2023).
4. *Survey: 8 in 10 Hospital Leaders Say Predictive Analytics is Important to Future Yet Only One-Third Use It*. URL: <https://chimecentral.org/survey-8-10-hospital-leaders-say-predictive-analytics-important-future-yet-one-third-use/> (access date: 20.09.2023).
5. *Helsi*. URL: <https://helsi.me/about> (access date: 02.11.2023).
6. *Medeür*. URL: <https://e-life.com.ua/prod-tech/medejr-uk/> (access date: 02.11.2023).
7. *Медстар Солошенс*. URL: <https://medstar.ua/ua/> (access date: 02.11.2023).
8. *Doctor Eleks*. URL: <https://doctor.eleks.com/ua/> (access date: 02.11.2023).
9. *Health 24*. URL: <https://h24.ua/> (access date: 02.11.2023).
10. Копняк К. *Оцінювання ефективності впровадження медичних інформаційних систем*. URL: <https://jeou.donnu.edu.ua/article/view/4784> (access date: 04.03.2024).

11. Gillis A. S. *What is a thin client (lean client)?* URL: <https://www.techtarget.com/searchnetworking/definition/thin-client> (access date: 09.04.2024).
12. Mărcuță C., *Beyond the Hype Practical Use Cases for MongoDB in Real-world Scenarios*. URL: <https://moldstud.com/articles/p-beyond-the-hype-practical-use-cases-for-mongodb-in-real-world-scenarios> (access date: 12.04.2024).
13. Amengual F. M. *The Dual Journey: Healthcare Interoperability and Modernization*. URL: <https://www.mongodb.com/blog/post/dual-journey-healthcare-interoperability-modernization> (access date: 12.04.2024).
14. *IoT Communication Protocols with measurements for NB-IoT - Expert Guide* URL: <https://avsystem.com/blog/iot/iot-communication-protocols-with-calculations> (access date: 14.06.2024).

References (transliterated)

1. Brown D. E. *Introduction to data mining for medical informatics*. Available at: <https://pubmed.ncbi.nlm.nih.gov/18194716/> (accessed: 12.09.2023).
2. Comak E., Polat K., Güneş S., Arslan A. *A new medical decision making system: least square support vector machine (LSSVM) with fuzzy weighting pre-processing*. Available at: https://journals.scholarsportal.info/details/09574174/v32i0002/409_anmdmsvmwfw.xml&sub=all (accessed: 18.09.2023).
3. *7 Ways Data Science Is Reshaping Healthcare*. Available at: <https://www.altexsoft.com/blog/datascience/7-ways-data-science-is-reshapinghealthcare/> (accessed: 20.09.2023).
4. *Survey: 8 in 10 Hospital Leaders Say Predictive Analytics is Important to Future Yet Only One-Third Use It*. Available at: <https://chimecentral.org/survey-8-10-hospital-leaders-say-predictive-analytics-important-future-yet-one-third-use/> (accessed: 20.09.2023).
5. *Helsi*. Available at: <https://helsi.me/about> (accessed: 02.11.2023).
6. *Medeür*. Available at: <https://e-life.com.ua/prod-tech/medejr-uk/> (accessed: 02.11.2023).
7. *Medstar Solutions*. Available at: <https://medstar.ua/ua/> (accessed: 02.11.2023).
8. *Doctor Eleks*. Available at: <https://doctor.eleks.com/ua/> (accessed: 02.11.2023).
9. *Health 24*. Available at: <https://h24.ua/> (accessed: 02.11.2023).
10. Копняк К. *Otsiniuvannia efektyvnosti vprovadzhennia medychnykh informatsiynykh system*. Available at: <https://jeou.donnu.edu.ua/article/view/4784> (accessed: 04.03.2024).
11. Gillis A. S. *What is a thin client (lean client)?* Available at: <https://www.techtarget.com/searchnetworking/definition/thin-client> (accessed: 09.04.2024).

12. Mărcuță C., *Beyond the Hype Practical Use Cases for MongoDB in Real-world Scenarios*. Available at: <https://moldstud.com/articles/p-beyond-the-hype-practical-use-cases-for-mongodb-in-real-world-scenarios> (accessed: 12.04.2024).
13. Amengual F. M. *The Dual Journey: Healthcare Interoperability and Modernization*. Available at: <https://www.mongodb.com/blog/post/dual-journey-healthcare-interoperability-modernization> (accessed: 12.04.2024).
14. *IoT Communication Protocols with measurements for NB-IoT - Expert Guide*. Available at: <https://avsystem.com/blog/iot/iot-communication-protocols-with-calculations> (accessed: 14.06.2024).

Received 05.11.2024

УДК 004.9

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>

І. В. ЛЮТЕНКО, кандидат технічних наук (PhD), доцент, Національний технічний університет «Харківський політехнічний інститут», доцент кафедри програмної інженерії та інтелектуальних технологій управління, м. Харків, Україна; e-mail: cherliv68@gmail.com; ORCID: <https://orcid.org/0000-0003-4357-1825>

В. В. ЯМБУРЕНКО, Національний технічний університет «Харківський політехнічний інститут», аспірант, м. Харків, Україна; e-mail: yamburenko@gmail.com; ORCID: <https://orcid.org/0009-0005-4797-8684>

А. А. ПАШНЄВ, кандидат технічних наук (PhD), старший науковий співробітник, Національний технічний університет «Харківський політехнічний інститут», доцент кафедри інформаційних систем та технологій, м. Харків, Україна; e-mail: pashniev@email.ua; ORCID: <https://orcid.org/0000-0002-9150-6108>

ПРОГРАММЕ РІШЕННЯ ДЛЯ ЗБОРУ ТА ОБРОБКИ В РЕАЛЬНОМУ ЧАСІ МЕДИЧНИХ ДАНИХ ХВОРИХ НА ЕПІЛЕПСІЮ

Стрімкий розвиток комп'ютерних технологій суттєво вплинув на численні сфери суспільного життя, включаючи охорону здоров'я. Можливість збору, обробки та візуалізації медичних даних у режимі реального часу набуває дедалі більшого значення, особливо в контексті управління хронічними захворюваннями, зокрема епілепсією. У статті представлено веб-додаток, створений для моніторингу показників здоров'я у реальному часі, який дозволяє медичним працівникам ефективно відстежувати дані пацієнтів. Система автоматизує процес збору даних з фітнес-трекерів, їх передачі через мобільний пристрій на сервер та подальшої візуалізації у веб-інтерфейсі. Архітектура системи побудована на основі моделі тонкого клієнта, що використовує Node.js для серверної логіки та React.js для розробки користувацького інтерфейсу, забезпечуючи при цьому високу масштабованість і адаптивність. Основні функціональні можливості системи включають візуалізацію даних у режимі реального часу, аналіз історичних тенденцій та можливість експорту ключових показників здоров'я для їх подальшого дослідження. Архітектурна модель системи передбачає модульний підхід із чітким розмежуванням відповідальностей між клієнтською частиною, серверною стороною та компонентами бази даних. У якості бази даних використовується MongoDB, що забезпечує високу гнучкість в управлінні великими обсягами медичних даних. Система була піддана комплексному тестуванню у два етапи. На першому етапі збір реальних даних продемонстрував середній час передачі, що не перевищує 112 мс, що відповідає вимогам функціонування у реальному часі. На другому етапі, під час стрес-тестування за участю до 100 одночасних користувачів, було зафіксовано середній час відповіді сервера на рівні 145,8 мс, а 95-й перцентиль часу відповіді склав 167,1 мс. Отримані результати підтверджують високу надійність системи та її готовність до впровадження у медичних установах. Подальший розвиток системи передбачає інтеграцію розширених механізмів сповіщення у реальному часі та додаткових показників здоров'я, таких як рівень насиченості киснем та рівень фізичної активності, що дозволить забезпечити комплексний моніторинг стану пацієнтів. Запропоноване рішення демонструє перспективи інтеграції сучасних веб-технологій в охорону здоров'я, сприяючи підвищенню ефективності медичних процесів та покращенню результатів лікування пацієнтів.

Ключові слова: моніторинг у реальному часі, обробка медичних даних, автоматизація, показники здоров'я, програмне забезпечення, архітектура системи, технології розробки.

Повні імена авторів / Author's full names

Автор 1 / Author 1: Копп Андрій Михайлович / Kopp Andrii Mykhailovych

Автор 2 / Author 2: Лютенко Ірина Вікторівна / Liutenko Iryna Viktorivna

Автор 3 / Author 3: Ямбуренко Віктор Вікторович / Yamburenko Viktor Viktorovich

Автор 4 / Author 4: Пашнєв Андрій Анатолійович / Pashniev Andrii Anatoliyovich

М. М. КОЗУЛЯ, кандидат технічних наук (PhD), доцент, доцент кафедри програмної інженерії та інтелектуальних технологій управління, Національний технічний університет «Харківський політехнічний інститут», м. Харків, Україна; e-mail: mariia.kozulia@kphi.edu.ua; ORCID: <https://orcid.org/0000-0002-4090-8481>

О. В. СОЛДАТКО, студент, Національний технічний університет «Харківський політехнічний інститут», м. Харків, Україна; e-mail: oleksandr.soldatko@cs.kphi.edu.ua; ORCID: <https://orcid.org/0009-0002-2802-8980>

ПРОЄКТУВАННЯ ІНФОРМАЦІЙНОЇ CRM-СИСТЕМИ У СФЕРІ СТОМАТОЛОГІЧНОГО АДМІНІСТРУВАННЯ

У статті досліджуються особливості оптимізації та автоматизації адміністративних послуг у стоматологічних клініках. На основі аналізу сучасних підходів до управління медичними закладами визначено основні напрями вдосконалення роботи адміністративного персоналу, включаючи впровадження CRM-систем, використання електронного документообігу та цифрових рішень для управління даними пацієнтів. Важливим аспектом є забезпечення високої якості адміністративних послуг за ключовими критеріями, такими як оперативність, точність та доступність інформації. Дослідження показують, що ефективність роботи клініки багато в чому залежить від організації процесів, що стосуються запису пацієнтів, управління розкладом лікарів, зберігання медичних даних та комунікації з клієнтами. Разом із тим, у процесі роботи виникають численні виклики. На рівні організації основними проблемами є помилки адміністративного персоналу та недостатня інтеграція сучасних технологій. На технічному рівні зустрічаються проблеми, пов'язані з затримками передачі даних, несумісністю програмного забезпечення або ризиками втрати інформації через технічні збої. Для розв'язання цих проблем впроваджено інтегровані системи управління та вдосконалено існуючі процеси. Водночас важливо дотримуватися балансу між складністю систем і їхньою функціональністю. Надмірно складні рішення можуть створювати бар'єри для персоналу, а надмірно прості – не забезпечувати необхідний рівень ефективності. У статті пропонується підхід до проєктування адміністративних послуг, який поєднує використання сучасних технологій, таких як CRM-системи, автоматизація обліку пацієнтів та аналіз робочих процесів. Це дозволяє підвищити ефективність управління та забезпечити якісне обслуговування пацієнтів.

Ключові слова: програмне забезпечення, картки клієнтів, проєктування системи, бази даних, стоматологічне адміністрування, CRM.

Вступ. Інформатика в стоматології представляє застосування комп'ютерів та інформатики для покращення стоматологічної практики, досліджень, освіти та управління. Останнім часом стоматологічні клініки все частіше використовують електронні комп'ютеризовані медичні записи пацієнтів, які замінили попереднє ручне управління [1]. Комп'ютеризоване управління клінікою складається з двох елементів: дослідження та практичного застосування в клініці. Елемент дослідження є наслідком аналізу та оцінки даних, створених у базі під час якісного застосування системи на практиці. У порівнянні з комп'ютеризованою системою, ручне управління клінікою має численні недоліки, хоча також існують певні переваги.

Переваги комп'ютеризованого управління включають швидкість, низьку ціну, ефективність, документовані рекомендації, мінімізацію зайнятості, одночасний доступ для кількох лікарів або користувачів, асинхронність [1]. Основними потенційними недоліками є необхідність навчання персоналу та технічні вимоги. Хоча перевага у швидкості цифрових систем є очевидною, у нашій країні немає клінічних досліджень, які б показували точний час, зекономлений завдяки застосуванню цих систем у порівнянні з традиційними ручними методами [2–4].

Економія часу є прямим фінансовим зиском для стоматологічної клініки. Оскільки використання комп'ютерів у стоматологічній практиці стало більш поширеним, але все ще далеким від розповсюдженої практики (у 2007 році лише 25 % стоматологів у Сполучених Штатах використовували комп'ютери для управління своїми клініками) [5]. Необхідно чітко ро-

зуміти переваги використання комп'ютеризованого управління з точки зору економії часу. Це можна легко розрахувати у вигляді робочих годин та заощаджених коштів, що стане додатковим стимулом для стоматологів застосовувати цю технологію у своїй роботі.

Отже, впровадження комп'ютеризованих систем управління у стоматологічних клініках є важливим кроком до підвищення ефективності та якості надання послуг. Аналіз економії часу, зменшення витрат на адміністративні процеси та забезпечення кращого доступу до інформації може слугувати основою для широкого використання таких систем. Окрім очевидних переваг для лікарів та пацієнтів, ці технології відкривають нові можливості для досліджень, аналізу даних та впровадження інновацій у сфері стоматології, що дозволить оптимізувати роботу клінік та покращити загальний рівень медичних послуг.

Мета та задачі дослідження. Метою даної статті є аналіз методів оптимізації адміністративних послуг у сфері стоматології, розробка моделі інформаційної системи для ефективного управління клінічними процесами та покращення обслуговування пацієнтів.

Для досягнення мети поставлені наступні задачі дослідження:

- 1) провести аналіз існуючих перешкод у впровадженні інформаційних систем для управління адміністративними послугами в стоматологічних клініках;
- 2) дослідити методи усунення цих перешкод і забезпечення ефективного використання інформаційних технологій;

© Козуля М. М., Солдатко О. В., 2024



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.



3) розробити модель інформаційної системи, яка враховуватиме специфіку адміністрування в сфері стоматологічних послуг та сприятиме їх оптимізації.

Аналіз та вирішення проблем стоматологічного адміністрування. У процесі ідентифікації інцидентів, порівняння та виділення тем було визначено шість основних категорій, кожен з яких представляє передбачуваний критичний фактор успіху (CSF): «зручність використання системи», «виникаюча поведінка», «управління змінами», «організація проекту», «навчання» та «аналіз вимог» [6]. Зручність використання інформаційних систем з точки зору кінцевих користувачів є загальним показником, який використовується для визначення успіху впровадження. Сфери занепокоєння, навколо яких учасники висловлювали схожі думки, включали: відсутність адаптації системи до потреб стоматологічного відділення; трудомісткі процеси; надмірну кількість кліків; додаткове навантаження на клініцистів; незручність у користуванні; період адаптації користувачів; опір користувачів; відсутність довіри до системи та проблеми з доступом до даних.

З точки зору власника продукту та постачальника, аналіз вимог до системи був визначений як критичний фактор успіху та водночас виклик, із труднощами, які спостерігалися на етапах визначення цілей перед впровадженням, ігнорування потреб кінцевих користувачів та невідповідності у вимогах [6].

Навчання відіграє фундаментальну роль у впровадженні інформаційних систем, оскільки спосіб, у який люди вперше знайомляться з новою системою та вчать її використовувати, впливає на їхню залученість, мотивацію та готовність до змін.

Для успішного впровадження інформаційних технологій у стоматологічних клініках важливо враховувати технічні, організаційні й адміністративні аспекти. Передусім, адаптація системи до потреб клініки має базуватися на залученні співробітників, які працюватимуть із нею щодня. Це дозволить створити програмне забезпечення, що відповідає реальним викликам та задачам.

Особливої уваги потребує підготовка персоналу. Проведення тренінгів і розробка простих інструкцій допоможуть працівникам освоїти нові інструменти без зайвих ускладнень. При цьому навчання має враховувати різний рівень володіння технологіями, щоб усі співробітники почувалися впевнено.

Також необхідно чітко координувати сам процес впровадження. Планування етапів реалізації, розподіл ролей серед учасників та регулярні зустрічі для обговорення прогресу дозволять уникнути непорозумінь. Постійна комунікація між адміністрацією, IT-спеціалістами та медичним персоналом значно підвищить ефективність проекту.

Додатково потрібно враховувати фінансову складову. Можна розглядати можливість залучення грантів або партнерів, готових інвестувати у розвиток клініки. Це полегшить впровадження та дозволить використовувати сучасні рішення [7].

Для забезпечення безпеки даних необхідно впровадити шифрування, регулярно створювати резервні

копії та навчати працівників основам захисту інформації. Це допоможе уникнути витоків даних та зберегти конфіденційність.

Нарешті, важливо дотримуватись законодавчих норм і стандартів у сфері охорони здоров'я. Співпраця з регуляторними органами дозволить уникнути штрафів і забезпечити повну відповідність нових рішень встановленим вимогам.

Серед існуючих рішень є наступні інформаційні системи:

1. Система управління стоматологічними клініками Dentrix, яка надає можливості для ведення карток пацієнтів, фінансового обліку та інтеграції з іншими сервісами, також має недоліки: висока вартість ліцензії, складність у навчанні персоналу [8].
2. Відкрите програмне забезпечення Open Dental з широким функціоналом та можливістю адаптації, однак з обмеженою підтримкою інтеграції з іншими платформами, ускладнює роботу великих клінік [9].
3. Платформа Eaglesoft з потужним функціоналом для управління записами пацієнтів та фінансами має недоліки у вигляді частих технічних збоїв [10].
4. CRM-система Salesforce з централізованим управлінням, але залежить від інтернет-з'єднання [11].

Таким чином, необхідне впровадження комплексного підходу до впровадження технологій, що дозволить не лише оптимізувати роботу клініки, але й суттєво покращити якість послуг для пацієнтів.

Модель системи для оптимізації адміністрування стоматологічних послуг. Система має забезпечувати маніпулювання даними у картках клієнтів [12]. Основними функціями маніпулювання даними є створення, доповнення, зміна й видалення карток клієнтів, а також пошук та перегляд. Усі дані карток клієнтів повинні зберігатися у базі даних. Дані, що повинні зберігатися: ПІБ, номер телефону, надані послуги, нумерація зубів, дата прийому, ціна послуг, стан послуг.

Також у програмному забезпеченні потрібно створити календар, де адміністратор зможе записувати пацієнтів на прийом, а потім переглядати інформацію з якого та по який час буде вестися прийом. Потрібно також реалізація видалення даних з календаря.

У системі не повинен бути перенавантажений інтерфейс, він повинен бути зрозумілим звичайному не підготовленому користувачу, це зменшить час навчання персоналу, що спростить впровадження.

Вимоги під час обговорення із замовником, командою, для власного аналізу, під час розробки бажано представляти у графічному вигляді. Зручним для цього є діаграма вимог. Діаграма вимог дозволяє візуально моделювати вимоги, показуючи, як вимоги пов'язані між собою і як вони пов'язані з іншими елементами моделі. На рис. 1 зображена діаграма вимог для розроблюваної системи, що оптимізує адміністрування стоматологічних послуг.

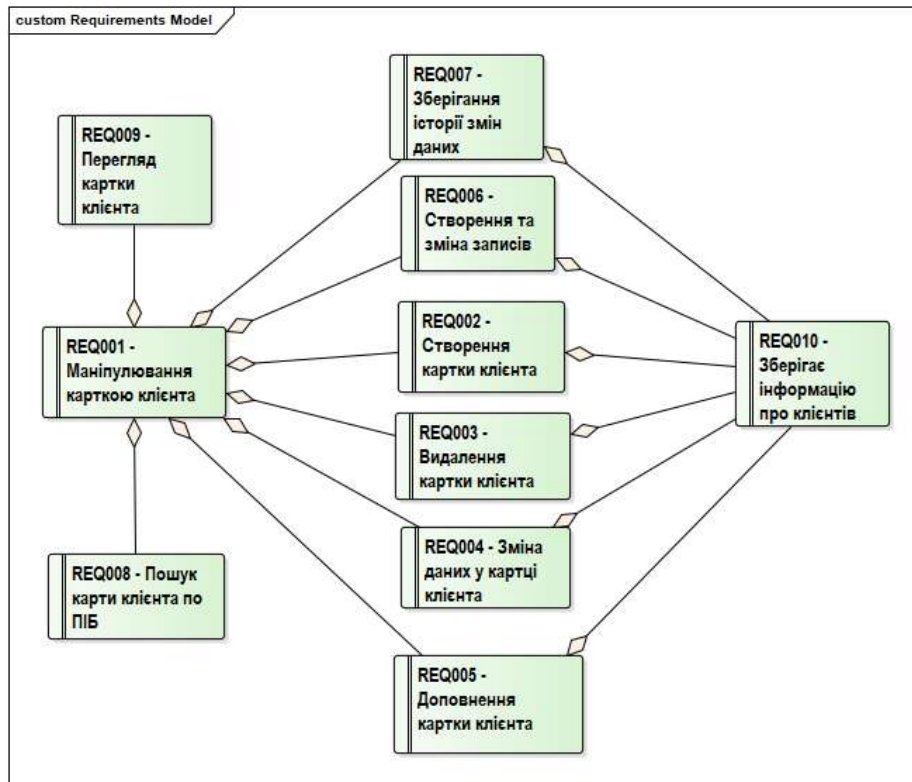


Рис. 1. Діаграма вимог для системи, що оптимізує адміністрування стоматологічних послуг

Висновки. Використання сучасних інформаційних систем у стоматологічних клініках стає ключовим фактором у підвищенні ефективності їхньої роботи. Ці технології дозволяють автоматизувати значну частину адміністративних процесів, що зменшує навантаження на персонал, економить час і забезпечує високий рівень обслуговування пацієнтів. Серед основних переваг таких систем можна виділити швидкість і точність обробки даних, зручний доступ до інформації, а також можливість інтеграції з іншими внутрішніми процесами клініки.

Однак упровадження інформаційних систем супроводжується низкою труднощів. Часто виникає необхідність у додатковому навчанні персоналу, технічній адаптації системи до специфіки медичного закладу та розв'язанні питань, пов'язаних із кібербезпекою. Крім того, нововведення можуть стикатися з опором з боку співробітників, які звикли до традиційних методів роботи. Для успішного впровадження таких систем важливо забезпечити якісну підготовку персоналу, узгодити нові рішення з існуючими процесами та ретельно продумати питання захисту персональних даних.

Правильно сплановане впровадження інформаційних систем значно покращує якість управління клінікою. Це досягається завдяки ефективному обміну інформацією між усіма учасниками процесу, чіткому розподілу обов'язків і орієнтації на потреби пацієнтів. Такий підхід сприяє покращенню медичних послуг, підвищенню довіри пацієнтів до клініки й дозволяє закладу залишатися конкурентоспроможним на ринку.

Отже, сучасні інформаційні технології відкривають нові перспективи для стоматологічної галузі.

Їхнє впровадження дає змогу не лише автоматизувати рутинні процеси, а й сприяє стратегічному розвитку клінік, формуванню позитивного іміджу та вдосконаленню системи управління медичними послугами.

Список використаної літератури

1. Branko Mihailović, Zoran Lazić, Dragan Mladenović, Goran Tošić, Aleksandar Janković, Miloš Duka. The advantages of speed of computerized management of a dental clinic. *ACTAFAC MED NAISS* 2009. Vol. 26, no. 4. P. 203–209.
2. Gillette J. Striving for excellence with evidence-based dentistry. *Journal of Evidence Based Dental Practice*. 2009 Vol.9, no. 3. P. 125–128.
3. Gillette J, Matthews JD, Frantsve-Hawley J, Weyant RJ. The benefits of evidence-based dentistry for the private dental office. *Dental Clinics of North America*. 2009. Vol. 53, no. 1. P. 33–45. DOI: 10.1016/j.cden.2008.09.002.
4. Aravamudhan K, Frantsve-Hawley J. American Dental Association's Resources to Support Evidence-Based Dentistry *Journal of Evidence Based Dental Practice*. 2009. Vol. 9, no. 3. P. 139–144. DOI: 10.1016/j.jebdp.2009.06.011.
5. Schleyer T, Spallek H, Hernández P. A qualitative investigation of the content of dental paper-based and computer-based patient record formats *Journal of the American Medical Informatics Association*. 2007. Vol. 14, no. 4. P.515–526. DOI:10.1197/jamia.M2335.
6. Yusof H S., Jorge T. M. Perceived critical success factors of electronic health record system implementation in a dental clinic context: An organisational management perspective. *International Journal of Medical Informatics*. 2017. Vol. 107. P. 88–100. DOI:10.1016/j.ijmedinf.2017.08.007.
7. Viktor Arvidsson, Jonny Holmström, Kalle Lyytinen. Information systems use as strategy practice: A multi-dimensional view of strategic information system implementation and use. *The Journal of Strategic Information Systems*. 2014. Vol. 23, is. 1. P. 45–61.
8. Thankam P., Monaco V., Bindu H., Schleyer T. A usability evaluation of four commercial dental computer-based patient record systems. *The Journal of the American Dental Association*. 2008. Vol. 139. No. 12. P. 1632–1642.

- 9 Sihombing, D. J. C. (2024). Developing Customer Relationship Management Application for Dental Clinic Using Agile Approach to Enhance Efficiency and Customer Satisfaction. *Jurnal Ekonomi*. 2024. Vol. 13, no. 01. P. 2558–2569.
- 10 Raquel R. *How to Use EagleSoft: a clinical dental software*. Instructional Design and Technology. 2024. 6 p. URL: <http://raqueleportfolio.pbworks.com/f/EDTC+6321-Revised+IU4-Instructional+Strategy.pdf>. (дата звернення: 12.11.2024).
- 11 Борисюк, І. Ю., Чаварга, М. І., & Павленко, І. А. Роль Salesforce у підвищенні якості медичної освіти України: впровадження системи в загальну медицину. *Педагогічна Академія: наукові записки*. 2024. № 7. DOI: 10.5281/zenodo.12582453.
- 12 Kharchenko B. *Analyzing and evaluating existing dental practice management software: A comprehensive study to identify gaps and opportunities for improvement*. Universitat Politècnica de València. 2023. 88 p. URL: <http://hdl.handle.net/10251/197987>. (дата звернення 12.11.2024).
- 6 Yusof H S., Jorge T. M. Perceived critical success factors of electronic health record system implementation in a dental clinic context: An organisational management perspective. *International Journal of Medical Informatics*. 2017, vol. 107, pp. 88–100. DOI:10.1016/j.ijmedinf.2017.08.007
- 7 Viktor Arvidsson, Jonny Holmström, Kalle Lyytinen. Information systems use as strategy practice: A multi-dimensional view of strategic information system implementation and use. *The Journal of Strategic Information Systems*. 2014, vol. 23, is. 1, pp. 45–61.
- 8 Thankam P., Monaco V., Bindu H., Schleyer T. A usability evaluation of four commercial dental computer-based patient record systems. *The Journal of the American Dental Association*. 2008, vol. 139, no. 12, pp. 1632–1642.
- 9 Sihombing, D. J. C. (2024). Developing Customer Relationship Management Application for Dental Clinic Using Agile Approach to Enhance Efficiency and Customer Satisfaction. *Jurnal Ekonomi*. 2024, vol. 13, no. 01, pp. 2558–2569.

References (transliterated)

1. Branko Mihailović, Zoran Lazić, Dragan Mladenović, Goran Tošić, Aleksandar Janković, Miloš Duka. The advantages of speed of computerized management of a dental clinic. *ACTAFAC MED NAISS* 2009, vol. 26, no. 4, pp. 203–209.
2. Gillette J. Striving for excellence with evidence-based dentistry. *Journal of Evidence Based Dental Practice*. 2009, vol. 9, no. 3, pp. 125–128.
3. Gillette J, Matthews JD, Frantsve-Hawley J, Weyant RJ. The benefits of evidence-based dentistry for the private dental office. *Dental Clinics of North America*. 2009, vol. 53, no. 1, pp. 33–45. DOI: 10.1016/j.cden.2008.09.002.
4. Aravamudhan K, Frantsve-Hawley J. *Journal of Evidence Based Dental Practice*. 2009, vol. 9, no. 3, pp. 139–144. DOI: 10.1016/j.jebdp.2009.06.011.
5. Schleyer T, Spallek H, Hernández P. A qualitative investigation of the content of dental paper-based and computer-based patient record formats *Journal of the American Medical Informatics Association*. 2007, vol. 14, no. 4. pp.515–526. DOI: 10.1197/jamia.M2335.
- 10 Raquel R. *How to Use EagleSoft: a clinical dental software*. Instructional Design and Technology, 2024. 6 p. URL: <http://raqueleportfolio.pbworks.com/f/EDTC+6321-Revised+IU4-Instructional+Strategy.pdf>. (accessed: 12.11.2024).
- 11 Borysiuk, I. Y., Chavarga, M. I., & Pavlenko, I. A. Role of Salesforce in improving the quality of medical education in Ukraine: Implementation of the system in general medicine. *Pedagogical Academy: Scientific Notes*. 2024, no. 7. DOI: 10.5281/zenodo.12582453.
- 12 Kharchenko B. *Analyzing and evaluating existing dental practice management software: A comprehensive study to identify gaps and opportunities for improvement*. Universitat Politècnica de València, 2023. 88 p. Available at: <http://hdl.handle.net/10251/197987> (accessed 12.12.2024).

Надійшла (received) 14.12.2024

UDC 004.75

M. M. KOZULIA, Candidate of Technical Sciences (PhD), Docent, National Technical University "Kharkiv Polytechnic Institute", Associate Professor at the Department of SEMIT, Kharkiv, Ukraine; e-mail: mariia.kozulia@kphi.edu.ua; ORCID: <https://orcid.org/0000-0002-4090-8481>

O. V. SOLDATKO, National Technical University "Kharkiv Polytechnic Institute", Student, Kharkiv, Ukraine; e-mail: oleksandr.soldatko@cs.kphi.edu.ua; ORCID: <https://orcid.org/0009-0002-2802-8980>

DESIGN OF CRM INFORMATION SYSTEM IN THE FIELD OF DENTAL ADMINISTRATION

The article examines the features of optimization and automation of administrative services in dental clinics. Based on the analysis of modern approaches to the management of medical institutions, the main directions for improving the work of administrative personnel have been identified, including the implementation of CRM systems, the use of electronic document management and digital solutions for managing patient data. An important aspect is ensuring high quality administrative services according to key criteria, such as efficiency, accuracy and availability of information. Studies show that the efficiency of the clinic largely depends on the organization of processes related to patient registration, doctor's schedule management, medical data storage and communication with clients. However, numerous challenges arise in the process of work. At the organizational level, the main problems are errors of administrative personnel and insufficient integration of modern technologies. At the technical level, there are problems associated with data transmission delays, software incompatibility or risks of information loss due to technical failures. To solve these problems, integrated management systems have been implemented and existing processes have been improved. At the same time, it is important to maintain a balance between the complexity of the systems and their functionality. Overly complex solutions can create barriers for staff, and overly simple ones may not provide the necessary level of efficiency. The article proposes an approach to the design of administrative services that combines the use of modern technologies, such as CRM systems, patient registration automation and workflow analysis. This allows you to increase management efficiency and ensure high-quality patient care.

Keywords: software, customer cards, system design, databases, dental administration, CRM.

Повні імена авторів / Author's full names

Автор 1 / Author 1: Козуля Марія Михайлівна / Kozulia Mariia Mykhailivna

Автор 2 / Author 2: Солдатко Олександр Володимирович / Soldatko Oleksandr Volodymyrovych

МАТЕМАТИЧНЕ І КОМП'ЮТЕРНЕ МОДЕЛЮВАННЯ

MATHEMATICAL AND COMPUTER MODELING

DOI: 10.20998/2079-0023.2024.02.07

УДК 004.9

К. Е. ПЕТРОВ, доктор технічних наук, професор, Харківський національний університет радіоелектроніки, завідувач кафедри інформаційних управляючих систем, м. Харків, Україна; e mail: kostiantyn.petrov@nure.ua; ORCID: <https://orcid.org/0000-0003-1973-711X>

Т. В. ЧАЛИЙ, Харківський національний університет радіоелектроніки, аспірант кафедри інформаційних управляючих систем, м. Харків, Україна; e mail: taras.chalyi@nure.ua; ORCID: <https://orcid.org/0009-0002-3796-5810>

СИТУАЦІЙНА МОДЕЛЬ МЕДИЧНОГО БІЗНЕС-ПРОЦЕСУ

Предметом дослідження є медичні бізнес-процеси. Мета роботи полягає у розробці ситуаційного підходу до моделювання узагальненого медичного бізнес-процесу. Досягнення мети дає можливість забезпечити заданий рівень медичних послуг при наявних ресурсних обмеженнях медичного бізнес-процесу. Для досягнення мети вирішуються такі задачі: структуризація медичних бізнес-процесів з урахуванням відмінностей лікування в клініці, догляду, та повного процесу лікування пацієнтів; розробка ситуаційної моделі медичного бізнес-процесу. Виконано структуризацію медичних бізнес-процесів. Показано, що такі процеси складаються із клінічного шляху, шляху догляду та шляху лікування захворювання. Клінічний шлях реалізується в межах одного медичного закладу. Шлях догляду визначає комплексний опис послідовності догляду та лікування. Повний шлях лікування захворювання інтегрує різні заклади охорони здоров'я. Шлях лікування включає клінічний шлях та шлях догляду в якості підпроцесів. Запропоновано ситуаційну модель узагальненого медичного бізнес-процесу, що складається із послідовності ситуацій. Кожна ситуація передбачає вибір з урахуванням ресурсних, фінансових та темпоральних обмежень, та подальше виконання послідовності дій медичного бізнес-процесу. Модель, у відповідності до представленої структуризації медичних бізнес-процесів, на верхньому рівні представлення містить фази первинної медичної допомоги, амбулаторного лікування, клінічного шляху та виписки й реабілітації. Модель створює умови для вибору індивідуалізованого процесу лікування з урахуванням потреб пацієнта та темпоральних і ресурсних обмежень. Типова послідовність дій медичного бізнес-процесу визначається на рівні множини ситуацій, а вибір альтернатив здійснюється в межах ситуації. Послідовність ситуацій задає загальний стандарт догляду або лікування, а індивідуалізація виконується в рамках окремих ситуацій з урахуванням потреб пацієнта та фінансових обмежень.

Ключові слова: медичний бізнес-процес, ситуаційна модель, клінічний шлях, шлях догляду, шлях лікування захворювання, ресурсні обмеження, темпоральні обмеження, індивідуалізоване лікування.

Вступ. Бізнес-процес представляє собою послідовність взаємопов'язаних процедур, що створюють продукт (послугу), яка має цінність для споживача [1].

Управління бізнес-процесами, або процесний підхід до управління включає аналіз діяльності організації, побудову та імплементацію моделей бізнес-процесів, а також моніторинг процесів та подальше удосконалення процесних моделей [2]. Моделі бізнес-процесів відображають взаємозв'язки між процедурами (діями) процесу з урахуванням наявних ресурсів.

Медичні бізнес-процеси (МБП) інтегрують багатаспектні дії з надання медичної допомоги, починаючи від реєстрації пацієнта та до виписки й подальшого спостереження за пацієнтом [3]. МБП включають діагностичні, лікувальні, реабілітаційні та адміністративні процедури, що орієнтовані на забезпечення якісної медичної допомоги. Особливість медичних бізнес-процесів полягає у необхідності врахувати індивідуальні особливості кожного пацієнта, що потребує їх адаптації у процесі лікування [4].

Побудова МБП з позицій якості послуг орієнтована на зменшення медичних помилок, скорочення

часу очікування послуг пацієнтами, підтримку координації між медичними закладами та підвищення рівня задоволеності пацієнтів. Моделювання МБП у фінансовому аспекті орієнтовано на вирішення задачі оптимізації використання ресурсів закладу охорони здоров'я, зменшення операційних витрат, підвищення ефективності роботи персоналу та покращення фінансових показників в цілому [5].

Таким чином, побудова та подальше використання моделей медичних бізнес-процесів дає можливість підвищити ефективність надання медичних послуг, узгодивши вимоги користувача та можливості медичного закладу.

Аналіз останніх досліджень і публікацій.

Сучасні підходи до моделювання бізнес-процесів орієнтовані на формування набору можливих послідовностей дій, що забезпечують досягнення результатів процесу. Послідовність дій бізнес-процесу зазвичай моделюється як набір активностей (процедур) та підпроцесів. Життєвий цикл процесу включає визначення цілей, планування, виконання, тестування, впровадження та моніторинг [6]. Для візуалізації процесів ви-

© К.Е. Петров, Т.В. Чалий, 2024



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.



користовується нотація BPMN, яка представляє собою стандартизовану мову для опису складних процесів [7].

При моделюванні послідовностей процедур бізнес-процесів зазвичай виділяють базовий шлях, альтернативні послідовності дій та варіанти винятків. Базовий шлях відображає типову стандартизовану послідовність дій, яка буде реалізована у більшості імплементацій бізнес-процесу. Альтернативні шляхи описують типові відхилення від стандартного процесу. Варіанти винятків описують разові відхилення [8].

Для оцінки ситуацій із альтернативними сценаріями та їх потенційних результатів при побудові моделей бізнес-процесів використовуються контрфакти, тобто альтернативні факти, які можуть бути розглянуті як вхідні та проміжні дані бізнес-процесу. Контрфактичний аналіз дає можливість виявити причини вибору послідовності дій процесу [9].

Медичні бізнес-процеси забезпечують стандартизацію процесів лікування межах одного закладу, стандартизацію догляду та стандартизацію процесу лікування, що відбувається у різних медичних закладах [2, 5, 10].

Проведений в [2] аналіз показав важливість вдосконалення МБП з урахуванням ресурсних та фінансових обмежень.

Декларативний підхід до моделювання МБП з використанням технології process mining, запропонований в [1], показав можливість адаптації процесних моделей з урахуванням темпоральних та ресурсних обмежень.

Таким чином, наявне протиріччя між потребою забезпечення якісної охорони здоров'я з урахуванням фінансових та ресурсних обмежень та можливостями існуючих методів з побудови жорстко визначених послідовностей дій МБП потребує розробки гнучких ситуаційних моделей таких процесів з можливостями оцінки та динамічної зміни послідовності дій на основі контрфактичного аналізу. Зазначене свідчить про актуальність теми даного дослідження.

Мета та задачі дослідження. Мета роботи полягає у розробці ситуаційного підходу до моделювання узагальненого медичного бізнес-процесу з тим, щоб врахувати можливості вибору альтернативних рішень при клінічному лікуванні, догляді та у процесі лікування захворювання в цілому.

Вирішення цієї задачі орієнтовано на вибір раціонального рішення, що дає можливість забезпечити заданий рівень медичних послуг при наявних ресурсних обмеженнях медичного бізнес-процесу.

Для досягнення мети вирішуються наступні задачі:

- структуризація медичних бізнес-процесів з урахуванням відмінностей лікування в клініці, догляду та повного процесу лікування пацієнтів;
- розробка ситуаційної моделі медичного бізнес-процесу.

Структуризація медичних бізнес-процесів.

Бізнес-процеси у сфері охорони здоров'я включають: клінічний шлях (clinical pathway); шлях догляду (care pathway); шлях лікування захворювання (transmural pathway). Ці бізнес-процеси реалізують різні за

масштабом підходи до надання медичної допомоги. Порівняльна характеристика цих бізнес-процесів представлена в табл. 1.

Клінічний шлях визначає та реалізує мультидисциплінарний план догляду для конкретних клінічних проблем в межах одного медичного закладу. Цей процес містить стандартизовану послідовність клінічних процедур для конкретного захворювання та є короткотривалим. Клінічний шлях реалізується в межах одного медичного закладу, такого як лікарня або клініка.

Шлях догляду визначає комплексне рішення щодо послідовності догляду та організації процесів медичного огляду, включаючи амбулаторне лікування, реабілітацію та подальший нагляд.

Шлях догляду є довгостроковим процесом та реалізується в ряді закладів, включаючи лікарні, амбулаторні заклади та заклади реабілітації.

Таблиця 1 – Відмінності клінічного шляху, шляху догляду та шляху лікування захворювання

Характеристика	Опис
1. Заклади, де виконується МБП	1) клінічний шлях – в межах одного медичного закладу, наприклад лікарня, клініка; 2) шлях догляду – в межах декількох медичних закладів – лікарня, амбулаторні заклади, заклади реабілітації; 3) шлях лікування захворювання – медичні заклади та домашнє лікування: лікарня, амбулаторні заклади, заклади реабілітації, первинна медична допомога, домашній догляд.
2. Охоплення на рівні підпроцесів	1) клінічний шлях – конкретний клінічний епізод; 2) шлях догляду – повний процес догляду за пацієнтом; 3) шлях лікування захворювання – міжорганізаційний процес, що інтегрує догляд між різними рівнями медичної допомоги.
4. Тривалість	1) клінічний шлях – короткострокова; 2) шлях догляду – довгострокова; 3) шлях лікування захворювання – довгострокова.
5. Загальне призначення	1) клінічний шлях – стандартизація лікування в межах медичного закладу; 2) шлях догляду – підтримка координації та комунікації впродовж медичного обслуговування; 3) шлях лікування захворювання – інтеграція догляду між різними закладами охорони здоров'я.

Шлях лікування захворювання представляє собою інтегрований шлях догляду, що охоплює різні заклади охорони здоров'я. Цей підхід забезпечує безперервність догляду між різними рівнями медичної допомоги, включаючи первинну, вторинну та третинну ланки. Шлях лікування є довгостроковим процесом та харак-

теризується організацією міждисциплінарної співпраці. Він реалізується в таких закладах, як лікарні, амбулаторні заклади, заклади реабілітації, ланки первинної медичної допомоги, а також включає домашній догляд.

Ситуаційна модель узагальненого медичного бізнес-процесу.

Схему послідовності процедур узагальненого медичного бізнес-процесу наведено на рис. 1. Такий процес включає шлях лікування захворювання, що містить шлях догляду в якості підпроцесу.

В свою чергу, шлях догляду в якості підпроцесу містить клінічний шлях. Останній включає передопераційний, операційний на післяопераційний етапи.

Відповідно до наведеної структури МБП, процес на верхньому рівні представлення містить 4 ситуації: первинна медична допомога (1), амбулаторне лікування (2), клінічний шлях (3) та реабілітація (4).

Ситуаційна модель B узагальненого МБП на першому (верхньому) рівні представлення складається з упорядкованого набору ситуацій S_i^1 , що мають задовольняти обмеженням Φ , а також множини даних про пацієнта до початку лікування X та після лікування Y :

$$B = \{X, \langle S_1^1, S_2^1, S_3^1, S_4^1 | \Phi \rangle, Y\}. \quad (1)$$

Обмеження Φ включають темпоральні T , ресурсні R та фінансові F складові:

$$\Phi = \{T, R, F\}. \quad (2)$$

При деталізації МБП кожна з наведених ситуацій може бути представлена у вигляді упорядкованого набору ситуацій нижнього рівня $\{S_i^2\}$.

Ключова відмінність такого ситуаційного представлення від традиційного workflow – опису полягає в тому, що типова послідовність дій визначається на рівні ситуацій, а вибір альтернатив здійснюється в межах ситуації. Тоді послідовність ситуацій задає загальний стандарт догляду або лікування, а індивідуалізація виконується на рівні ситуацій з урахуванням наявного досвіду, потреб пацієнта та фінансових обмежень. Кожна ситуація може бути представлена як набір послідовностей D_i дій $d_{i,k}$ з оцінкою потреб в ресурсах W_i за умови темпоральних обмежень T_i :

$$S_i^1 = \{D_i : (\forall l) \exists W_l | T_l\}. \quad (3)$$

Розглянемо приклад використання даної моделі для фрагменту стоматологічного МБП. В даному випадку вхідні дані будуть представлені характеристиками стану зубів пацієнта, що потребують лікування або протезування. Ситуація містить альтернативи D_i , пов'язані із протезуванням або лікуванням різної вартості, наприклад з використанням технологій різних постачальників (коронки, імпланти тощо). Кожна технологія передбачає свою послідовність дій в рамках визначеної ситуації. Для кожної технології може бути надано оцінку вартості та часу. Вибір варіанту ліку-

вання здійснюється на основі оцінки вартості з урахуванням обмежень на час лікування, які задовольняють пацієнта та які може реалізувати клініка.

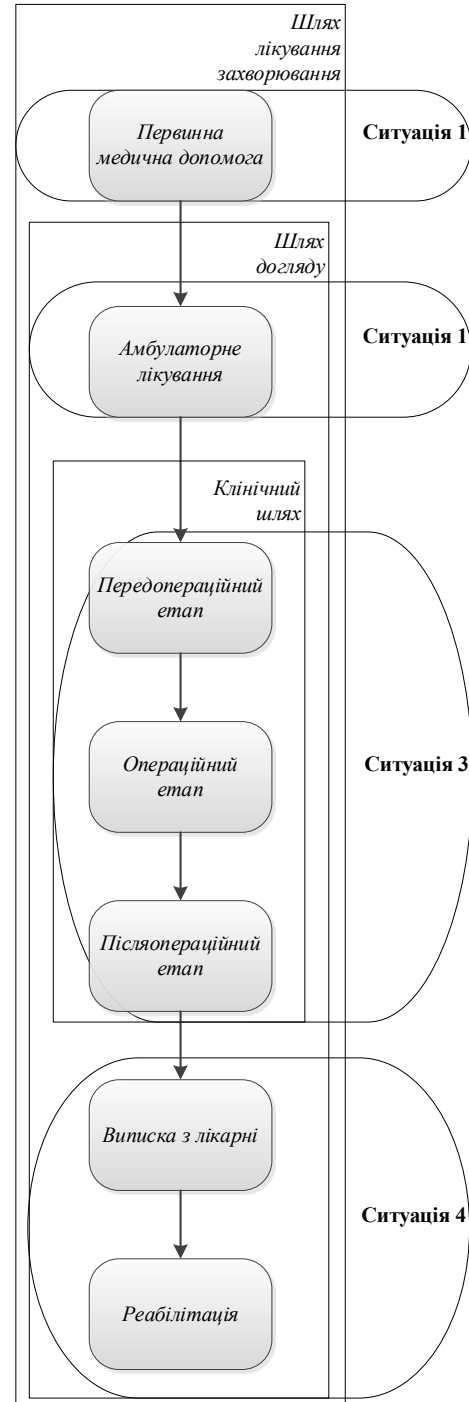


Рис. 1. Ситуаційне представлення та послідовність процедур узагальненого медичного бізнес-процесу

Висновки. Виконано структурування медичних бізнес-процесів. Показано, що такі процеси складаються із клінічного шляху, шляху догляду та шляху лікування захворювання, причому останній включає клінічний шлях та шляху догляду як підпроцеси.

Запропоновано ситуаційну модель узагальненого медичного бізнес-процесу, що складається із послі-

довності ситуацій. Кожна ситуація передбачає вибір з урахуванням ресурсних, фінансових та темпоральних обмежень, та подальше виконання послідовності дій медичного бізнес-процесу. Запропонована модель, у відповідності до представленої структуризації медичних бізнес-процесів, містить фази первинної медичної допомоги, амбулаторного лікування, клінічного шляху та виписки й реабілітації. Модель створює умови для вибору індивідуалізованого процесу лікування з урахуванням потреб пацієнта та з урахуванням темпоральних та ресурсних обмежень закладу охорони здоров'я.

Список використаної літератури

1. Patharkar A., C., Cai F., B., Zardiniy A., Al-Hindawi F., Wu T. Predictive modeling of biomedical temporal data in healthcare applications: review and future directions. *Frontiers in Physiology 15. Sec. Computational Physiology and Medicine*, 2024. Vol. 15. DOI: 10.3389/fphys.2024.1386760.
2. Van der Aalst W. Business Process Management: A Comprehensive Survey. *ISRN Software Engineering*, 2013. P. 1-37. DOI:1155/2013/507984.
3. Rasoli N., Jolai F. Mehdi M., Tehranian A. BPM application in clinical process improvement: a women 'hospital case study. *Business Management Journal*, 2024. Vol. 30(3). P. 986-1011. DOI: 10.1108/BPMJ-07-2023-0560.
4. Petrocelli John V. Pitfalls of Counterfactual Thinking in Medical Practice: Preventing Errors by Using More Functional Reference Points. *Journal of Public Health Research*, 2013. Vol. 2:e24. P. 136-143.
5. De Ramón Fernández A., Ruiz Fernández D., Sabuco García Y. Business Process Management for optimizing clinical processes: A systematic literature review. *Health Informatics Journal*, 2020. Vol. 26(1). P. 1305-1320. DOI: 10.1177/1460458219877092.
6. Dumas M., La Rosa M., Mendling J., Reijers H.A. *Fundamentals of Business Process Management*. Berlin: Springer, 2018. 527 p.
7. Chinosi M., Trombetta A. BPMN: An introduction to the standard. *Computer Standards & Interfaces*, 2012. Vol. 34. P. 124-134.

8. Van der Aalst W. *Process Mining: Data Science in Action*. 2nd ed. Berlin: Springer, 2016. 467 p.
9. Pearl J., Mackenzie D. *The Book of Why: The New Science of Cause and Effect*. New York: Basic Books, 2018. 432 p.
10. Byscher A., Kugler J. The effectiveness of clinical pathways in inpatient settings - an umbrella review. *Journal of Public Health*, 2024. P. 1-15. DOI: 10.1007/s10389-024-02227-w.

References (transliterated)

1. Patharkar A., C., Cai F., B., Zardiniy A., Al-Hindawi F., Wu T. Predictive modeling of biomedical temporal data in healthcare applications: review and future directions. *Frontiers in Physiology 15. Sec. Computational Physiology and Medicine*. 2024, vol. 15. DOI: 10.3389/fphys.2024.1386760.
2. Van der Aalst W. Business Process Management: A Comprehensive Survey. *ISRN Software Engineering*. 2013, pp. 1-37. DOI:1155/2013/507984.
3. Rasoli N., Jolai F. Mehdi M., Tehranian A. BPM application in clinical process improvement: a women 'hospital case study. *Business Management Journal*. 2024, vol. 30(3), pp. 986-1011. DOI: 10.1108/BPMJ-07-2023-0560.
4. Petrocelli John V. Pitfalls of Counterfactual Thinking in Medical Practice: Preventing Errors by Using More Functional Reference Points. *Journal of Public Health Research*. 2013, vol. 2:e24, pp. 136-143.
5. De Ramón Fernández A., Ruiz Fernández D., Sabuco García Y. Business Process Management for optimizing clinical processes: A systematic literature review. *Health Informatics Journal*. 2020, vol. 26(1), pp. 1305-1320. DOI: 10.1177/1460458219877092.
6. Dumas M., La Rosa M., Mendling J., Reijers H.A. *Fundamentals of Business Process Management*. Berlin: Springer, 2018. 527 p.
7. Chinosi M., Trombetta A. BPMN: An introduction to the standard. *Computer Standards & Interfaces*. 2012, vol. 34, pp. 124-134.
8. Van der Aalst W. *Process Mining: Data Science in Action*. 2nd ed. Berlin: Springer. 2016. 467 p.
9. Pearl J., Mackenzie D. *The Book of Why: The New Science of Cause and Effect*. New York: Basic Books. 2018. 432 p.
10. Byscher A., Kugler J. The effectiveness of clinical pathways in inpatient settings - an umbrella review. *Journal of Public Health*. 2024, pp. 1-15. DOI: 10.1007/s10389-024-02227-w.

Надійшла (received) 31.10.2024

UDC 004.9

K. E. PETROV, Doctor of Technical Sciences, Professor, Kharkiv National University of Radio Electronics, Head of the Department of Information Control Systems, Kharkiv, Ukraine; e-mail: kostiantyn.petrov@nure.ua; ORCID: <https://orcid.org/0000-0003-1973-711X>

T. V. CHALYI, Kharkiv National University of Radio Electronics, PhD student of the Department of Information Control Systems, Kharkiv, Ukraine; e mail: taras.chalyi@nure.ua; ORCID: <https://orcid.org/0009-0002-3796-5810>

SITUATIONAL MODEL OF A MEDICAL BUSINESS PROCESS

The subject of the research is medical business processes. The aim of the work is to develop a situational approach to modeling a generalized medical business process. Achieving this goal makes it possible to provide a specified level of medical services within the existing resource constraints of the medical business process. To achieve the goal, the following tasks are solved: structuring medical business processes considering the differences in treatment in the clinic, care, and the full process of patient treatment; development of a situational model of the medical business process. The structuring of medical business processes has been carried out. It is shown that such processes consist of a clinical pathway, care pathway, and disease treatment pathway. The clinical pathway is implemented within a single medical institution. The care pathway defines a comprehensive description of the sequence of care and treatment. The complete disease treatment pathway integrates various healthcare institutions. The treatment pathway includes the clinical pathway and care pathway as subprocesses. A situational model of a generalized medical business process is proposed, consisting of a sequence of situations. Each situation involves a choice considering resource, financial, and temporal constraints, and the subsequent execution of a sequence of actions of the medical business process. The model, in accordance with the presented structuring of medical business processes, at the top level of representation contains phases of primary medical care, outpatient treatment, clinical pathway, and discharge and rehabilitation. The model creates conditions for choosing an individualized treatment process considering the patient's needs and temporal and resource constraints. The typical sequence of actions of the medical business process is determined at the level of a set of situations, and the choice of alternatives is made within the situation. The sequence of situations sets the general standard of care or treatment, and individualization is performed within individual situations, taking into account the patient's needs and financial constraints.

Keywords: medical business process, situational model, clinical pathway, care pathway, transmur pathway, resource constraints, temporal constraints, individualized treatment.

Повні імена авторів / Author's full names

Автор 1 / Author 1: Петров Костянтин Едуардович / Petrov Kostiantyn Eduardovych

Автор 2 / Author 2: Чалий Тарас Володимирович / Chalyi Taras Volodymyrovych

ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ

INFORMATION TECHNOLOGY

DOI: 10.20998/2079-0023.2024.02.08
UDC 004.94

A. M. KOPP, Doctor of Philosophy (PhD), Docent, National Technical University "Kharkiv Polytechnic Institute", Head of Software Engineering and Management Intelligent Technologies Department, Kharkiv, Ukraine; e mail: andrii.kopp@kphi.edu.ua; ORCID: <https://orcid.org/0000-0002-3189-5623>

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

I. P. GAMAYUN, Doctor of Technical Sciences, Professor, National Technical University "Kharkiv Polytechnic Institute", Full Professor of Software Engineering and Management Intelligent Technologies Department, Kharkiv, Ukraine; e-mail: ihor.hamaiun@kphi.edu.ua; ORCID: <https://orcid.org/0000-0003-2099-4658>

I. V. SAPOZHNYKOV, Student, National Technical University "Kharkiv Polytechnic Institute", Kharkiv, Ukraine; e-mail: illia.sapozhnykov@cs.kphi.edu.ua; ORCID: <https://orcid.org/0009-0003-2802-1807>

SOFTWARE DEVELOPMENT AND RESEARCH FOR MACHINE LEARNING-BASED STRUCTURAL ERRORS DETECTION IN BPMN MODELS

The most important tool for process management is business process modeling. Business process models allow to graphically represent the sequences of events, activities, and decision points that make up business processes. However, models that contain errors in depicting the business process structure can lead to misunderstanding of a business process, errors in its execution, and associated expenses. Thus, the aim of this study is to ensure the comprehensibility of business process models by detecting structural errors in business process models and their subsequent correction. During the analysis of the Business Process Management (BPM) lifecycle, it was found that the created business process models do not have a stage of control for the presence of errors in them. Therefore, the paper analyzes and improves the BPM lifecycle using the proposed approach. In the improved BPM lifecycle, it is proposed to take into account the correctness validation stage of business process models using the developed software. The paper proposes to process created BPMN (Business Process Model and Notation) models as connected directed graphs. To detect errors in business process models, one of the Machine Learning methods, K-Nearest Neighbors, is chosen, which is a fairly simple and effective classification method. The study also includes the software design and development, its performance validation, and usage to solve the given problem. To analyze the obtained results, the confusion matrix was used and the corresponding quality metrics were calculated. The obtained results confirm the suitability of the developed software for detecting structural errors in business process models. This web application, which is based on the created classification model, allows all interested users to upload business process models in BPMN 2.0 format, view the uploaded models, and analyze them for structural errors.

Keywords: business process models, structural errors, BPMN structural analysis, machine learning.

Introduction. Modern organizations are constantly improving their business processes to make them as efficient and cost-effective as possible. Business process modeling is a tool that businesses use to assess their current operations. In addition, business process modeling allows analysts to visualize a more efficient business process planned for implementation in the future. Business process modeling provides opportunities for improving the activities of the entire enterprise [1]. Business process modeling involves creating a visual representation of a business process. This is usually achieved with the help of business process modeling tools, such as flowcharts or specialized business process modeling notations, such as Business Process Model and Notation (BPMN) [1]. Business process modeling is used to identify improvements in business processes. This is done by depicting two different versions of a particular business process: the process as it exists now without changes (AS-

IS); the process as it will be after improvements are made (TO-BE) [1].

Business process models containing structural errors may indicate errors in the actual processes they describe [1]. Timely detection and elimination of such errors will allow avoiding both financial losses and much more dangerous consequences of errors in critical business processes that may pose a threat to society, business, the state, and the environment. Thus, this study solves the relevant problem of identifying structural errors in business process models to ensure their quality and suitability for use.

The rest of this paper is organized as following: Section 2 includes literature analysis of recent studies, limitations, and approaches; Section 3 describes the object, subject, and considered research methods; Section 4 outlines the research results and analysis, Section 5 includes conclusions on the study.

© Kopp A. M., Orlovskiy D. L., Gamayun I. P., Sapozhnykov I. V., 2024



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.



Related work. Recent studies focused on business process model quality consider different structural metrics [2], application of fuzzy logic [3], clustering techniques based on complexity metrics [4], and verbal business process modeling guidelines extended by the various metrics and corresponding threshold values [5]. The IDEF0

corresponding indicator values are produced. The BPM lifecycle is shown in more detail in the decomposition diagram in fig. 2.

The demonstrated decomposition diagram (fig. 2) complements the inputs and outputs with a cyclic dependence – the obtained values of business process

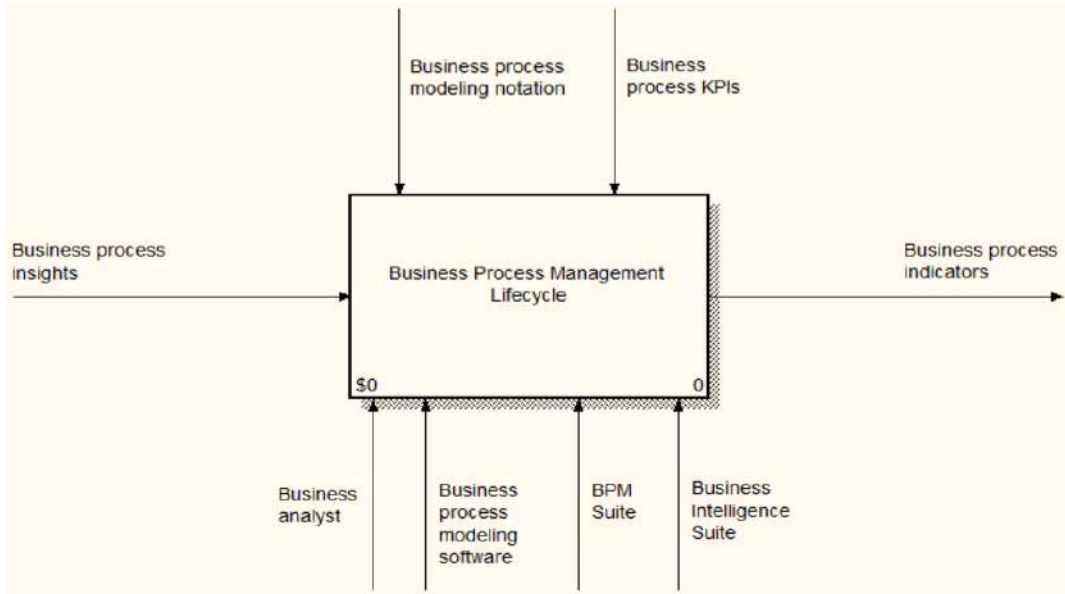


Fig. 1. The context diagram of the BPM lifecycle

methodology [6] is used to analyze the BPM lifecycle. The analysis of the “AS-IS” BPM lifecycle state starts with the IDEF0 context diagram (fig. 1), which defines inputs, outputs, controls (regulations, rules, policies, standards, etc.), and mechanisms (performers of activities and the required tools for this).

According to the presented context diagram (fig. 2), information about the business process is used as an input, and as a result of the business process execution, the

indicators are used to analyze its efficiency and, accordingly, create a new model of its future improved state. Moreover, fig. 2 shows the activity on creating business process models, business processes implementation and monitoring. The governance and mechanisms (e.g., business analyst, modeling notation, etc.) that regulate and support the relevant stages of the BPM lifecycle are defined (fig. 2).

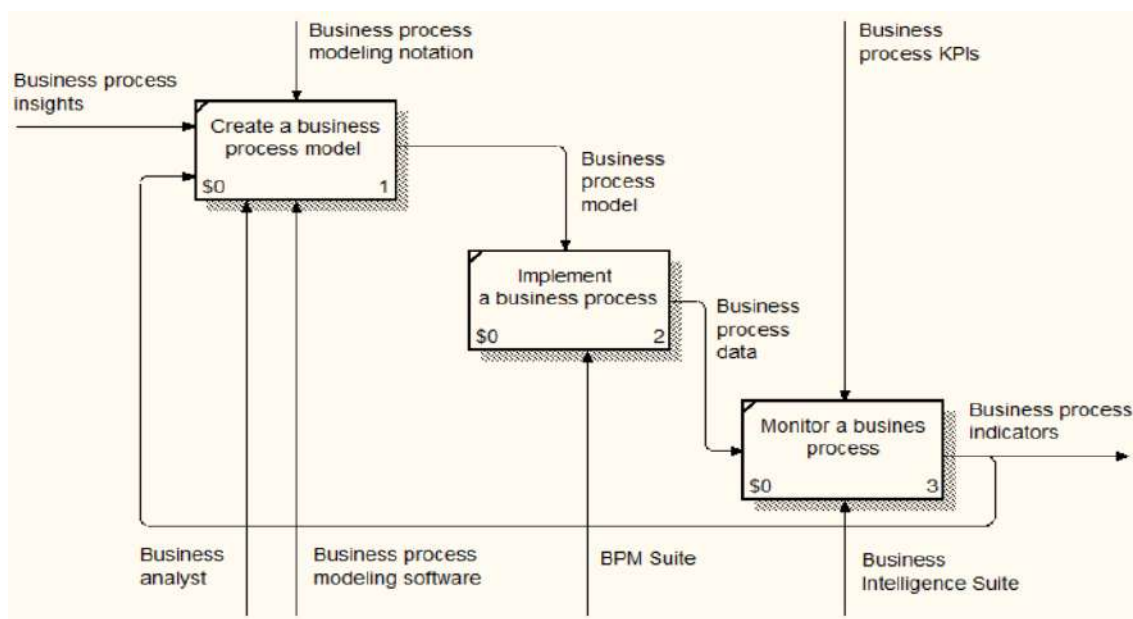


Fig. 2. The decomposition diagram of the BPM lifecycle

A detailed examination of the decomposition diagram (fig. 2) of the BPM lifecycle allows us to identify a “weakness” – the created models are immediately used to implement new business processes or planned changes in existing business processes without first checking for errors in these models (fig. 3).

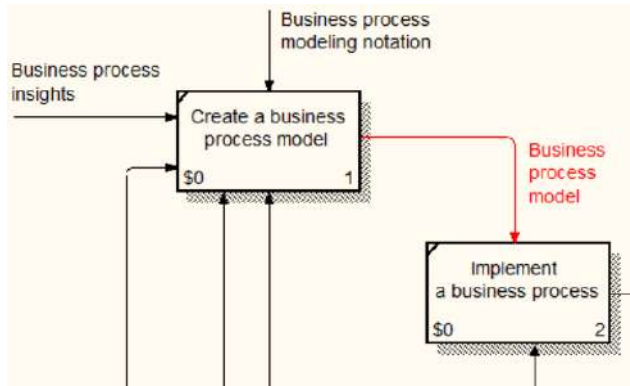


Fig. 3. The identified “weakness” of the BPM lifecycle

Thus, it is necessary to develop the algorithm and software to solve the problem of detecting structural errors in business process models.

Formally, a business process model given using the BPMN notation can be represented by a connected directed graph [7]:

$$BP = (N, A), \tag{1}$$

where:

N is the set of nodes – elements of a business process (events $E \subseteq N$, tasks $T \subseteq N$, gateways $G \subseteq N$, etc.);

A is the set of arcs – sequence flows of a business process that connect elements.

Since BPMN business process models are actually XML (eXtensible Markup Language) documents [8] created according to a special scheme, these models can be processed by computers relatively easily. Thus, the transformation of the BPMN model into a connected oriented graph (1) is possible by considering the “sequenceFlow” tags as graph arcs, and the “task” tags for actions, “event” for events, and “gateway” for gateways as graph nodes. A corresponding example of processing a BPMN file of the order processing business process model is shown in fig. 4.

High-quality business process models that do not contain errors must meet the following requirements [9]:

- use as few elements as possible in the business process model – if the business process model contains more than 50 elements, it should be decomposed into several simpler models;
- use one initial event;
- use one final event;
- do not use OR gateways.

Hence, formally, these rules of business process modeling can be represented as follows:

$$\theta = \begin{cases} 1, & |N| > 50 \wedge |E^s| \neq 1 \wedge |E^e| \neq 1 \wedge |G^{or}| > 0, \\ 0, & \text{else.} \end{cases} \tag{2}$$

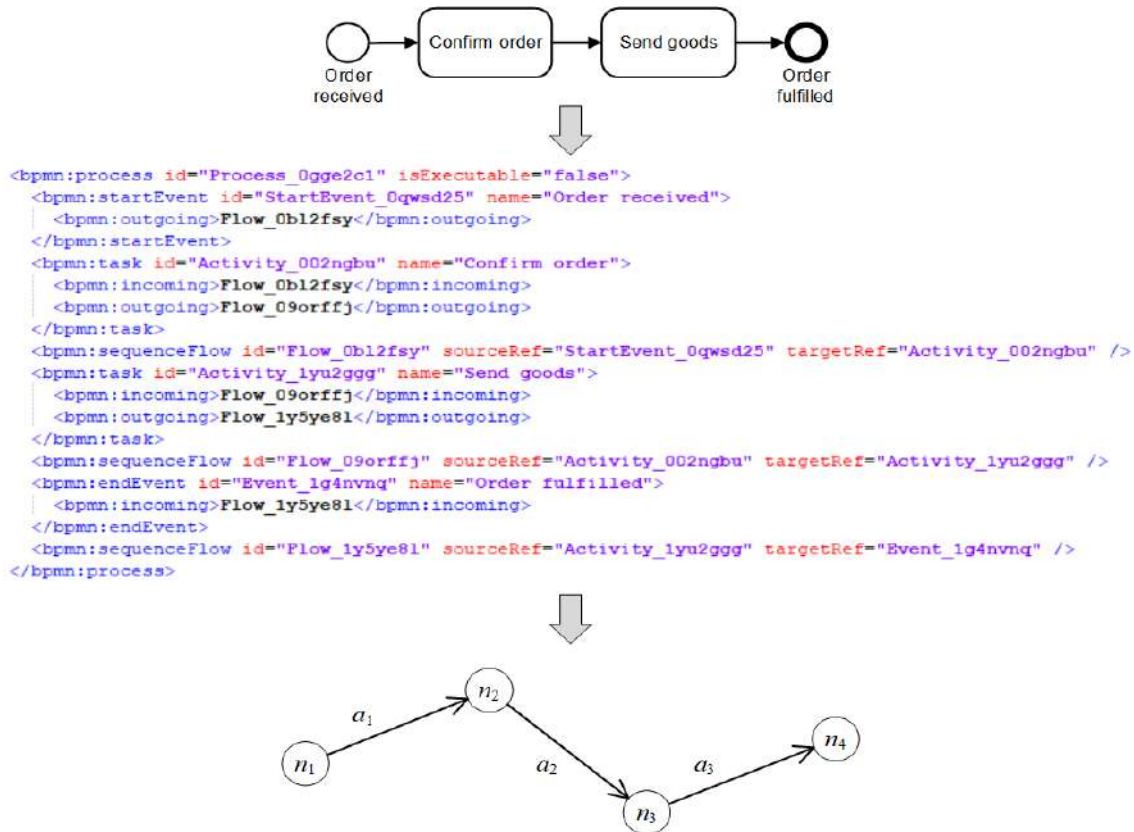


Fig. 4. The graph-based representation of the BPMN business process model

where:

- E^s is the set of start events of a business process, $E^s \subseteq E$;
- E^e is the set of end events of a business process, $E^e \subseteq E$;
- G^{or} is the set of “OR” logic gateways (inclusive) of a business process, $G^{or} \subseteq G$.

Therefore, if a business process model violates these rules (2), such a model can be classified as containing structural errors that make it difficult to understand and use by stakeholders. Furthermore, it can lead to errors in the business process itself, which is why these modeling errors must be detected and fixed.

Thus, each BPMN business process model from the training dataset can be assigned to one of the classes “has errors” ($\theta = 1$) or “has no errors” ($\theta = 0$) according to (2). It is proposed to detect structural errors in business process models using one of the well-known Machine Learning (ML) methods – K-nearest neighbors [10].

K-Nearest Neighbors (K-NN) is one of the simplest ML methods based on the supervised learning technique. The K-NN method determines the similarity between a new object and existing objects by adding the new object to the category that contains the most similar existing objects. The K-NN method stores all existing data and categorizes the new data object based on similarity. This means that when new data becomes available, it can be easily categorized using the K-NN method. The K-NN method can be used for both regression and classification, but it is mainly used for classification tasks. The K-NN method is based on a lazy learning method, because during the training phase it simply stores the dataset, and only when it receives new data for classification does K-NN assign it to the category that is most similar to the new data [10].

To apply the K-NN method to the problem of detecting structural errors in business process models, it is proposed to take as a basis the essential structural metrics of business process models:

- $|N|$ is the number of elements of a business process model (events, tasks, etc.);
- $|A|$ is the number of sequence flows that connect business process elements.

Hence, the distance between business process models is proposed to be measured using the following expression based on the Minkowski metric (3), which at $p = 2$ turns into the Euclidean distance, and standardization [11]:

$$d(BP^*, BP_i) = \left[\left(\frac{|N^*| - \mu(|N_1|, \dots, |N_m|)}{\sigma(|N_1|, \dots, |N_m|)} \right)^p + \left(\frac{|A^*| - \mu(|A_1|, \dots, |A_m|)}{\sigma(|A_1|, \dots, |A_m|)} \right)^p \right]^{\frac{1}{p}}, i = \overline{1, m}, \quad (3)$$

where:

- BP^* is the graph of the new BPMN model to be classified;
- BP_i is the graph of the i -th BPMN model from the training data set, $i = \overline{1, m}$;
- N^* is the set of elements of the new BPMN model;
- N_i the set of elements of the i -th model from the training set, $i = \overline{1, m}$;
- A^* is the set of sequence flows of the new BPMN model;
- A_i is the set of sequence flows of the i -th model from the training set, $i = \overline{1, m}$;
- μ is the arithmetic mean (4):

$$\mu(X) = \frac{1}{m} \sum_{i=1}^m x_i; \quad (4)$$

- σ is the standard deviation (5):

$$\sigma(X) = \sqrt{\frac{1}{m} \sum_{i=1}^m (x_i - \mu(X))^2}; \quad (5)$$

- m is number of business process models in the training set.

The algorithm for detecting structural errors in business process models based on the K-NN method is shown in fig. 5 [10].

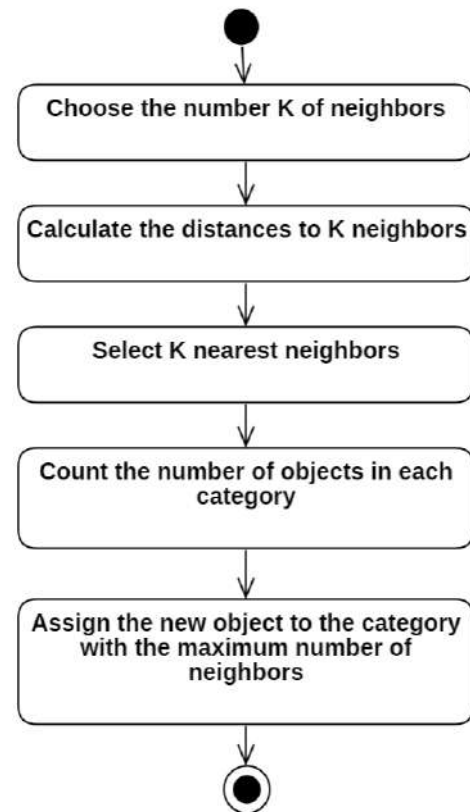


Fig. 5. The K-NN algorithm for detecting structural errors in business process models

Regarding the choice of the parameter, it is usually taken as $K = 5$ [10]. Thus, the algorithm based on the K-nearest neighbors method (fig. 5) is proposed to solve the problem of detecting structural errors in business process models and fixing them.

Materials and methods. This paper addresses a relevant problem of detecting structural errors in business process models to ensure their quality and suitability for use in organizational improvement or software development projects. Business process models that contain structural errors can cause difficulties in understanding them by stakeholders, which can result in the loss of all the benefits of modeling, analyzing, and improving of the real business processes.

The aim of this study is to ensure the comprehensibility of business process models by detecting structural errors in business process models and fixing them.

The research object is the process of detecting structural errors in business process models.

The research subject is the algorithm and software for detecting structural errors in business process models.

The decomposition diagram of the BPM lifecycle in the “TO-BE” state, which takes into account the need to control the quality of business process models and improve them using the proposed algorithm and software for solving the problem of detecting structural errors in business process models, is shown in fig. 6.

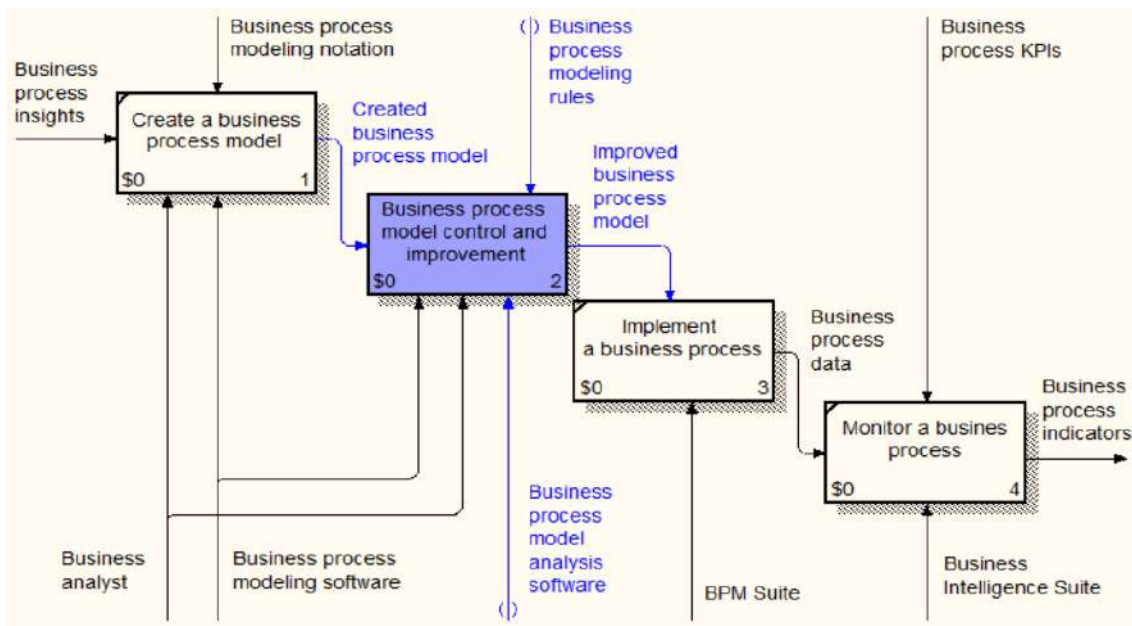


Fig. 6. The proposed improvement of the BPM lifecycle

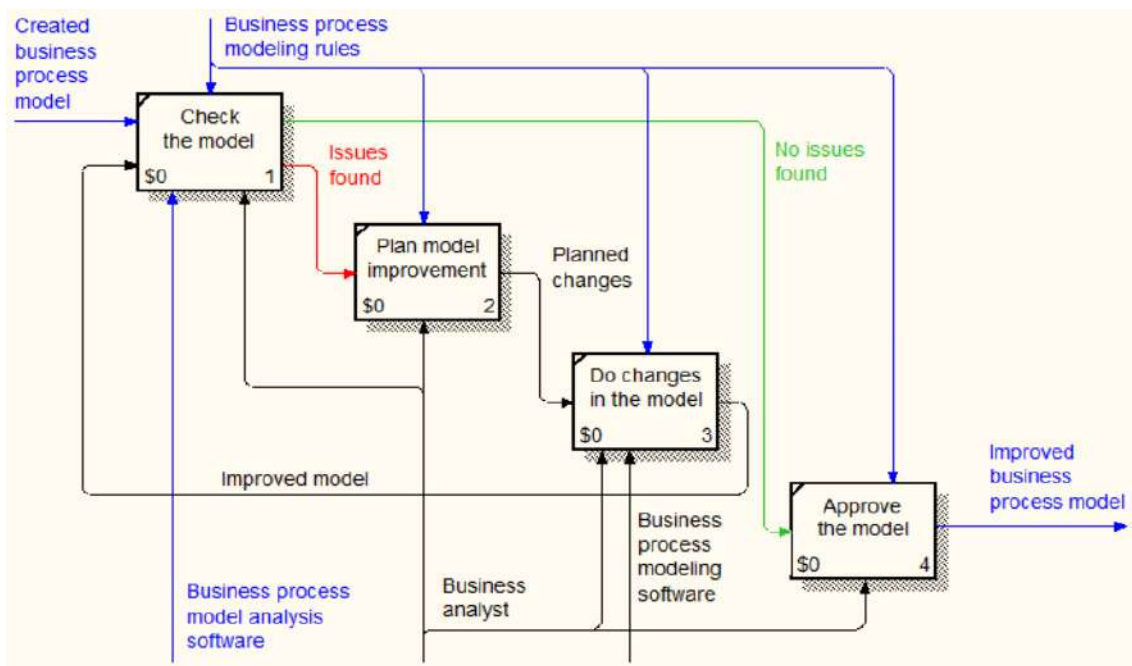


Fig. 7. The decomposition diagram of the business process model validation and improvement stage

The proposed changes (fig. 6) provide the intermediate stage to the BPM lifecycle, related to the validation of business process models and their improvement in case of structural errors. The decomposition diagram of the business process model validation and improvement stage (fig. 6) is shown in fig. 7.

Such validation is proposed to be carried out using a formal representation of BPMN models as connected directed graphs (1). The models from the training sample should be checked based on the rules (2) [9], and the rest of the business process models that will be checked by the software tool later, should be checked using the K-nearest neighbors method (fig. 5) to classify these models as having structural errors or not.

The proposed stage involves the following activities:

- analyze business process models using the developed software;
- plan changes in the business process model according to the identified errors;
- make changes to the business process model and re-analyze it;
- approve the business process model for further use if no errors were found during the last check.

users to allow them validate uploaded BPMN models and detect possible errors.

An activity diagram [12] demonstrating the process of user interaction with the software to detect structural errors in business process models is shown in fig. 9.

Let us demonstrate the results of using the created software “BPMN Assistant” to detect structural errors in business process models.

Results and discussion. The user starts working with the software by uploading a BPMN 2.0 model file that contains a description of the business process.

As an example, let us use the goods dispatch business process model, which is shown in fig. 10. This BPMN model was obtained from the public Camunda’s GitHub repository, which contains several thousands of business process models for research purposes [13].

The initial version of the considered business process model (fig. 10) shows that it contains certain structural flaws:

- 1) branching is used incorrectly (A, B);
- 2) the task is not connected to the process (C);
- 3) there are no end events, instead the workflow is completed by tasks (D, E, F).

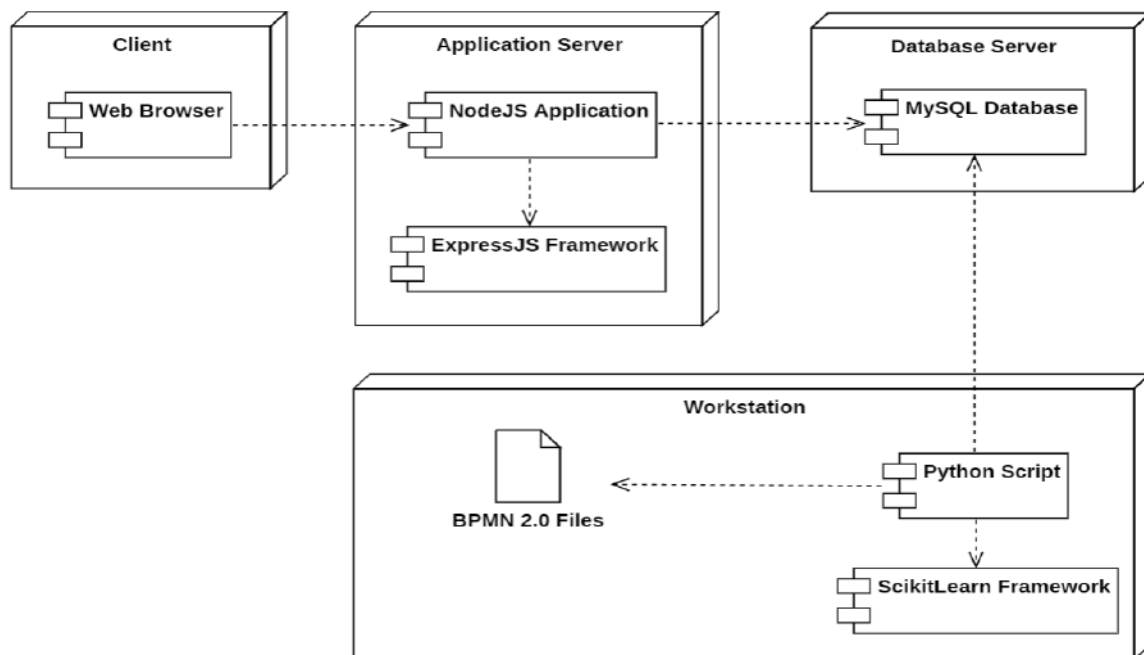


Fig. 8. The deployment diagram of the developed software components

The application “BPMN Assistant” is developed for end-users, responsible for BPMN models design and quality assurance. A deployment diagram of the software components [12], designed to validate business process models is shown in fig. 8.

As demonstrated in fig. 8, Python and Scikit Learn are used to detect structural errors in BPMN models using the K-NN algorithm and evaluate the obtained results, BPMN.IO is used to process BPMN 2.0 files, MySQL is used to store required data, Node JS and Express JS are used to build a client-server web application for business

The comparison of primary metrics of the initial and modified versions of the considered BPMN model is given in table 1.

Table 1 – The comparison of initial and modified BPMN model versions

BPMN model's version	$ N $	$ E^s $	$ E^e $	$ G^{or} $	$ A $
Initial	10	1	0	0	8
Modified	13	1	1	0	14

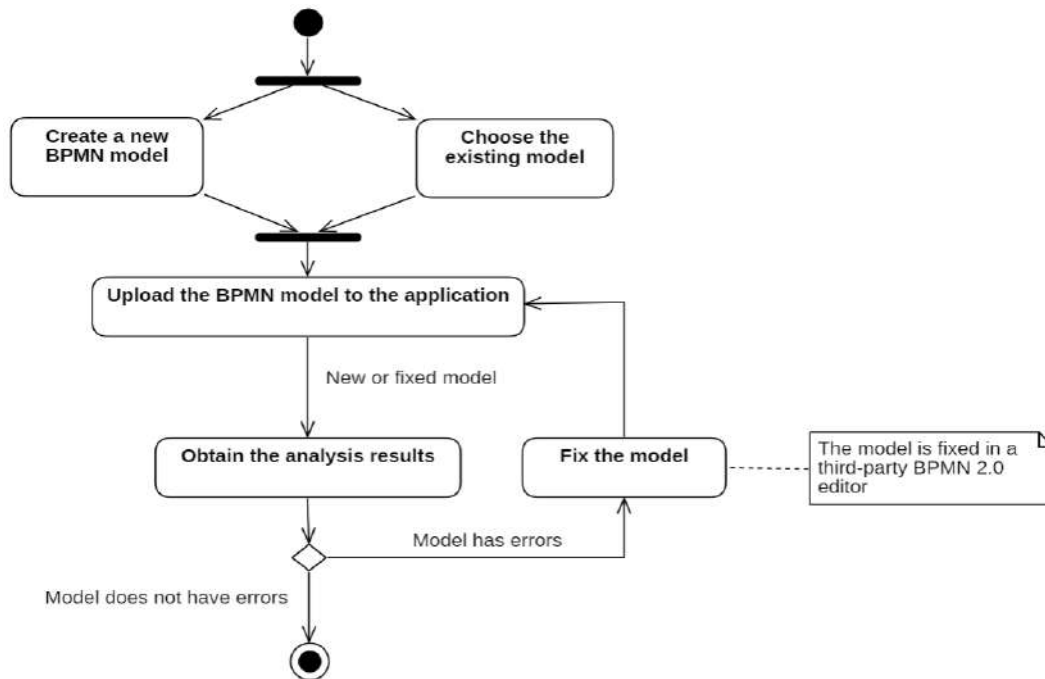


Fig. 9. The activity diagram of the user interaction with the developed software

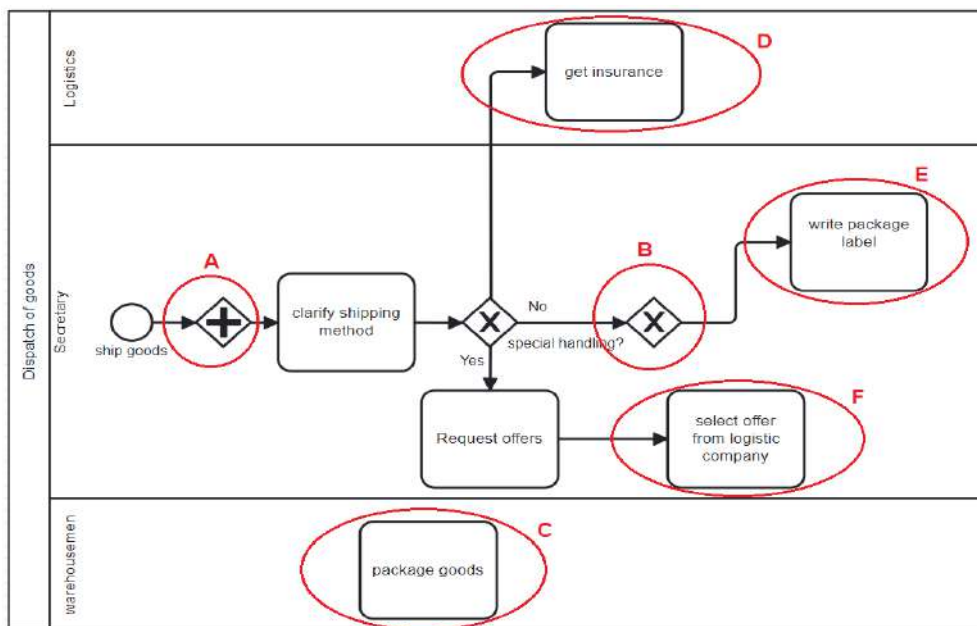


Fig. 10. The goods dispatch business process model with the detected structural errors

The modified version of the considered BPMN model is given in fig. 11.

To validate the performance of the developed software for detecting structural errors in business process models, we used the dataset of the Business Process Management Academic Initiative (BPMAI) project [14], which contains descriptions of 18812 BPMN models of various business processes.

Thus, using the data about business process models and their features, the possible presence of errors for each

BPMN model was determined using formula (2). The original dataset was divided into 2 samples – training and test sets, which contain 75% (14109 models) and 25% (4703 models) of records respectively (fig. 12).

Further, using the Python programming language [15] and the Scikit Learn library [16], a model was built to detect structural errors in business process models based on the K-nearest neighbors classification method based on the BPMAI project training data set [14]. According to formula (3), the primary structural metrics of business process

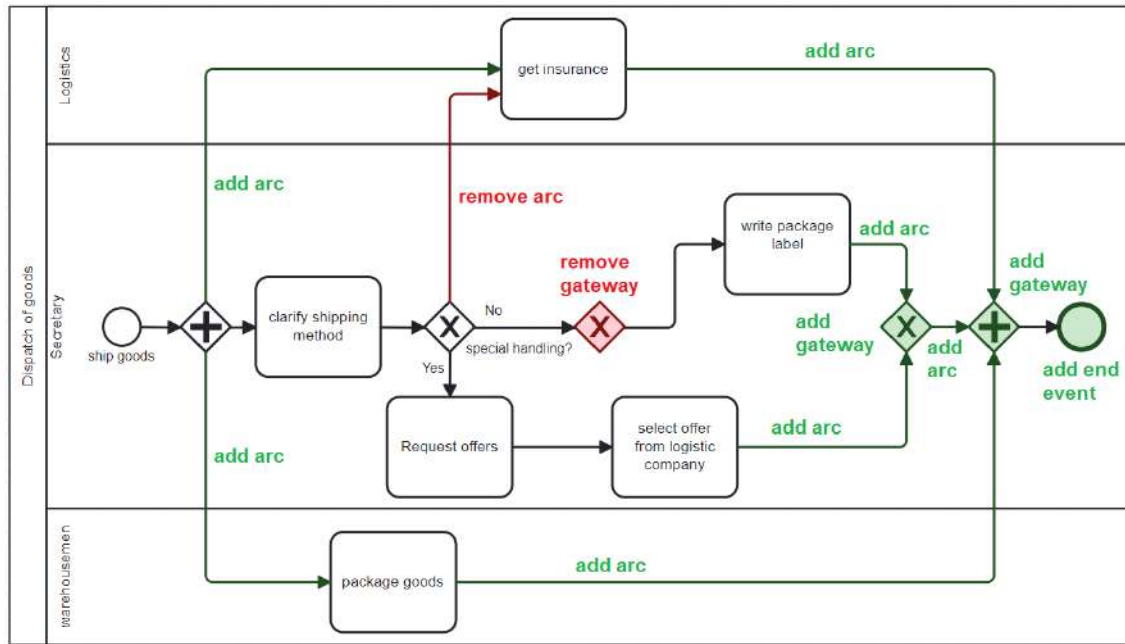


Fig. 11. The goods dispatch business process model with the modifications to eliminate structural errors

models used for classification are the number of elements $|N|$ and the number of sequence flows $|A|$.

The classification results are shown in fig. 13. The green color (category “1”) represents BPMN models that have been identified as containing structural errors. To analyze the performance of the developed software for detecting structural errors in business process models, we will use the confusion matrix (table 2) [17].

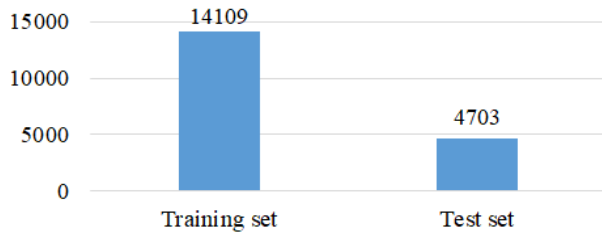


Fig. 12. The distribution of BPMN model data by training and test sets

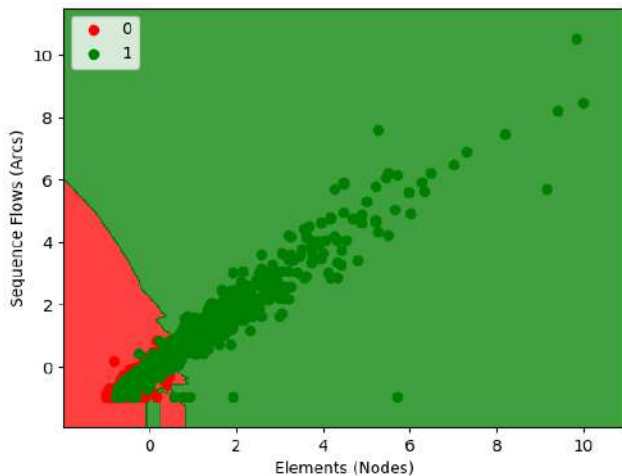


Fig. 13. The results of BPMN models classification using the K-NN method

Based on the obtained confusion matrix (table 2), the next measures were found [17]:

- true-positive (TP) results – 2716 models;
- false-positive (FP) results – 379 models;
- false negative (FN) results – 476 models;
- true-negative (TN) results – 1132 models.

Therefore, the following quality metrics [18] of the trained Machine Learning model were computed:

- precision (the share of correct answers within a class):

$$Precision = \frac{TP}{TP + FP} = 0.88; \tag{6}$$

- recall (the share of true-positive classifications):

$$Recall = \frac{TP}{TP + FN} = 0.85; \tag{7}$$

- accuracy (share of correct answers):

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} = 0.82; \tag{8}$$

- summarized performance metric:

$$F - score = 2 \frac{Precision \times Recall}{Precision + Recall} = 0.88. \tag{9}$$

Table 2 – The confusion matrix for K-NN performance analysis

Does a business process model have structural errors?	Correct answers	Yes	No
Classification results	Yes	2716	379
	No	476	1132

The obtained values of precision (4) and recall (5), as well as accuracy (6) and F-measure (7) are quite high, which makes it possible to consider the developed algorithm and software as suitable for use in detecting structural errors in business process models.

Conclusions and future work. This study addresses the relevant problem of ensuring the correctness of business process models by detecting structural errors in business process models. Thus, the following tasks were solved in the paper:

The stages of the BPM lifecycle are analyzed and the “weak spot” is identified – the created BPMN models tend to be immediately used to implement new business processes or planned changes in existing business processes without first checking for errors in these models

The algorithm for solving the problem of detecting structural errors in business process models is proposed – it is based on the K-NN Machine Learning method, taking into account primary BPMN model’s size features: the number of elements (events, tasks, etc.) $|N|$ and the number of sequence flows (i.e. arcs) $|A|$.

The appropriate changes to the BPM lifecycle are proposed – it is proposed to be appended with the intermediate stage, provided to validate business process models and assume their improvement in case of structural errors detection.

The software architecture is designed and respective software components are developed using Python and Scikit Learn to use K-NN method, BPMN.IO to process BPMN 2.0 files, MySQL to manage data, and Node JS with Express JS to create the client-server web application for business users. The use of the developed software called “BPMN Assistant” to solve the problem of detecting structural errors in BPMN models is demonstrated – as a result of the performance validation of the developed algorithm and software, a sufficiently high quality of the K-NN classifier: *Precision* = 0.88 (6), *Recall* = 0.85 (7), *Accuracy* = 0.82 (8), and *F – score* = 0.88 (9) was obtained, which makes it possible to consider the proposed software solution as suitable for use in detecting structural errors in BPMN 2.0 business process models.

Future work assumes using other metrics and ML algorithms to achieve more accurate results of structural shortcomings detection in BPMN business process models.

References

1. Reijers H. A. *Business Process Management: The evolution of a discipline*. URL: <https://doi.org/10.1016/j.compind.2021.103404> (access date: 17.09.2024).
2. Haj Ayeche B., Ghannouchi S. A., El Hadj Amor E. A. *Extension of the BPM lifecycle to promote the maintainability of BPMN models*. URL: <https://doi.org/10.1016/j.procs.2021.01.239> (access date: 17.09.2024).
3. Yahya F., Boukadi K., Ben-Abdallah H. *Improving the quality of Business Process Models: Lesson learned from the State of the Art*. URL: <https://doi.org/10.1108/BPMJ-11-2017-0327> (access date: 17.09.2024).
4. Fotoglou C. et al. *Complexity clustering of BPMN models: initial experiments with the K-means algorithm*. URL: https://doi.org/10.1007/978-3-030-46224-6_5 (access date: 17.09.2024).

5. Corradini F. et al. *Correctness checking for BPMN collaborations with sub-processes*. URL: <https://doi.org/10.1016/j.jss.2020.110594> (access date: 17.09.2024).
6. *IDEFO Function Modeling*. URL: https://doi.org/10.1007/978-3-540-89556-5_5 (access date: 17.09.2024).
7. O'Regan G. *Graph Theory*. URL: https://doi.org/10.1007/978-3-031-26212-8_7 (access date: 17.09.2024).
8. *BPMN, Version 2.0*. URL: <https://www.omg.org/spec/BPMN/2.0/PDF> (access date: 17.09.2024).
9. Avila D. T., dos Santos R. I., Mendling J., Thom L. H. *A systematic literature review of process modeling guidelines and their empirical support*. URL: <https://doi.org/10.1108/BPMJ-10-2019-0407> (access date: 17.09.2024).
10. *K-Nearest Neighbors*. URL: https://doi.org/10.1007/978-3-642-38652-7_2 (access date: 17.09.2024).
11. *Distance Measures*. URL: https://doi.org/10.1007/978-3-642-04898-2_626 (access date: 17.09.2024).
12. Unhelkar B. *Software Engineering with UML*. URL: <https://doi.org/10.1201/9781351235181> (access date: 17.09.2024).
13. *BPMN for research*. URL: <https://github.com/camunda/bpmn-for-research> (access date: 17.09.2024).
14. *Model Collection of the Business Process Management Academic Initiative*. URL: <https://doi.org/10.5281/zenodo.3758705> (access date: 17.09.2024).
15. *Python*. URL: <https://www.python.org/> (access date: 17.09.2024).
16. *Scikit-learn*. URL: <https://scikit-learn.org/> (access date: 17.09.2024).
17. *Confusion Matrix*. URL: https://doi.org/10.1007/978-0-387-30164-8_157 (access date: 17.09.2024).
18. Ting K. M. *Encyclopedia of Machine Learning*. URL: https://doi.org/10.1007/978-0-387-30164-8_652 (access date: 17.09.2024).

References (transliterated)

1. Reijers H. A. *Business Process Management: The evolution of a discipline*. Available at: <https://doi.org/10.1016/j.compind.2021.103404> (accessed: 17.09.2024).
2. Haj Ayeche B., Ghannouchi S. A., El Hadj Amor E. A. *Extension of the BPM lifecycle to promote the maintainability of BPMN models*. Available at: <https://doi.org/10.1016/j.procs.2021.01.239> (accessed: 17.09.2024).
3. Yahya F., Boukadi K., Ben-Abdallah H. *Improving the quality of Business Process Models: Lesson learned from the State of the Art*. Available at: <https://doi.org/10.1108/BPMJ-11-2017-0327> (accessed: 17.09.2024).
4. Fotoglou C. et al. *Complexity clustering of BPMN models: initial experiments with the K-means algorithm*. Available at: https://doi.org/10.1007/978-3-030-46224-6_5 (accessed: 17.09.2024).
5. Corradini F. et al. *Correctness checking for BPMN collaborations with sub-processes*. Available at: <https://doi.org/10.1016/j.jss.2020.110594> (accessed: 17.09.2024).
6. *IDEFO Function Modeling*. Available at: https://doi.org/10.1007/978-3-540-89556-5_5 (accessed: 17.09.2024).
7. O'Regan G. *Graph Theory*. Available at: https://doi.org/10.1007/978-3-031-26212-8_7 (accessed: 17.09.2024).
8. *BPMN, Version 2.0*. Available at: <https://www.omg.org/spec/BPMN/2.0/PDF> (accessed: 17.09.2024).
9. Avila D. T., dos Santos R. I., Mendling J., Thom L. H. *A systematic literature review of process modeling guidelines and their empirical support*. Available at: <https://doi.org/10.1108/BPMJ-10-2019-0407> (accessed: 17.09.2024).
10. *K-Nearest Neighbors*. Available at: https://doi.org/10.1007/978-3-642-38652-7_2 (accessed: 17.09.2024).
11. *Distance Measures*. Available at: https://doi.org/10.1007/978-3-642-04898-2_626 (accessed: 17.09.2024).
12. Unhelkar B. *Software Engineering with UML*. Available at: <https://doi.org/10.1201/9781351235181> (accessed: 17.09.2024).
13. *BPMN for research*. Available at: <https://github.com/camunda/bpmn-for-research> (accessed: 17.09.2024).
14. *Model Collection of the Business Process Management Academic Initiative*. Available at: <https://doi.org/10.5281/zenodo.3758705> (accessed: 17.09.2024).

15. Python. Available at: <https://www.python.org/> (accessed: 17.09.2024).
16. Scikit-learn. Available at: <https://scikit-learn.org/> (accessed: 17.09.2024).
17. Confusion Matrix. Available at: https://doi.org/10.1007/978-0-387-30164-8_157 (accessed: 17.09.2024).
18. Ting K. M. *Encyclopedia of Machine Learning*. Available at: https://doi.org/10.1007/978-0-387-30164-8_652 (accessed: 17.09.2024).

Received 15.11.2024

УДК 004.94

А. М. КОПП, доктор філософії (PhD), доцент, Національний технічний університет

«Харківський політехнічний інститут», завідувач кафедри програмної інженерії та інтелектуальних технологій управління, м. Харків, Україна; e-mail: andrii.kopp@khpri.edu.ua; ORCID: <https://orcid.org/0000-0002-3189-5623>

Д. Л. ОРЛОВСЬКИЙ, кандидат технічних наук (PhD), доцент, Національний технічний університет

«Харківський політехнічний інститут», професор кафедри програмної інженерії та інтелектуальних технологій управління, м. Харків, Україна; e-mail: dmytro.orlovskiy@khpri.edu.ua; ORCID: <https://orcid.org/0000-0002-8261-2988>

І. П. ГАМАЮН, доктор технічних наук, професор, Національний технічний університет «Харківський політехнічний інститут», професор кафедри програмної інженерії та інтелектуальних технологій управління, м. Харків, Україна; e-mail: ihor.hamaiun@khpri.edu.ua; ORCID: <https://orcid.org/0000-0003-2099-4658>

І. В. САПОЖНИКОВ, Національний технічний університет «Харківський політехнічний інститут», студент, м. Харків, Україна; e-mail: illia.sapozhnykov@cs.khpri.edu.ua; ORCID: <https://orcid.org/0009-0003-2802-1807>

РОЗРОБКА ТА ДОСЛІДЖЕННЯ ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ ДЛЯ ВИЯВЛЕННЯ СТРУКТУРНИХ ПОМИЛОК У BPMN-МОДЕЛЯХ НА ОСНОВІ МАШИННОГО НАВЧАННЯ

Найважливішим інструментом управління процесами є моделювання бізнес-процесів. Моделі бізнес-процесів дозволяють графічно представляти послідовності подій, дій і точок прийняття рішень, з яких складаються бізнес-процеси. Однак моделі, які містять помилки в представленні структури бізнес-процесу, можуть призвести до неправильного розуміння бізнес-процесу, помилок у його виконанні та пов'язаних з цим витрат. Таким чином, метою даного дослідження є забезпечення зрозумілості моделей бізнес-процесів шляхом виявлення структурних помилок у моделях бізнес-процесів та їх подальшого виправлення. Під час аналізу життєвого циклу управління бізнес-процесами (Business Process Management, BPM) виявлено, що для створених моделей бізнес-процесів не передбачено етапу контролю на наявність у них помилок. Таким чином, в статті аналізується та вдосконалюється життєвий цикл BPM, використовуючи запропонований підхід. У вдосконаленому життєвому циклі BPM пропонується врахувати етап перевірки коректності моделей бізнес-процесів за допомогою розробленого програмного забезпечення. У статті пропонується обробляти створені моделі BPMN (Business Process Model and Notation) у вигляді зв'язаних орієнтованих графів. Для виявлення помилок у моделях бізнес-процесів вибирається один із методів машинного навчання K-Nearest Neighbors, який є достатньо простим і ефективним методом класифікації. Дослідження також включає проектування та розробку програмного забезпечення, перевірку його продуктивності та застосування для розв'язання поставленої задачі. Для аналізу отриманих результатів було використано матрицю помилок та розраховано відповідні метрики якості. Отримані результати підтверджують придатність розробленого програмного забезпечення до виявлення структурних помилок у моделях бізнес-процесів. Веб-застосунок, який базується на побудованій моделі класифікації, дозволяє всім зацікавленим користувачам завантажувати моделі бізнес-процесів у форматі BPMN 2.0, переглядати завантажені моделі та аналізувати їх на наявність структурних помилок.

Ключові слова: моделі бізнес-процесів, структурні помилки, структурний аналіз BPMN, машинне навчання.

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

Автор 3 / Author 3: Гамаюн Ігор Петрович / Gamayun Igor Petrovych

Автор 4 / Author 4: Сапожников Ілля Віталійович / Sapozhnykov Illia Vitaliiovich

O. V. RUDSKYI, Student, National Technical University "Kharkiv Polytechnic Institute", Kharkiv, Ukraine;
e-mail: oleksandr.rudskyi@cs.khpi.edu.ua; ORCID: <https://orcid.org/0009-0001-1130-9957>

A. M. KOPP, Doctor of Philosophy (PhD), Docent, National Technical University "Kharkiv Polytechnic Institute",
Head of Software Engineering and Management Intelligent Technologies Department, Kharkiv, Ukraine;
e-mail: andrii.kopp@khpi.edu.ua; ORCID: <https://orcid.org/0000-0002-3189-5623>

T. Ye. GONCHARENKO, Candidate of Pedagogical Sciences (PhD), Docent, National Technical University
"Kharkiv Polytechnic Institute", Head of Foreign Languages Department, Kharkiv, Ukraine;
e-mail: tetiana.goncharenko@khpi.edu.ua; ORCID: <https://orcid.org/0000-0001-6630-307X>

I. P. GAMAYUN, Doctor of Technical Sciences, Professor, National Technical University "Kharkiv Polytechnic Institute",
Full Professor of Software Engineering and Management Intelligent Technologies Department, Kharkiv, Ukraine;
e-mail: ihor.hamaiun@khpi.edu.ua; ORCID: <https://orcid.org/0000-0003-2099-4658>

INTELLIGENT TECHNOLOGY FOR SEMANTIC COMPLETENESS ASSESSMENT OF BUSINESS PROCESS MODELS

In this paper, we present a method for comparing business process models with their textual descriptions, using a semantic-based approach based on the SBERT (Sentence-Bidirectional Encoder Representations from Transformers) model. Business process models, especially those created with the BPMN (Business Process Model and Notation) standard, are crucial for optimizing organizational activities. Ensuring the alignment between these models and their textual descriptions is essential for improving business process accuracy and clarity. Traditional set similarity methods, which rely on tokenization and basic word matching, fail to capture deeper semantic relationships, leading to lower accuracy in comparison. Our approach addresses this issue by leveraging the SBERT model to evaluate the semantic similarity between the text description and the BPMN business process model. The experimental results demonstrate that the SBERT-based method outperforms traditional methods, based on similarity measures, by an average of 31%, offering more reliable and contextually relevant comparisons. The ability of SBERT to capture semantic similarity, including identifying synonyms and contextually relevant terms, provides a significant advantage over simple token-based approaches, which often overlook nuanced language variations. The experimental results demonstrate that the SBERT-based approach, proposed in this study, improves the alignment between textual descriptions and corresponding business process models. This advancement is allowing to improve the overall quality and accuracy of business process documentation, leading to fewer errors, introducing better clarity in business process descriptions, and better communication between all the stakeholders. The overall results obtained in this study contribute to enhancing the quality and consistency of BPMN business process models and related documentation.

Keywords: business process modeling, BPMN, semantic similarity, SBERT, text comparison, business process optimization, natural language processing.

Introduction. In today's world, business process modeling plays an important role in improving management and optimizing organizational activities. However, creating appropriate business process models is a task that requires significant efforts and resources. Comparing business process models with their textual descriptions proves to be a crucial task, as it can help to ensure the accuracy of the model, identify discrepancies, and improve the quality of both the models and the textual descriptions of business processes [1].

In this context, the relevance of evaluating the alignment of business process models with their textual descriptions is evident. Business process modeling, especially using the BPMN standard, provides a tool for representing business processes in a graphical format, making them easier to understand and analyze. However, ensuring consistency between the model and the textual description is essential to avoid errors and inconsistencies in business processes [2].

Comparing business process models with their textual descriptions not only ensures accuracy and consistency but also helps to identify potential shortcomings and ambiguities in the textual descriptions, which can lead to improvements in the quality of business processes. Additionally, this approach fosters a shared understanding among all business process stakeholders, regardless of their level of expertise in process modeling [3].

Related work. A systematic literature review was used to explore current methods for text comparison.

A Systematic Literature Review (SLR) identifies, selects, and critically evaluates studies to answer a clearly formulated question. The systematic review must follow a well-defined protocol or plan that clearly outlines the criteria for conducting the review. It involves a comprehensive and transparent search, conducted across multiple databases and grey literature, which can be replicated by other researchers. This requires a well-thought-out search strategy aimed at answering a specific question. The review identifies the type of information that was searched, critiqued, and reported over a known period of time. Search terms, search strategies (including database names, platforms, search dates), and limitations must all be included in the review [4].

To answer the research questions, the following SLR objectives were defined:

1. Review articles to identify existing methods for text comparison;
2. Highlight weaknesses in the methods with the aim of addressing them through further research;
3. Gain new insights into text comparison methods that can be applied in future research.
4. The following search string was used for the study: ("allintitle:" + "text" + "similarity" + "site:" + "ieeexplore.ieee.org").

© Rudskyi O. V., Kopp A. M., Goncharenko T. Ye., Gamayun I. P., 2024



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.



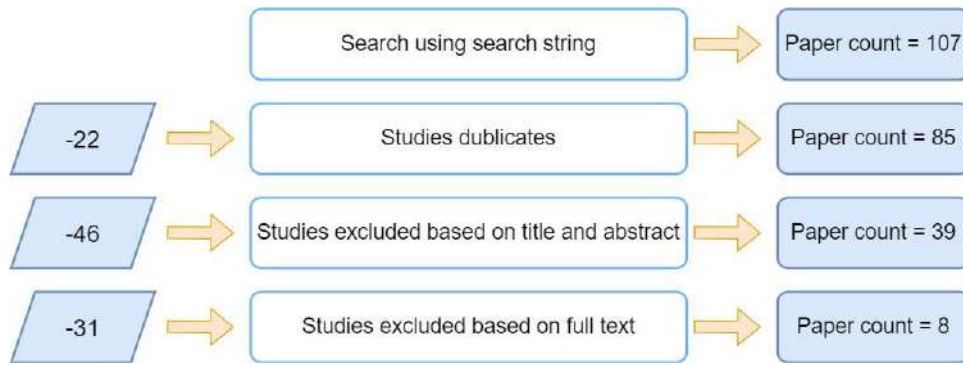


Fig. 1. General systematic literature review scheme

The initial search using only the keywords yielded 107 academic articles related to text similarity in various languages. After thoroughly reviewing the articles, those that were not directly related to the research topic but appeared due to keyword matches were excluded. Additionally, articles were excluded due to duplication, lack of full text, or if the research did not address any of the research questions. After all exclusions, 8 academic articles remained (fig. 1).

State-of-the-art. The first article analyzed was [5], which presents the results of applying various methods for measuring semantic text similarity. The goal of the article is to assess the degree of semantic equivalence of multi-word sentences [5].

One of the methods discussed in [5] is Bag-of-Words (BOW), a technique used to represent fixed-length vectors from which features are extracted for modeling. One of the drawbacks of this method is that the word order is lost, leading to identical vector representations for different sentences with the same words [5].

Another method presented is word2vec. The neural network model “Word2Vec” based on skip-gram predicts surrounding words in sentences without using hidden neurons. Here, the artificial neural network (ANN) is trained on word pairs extracted from documents, considering the window size as a critical parameter of the algorithm. The skip-gram neural network model consists of weights and biases that are updated with each iteration of the input data set, and training on a large set of words would be a time-consuming task [5]. The main idea behind the skip-gram-based Word2Vec algorithm is that a vector is initially randomly initialized for each word in the vocabulary. Then, for each position t , the central word at this position is determined as c , and its context word as o . To identify the con-

text words, a window size of m is defined, meaning that the model will consider words in positions from $t - m$ to $t + m$ as context (fig. 2).

To calculate the probability of a context word by a given central word, each word is represented by two sets of vectors: U_w and V_w . U_w is used when w is a context word, and V_w when w is a central word. Using these two vectors, the probability equation for the central word o and the context word c is as follows:

$$P(O = o | C = c) = \frac{\exp(u_o^T v_c)}{\sum_{w \in Vocab} \exp(u_w^T v_c)} \quad (1)$$

In the numerator (1), there is the dot product of words o and c , which reflects the similarity between these two vectors. The higher the similarity, the higher the probability. The denominator (1) normalizes the probability values across the entire vocabulary so that the overall sum equals 1.

The next article analyzed was [6]. This paper discusses a method called Word Vector Distance Decentralization (WVDD), which can handle complex semantic relations, including sentence components and word order [6]. Based on the popular Word2vec model, the WVDD method is proposed for transforming word vectors into sentence vectors and implementing the merging of word vectors to measure sentence similarity, taking into account word order, weighting parameters, and semantic relations. For text clustering, it suggests using the Apache Spark clustering algorithm, which employs the K-means algorithm on the Spark architecture for parallel computing to speed up the text clustering process [6].

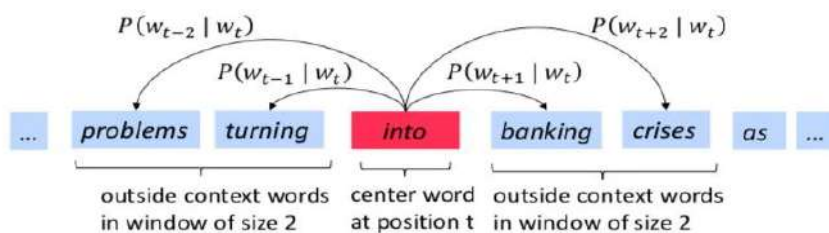


Fig. 2. Process of learning surrounding words in Word2Vec [5]

The following article was [7]. This paper examines the Siamese Neural Network (SNN) and the self2self-attention (S2SA), which is introduced into a Convolutional Neural Network (CNN) to build a new Siamese neural network, specifically the S2SA-SNN.

In S2SA-SNN, self2self-attention is used to learn the varying importance of words and complex syntactic features within a single sentence. Semantic text similarity at the sentence level involves having two sentences. With one sentence X and another sentence Y, the goal of the proposed model is to learn the semantic representations of sentences X and Y and compute a score to measure their similarity or obtain the output of the activation function through these semantic representations.

The next article analyzed was [8]. This paper proposes a short text clustering algorithm based on the fusion of BTM and GloVe similarity (BG & SLF-Kmeans). These are used to model pre-processed short texts. To calculate text similarity based on GloVe word vector modeling, an improved word weighting method (IWMD) is employed. Afterward, the two similarities are linearly combined and used as a distance function to implement clustering via the Kmeans method. The results indicate that BG & SLF-Kmeans significantly improves clustering accuracy compared to TF-IDF & Kmeans, BTM & Kmeans, and BTF & SLF-Kmeans [8].

The next article analyzed was [9]. This paper reviews the limitations of the traditional TF-IDF algorithm and proposes an improved PTF-IDF algorithm. Also, a text classification algorithm based on PTF-IDF and cosine similarity is proposed. Compared to the traditional TF-IDF algorithm, based on an experiment for finding the optimal keyword, the paper finds that text classification accuracy reaches a stable value when the category keywords reach a certain proportion [9].

The next article analyzed was [10]. This paper explores text similarity using a two-stage model for fine-tuning Bidirectional Encoder Representations from Transformers (BERT). Text similarity, as a vertical task in natural language processing, can achieve performance improvements through the two-stage model proposed in this paper [10].

BERT is an abbreviation for Bidirectional Encoder Representation from Transformers, which is a transformer-based machine learning technique for pre-training natural language processing (NLP) developed by Google [10].

BERT can be defined as a function:

$$B : P \rightarrow R^{N \times h}, \quad (2)$$

where:

- h is the size of the hidden level;
- $N=512$ is the maximum sequence length supported by the model.

As an output, BERT (2) receives a paragraph $\rho \in P$ and decomposes it into a sequence $q \in N$ tokens $(p^j)_{j=1}^q$. After that, the sequence (3) is supplemented with N elements by adding special CLS (Classification), SEP (Separator), and PAD (Padded) tokens [10].

This token sequence can be written in the form:

$$I^p = \left(CLS, (p^j)_{j=1}^q, SEP, \dots, PAD \right). \quad (3)$$

In BERT, all tokens are embedded using three functions: embedded tokens, positions, and segments, denoted as T , O , and G , respectively. Token embedding converts unique token values into intermediate vectors $T(I^p) \in R^{N \times h}$. Position embedding encodes the token positions into a single space, $O(I^p) \in R^{N \times h}$. Segment embedding is used to associate each token with one of two sequences $G(\{0,1\}^N) \in R^{N \times h}$ [10]. The block diagram of the BERT model is given in fig. 3.

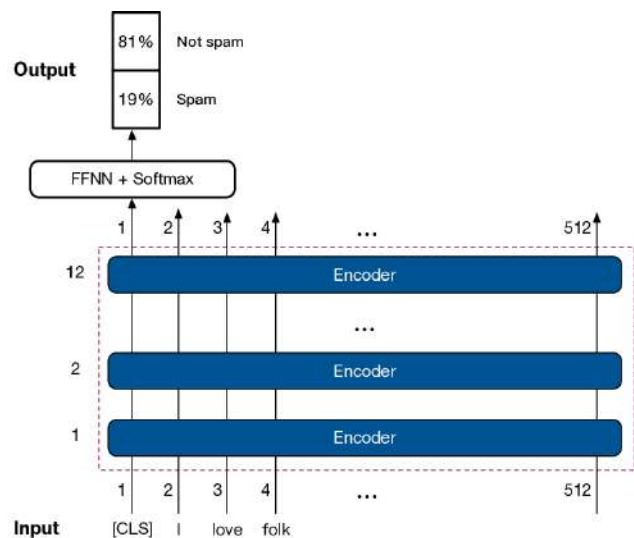


Fig. 3. Block diagram of the BERT model [10]

The next article analyzed was [11]. This paper presents two different models for article categorization. These models consist of two key components: text semantic representation and similarity calculation. First, they represent the text document (article) and then classify it into one of the predefined categories. Afterward, the models dynamically match the output category with the user-defined category. The first model uses TF-IDF features as the semantic representation method, a classifier trained on the BBC dataset, and GloVe to compute category similarity. The second model is an improvement of the first [11]. The GloVe model is an unsupervised learning method used to obtain vector representations of words. GloVe represents words in a multi-dimensional space, placing related words closer to each other in this space. As a result, GloVe implicitly models complex relationships between words in a large vector space. To compute the similarity between different words, GloVe uses cosine similarity and the vector difference between the given words; it associates more than one value for a word pair. The vector difference is needed to better differentiate between words [11].

The next article analyzed was [12]. This paper analyzes the relationship between the true similarity of words and the similarity obtained by various word

embedding methods. The following methods are analyzed in this paper:

1. Word2vec. Word2vec includes two different models: Continuous Bag Of Words (CBOW) and Skip-gram. Both of these methods are neural networks with a hidden layer of N neurons, where N is the dimensionality of the generated word embeddings. The first method, CBOW, is a neural network where the context of words serves as the input. The task is to predict the current word as the network's output. The second method, Skip-gram, is a neural network where the input is a one-hot encoding of a word, and the output is the predicted context of the word, i.e., the surrounding words [12].

2. FastText. The FastText model directly derives from the Skip-gram Word2Vec model. The authors claim that by using a clear vector representation for each word, the Skip-gram model ignores the internal structure of words. For this, they proposed a different scoring function that considers the internal structure. Their subword model represents each word as a bag of character n -grams. Special symbols $<$ and $>$ are added at the beginning and the end of words to distinguish prefixes and suffixes from other character sequences. The word is also included in its set of n -grams to learn a better representation of each word. This model allows sharing representations between words, thus enabling a more robust representation of rare words [12].

3. GloVe (Global Vectors for Word Representation). GloVe is a logarithmic bilinear regression model for unsupervised word representation learning, which combines the advantages of two families of models: global matrix factorization and local context window methods. The overall idea is that the relationship between any two words, i.e., the frequency of words co-occurring in each other's context, encodes information about the words. It captures meaningful linear substructures by effectively using global word co-occurrence statistics. The model is optimized so that the scalar product of any word pair vectors equals the ratio of the corresponding words' occurrences [12].

4. LexVec is based on the idea of factorizing the PPMI matrix using a reconstruction loss function. This loss function does not weigh all errors equally, unlike SVD, but penalizes frequent co-occurrence errors more heavily while also handling negative co-occurrence cases, unlike GloVe. The authors argue that the performance of word similarity and analogy tasks shows that LexVec compares favorably with state-of-the-art methods and often surpasses them in many of these tasks [12].

Algorithm based on the SBERT model. To solve the task of analyzing the alignment of business process models with their textual descriptions, the software application must generate texts T_1 and T_2 based on data extracted from the BPMN and text files. To generate text T_1 , the application must extract all the names of “task” elements and related action elements from the BPMN file:

- “Service Task” is a task that uses a service, which can be a web service or an automated application [13];
- “Send Task” is a simple task designed to send a message to an external participant. As soon as the message is sent, the task is completed [13];

- “Receive Task” is a simple task designed to wait for receiving a message from an external user [13];

- “User Task” is a typical task of a business process in which a human executor performs a task with the help of a software application and is scheduled through some task list manager [6];

- “Manual Task” is a task that is supposed to be performed without the help of any business process execution mechanism or any program [13];

- “Business Rule Task” is a task that involves a mechanical process to provide input data for the business rule mechanism and obtain the output data of calculations that the business rule mechanism can provide [13];

- “Script Task” is a task that is executed by the business process engine. When the task is ready to run, the engine will execute the script. After completing the script, the task will also be executed [13].

The following algorithm, presented in the UML activity diagram in fig. 4, can be used to generate text T_1 .

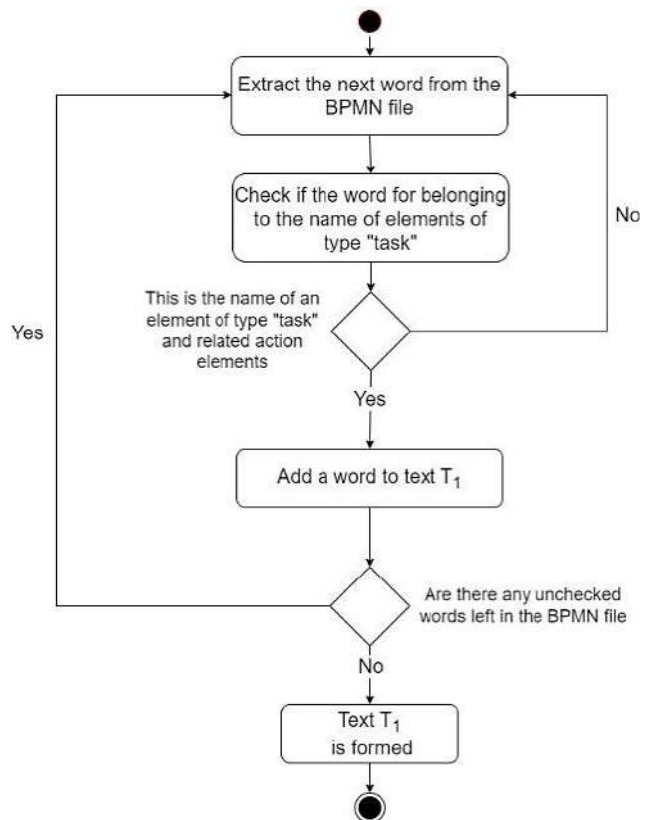


Fig. 4. Algorithm for generating text from the names of business process model tasks

Next, we will use Sentence-BERT (SBERT), a modification of a pre-trained BERT network to measure the degree of semantic textual similarity between two texts.

BERT is used to solve various tasks, such as sentiment analysis or question answering, and it is becoming increasingly popular for creating word embeddings – vector representations of words that reflect their semantic meanings [14].

Representing words as embeddings has provided a huge advantage, as machine learning algorithms cannot work with raw text but can work with vector embeddings.

This allows different words to be compared based on their similarity using standard metrics, such as Euclidean or cosine similarity [14].

Transformer-based models expect a sequence of tokens as input. Therefore, the very first step is to transform the input text into a sequence of tokens, or tokenization. BERT accepts the token [CLS] and two sentences separated by a special [SEP] token as input. Depending on the maximum token sequence length, which is predetermined, a set of [PAD] tokens will also be added after the [SEP] token. Depending on the model configuration, this information is processed 12 or 24 times by multi-head attention blocks. The output is then aggregated and passed to a simple regression model to produce the final label [15]. fig. 5 shows the architecture of the BERT model.

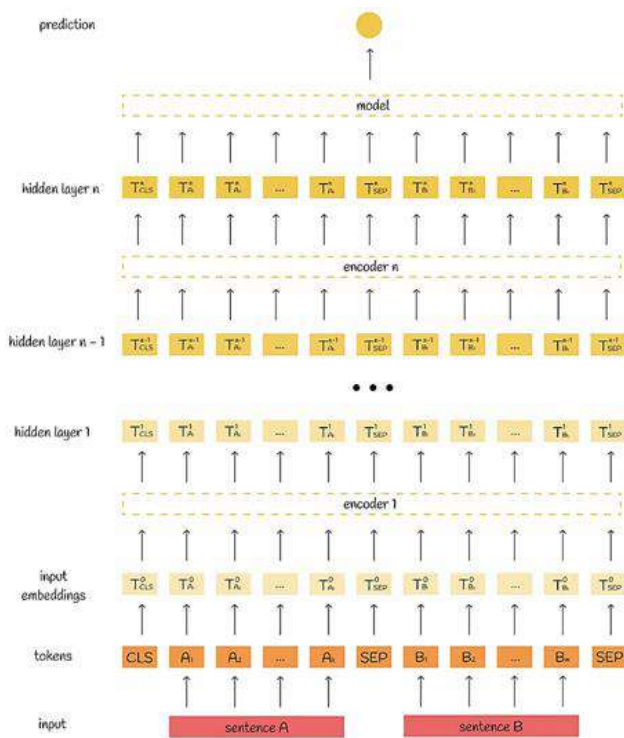
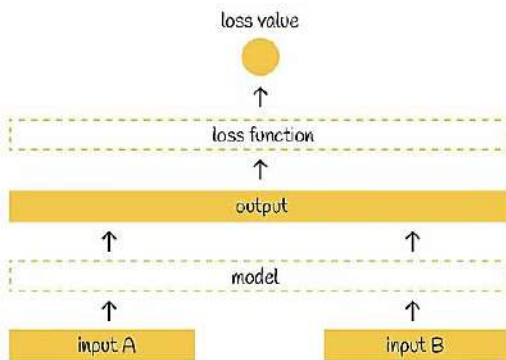


Fig. 5. Architecture of the BERT model [15]

The main problem with BERT is that whenever two sentences are passed and processed simultaneously, it complicates obtaining embeddings that independently represent only one sentence [14].



SBERT introduces the concept of a Siamese network, which means that two sentences are independently pass through the same BERT model each time. The Siamese network architecture allows splitting fixed-size vectors for the input sentences [14].

Fig. 6 shows a comparison between the non-Siamese and Siamese architectures. As can be seen in the figure, the key difference is that on the left, the model processes both inputs simultaneously, while on the right, the model processes both inputs in parallel, meaning the outputs are independent of each other.

After the sentence passes through BERT, a pooling layer is applied to the BERT embeddings to obtain a lower-dimensional representation: the initial 512 768-dimensional vectors are converted into a single 768-dimensional vector. Mean pooling is chosen for the pooling layer [14].

Once both sentences are passed through the pooling layers, we obtain two 768-dimensional vectors, u and v (fig. 7). After obtaining the vectors u and v , the similarity between them is directly computed using cosine similarity. The predicted similarity score is compared with the true value, and the model is updated using the MSE loss function [14]. Fig. 7 presents the SBERT architecture for calculating the similarity score.

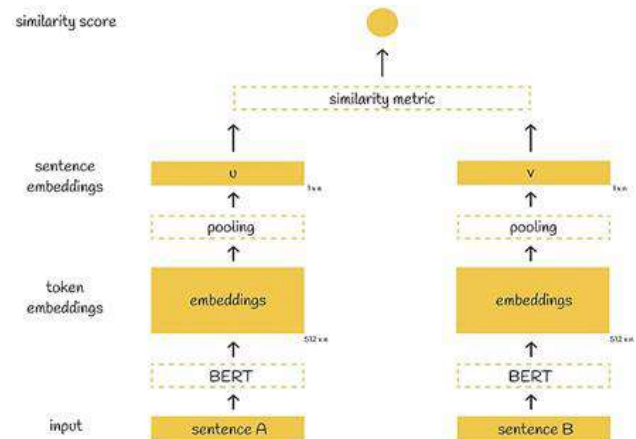


Fig. 7. SBERT architecture for similarity score calculation [15]

By using a similarity measure such as cosine similarity, semantically similar sentences can be found:

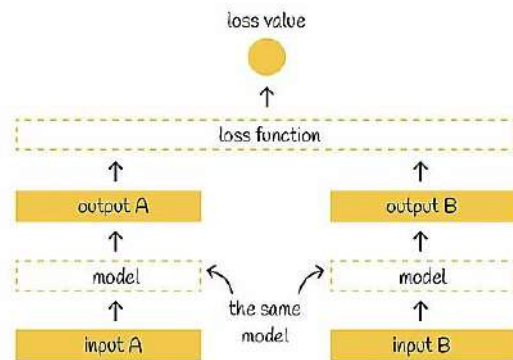


Fig. 6. Comparison of non-Siamese and Siamese architectures [15]

$$\cos(\theta) = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \cdot \sqrt{\sum_{i=1}^n B_i^2}}, \quad (4)$$

where A_i and B_i are coordinates of vectors A and B , respectively.

This similarity measure (4) can be efficiently computed on modern hardware, allowing SBERT to be used for both semantic similarity search and clustering [14].

Results and discussion. To evaluate the effectiveness of the proposed method, it was compared to the set similarity method, which consists of the following steps:

- tokenization;
- stop-word removal;
- word stemming.

The similarity of word sets in this method is calculated using the Jaccard coefficient.

The performance of these methods was tested on four business process models.

The first model considered is the business process called “Dispatch of goods” [16]. The text of this process is provided below, and the BPMN model of the business process is shown in fig. 8. First text: “If goods shall be shipped, the secretary clarifies who will do the shipping. If you have large amounts, special shipping will be necessary. In these cases, the secretary invites three logistic companies to make offers and she selects one of them. In case of small amounts, normal post shipment is used. Therefore, a package label is written by the secretary and a parcel insurance taken by the logistics department head if necessary. In the meantime, the goods can be already packaged by the warehousemen. If everything is ready, the packaged goods are prepared for being picked up by the logistic company”.

Based on this model, the following names of the tasks were defined:

- “Insure parcel”;
- “Write package label”;
- “Clarify shipment method”;
- “Get 3 offers from logistic companies”;
- “Select logistic company and place order”;
- “Package goods”;
- “Prepare for picking up goods”.

The first method showed a result of 38% similarity.

The proposed new method showed a result of 72% similarity.

The next model considered is “Credit Scoring Asynchronous” [16]. The text of this process is provided below, and the BPMN model of the business process is shown in fig. 9. Second text: “The sales clerks in a bank can use their software frontend to receive the credit-scoring for a certain customer. This starts a process in the banking system which communicates with the agency in the background. This process sends a scoring request to the agency right after the beginning. Then, the Agency does a first quick scoring (level 1). This will often lead to an immediate result which is then returned directly to the banking system within seconds. The banking process presents the result to the clerk sitting at the frontend. Sometimes the scoring cannot be determined immediately and takes longer. In this case the agency informs the banking process of the delay and then starts the level 2 scoring (which can take up to a couple of minutes). After the scoring result is determined, the information is sent back to the banking process. The banking process displays a message to the clerk when he receives information about the delay to check again later. As soon as the result arrives, it can be seen at the frontend”.

Based on this model, the following names of the tasks were defined:

- “Request credit score”;

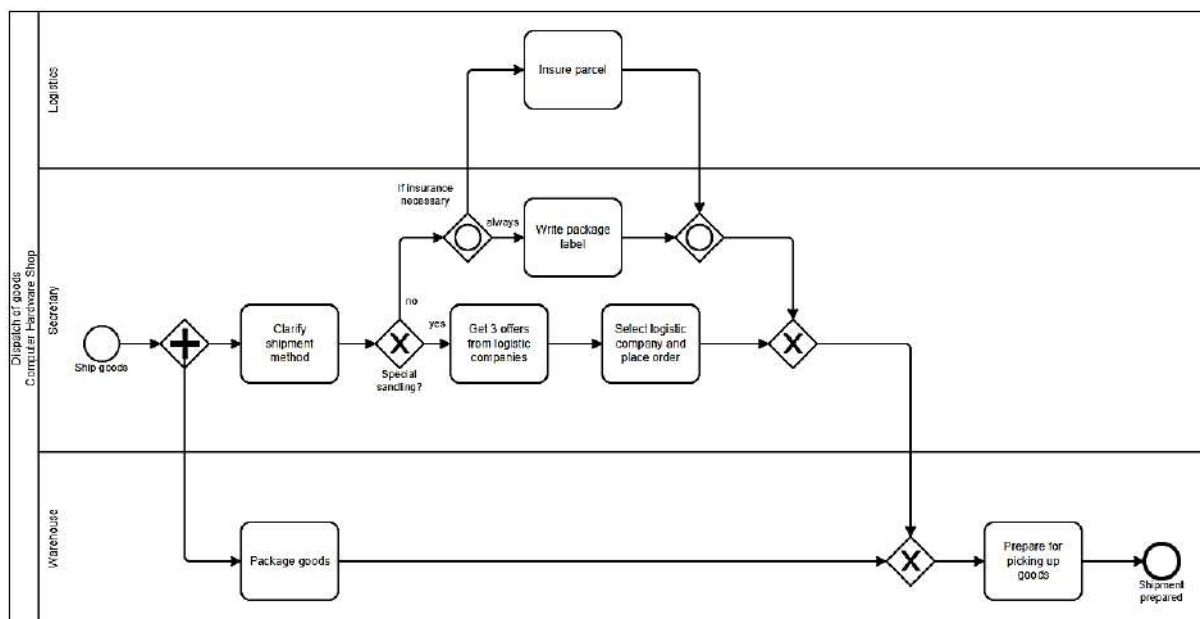


Fig. 8. Model 1 – “Dispatch of goods”

- “Send credit score”;
- “Report delay”;
- “Send credit score”;
- “Report delay”;
- “Send credit score”;
- “Compute credit score (level 2)”;
- “Compute credit score (level 1)”.

The first method showed a result of 12% similarity.

The proposed new method showed a result of 58% similarity.

The next model considered is “Recourse” [16]. The text of this process is provided below, and the BPMN model of the business process is shown in fig. 10. Third text: “If an insurant could be possibly subrogated against, I get information about that. I check that case and if the possibility is really there, I send a request for payment to the insurant and make me a reminder. If recourse is not possible, I close the case. When we receive the money, I make a booking and close the case. If the insurant disagrees with the recourse, I will have to check the reasoning of that.

If he is right, I simply close the case. If he is wrong, I forward the case to a collection agency. It the deadline for disagreement is reached and we have not received any money, I forward the case to the collection agency as well”.

Based on this model, the following names of the tasks were defined:

- “Check case”;
- “Send request for payment”;
- “Close case”;
- “Send reminder”;
- “Check reasoning”;
- “Close case”;
- “Hand over to collection agency”;
- “Make booking”;
- “Close case”.

The first method showed a result of 44% similarity.

The proposed new method showed a result of 61% similarity.

The next model considered is “Self Service Restaurant” [16]. The text of this process is provided

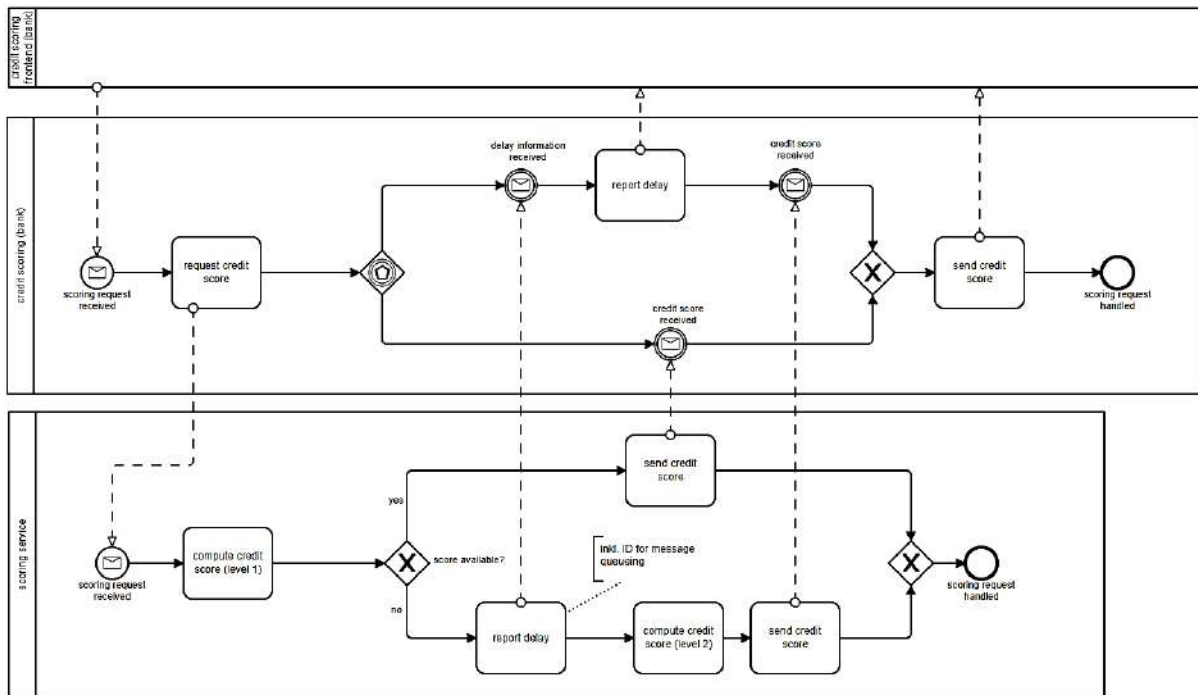


Fig. 9. Model 2 – “Credit Scoring Asynchronous”

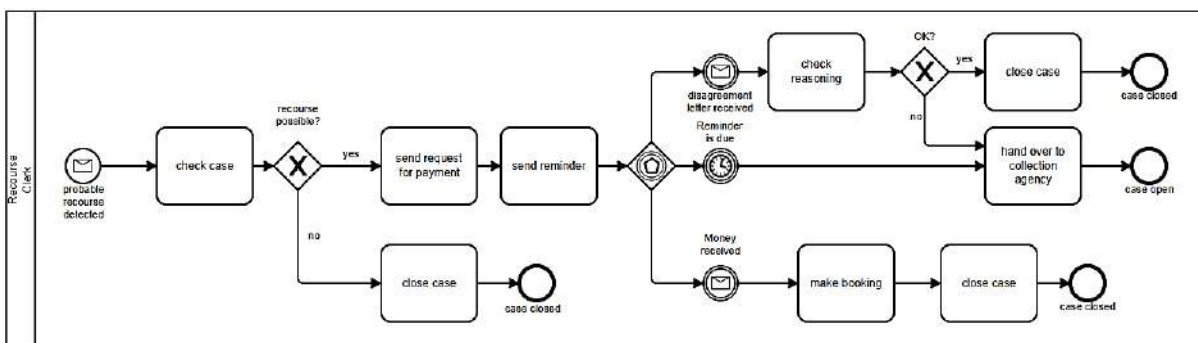


Fig. 10. Model 3 – “Recourse”

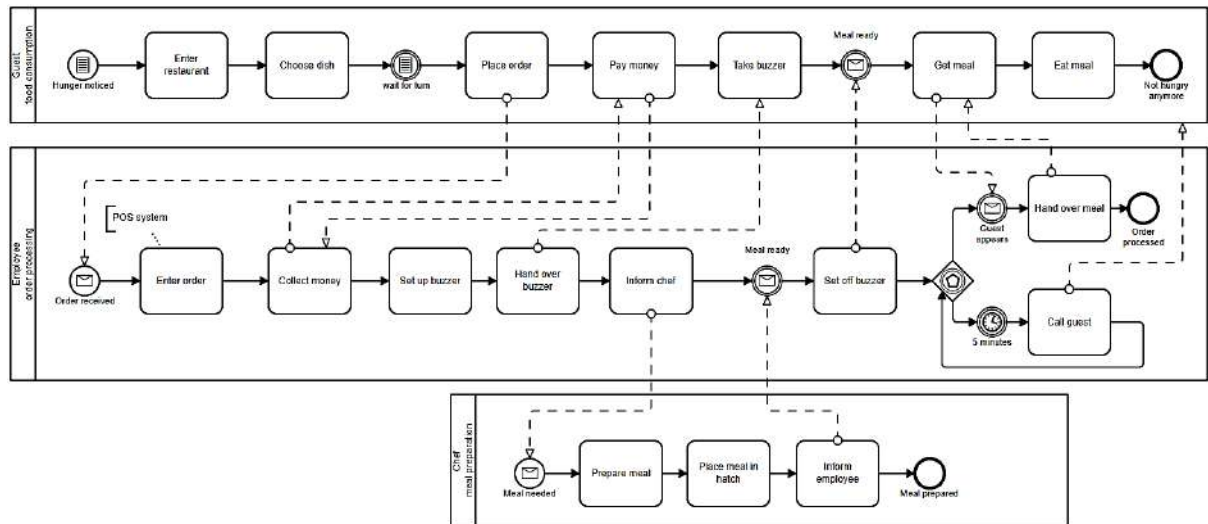


Fig. 11. Model 4 – “Self Service Restaurant”

below, and the BPMN model of the business process is shown in fig. 11. Fourth text: “A guest enters the restaurant when feeling hungry. He chooses a dish from the changing meal range and waits until it is his turn. Following this he places his order with the employee. The employee enters the order into the POS system and collects the money from the guest. After the payment, the employee sets up a buzzer and passes it on to the guest with the following information: When the buzzer rings, your dinner is ready. Afterwards the employee informs the chef of the new meal order. The chef prepares the meal and places it in the service hatch. He then informs the employee that he has placed the finished meal in the service hatch. As soon as the employee is aware that the meal is ready he sets off the guests buzzer. This is how the guest finds out that his meal is ready for collection. He can pick up his meal and eat it. As soon as the guest appears at the service hatch, the employee hands over his meal.

Should a guest not react to the buzzer, the employee calls for him after 5 minutes, if necessary several times in a row”.

Based on this model, the following names of the tasks were defined:

- “Enter restaurant”;
- “Choose dish”;
- “Place order”;
- “Pay money”;
- “Take buzzer”;
- “Get meal”;
- “Eat meal”;
- “Enter order”;
- “Collect money”;
- “Set up buzzer”;
- “Hand over buzzer”;
- “Inform chef”;
- “Set off buzzer”;
- “Hand over meal”;

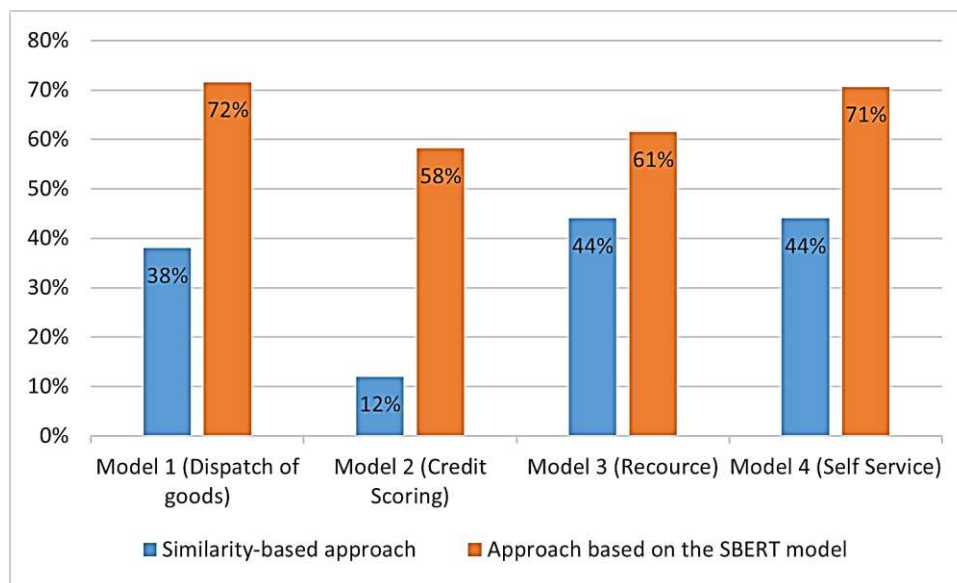


Fig. 12. Comparison results chart

- "Call guest";
- "Prepare meal";
- "Place meal in hatch";
- "Inform employee".

The first method showed a result of 44% similarity.

The proposed new method showed a result of 71% similarity.

Fig. 12 shows a bar chart with the comparison results.

As can be seen in fig. 12, the SBERT-based approach provides an average of 31% higher similarity compared to the set similarity approach. This is due to the fact that the set similarity approach cannot identify synonym words and semantic similarity, leading to a lower comparison score.

Conclusion and future work. In this paper, we have demonstrated the effectiveness of using a semantic-based approach for comparing business process models and their textual descriptions. The proposed method, based on SBERT, outperforms the traditional set similarity approach by an average of 31%, as shown in the comparative analysis of multiple business process models. The ability of SBERT to capture semantic similarity, including identifying synonyms and contextually relevant terms, provides a significant advantage over simple token-based approaches, which often overlook nuanced language variations.

The experimental results show that the SBERT-based approach improves the alignment of textual descriptions with business process models. This advancement enhances the overall quality and accuracy of business process documentation, leading to fewer errors, more clarity in process descriptions, and better communication between stakeholders.

In the future, we plan to improve our work with industry-specific terminology, which will allow for more accurate comparisons of models in specialized sectors. In addition, we plan to explore real-time applications of this method, such as using semantic analysis tools during the modeling process to provide immediate feedback on discrepancies between the BPMN model and its textual description.

References

1. Jošt G., Polančič G., Heričko M., Kocbek M. *Business process model and notation: The current state of affairs*. URL: <https://doi.org/10.2298/CSIS140610006K> (access date: 20.09.2024).
2. Von Rosing M., White S., Cummins F., De Man H. *Business process model and notation-BPMN*. URL: <https://doi.org/10.1016/B978-0-12-799959-3.00021-5> (access date: 20.09.2024).
3. Mroczek A., Wiśniewski P., Ligeza A. *Overview of Verification Tools for Business Process Models*. URL: <https://doi.org/10.15439/2017f308> (access date: 20.09.2024).
4. Ottensooser A., Fekete A., Reijers H., Mendling J., Menictas C. *Making sense of business process descriptions: An experimental comparison of graphical and textual notations*. URL: <https://doi.org/10.1016/j.jss.2011.09.023> (access date: 20.09.2024).
5. Qurashi A., Holmes V., Johnson A. *Document Processing: Methods for Semantic Text Similarity Analysis*. URL: <https://doi.org/10.1109/INISTA49547.2020.9194665> (access date: 20.09.2024).
6. Zhou S., Xu X., Liu Y., Chang R., Xiao Y. *Text Similarity Measurement of Semantic Cognition Based on Word Vector Distance Decentralization with Clustering Analysis*. URL: <https://doi.org/10.1109/ACCESS.2019.2932334> (access date: 20.09.2024).
7. Li Z., Chen H., Chen H. *Biomedical Text Similarity Evaluation Using Attention Mechanism and Siamese Neural Network*. URL: <https://doi.org/10.1109/ACCESS.2021.3099021> (access date: 20.09.2024).
8. Wu D., Zhang M., Shen C., Huang Z., Gu M. *BTM and GloVe Similarity Linear Fusion-Based Short Text Clustering Algorithm for Microblog Hot Topic Discovery*. URL: <https://doi.org/10.1109/ACCESS.2020.2973430> (access date: 20.09.2024).
9. Liu Y., Xu Q., Tang Z. *Research on Text Classification Method Based on PTF-IDF and Cosine Similarity*. URL: <https://doi.org/10.1109/ICIIBMS46890.2019.8991542> (access date: 20.09.2024).
10. Zhengfang H., MacHica I., Zhimin B. *Textual Similarity Based on Double Siamese Text Convolutional Neural Networks and Using*

References (transliterated)

1. Jošt G., Polančič G., Heričko M., Kocbek M. *Business process model and notation: The current state of affairs*. Available at: <https://doi.org/10.2298/CSIS140610006K> (accessed: 20.09.2024).
2. Von Rosing M., White S., Cummins F., De Man H. *Business process model and notation-BPMN*. Available at: <https://doi.org/10.1016/B978-0-12-799959-3.00021-5> (accessed: 20.09.2024).
3. Mroczek A., Wiśniewski P., Ligeza A. *Overview of Verification Tools for Business Process Models*. Available at: <https://doi.org/10.15439/2017f308> (accessed: 20.09.2024).
4. Ottensooser A., Fekete A., Reijers H., Mendling J., Menictas C. *Making sense of business process descriptions: An experimental comparison of graphical and textual notations*. Available at: <https://doi.org/10.1016/j.jss.2011.09.023> (accessed: 20.09.2024).
5. Qurashi A., Holmes V., Johnson A. *Document Processing: Methods for Semantic Text Similarity Analysis*. Available at: <https://doi.org/10.1109/INISTA49547.2020.9194665> (accessed: 20.09.2024).
6. Zhou S., Xu X., Liu Y., Chang R., Xiao Y. *Text Similarity Measurement of Semantic Cognition Based on Word Vector Distance Decentralization with Clustering Analysis*. Available at: <https://doi.org/10.1109/ACCESS.2019.2932334> (accessed: 20.09.2024).
7. Li Z., Chen H., Chen H. *Biomedical Text Similarity Evaluation Using Attention Mechanism and Siamese Neural Network*. Available at: <https://doi.org/10.1109/ACCESS.2021.3099021> (accessed: 20.09.2024).
8. Wu D., Zhang M., Shen C., Huang Z., Gu M. *BTM and GloVe Similarity Linear Fusion-Based Short Text Clustering Algorithm for Microblog Hot Topic Discovery*. Available at: <https://doi.org/10.1109/ACCESS.2020.2973430> (accessed: 20.09.2024).
9. Liu Y., Xu Q., Tang Z. *Research on Text Classification Method Based on PTF-IDF and Cosine Similarity*. Available at: <https://doi.org/10.1109/ICIIBMS46890.2019.8991542> (accessed: 20.09.2024).
10. Zhengfang H., MacHica I., Zhimin B. *Textual Similarity Based on Double Siamese Text Convolutional Neural Networks and Using*

- BERT for Pre-training Model.* Available at: <https://doi.org/10.1109/ICAIBD55127.2022.9820371> (accessed: 20.09.2024).
11. Dazhan G., Iskakov A., Kenzhegaliev M., Bui D. *Dynamic Text Modeling and Categorization Framework based on Semantics Extraction and Similarity Checking.* Available at: <https://doi.org/10.1109/CSCI58124.2022.00132> (accessed: 20.09.2024).
12. Toshevska M., Stojanovska F., Kalajdjieski J. *Comparative Analysis of Word Embeddings for Capturing Word Similarities.* Available at: <https://doi.org/10.5121/csit.2020.100402> (accessed: 20.09.2024).
13. *Business Process Model and Notation (BPMN), Version 2.0.* Available at: <https://www.omg.org/spec/BPMN/2.0/PDF> (accessed: 20.09.2024).
14. Reimers N., Gurevych I. *Sentence-BERT: Sentence embeddings using siamese BERT-Networks.* Available at: <https://doi.org/10.18653/v1/d19-1410> (accessed: 20.09.2024).
15. *Large Language Models: SBERT – Sentence-BERT.* Available at: <https://towardsdatascience.com/sbert-deb3d4aef8a4> (accessed: 20.09.2024).
16. *BPMN for research.* Available at: <https://github.com/camunda/bpmn-for-research> (accessed: 20.09.2024).

Received 15.11.2024

УДК 004.94

О. В. РУДСЬКИЙ, Національний технічний університет «Харківський політехнічний інститут», студент, м. Харків, Україна; e-mail: oleksandr.rudskyi@cs.khpi.edu.ua; ORCID: <https://orcid.org/0009-0001-1130-9957>

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

Т. Є. ГОНЧАРЕНКО, кандидат педагогічних наук (PhD), доцент, Національний технічний університет «Харківський політехнічний інститут», завідувач кафедри іноземних мов, м. Харків, Україна; e-mail: tetiana.goncharenko@khpi.edu.ua; ORCID: <https://orcid.org/0000-0001-6630-307X>

І. П. ГАМАЮН, доктор технічних наук, професор, Національний технічний університет «Харківський політехнічний інститут», професор кафедри програмної інженерії та інтелектуальних технологій управління, м. Харків, Україна; e-mail: ihor.hamaiun@khpi.edu.ua; ORCID: <https://orcid.org/0000-0003-2099-4658>

ІНТЕЛЕКТУАЛЬНА ТЕХНОЛОГІЯ ОЦІНЮВАННЯ СЕМАНТИЧНОЇ ПОВНОТИ МОДЕЛЕЙ БІЗНЕС-ПРОЦЕСІВ

У цій статті авторами представлено метод порівняння моделей бізнес-процесів з їх текстовими описами на основі використання семантичного підходу з використанням моделі SBERT (Sentence-Bidirectional Encoder Representations from Transformers). Моделі бізнес-процесів, зокрема створені за стандартом BPMN (Business Process Model and Notation), мають вирішальне значення для оптимізації організаційної діяльності. Забезпечення узгодженості між цими моделями та їхніми текстовими описами має важливе значення для підвищення точності та зрозумілості бізнес-процесів. Традиційні методи схожості множин, які покладаються на токенизацію та базове зіставлення слів, не можуть охопити глибші семантичні зв'язки, що призводить до нижчої точності порівняння. Запропонований підхід дозволяє розв'язати цю задачу, за рахунок використання моделі SBERT для оцінки семантичної подібності між текстовим описом і BPMN-моделлю бізнес-процесу. Експериментальні результати демонструють, що метод на основі SBERT перевершує традиційні методи, засновані на показниках подібності, в середньому на 31%, пропонуючи більш надійні та контекстуально відповідні порівняння. Здатність SBERT фіксувати семантичну схожість, включаючи ідентифікацію синонімів і контекстуально релевантних термінів, забезпечує значну перевагу перед більш простими підходами на основі токенизації, які часто не помічають нюансів мовних варіацій. Експериментальні результати демонструють, що підхід на основі SBERT, запропонований у цьому дослідженні, покращує узгодженість між текстовими описами та відповідними моделями бізнес-процесів. Таке удосконалення дозволяє підвищити загальну якість і точність документації бізнес-процесів, що призводить до зменшення помилок, запровадження кращої зрозумілості описів бізнес-процесів, а також кращої взаємодії між усіма зацікавленими сторонами. Загальні результати, отримані в цьому дослідженні, сприяють підвищенню якості та узгодженості моделей бізнес-процесів BPMN і відповідної документації.

Ключові слова: моделювання бізнес-процесів, BPMN, семантична подібність, SBERT, порівняння текстів, оптимізація бізнес-процесів, обробка природної мови.

Повні імена авторів / Author's full names

Автор 1 / Author 1: Рудський Олександр Вадимович / Rudskyi Oleksandr Vadymovych

Автор 2 / Author 2: Копп Андрій Михайлович / Kopp Andrii Mykhailovych

Автор 3 / Author 3: Гончаренко Тетяна Євгенівна / Goncharenko Tetiana Yevhenivna

Автор 4 / Author 4: Гамаюн Ігор Петрович / Gamaiun Igor Petrovych

V. I. ZIUZIUN, Candidate of Technical Sciences (PhD), Docent, Associate Professor at the Department of Management Technologies, Taras Shevchenko National University of Kyiv, Kyiv, Ukraine; e-mail: vadym.ziuziun@knu.ua; ORCID: <https://orcid.org/0000-0001-6566-8798>

N. A. PETRENKO, Student, Taras Shevchenko National University of Kyiv, Kyiv, Ukraine; e-mail: nikita.petrenko@knu.ua; ORCID: <https://orcid.org/0009-0006-3921-8412>

APPLICATION OF OPTICAL CHARACTER RECOGNITION AND MACHINE LEARNING TECHNOLOGIES TO CREATE AN INFORMATION SYSTEM FOR AUTOMATIC VERIFICATION OF OFFLINE TESTING

During the learning process in any field, testing and monitoring the knowledge of students or other learners is an essential part. Teachers often spend considerable time grading large volumes of standardized tests. While online testing systems have been developed to streamline this process, offline paper tests remain popular as they do not require access to computers, electricity, or a stable internet connection. Offline testing is often considered one of the most representative methods for assessment, but it leads to repetitive work for teachers during the grading process. To save time, some educators use test sheets to structure responses, simplifying grading tasks. Consequently, developing a system that automates the grading of offline tests has become increasingly relevant. The purpose of this research was to develop an information system (web platform) that simplifies the offline test grading process using optical character recognition technologies powered by machine learning algorithms. The object of this research is the processes and functionality involved in creating an information system for the automated grading and evaluation of offline tests. The scientific novelty lies in integrating machine learning algorithms with modified image processing algorithms to create a system capable of analyzing and grading a wide range of offline test tasks, including open-ended, closed-ended, sequence identification, and multiple-correct-answer questions. The practical significance of this research is the development of a web platform to automate offline test grading through optical character recognition and machine learning technologies, reducing teachers' time spent on grading, enabling analysis and improvement of educational programs, supporting various test types, and promoting scientific and technological advancement in education. The developed system can recognize handwritten text from photos, create an array of responses, and compare them to the answers provided by the teacher. This approach significantly reduces the time teachers spend on grading tests. For user convenience, a minimalist interface was created, granting access to all main system functions with intuitive controls. A detailed description of the developed algorithms and machine learning models is provided. This project offers broad potential for further development, including integration with other educational platforms, enhancements in recognition technology, and system scalability.

Keywords: information system, web platform, IT project, machine learning, neural networks, algorithm, IAM dataset, optical character recognition, testing, educational process.

Introduction. Before starting the development of the information system, it is helpful to conduct a brief analysis of software solutions aimed at simplifying the testing process. It is important to note that while there is a wide range of online testing systems, there are relatively few programs specifically specializing in handwritten answer recognition. Among the most popular and successful ones are ZipGrade, GradeCam, and Essay Grader.

The ZipGrade mobile app is designed to automate the test grading process and offers users a range of functionalities [1]. At first glance, it is a convenient and efficient tool, but it has limitations in its free version. While the system is user-friendly, it has several drawbacks: limited flexibility in creating tests; no option to assign specific point values to questions; the need to print answer sheets for the entire group; restrictions on test types (inability to create closed-ended tests, multiple-answer tests, matching tests, or sequence-based tests) and limitations on the number of test versions available.

GradeCam is a web application that offers a range of functionalities beyond those of the previous product. Like ZipGrade, its primary function is to automate the test grading process. However, GradeCam distinguishes itself with a more appealing analytics interface and the ability to recognize text [2]. This system is designed to automate the entire educational process – from creating tests and storing them in a database to assigning grades, monitoring student performance, and generating reports for administration, parents, and students.

Among the advantages of GradeCam are the following: it is easier to create tests on a computer than on a mobile device; the response format is adapted for various subjects; it allows scanning of student work using any camera: on a mobile device, laptop, or computer, as well as an external camera; there is an option to save images of questions for subsequent manual grading by the instructor; it recognizes handwritten text in open-ended answers; all grades and student work are stored; it generates reports on group performance and individual student success; the system automatically assigns final grades based on test results.

The drawbacks of this system include: limitations in that it only works with forms created within the system itself; the need to print a large number of answer sheets; the requirement to input open-ended questions in printed letters, placed separately in each cell; lack of support for languages other than English; high product prices, which depend on the number of students or the volume of assessed work.

Although the system offers numerous useful features, instructors are still required to print specialized answer sheets, resulting in additional time and financial costs. Moreover, this complicates the administration of tests without prior preparation, such as if there is free time during a class. For individual instructors, using this system can become costly if their university or educational institution does not collaborate with the platform.

© Ziuziun V. I., Petrenko N. A., 2024



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.



Essay Grader is an application designed to simplify the process of grading written works, particularly comprehensive handwritten texts rather than just closed-ended tests. This system allows educators to upload images of essays, check for plagiarism, identify grammatical errors, and provides an interface for digital editing of sentences, along with the ability to add comments and send results to students [3]. Although the system includes handwriting recognition, it is not intended for automated test grading and is exclusively focused on assisting in the grading of essays. This application does not have features for reporting, storing student grades, or analyzing statistics. Therefore, while this system shares a similar theme, it is not an equivalent to the project being developed.

Analysis of recent research and publications. After analyzing well-known solutions in the field of test grading, it can be noted that the issue of effective and rapid grading of test items remains relevant.

However, many existing software products aimed at addressing this issue have rigid logic that limits users in customizing tests to their needs. For example, they may not allow the creation of open-ended tests, matching questions, or sequencing questions, and they may restrict the ability to change the point value for specific items. Even when a system permits flexible test creation, it may still impose limitations on the ways in which answers can be entered.

The process of handwriting recognition consists of the following subtasks [4]: extracting the page from the photograph and establishing text boundaries; processing the image by enhancing contrast and line thickness; compensating for lighting by illuminating dark areas; aligning the text tilt; aligning the line tilt; segmenting the page into lines; segmenting lines into words; recognizing words; and correcting recognized words and sentences using natural language processing techniques.

There are algorithms [5] that allow for the immediate identification and recognition of words on a page, bypassing the corresponding steps outlined above. In such cases, the technique of «transfer learning» can be applied, meaning that a model is initially trained on one dataset and then fine-tuned on real data. However, since the data collection process is time-consuming and labor-intensive, it is essential to choose more robust algorithms. Thus, statistical approaches will be used to address the problem of text segmentation. During the analysis of existing text segmentation solutions, a module was found that implements the methodologies from the articles referenced [6, 7]. Despite being published in 2007 and 1999, respectively, these articles remain competitive, yielding good results and operating relatively quickly. The module also includes implementations of algorithms for document scanning, finding the contour of an image using a four-point transformation; binarization techniques such as Niblack, Sauvola, and Wolf; lighting compensation according to the work materials; segmentation based on the algorithms referenced [8, 9]; and text tilt alignment according to the referenced methodology.

Additionally, all of the aforementioned algorithms are implemented in C++ and utilize OpenCV, allowing them to operate much more quickly.

Since there are no ready-made modules for handwriting recognition, an analysis of machine learning algorithms was conducted. Over the past several decades, researchers have developed various models for this task. Historically, this problem has been viewed as a sequence alignment task: a sequence of features derived from the input image is compared to a sequence of letters, Hidden Markov Models were used for this purpose. However, the primary drawback of such models is the limitation in utilizing the full context of the data due to the nature of Markov models, where each observation depends only on the current state and is independent of previous states [9]. This limitation has been overcome through the use of networks such as RNNs and LSTMs, which can «remember» long sequences. A true breakthrough occurred with the invention of Connectionist Temporal Classification (CTC) [10] and its combination with recurrent networks.

As noted in the research [11, 12], models composed of layers such as convolutional layers (CNNs), recurrent layers (RNNs), specifically the LSTM layer, and the Connectionist Temporal Classification (CTC) layer are the most effective. Two metrics were used to determine effectiveness: Character Error Rate (CER) and Word Error Rate (WER). Therefore, for the implementation of this IT project, a model was chosen that demonstrates the highest efficiency in text recognition.

Based on the conducted research [13, 14], it can be concluded that the IAM (IAM OnLine Handwriting Database) dataset is the best choice for training the model.

The IAM is a collection of handwritten English text on a white background. It features the handwriting of 657 writers who contributed to the creation of the dataset. The database contains a total of 1,539 scanned pages of text, 5,685 isolated and annotated sentences, 13,353 isolated and annotated lines of text, as well as 115,320 individual words [15].

In addition to the aspects of direct IT product development, it is also important to analyze the approaches to managing IT projects and making managerial decisions. These issues are discussed in the works of [16, 17].

Presentation of the main material. The information system is being developed to optimize the quality control of education, and it will include the following key elements:

- Participants (teachers (instructors), students, learners).
- Methods of knowledge assessment (closed tests, open tests, and matching questions).
- Information system (web pages, database, test grading module).
- Reporting documents (logs, records).

As mentioned earlier, the proposed information system is considered in the context of quality control of knowledge, specifically to optimize this process during the grading of offline tests. Therefore, let us first examine the mathematical processes involved in grading tests manually and automatically.

Let's define the variables:

- Set of instructors (1):

$$T = \{t_1, t_2, \dots, t_n\}, \quad (1)$$

where t_1, t_2, \dots, t_n are the instructors.

- Set of courses (2):

$$C = \{c_1, c_2, \dots, c_n\}, \quad (2)$$

where c_1, c_2, \dots, c_n are the courses.

- Set of questions (3):

$$Q = \{q_1, q_2, \dots, q_n\}, \quad (3)$$

where q_1, q_2, \dots, q_n are the questions.

- Set of students (4):

$$S = \{s_1, s_2, \dots, s_n\}, \quad (4)$$

where s_1, s_2, \dots, s_n are the students.

- Set of groups (5):

$$G = \{g_1, g_2, \dots, g_n\}, \quad (5)$$

where g_1, g_2, \dots, g_n are the groups.

Let's describe the constraints:

- Constraints for the category of instructors (6):

$$\text{teaches: } C \rightarrow T, \text{ then } \forall c \in C, \exists! t \in T: \text{teaches}(c) = t \quad (6)$$

- Constraints for the category of questions (7):

correct.answer: $Q \rightarrow A$, then

$$\forall q \in Q, \exists! a \in A : \text{correct.answer}(q) = a;$$

$$|\text{correct.answers}(q)| = 1, \forall q \in Q. \quad (7)$$

Let's describe mathematically the test grading functions using the information system and manually.

Common:

- Function (8) associates a combination of instructor and course with a set of tests:

$$f_{\text{test}} : T \times C \rightarrow \text{Tests}. \quad (8)$$

- Function (9) maps a test to a set of questions:

$$f_{\text{questions}} : \text{Tests} \rightarrow 2^Q, \quad (9)$$

where 2^Q is the power set of Q , representing all possible combinations of questions.

- Function (10) maps a student and a test to a real number representing the student's score on that test:

$$f_{\text{marks}} : S \times \text{Tests} \rightarrow R. \quad (10)$$

Manual grading of assignments by the instructor.

Let $T_m(q,s)$ be the time the instructor spends grading question q for a single student s . The total time for manual grading of all questions for all students is calculated using formulas (11):

$$T_{m(\text{total})} = \sum(T_m(q,s), \forall q \in Q, \forall s \in S). \quad (11)$$

Grading using the information system. Let $T_a(q)$ represent the time the automated system takes to grade question q . This time should be significantly less than

$T_m(q,s)$, as the system should maintain a consistent (or negligibly variable) grading time regardless of question complexity or answer length. The total grading time by the automated system for all questions for all students is calculated using formulas (12):

$$T_{a(\text{total})} = \sum(T_a(q), \forall q \in Q). \quad (12)$$

Thus, we have provided formulas for calculating the time required for grading tests both manually and with the application. Now, let us present the formulas for determining the efficiency gain. The efficiency gain G can be determined using formulas (13):

$$G = \left(\frac{E_m - E_a}{E_m} \right) \cdot 100\%, \quad (13)$$

where E_m – efficiency of manual test grading;

E_a – efficiency of grading using the application.

The efficiency of manual test grading can be determined using formulas (14):

$$E_m = \left(\frac{T_m}{N} \right), \quad (14)$$

where T_m – time for manual test grading;

N – number of tests graded.

The efficiency of test grading using the application can be determined using formulas (15):

$$E_a = \left(\frac{T_a}{N} \right), \quad (15)$$

where T_a – time for grading tests using the application;

N – number of tests graded.

Formulas (16) can be used to calculate the time saved by using the application:

$$E = \sum(T_m(q,s) - |S|) \cdot \sum(T_a(q), \forall q \in Q, \forall s \in S). \quad (16)$$

Thus, we have obtained all the necessary formulas that describe the processes of grading tests manually and automatically. We have also derived the formulas to calculate the efficiency gain from using the application, as well as the formula for calculating the time saved.

Now, let us mathematically define the task of the application, specifically to automatically grade and assign scores for the test. Formulas (17) provides a mathematical description of how to calculate the scores based on the responses:

$$\text{score}(A_{s(\text{test})} Q_{\text{test}}) = \sum(\text{points}(q) \cdot \delta(a_q, A_{s(\text{test})}[q])), \quad (17)$$

where $A_{s(\text{test})}$ – the set of answers provided by the student for a specific test;

$\text{points}(q)$ – represents the number of points for question q ;

Q_{test} – the set of questions included in the specific test;

δ – the Kronecker delta function [18].

Database of the system. Since the system being developed will work with a large amount of unstructured data, MongoDB was chosen as the database. MongoDB is

a document-oriented NoSQL database [19]. Here's why MongoDB is the best fit:

- Document-oriented storage. Since MongoDB is a NoSQL database that stores data in a document-oriented format similar to JSON (known as BSON), it is well-suited for storing and manipulating large volumes of text data. This creates a natural correspondence between the data storage format and the data itself.

- Schema-less nature. MongoDB does not enforce schemas, meaning you can store documents without needing to define their structure in advance. This flexibility is particularly beneficial for text data, which may have varying structures, allowing different documents to be stored in the same collection without adhering to a strict schema.

- Designed for scalability. MongoDB is built for scalability, supporting sharding (distributing data across multiple servers) and replication (creating copies of data on different servers). This ensures that the database can handle increasing volumes of text data and user queries without significant performance degradation.

- Aggregation framework. MongoDB provides a powerful aggregation framework that allows for the processing and analysis of large amounts of text data in the database. This can be utilized for tasks such as text analysis, data transformation, and generating summaries or reports based on text data.

- Development efficiency. Given the popularity of JSON in modern web applications, using MongoDB can simplify the development process. Developers can work with a familiar format, reducing the complexity of data conversion between the application and the database.

Structural diagram of the system. The software will feature a client-server architecture, as depicted in fig. 1.

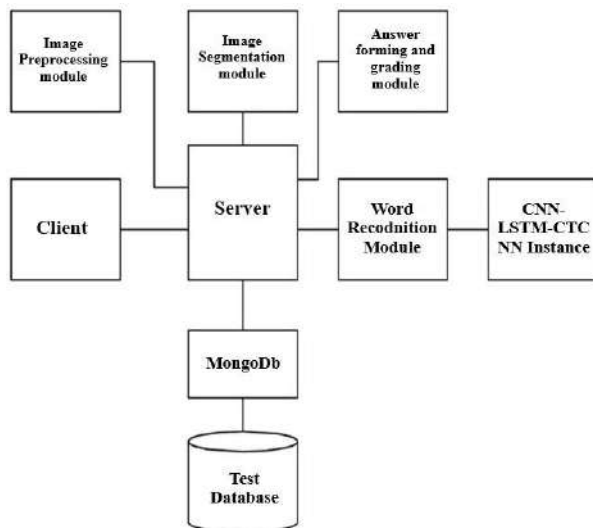


Fig. 1. Structural diagram of the system

The client provides the user interface for interacting with the server, sends requests to the server, receives responses from it, and loads the results in a web browser. The server receives requests from the client, processes them, generates a webpage, and returns it to the client.

In addition, it performs the following functions:

- Interacts with the database containing information about user-created tests.
- Interacts with the image preprocessing module.
- Interacts with the text segmentation module.
- Interacts with the text recognition module.
- Interacts with the response generation and grading module.

Algorithm for processing incoming images. Since the images in the dataset selected for training the neural network contain clear, high-contrast handwritten text with large spaces between lines and words, it is necessary to transform each image to resemble the images from the training dataset as closely as possible. To test this part of the project, data from real English language tests were collected. Unfortunately, only previously graded work was available, which contains corrections in red ink. At this stage, it was necessary to remove the red color from the images using OpenCV tools. The image processing algorithm is outlined below (fig. 2).

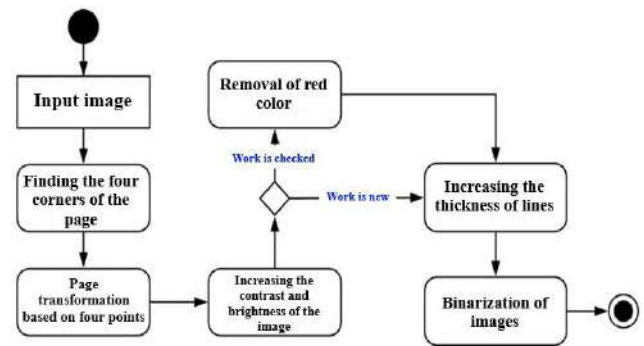


Fig. 2. Image preprocessing algorithm

The image processing algorithm shown in Figure 2 includes the following steps:

- Page boundary detection and corner coordinates fixing. The page boundary is identified by finding the convex hull using OpenCV's convexHull function, which is then approximated to a rectangle with the approxPolyDP function.

- Four-point perspective transformation. The page is transformed based on the four detected corner points (as mentioned above) to obtain an image as if taken at a 90° angle to the page.

- Contrast and brightness enhancement. This step clarifies the distinction between text and paper while reducing potential background defects, such as paper creases, grid markings, and text bleed-through from the opposite side.

- Red color removal. This is applied only to graded papers with red marks or those with a red margin line separating the page borders.

- Line thickening. After previous steps, some lines may appear thin or broken. They are thickened to restore word integrity and match the training data as closely as possible.

- Image binarization. Wolf's binarization method is used to convert the image into a 2D array of zeros and ones, which serves as the input vector for the segmentation algorithm.

Four-point transformation algorithm. If a user takes a photo of a handwritten document at an angle other than 90° to the page, the word recognition algorithm will not function correctly, as the computer perceives the same word at different angles as entirely different words. This issue can be addressed during the training data preparation stage by introducing all possible distortions to the images, allowing the machine learning algorithm to adapt to such variations and accurately classify even highly skewed words.

However, a much faster and more efficient approach is to use perspective transformation. This transformation was implemented using functions from the OpenCV library.

To apply the transformation, it is necessary to specify the coordinates of the four corners of the page (fig. 3), which are identified during the initial processing stage, as well as the coordinates of the new image's corners, which are defined as follows.

Let *A* be the top-left corner of the page, *B* the top-right, *C* the bottom-left, and *D* the bottom-right, as shown in fig. 3. Their coordinates will be as follows (18):

$$(x_a, y_a), (x_b, y_b), (x_c, y_c), (x_d, y_d). \quad (18)$$

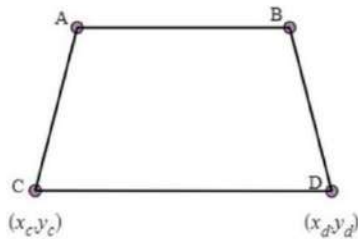


Fig. 3. Schematic representation of a page photographed at an angle

We find the side lengths using formulas (19-22):

$$AB = \sqrt{(x_b - x_a)^2 + (y_b - y_a)^2}. \quad (19)$$

$$CD = \sqrt{(x_d - x_c)^2 + (y_d - y_c)^2}. \quad (20)$$

$$CA = \sqrt{(x_a - x_c)^2 + (y_a - y_c)^2}. \quad (21)$$

$$BD = \sqrt{(x_d - x_b)^2 + (y_d - y_b)^2}. \quad (22)$$

We calculate the width and height of the new image using formulas 23 and 24:

$$W = \max(AB, CD). \quad (23)$$

$$H = \max(CA, BD). \quad (24)$$

The coordinates of the corners of the new image are then defined as follows: the top-left corner has coordinates (0, 0), the top-right corner has coordinates (W, 0), the bottom-left corner has coordinates (0, H), and the bottom-right corner has coordinates (W, H). The `getPerspectiveTransform` function [20] is then applied to obtain the transformation matrix, followed by the

`warpPerspective` function to apply this matrix and generate the new image. The algorithm's schematic is shown in fig. 4.

Word recognition algorithm for images. The machine learning model used for image classification is a neural network consisting of five convolutional layers and two recurrent layers, followed by a CTC (Connectionist Temporal Classification) layer. Formally, this model can be represented as a function that maps an image – specifically, a matrix of size $W \times H$ – to a sequence of characters (c_0, c_1, \dots, c_L) with a length ranging from 0 to *L*.

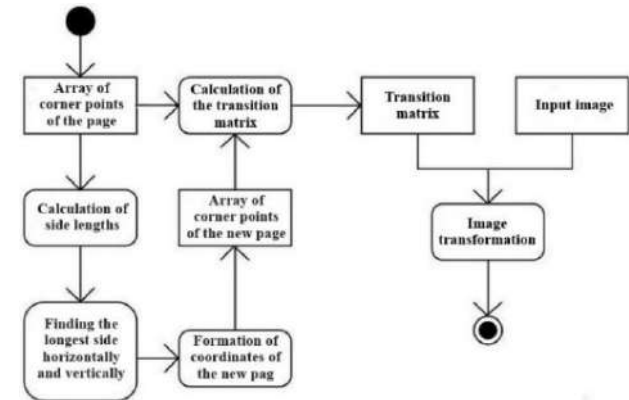


Fig. 4. Four-point transformation algorithm

Thus, text recognition occurs character by character, enabling the classification of words that were not present in the training data, as well as names and misspelled words. The architecture of the neural network is schematically illustrated in fig. 5.

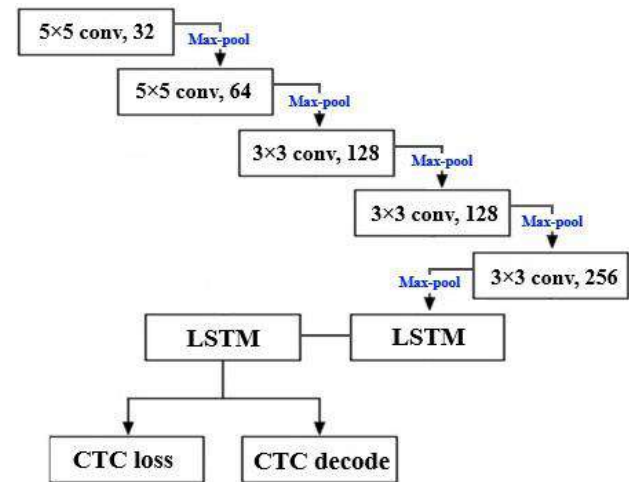


Fig. 5. Neural network architecture

The algorithm takes a binary image of size 128×32 as input. Since not all word images will have this exact size, the image is resized until either the height reaches 32 pixels or the width reaches 128 pixels. After resizing, the image is padded with white pixels to make it 128×32. The image then enters the convolutional layers, which consist of five layers trained to identify relevant features in the image. Each layer includes three operations: first, a filter of size 5×5 is applied in the first two layers and 3×3 in the last three layers. Then, the nonlinear activation function ReLU, defined by formulas (25), is applied:

$$f(x) = \max(0, x), \quad (25)$$

where x – input value of the neuron.

Next is the pooling layer, which reduces the dimensionality of the input image. In this model, max-pooling is used, which retains only the maximum value within each 4×4 region, thus capturing a generalized version of the input image's features. This is highly beneficial, as minor variations in features identified by the convolutional layer won't affect the resulting feature matrix. After passing through these five layers, we obtain a feature matrix of size 32×256 , which is then fed into the recurrent layers. This matrix contains 32 sequences, each with 256 features per time step.

LSTM layers were chosen as the recurrent layers because they can propagate information over long distances and learn long sequences of characters. The output of the recurrent layers is a matrix of size 32×80 , where 32 represents the number of time steps, and 80 represents the total number of unique characters in the IAM dataset (79) plus 1 additional symbol required for Connectionist Temporal Classification (CTC). During training, CTC receives the 32×80 matrix along with the correct transcription of the text in the image and calculates the CTC loss. During prediction, CTC only receives the feature matrix and decodes it into text. RMSProp was chosen as the optimizer. The trained model achieved a Character Error Rate (CER) of 10.63% on the validation dataset.

The answer array generation algorithm. To facilitate comparison with the reference answer, an array is created containing answers for each question in a row format. After processing the image, the image is segmented into rows. Each row is then split into words. At this stage, we know the order of the rows in the text and the order of words within each row.

After obtaining an array of images for each word in the current line, we begin recognition. If the word in the image is a number and appears first in the line, it is assumed to indicate the question number for which this is the beginning of an answer. If it appears second or later in the line, it is added to the answer as regular text.

If a new answer start is identified, all subsequent words following this number are part of the answer until another number is found at the beginning of a line.

The algorithm produces an array where each element represents the student's answer to a specific question, which will then be used for test grading in the next step. The algorithm for generating the answer array is shown in fig. 6.

Answer verification algorithm. After image processing, text segmentation, word recognition, and answer array generation, the test grading begins. The system supports the following question types: open-ended, single-answer multiple-choice, sequence ordering, and multiple-answer (or set) questions. The «set» question type can be further divided into two cases: where points are awarded for each correct answer, or where points are only awarded if all answers are correct.

The simplest case – single-answer multiple-choice – is graded by comparing letters written in lowercase.

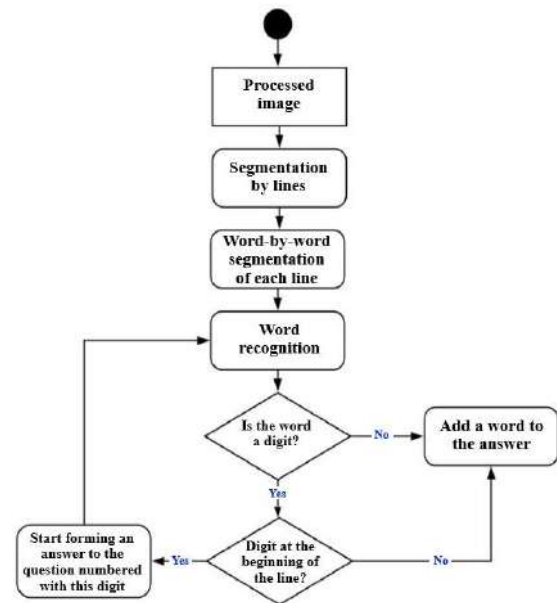


Fig. 6. Answer array generation algorithm

The simplest case – single-answer multiple-choice – is graded by comparing letters written in lowercase.

If a multiple-answer question (set type) contains several correct answers, there are two possible grading scenarios. If the instructor specified that all correct answers must be identified, the following steps are performed:

- Convert letters to lowercase.
- Sort answers alphabetically.
- Concatenate into a single string.
- Compare with the correct answer.

If the instructor intends to award points for each correct answer identified, the grading process is as follows:

- Convert letters to lowercase.
- Calculate the intersection between the sets of correct answers and the student's answers.
- Count the number of correct answers.

If the question requires sequence ordering, it is graded similarly to a set-type question where all correct answers must be identified. After checking multiple-choice questions, the student is awarded the points specified by the instructor for these questions. For grading open-ended text responses, the «cosine similarity» metric and Tf-idf vectorization from the sklearn package are used. The verification algorithm proceeds as follows:

- Create an instance of the TfidfVectorizer class.
- Use the fit_transform function to obtain two vectors of length n , representing the correct answer and the student's response.
- Construct a matrix $A_{2 \times n}$, where the first row is the vector for the correct answer, and the second row is the vector for the actual response.
- Calculate a new matrix as the product of matrix A and the transpose of matrix A using formulas (26):

$$C = A \cdot A^T. \quad (26)$$

Matrix C has the following form (27):

$$\begin{pmatrix} 1 & s \\ s & 1 \end{pmatrix}, \quad (27)$$

where s – measure of similarity between the correct answer and the actual response. If the answers match completely, $s = 1$. If there is no overlap between them, $s = 0$. Open-ended responses are not graded; however, the instructor is informed of the similarity level.

Peculiarities of server-side implementation. The server side of the application is implemented using the Flask microframework, which has minimal out-of-the-box functionality and allows the developer to extend it as needed. In this case, the MVC (Model-View-Controller) model is implemented. This architectural design pattern for user interfaces aims to separate the internal data structure from its external representation and increase code modularity.

The model's module is the central component of MVC [21]. It is responsible for managing the data of the web platform using the MongoEngine ORM library, which allows manipulation of data from the MongoDB database. This module contains the definitions of all classes corresponding to the collections and documents in the MongoDB database.

The controller is a specific abstraction composed of three parts:

- Initialization. This involves creating an instance of the Flask application, configuring the necessary settings, and connecting to the database.
- Routing. This defines the URL paths through which the server loads specific pages using decorators in Python. Each route is associated with a controller, specifically with a particular action of the controller. During the execution of this action, models are used to retrieve the necessary data from the database, which is then sent to the views module for rendering.
- Execution. This is the process of launching the web application

The routes.py file contains the routing methods associated with the Model and View.

The views module is an abstraction responsible for the user interface. It receives data from the model via the controller and determines how to present it. For this purpose, the Flask framework provides the Jinja2 template engine for generating HTML pages, which also allows for template inheritance. The base template is the top navigation bar defined in base.html, which is inherited by the left navigation bar in dashboard.html, and all other templates inherit from it as well.

Design and Page Content. After registering and logging into the system, the user sees the main navigation panel, described in dashboard.html, which inherits from the authorization panel page, base.html. All other pages in the system inherit from dashboard.html, creating a four-level template inheritance structure. This approach maintains a consistent page style and reduces the amount of code.

The navigation panel contains essential controls, helping to minimize the number of intermediate steps required to work within the system. The menu items include links to the test page, group lists, statistics page, and new test creation page, along with graphical elements like Feather icons for each tab to aid in user navigation.

The group list page includes a form for uploading a new list from a CSV file, as well as a list of all groups that

have already been created. The group import form consists of a text field for entering the group number and a file upload field. If the current user hasn't created any groups yet, they will see a message in place of the group list indicating this.

If the user attempts to upload a file in any format other than CSV, an error message will appear indicating a file read issue. Once a group is added, it automatically appears in the list as a link. By clicking on this link, the user is directed to a page showing the current grades for that group across all tests.

The new test creation page contains a form with fields for the course name, test description, and a section for creating new questions (fig. 7). This section includes the following fields:

- Question Type. A dropdown menu with options such as single-choice, sequence, set, and text, which correspond to valid values in the database.
- Question Text. An optional text field.
- Correct Answer. A required text field.
- Question Score. A required numeric field, with a default value of 1.
- Add Question Button.

Number	Type	Question	Answer	Points
1	Text	By two o'clock teacher...all the students	had examined	1
2	Text	[translate]	On my way to school I remembered that I had left my report at home	2

Fig. 7. New test creation form

The question list appears in a table format, with columns corresponding to the question's properties and rows for each question. Upon clicking the «Add Question» button, the new question immediately displays in the list below the form, using JavaScript, and the Save Test button shifts down by the height of one table row. When the Save Test button is clicked, or if the user attempts to leave the page, a confirmation dialog appears. The user must select the desired option to proceed with the system.

The tests page displays a list of all courses created by the current user. If the user did not specify a course name when creating a test, those tests are grouped under the category «Other Tests».

Each course item in the list is an active link to the test view page for the selected course, which also includes a list of created tests.

When hovering over a list item, it changes color, and clicking it opens the test view page (fig. 8). This page displays the course name, test description, a table of questions, and a «Start Checking» button, which leads to the grading page for the selected test.

The Test Grading Page (fig. 9) includes the following fields:

- Group. A dropdown menu containing all groups imported by the current user.
- Student. A dropdown list of student names in the selected group. The values in this field update dynamically based on the selection in the first dropdown menu.
- File. An upload button for adding an image of the student's work from the file system.
- Recommended Score. Automatically populated by the system after the uploaded file is graded.
- Score. The actual grade that the instructor can assign based on the recommended score.
- Save Score Button.

Number	Type	Question	Answer	Points
1	Text	By two a clock teacher, all the students	had examined	1
2	Text	[translation]	On my way to school I remembered that I had left my report at home	2
3	Set	[selection]	A C	1
4	Text	Poor Oliver (to be) unconscious on the spot where Siles (to see) him	Poor Oliver lay unconscious on the spot where Siles had left him	2
5	Sequence	[sequence]	C D A B	2
6	Text	All the passen(per) see) at once that the old man(to travel) a green tree in his life	saw, had travelled	2
7	Single choice		C	1
8	Text	[translation]	During the holidays my friend visited the village where he had lived in his childhood	3
9	Text	[translation]	When they entered the hall, the performance had already begun	2
10	Text	[translation]	When I came home, my mother told me that she had received a letter from grandfather	2

Fig. 8. Test view page

Student's answers	Correct answers
1. had examined	1. had examined
2. On my way to school I remembered that I had left my report at home	2. On my way to school I remembered that I had left my report at home
3. A C	3. A C
4. Poor Oliver lied unconscious on the spot where Siles had left him	4. Poor Oliver lay unconscious on the spot where Siles had left him
5. C D A B	5. C D A B
6. saw, had travelled	6. saw, had travelled
7. C	7. C
8. During the holidays my friend visited the village where he had lived in his childhood	8. During the holidays my friend visited the village where he had lived in his childhood
9. When they entered the hall, the performance had already begun	9. When they entered the hall, the performance had already begun
10. When I came home, my mother told me that she had received a letter from grandfather	10. When I came home, my mother told me that she had received a letter from grandfather

Fig. 9. Test grading page

After uploading the image, the processed black-and-white version of the work appears below this form, along with a column displaying the student's answers obtained from the image recognition process and a column with the correct answers. Each correct student answer is highlighted in green, while incorrect answers are shown in red. If the user tries to enter anything other than a simple whole number in the score field, an error message will be displayed.

The Reports page contains three dropdown menus:

- the type of chart (histogram or boxplot);
- group (all groups for the current instructor, as well as an «All» option for viewing statistics across all groups simultaneously);

- test (all tests for the current instructor, along with an «All» option for viewing statistics for all tests).

Usage recommendations for the information system. When using the web platform, teachers or instructors are advised to follow several guidelines. First, it is best to take photos in well-lit areas and use a high-quality camera, as inadequate lighting or camera quality may lead to unpredictable word recognition results. Ideally, the page with the work should be a neutral color, free from additional markings, drawings, etc. Regular lined or grid paper is acceptable.

Second, when designing the test structure, it is recommended to avoid nested tasks. The test can include numerous questions, but grouping them into separate tasks, sections, or subsections should be avoided.

Third, students should be reminded to write words with sufficient spacing between them. For closed-ended tests with multiple answers, responses should be written in a single line, without numbering, and with adequate spacing between each answer.

Additionally, students should write answers sequentially in a single column, avoid flipping the answer sheet perpendicularly, and refrain from drawing sequence arrows. Corrections, cross-outs, drawings, or any elements differing from words and numbers may significantly impair the image recognition quality.

Recommendations for further improvement. During the development of this software, a machine learning model was trained, achieving satisfactory results on the IAM dataset. However, real-world images differ significantly from those in the training data, leading to occasional issues with page segmentation and incorrect word classification by the algorithm. For future enhancement, it is necessary to collect and annotate a custom training dataset. This step was not performed during development due to the substantial time required [22]. One way to collect training data is through crowdsourcing, which would be a viable method for future use.

It should be noted that it is not necessary to discard the IAM dataset and the existing model. Instead, the transfer learning technique could be applied, whereby a model is initially trained on one dataset, and then its ability to learn low-level data abstractions in the early stages is leveraged to train on a different dataset. This approach is commonly used in computer vision and has a strong potential for success in handwritten text recognition.

Additionally, recognition accuracy could be improved by adding more layers to the neural network and increasing the input vector size to enable the recognition of whole words, sentences, and paragraphs. This approach would require more training data.

During the development process, the machine learning model was designed with limited depth because training such models is time-intensive. To achieve better results, it is recommended to both deepen the neural network and perform training on multiple GPUs instead of a CPU. It is worth considering that cloud services like AWS and Google Cloud offer access to their GPUs, though these are paid services, and training may incur significant costs.

To improve the evaluation of open-ended questions, an additional algorithm using NLP technologies could be

developed. This would better align two answers and determine whether the text contains an answer to the question.

Conclusions. As a result of this research, an information system (web platform) was developed to automate the grading of offline tests, significantly easing the assessment process for educators through optical character recognition (OCR) technology based on machine learning algorithms.

At the outset, the problem domain was defined, with an analysis of existing research and publications on potential solutions, along with a review of existing software solutions. The research problem was then formulated mathematically, and a mathematical rationale for the development's usefulness was provided.

The study includes a brief overview of the database structure, which facilitates the storage and management of large volumes of data, as well as examples of the software in operation that demonstrate the system's capabilities. The developed system can recognize handwritten text from photos, create an array of responses, and compare them with the answers provided by the teacher. This approach significantly reduces the time teachers spend on test grading.

To enhance usability, a minimalist user interface was designed, providing access to all key system functions with intuitive controls. A detailed description of the developed algorithms and machine learning models is included.

This IT project has considerable potential for future development, including integration with other educational platforms, improvement of recognition technologies, and system scalability.

Thus, the web platform not only reduces teachers' time spent on grading tests but also supports the analysis and enhancement of educational programs, accommodates various types of tests, and fosters scientific and technological progress in the education sector. The system can be adapted for use in different educational institutions, making it a universal tool for automating knowledge assessment for learners.

References

1. *ZipGrade: Touchless grading when in class, online & remote for students that are not.* URL: <https://zipgrade.com> (access date: 04.10.2024).
2. *GradeCam: Assessment made easy.* URL: <https://gradecam.com> (access date: 03.10.2024).
3. *Essay Grader for Quality Assessment.* URL: <https://edubirdie.com/essay-grader> (access date: 04.10.2024).
4. Klishch D., Fedorchenko V. *Analysis of approaches to solving the problem of picture recognition using artificial intelligence. Management, Navigation, and Communication Systems.* URL: <https://journals.nupp.edu.ua/sunz/article/view/2835/2243> (access date: 06.10.2024).
5. Axler G., Wolf L. *Toward a Dataset-Agnostic Word Segmentation Method.* URL: <https://ieeexplore.ieee.org/document/8451570> (access date: 06.10.2024).
6. Arivazhagan M., Srinivasan H., Srihari S. N. *A statistical approach to line segmentation in handwritten documents.* URL: <https://cedar.buffalo.edu/~srihari/papers/SPIE-2007-lineSeg.pdf> (access date: 06.10.2024).
7. Manmatha R., Srimal N. *Scale space technique for word segmentation in handwritten documents.* URL: https://works.bepress.com/r_manmatha/11/ (access date: 08.10.2024).

8. Azmi R., Mohseni A., Maleki S., Layeghi K. *Handwriting Recognition: A Comprehensive Review.* URL: <https://civilica.com/doc/924197> (access date: 08.10.2024).
9. Awad M. & Khanna R. *Hidden Markov Model.* URL: https://doi.org/10.1007/978-1-4302-5990-9_5 (access date: 09.10.2024).
10. Graves A., Fernández S., Gomez F., Schmidhuber J. *Connectionist temporal classification: Labelling unsegmented sequence data with recurrent neural networks.* URL: https://www.cs.toronto.edu/~graves/icml_2006.pdf (access date: 10.10.2024).
11. Chowdhury A., Vig, L. *An Efficient End-to-End Neural Model for Handwritten Text Recognition.* URL: <https://arxiv.org/abs/1807.07965> (access date: 05.10.2024).
12. Puigcerver J. *Are Multidimensional Recurrent Layers Really Necessary for Handwritten Text Recognition?* URL: <https://doi.org/10.1109/ICDAR.2017.20> (access date: 10.10.2024).
13. Zhang Hui, Yao Quanming, Kwok James, Bai Xiang. *Searching a High-Performance Feature Extractor for Text Recognition Network.* URL: <https://doi.org/10.48550/arXiv.2209.13139> (access date: 12.10.2024).
14. *IAM Handwriting Database.* URL: <https://fki.tic.heia-fr.ch/databases/iam-handwriting-database> (access date: 10.10.2024).
15. Ziuziun V. *Analysis of the impact of information technologies for making management decisions, including project ones.* URL: https://www.researchgate.net/publication/371492759_Analysis_of_Aspects_of_Increasing_the_Efficiency_of_IT_Project_Management (access date: 10.10.2024).
16. Ziuziun V. *Substantiation of the importance of the role of using information technologies in business process reengineering.* URL: <https://doi.org/10.46299/ISG.2023.1.32> (access date: 10.10.2024).
17. Kronecker Delta. URL: <https://mathworld.wolfram.com/KroneckerDelta.htm> (access date: 10.10.2024).
18. Ziuziun V., Petrenko N. *Formation of Conceptual and Logical Databases for the Project of Creating a Web Platform for Automated Verification and Evaluation of Written Tests.* URL: https://www.researchgate.net/publication/Formation_of_Conceptual_and_Logical_Database (access date: 13.10.2024).
19. *OpenCv Perspective Transformation.* URL: <https://medium.com/analytics-vidhya/opencv-perspective-transformation-9edffefb2143> (access date: 15.10.2024).
20. *MVC Design Pattern.* URL: <https://www.geeksforgeeks.org/mvc-design-pattern> (access date: 15.10.2024).
21. *Machine Learning Mastery: A Gentle Introduction to Transfer Learning for Deep Learning.* URL: <https://machinelearningmastery.com/transferlearning-for-deep-learning> (access date: 15.10.2024).

References (transliterated)

1. *ZipGrade: Touchless grading when in class, online & remote for students that are not.* Available at: <https://zipgrade.com> (accessed 04.10.2024).
2. *GradeCam: Assessment made easy.* Available at: <https://gradecam.com> (accessed 03.10.2024).
3. *Essay Grader for Quality Assessment.* Available at: <https://edubirdie.com/essay-grader> (accessed 04.10.2024).
4. Klishch D., Fedorchenko V. *Analysis of approaches to solving the problem of picture recognition using artificial intelligence. Management, Navigation, and Communication Systems.* Available at: <https://journals.nupp.edu.ua/sunz/article/view/2835/2243> (accessed 06.10.2024).
5. Axler G., Wolf L. *Toward a Dataset-Agnostic Word Segmentation Method.* Available at: <https://ieeexplore.ieee.org/document/8451570> (accessed 06.10.2024).
6. Arivazhagan M., Srinivasan H., Srihari S. N. *A statistical approach to line segmentation in handwritten documents.* Available at: <https://cedar.buffalo.edu/~srihari/papers/SPIE-2007-lineSeg.pdf> (accessed 06.10.2024).
7. Manmatha R., Srimal N. *Scale space technique for word segmentation in handwritten documents.* Available at: https://works.bepress.com/r_manmatha/11/ (accessed 08.10.2024).
8. Azmi R., Mohseni A., Maleki S., Layeghi K. *Handwriting Recognition: A Comprehensive Review.* Available at: <https://civilica.com/doc/924197> (accessed 08.10.2024).

9. Awad M. & Khanna R. *Hidden Markov Model*. Available at: https://doi.org/10.1007/978-1-4302-5990-9_5 (accessed 09.10.2024).
10. Graves A., Fernández S., Gomez F., Schmidhuber J. *Connectionist temporal classification: Labelling unsegmented sequence data with recurrent neural networks*. Available at: https://www.cs.toronto.edu/~graves/icml_2006.pdf (accessed 10.10.2024).
11. Chowdhury A., Vig, L. *An Efficient End-to-End Neural Model for Handwritten Text Recognition*. Available at: <https://arxiv.org/abs/1807.07965> (accessed 05.10.2024).
12. Puigcerver J. *Are Multidimensional Recurrent Layers Really Necessary for Handwritten Text Recognition?* Available at: <https://doi.org/10.1109/ICDAR.2017.20> (accessed 10.10.2024).
13. Zhang Hui, Yao Quanming, Kwok James, Bai Xiang. *Searching a High-Performance Feature Extractor for Text Recognition Network*. Available at: <https://doi.org/10.48550/arXiv.2209.13139> (accessed 12.10.2024).
14. *IAM Handwriting Database*. Available at: <https://fki.tic.heia-fr.ch/databases/iam-handwriting-database> (accessed 10.10.2024).
15. Ziuziun V. *Analysis of the impact of information technologies for making management decisions, including project ones*. Available at: https://www.researchgate.net/publication/371492759_Analysis_of_Aspects_of_Increasing_the_Efficiency_of_IT_Project_Management (accessed 10.10.2024).
16. Ziuziun V. *Substantiation of the importance of the role of using information technologies in business process reengineering*. Available at: <https://doi.org/10.46299/ISG.2023.1.32> (accessed 10.10.2024).
17. *Kronecker Delta*. Available at: <https://mathworld.wolfram.com/KroneckerDelta.htm> (accessed 10.10.2024).
18. Ziuziun V., Petrenko N. *Formation of Conceptual and Logical Databases for the Project of Creating a Web Platform for Automated Verification and Evaluation of Written Tests*. Available at: https://www.researchgate.net/publication/Formation_of_Conceptual_and_Logical_Database (accessed 13.10.2024).
19. *OpenCv Perspective Transformation*. Available at: <https://medium.com/analytics-vidhya/opencv-perspective-transformation-9edffeb2143> (accessed 15.10.2024).
20. *MVC Design Pattern*. Available at: <https://www.geeksforgeeks.org/mvc-design-pattern> (accessed 15.10.2024).
21. *Machine Learning Mastery: A Gentle Introduction to Transfer Learning for Deep Learning*. Available at: <https://machinelearningmastery.com/transferlearning-for-deep-learning> (accessed 15.10.2024).

Received 29.10.2024

УДК 004.89:004.932:004.4

В. І. ЗЮЗІОН кандидат технічних наук (PhD), доцент, Київський національний університет імені Тараса Шевченка, доцент кафедри технологій управління, м. Київ, Україна; e-mail: vadym.ziuziun@knu.ua; ORCID: <https://orcid.org/0000-0001-6566-8798>

Н. А. ПЕТРЕНКО, Київський національний університет імені Тараса Шевченка, студент, м. Київ, Україна; e-mail: nikita.petrenko@knu.ua; ORCID: <https://orcid.org/0009-0006-3921-8412>

ЗАСТОСУВАННЯ ТЕХНОЛОГІЙ ОПТИЧНОГО РОЗПІЗНАВАННЯ СИМВОЛІВ ТА МАШИННОГО НАВЧАННЯ ДЛЯ СТВОРЕННЯ ІНФОРМАЦІЙНОЇ СИСТЕМИ АВТОМАТИЧНОЇ ПЕРЕВІРКИ ОФЛАЙН-ТЕСТУВАННЯ

Під час навчання в будь-якій сфері, тестування та моніторинг знань учнів, студентів, або здобувачів освіти є невід'ємною частиною процесу. Викладачі часто витрачають багато часу на перевірку великої кількості типових тестів. Хоча системи онлайн-тестування були розроблені для полегшення цього процесу, паперові офлайн-тести залишаються популярними, оскільки не вимагають доступу до комп'ютерів, електроенергії та надійного інтернет-з'єднання. Офлайн-тестування є найбільш репрезентативним видом перевірки, але призводить до монотонної роботи для викладачів під час перевірки. Деякі викладачі використовують бланки для тестів, щоб структурувати відповіді здобувачів знань та зекономити час. У зв'язку з цим стає актуальним створення системи, яка автоматизує перевірку офлайн-тестів. Метою дослідження стала розробка інформаційної системи (вебплатформи), яка спростить процес перевірки офлайн-тестів за допомогою технологій оптичного розпізнавання символів на основі алгоритмів машинного навчання. Об'єктом дослідження стали процеси та функціональні можливості створення інформаційної системи для автоматизованої перевірки та оцінювання офлайн-тестів. Наукова новизна полягає у інтеграції та використанні алгоритмів машинного навчання разом з модифікаціями алгоритмів оброблення зображень, для створення інформаційної системи, яка дозволить аналізувати та оцінювати широкий спектр тестових офлайн завдань (відкриті, закриті, визначення послідовності, закриті з декількома правильними відповідями). Практичне значення дослідження полягає в розробці вебплатформи для автоматизації перевірки офлайн-тестів за допомогою технологій оптичного розпізнавання символів та машинного навчання, що зменшує витрати часу викладачів, дозволяє аналізувати та вдосконалювати навчальні програми, підтримує різноманітні типи тестів і сприяє науковому та технологічному прогресу у сфері освіти. Розроблена система здатна розпізнавати рукописний текст з фотографій, формувати масив відповідей та порівнювати їх з відповідями, внесеними викладачем. Це значно зменшує час, який викладачі витрачають на перевірку тестів. Для зручності використання був створений мінімалістичний інтерфейс користувача, який надає доступ до всіх основних функцій системи та забезпечує інтуїтивно зрозуміле управління. Було надано детальний опис розроблених алгоритмів та моделей машинного навчання. Даний IT проєкт має широкі перспективи для подальшого розвитку, зокрема інтеграція з іншими освітніми платформами, покращення технологій розпізнавання та масштабування системи.

Ключові слова: інформаційна система, вебплатформа, IT проєкт, машинне навчання, нейронні мережі, алгоритм, IAM датасет, оптичне розпізнавання символів, тестування, освітній процес.

Повні імена авторів / Author's full names

Автор 1 / Author 1: Зюзюн Вадим Ігорович / Ziuziun Vadym Ihorovych

Автор 2 / Author 2: Петренко Нікіта Андрійович / Petrenko Nikita Andriiovych

V. I. ZIUZIUN, Candidate of Technical Sciences (PhD), Docent, Associate Professor at the Department of Management Technologies, Taras Shevchenko National University of Kyiv, Kyiv, Ukraine; e-mail: vadym.ziuziun@knu.ua; ORCID: <https://orcid.org/0000-0001-6566-8798>

D. S. OSOKA, Student, Taras Shevchenko National University of Kyiv, Kyiv, Ukraine; e-mail: osokadania@knu.ua; ORCID: <https://orcid.org/0009-0004-8603-737X>

MATHEMATICAL RATIONALE FOR CREATING AN APPLICATION FOR CONDUCTING RANDOM MEETINGS «COFFEE BREAK»

Modern society is facing an increasing trend of social isolation, as people increasingly rely on social media for interaction instead of face-to-face communication. This lack of in-person contact often leads to feelings of loneliness and disconnection. This study proposes the concept of a mobile application, CoffeeBreak, designed to counteract these trends by offering users a platform to arrange brief, in-person meetings, such as a quick coffee chat. By encouraging users to meet in real life, the application aims to foster meaningful social connections and combat the sense of isolation prevalent in today's digital world. The core innovation of CoffeeBreak lies in its unique approach to matchmaking. Instead of presenting users with an overwhelming array of choices, the app offers a single match within a specified timeframe, thus addressing the common issue of decision paralysis that can arise when users are presented with too many options. By simplifying the process, CoffeeBreak allows users to spend less time making selections and more time connecting with others. This approach is inspired by practices adopted within large companies, where employees use bots in work chat groups to find a partner for a short meeting. These interactions help raise awareness about the activities in other departments and foster informal and professional connections. Expanding this practice to a broader societal level, CoffeeBreak is intended to provide individuals with the opportunity to network beyond their immediate professional circles. This research has established a conceptual system model and developed the mathematical frameworks necessary to support this type of meeting arrangement. Specifically, the study has defined the concept of the CoffeeBreak mobile application, outlined the system model with detailed subsystems and environment interactions, and formulated mathematical models to form the basis of the candidate selection algorithm. The model ensures that users are matched in a way that promotes engagement, as each participant can be assured that their matched partner is equally motivated for the encounter. As the application continues to evolve, it can incorporate additional scheduling criteria to enhance the quality of matches and distribution. For example, if a user attends a meeting within the first two days, they could unlock the potential for additional matches by the end of the week. Ultimately, CoffeeBreak aims to broaden users' horizons, help them form new professional and informal connections, and enhance their social skills. This study's findings lay the groundwork for a new tool that encourages in-person interactions, enabling individuals to expand their social networks in a balanced and purposeful manner.

Keywords: mobile app facilitating random meetings, socialization, networking, information system, conceptual model, mathematical model.

Introduction. In today's world, humanity is increasingly immersed in online life and social media. Although these platforms were initially created with the purpose of building social connections and fostering societal unity, over the years, their actual impact has shown somewhat of a reverse effect – an increase in social isolation [1]. Additionally, the loss of social skills has been exacerbated by the relatively recent COVID-19 pandemic. As a result, the need for tools to facilitate the search for new acquaintances in the real world is now more urgent than ever.

In addition to the new challenges brought about by technological advancements, old issues persist. Most of a person's social interactions still occur with individuals from their own or adjacent professional fields and social groups. This significantly limits the expansion of one's horizons, a goal often pursued by those seeking new connections. Meeting people from different perspectives can foster powerful collaborations for solving various problems, as individuals from diverse backgrounds approach issues from different angles [2]. Many ambitious people actively seek such acquaintances to generate new ideas in areas as diverse as solving professional challenges, addressing personal issues, or brainstorming business concepts.

Therefore, there is a growing need for a new and effective tool that can help people broaden their horizons,

find both professional and informal connections, and improve their social skills. This mobile app facilitating random meetings should offer society a novel approach to expanding their social networks and should also encourage users to engage in real, face-to-face communication.

Analysis of recent research and publications.

Currently, there are numerous dating services available on the market [3-6]. The most popular ones are primarily focused on finding a romantic partner, such as apps like Tinder, Badoo, Blendr, Bumble, and Wink. All of these apps operate on the principle of browsing through a list of profiles and voting «like» or «dislike». If both users react positively to each other's profiles, the app provides contact information for further communication. A similar concept is employed by other apps that have more specific themes, for example:

- Hey! VINA – primarily aimed at women looking for new female friendships;
- Peanut – designed for mothers who want to connect with other moms;
- ATLETO – useful for those seeking a partner or opponent for team or competitive sports activities.

There are also other, more distinctive solutions:

- Down – bases the list of potential meeting candidates on the user's social media friends list.
- Happn – shows people with whom the user has crossed paths in the city.

© Ziuziun V. I., Osoka D. S., 2024



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.



All of the aforementioned solutions have their own advantages, each offering a unique user experience and catering to a specific audience. However, they share a common feature: each app provides the ability to choose from a vast number of potential candidates for dating [7]. While having a large selection might seem beneficial, it can actually have the opposite effect. A large number of options can lead to decision paralysis and reduce satisfaction with the choices made. Additionally, users often spend a significant amount of time filtering candidates based on photos, age, or an inappropriate joke in the profile description. Even when a match is achieved, it does not guarantee that the person on the other side is interested in meeting or engaging in real-life communication.

The aim and tasks of the research. The aim of this research is to develop conceptual and mathematical models for a system that provides an innovative approach to finding new acquaintances through a mobile application.

To achieve this aim, the following tasks must be completed:

- Define the concept of the future system.
- Create a conceptual model of the future system.
- Develop the mathematical models for the meeting arrangement information system (mobile app facilitating random meetings).

Presentation of the main material. The project idea is based on a widely adopted practice among employees of large companies. In work chat groups, bots are created where employees can submit requests to find a partner for a brief meeting. These sessions help raise awareness among employees of large corporations about what is happening in other departments and the activities of colleagues in different areas. Employees thus have the opportunity to establish new informal and professional connections. Simply put, this practice broadens employees' horizons and expands their social networks within the company.

If we extend this practice to a global society, such a solution has the potential to address both the problem of limited social interaction and the challenges of networking. Individuals will be able to expand their social networks beyond just a single company workforce. Importantly, this solution avoids the «overchoice» problem found in many leading apps on the online networking and dating market. Users will have only one match within a set timeframe, encouraging them to take action, knowing that they won't have another candidate for a noticeable period. Additionally, they can rest assured that the person on the other side is in exactly the same situation. Of course, as the product develops, it will be enhanced with additional features, but the core concept of limiting the number of potential matches within a specific timeframe will remain. For example, if a user attends a meeting within the first two days, they could unlock the possibility of up to two additional matches by the end of the week.

Formation of the concept. The idea for this system is based on a widely adopted practice among large companies' staff. In workplace chat environments, bots are created where employees can submit requests to find a candidate for a brief meeting. These practices help improve employees' awareness of what is happening in other

departments of their company, what activities other people might be engaged in, and so on. Employees have the opportunity to form new informal and professional connections. In simpler terms, this practice broadens employees' horizons and enhances their social networks.

If this practice is extrapolated to a global society, it suggests that such a solution could address issues of social interaction and networking. People would be able to expand their social networks beyond just the group of workers in their company. Additionally, this solution would avoid the problem of choice overload that is common in many leading online dating apps. Users would only have one match at a given time, which would encourage them to take action, as there would not be another candidate available for a noticeable period, and they would understand that the person on the other side is in a similar situation. As the product develops, it will be enhanced with new features, but the core concept of limiting the number of candidates for a given period will be maintained. For example, if a user goes on a meeting within the first two days, they will be able to get two more matches by the end of the week [8].

Since modern people always have their smartphones with them, it makes sense to develop this system as a mobile application. We will call it CoffeeBreak. The interface should be as simple as possible: fill out a profile, press a button to start searching, and wait for a candidate for a meeting.

Development of a conceptual model. To understand all the components of the future application, it is necessary to identify all subsystems of the CoffeeBreak product, including the internal and external environments, as well as the overarching system within which CoffeeBreak operates. Together, these components and their interactions form the system composition model [9].

First, we will outline the subsystems of the CoffeeBreak application. At this stage, it is crucial to have well-defined functional requirements for the application, as the subsystems will reflect how the system meets these requirements. The subsystems of CoffeeBreak include:

- **User Management.** Responsible for managing user accounts and profiles. Key functions of this subsystem include: registration of new users (user authentication and authorization); profile management (updating information, adding interests, changing settings); ensuring user data security (encryption, password protection).

- **Random Meeting Generation.** Responsible for organizing random meetings between users. Key functions of this subsystem include: Randomly pairing users for meetings; considering time and geographic constraints when generating meetings.

- **Meeting Management.** Responsible for managing all aspects of organizing and conducting meetings. Key functions include: creating, editing, and canceling meetings; managing meeting locations; sending reminders and notifications about changes in meeting schedules; integrating with mapping services to provide navigation to the meeting location.

- **Messaging.** Handles communication with users and sending various types of notifications. Key functions include: sending notifications about new meetings,

reminders, and updates; supporting push notifications and email; informing users about new features, app updates, and other important events.

- **Analytics and Reporting.** Ensures the collection, analysis, and visualization of data related to app performance and user interactions. Key functions include: collecting statistical data on app usage; analyzing user activity and the effectiveness of app features; generating reports for administrators and developers.

- **Partner Management.** This subsystem is designed for collaboration with partner cafes that will offer their services for meetings. Key functions include: registering and managing profiles of partner cafes; integrating with cafe booking systems; providing cafes with access to information about scheduled meetings; allowing cafes to offer special deals and discounts to CoffeeBreak users; tracking user reviews and ratings of partner cafes.

- **Chat.** Provides the ability for users to exchange messages before and after meetings. Key functions include: sending text messages and media files between users; storing chat history; notifying users of new messages from other users.

The internal environment of the system includes Administrators, who can assist users in case of unusual situations. Administrators will have access to the «Analytics and Reporting», «Meeting Management», «User Management», and «Partner Management» subsystems.

In the external environment, we should consider Partners and Competitors. Partners have access to the «Partner Management» subsystem to monitor and manage the special partner functionalities. Competitors, on the other hand, do not have direct influence on the application but may affect users and the development progress of the application.

The overarching system in the original concept is the global society; however, this solution could also be applied within large corporations that wish to implement such a system within their company, thereby providing a more convenient tool than a bot in a workplace chat.

Thus, the results of the model creation can be presented graphically, clearly indicating the relationships between the components of the model (fig. 1).

Let us also briefly outline the tools that will be used for the app development. Flutter was selected for the app development. This modern open-source framework enables seamless deployment across both mobile platforms, Android and iOS, eliminating the need to write separate applications for each platform [10].

For servers and web panels, Python [11] will be used in combination with the Flask framework. This setup provides a straightforward and widely adopted toolkit, allowing for the rapid and efficient development of a modern solution.

MySQL was chosen as the database management system (DBMS), as its simplicity and reliability make it ideal for creating the CoffeeBreak product [12].

Development the conceptual model of a mobile app facilitating random meetings. To develop the CoffeeBreak mobile application, it is necessary to formalize the main processes and tasks in mathematical terms. This will help

define the algorithms that will underpin the application's operation.

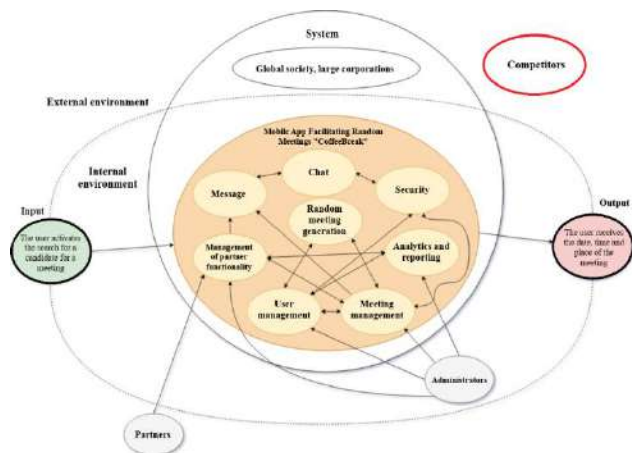


Fig. 1. Conceptual model mobile app facilitating random meetings CoffeeBreak

The primary objective of the application is to maximize the number of scheduled meetings between users (1).

Let:

$$U = \{u_1, u_2, \dots, u_n\}, \quad (1)$$

where U – set of users.

We will introduce a binary variable x_{ij} , which equals 1 if a user u_i is assigned to a meeting with user u_j , otherwise the variable will be 0.

The formulation of the problem is as follows (2):

$$\sum_{i=1}^n \sum_{j=1}^n x_{ij} \longrightarrow \max, \quad (2)$$

with conditions (3-5):

$$\sum_{j=1}^n x_{ij} \leq 1, \forall i \in [1, n]. \quad (3)$$

$$\sum_{i=1}^n x_{ij} \leq 1, \forall j \in [1, n]. \quad (4)$$

$$x_{ij} \in \{0, 1\}, \forall i, j \in [1, n]. \quad (5)$$

The objective function maximizes the number of assigned meetings, and the constraints ensure that each user can only participate in one meeting at a time.

As an extension of the basic functionality for selecting candidates for a meeting, it would be useful to provide the ability to set location constraints where it is convenient for the user to hold a meeting, as well as time constraints. This way, the system will only select candidates for the user who are geographically convenient and available at a mutually suitable time (6-7).

Let:

$$T = \{t_1, t_2, \dots, t_m\}, \quad (6)$$

where T – set of time intervals;

$$L = \{l_1, l_2, \dots, l_k\}, \quad (7)$$

where L – set of locations.

To take into account time and geographical constraints, we introduce the user availability function $A\{u_i, t, l\}$, which is equal to 1, if the user u_i is available in the time interval t at the location l , and 0 otherwise.

A user's accessibility function can be defined as:

$$A\{u_i, t, l\} = \begin{cases} 1, & \text{if the user } u_i \text{ is available in } t \\ & \text{at the location } l, \\ 0, & \text{otherwise.} \end{cases} \quad (8)$$

Let's enter the binary variable x_{ijtl} , which is equal to 1, if the user u_i is scheduled to meet with the user u_j in the time interval t at the location l , and 0 otherwise.

Then the new formulation of the problem will have the following form:

$$\sum_{i=1}^n \sum_{j=1}^n \sum_{t=1}^m \sum_{l=1}^k A(u_i, t, l) \cdot x_{ijtl} \longrightarrow \max, \quad (9)$$

with conditions (10-12):

$$\sum_{j=1}^n \sum_{t=1}^m \sum_{l=1}^k x_{ijtl} \leq 1, \quad \forall i \in [1, n]. \quad (10)$$

$$\sum_{i=1}^n \sum_{t=1}^m \sum_{l=1}^k x_{ijtl} \leq 1, \quad \forall j \in [1, n]. \quad (11)$$

$$x_{ijtl} \in \{0, 1\}, \quad \forall i, j, t, l \in [1, n]. \quad (12)$$

In general, the objective function is the same as in the base model, the number of assigned meetings, and the constraints ensure that each user can only participate in one meeting at a time.

Application of interval distribution. In order to make it easier for users to start communicating during a meeting, you can expand the functionality of the application and allow users to add categories of interests and views on life to their profile, thus the system will prioritize candidates for the meeting that best match each other according to a set of categories. For this, we introduce the similarity function (13):

$$S(u_i, u_j) = \frac{|I_i \cap I_j|}{|I_i \cup I_j|}, \quad (13)$$

where I_i – set of user interests u_i ;

I_j – set of user interests u_j .

As you can see, this function will give a value between 0 and 1, depending on how similar the sets of users are. Now let's apply this function to modify our problem formulation.

Then the new objective function will have the following form (14):

$$\sum_{i=1}^n \sum_{j=1}^n \sum_{t=1}^m \sum_{l=1}^k S(u_i, u_j) \cdot A(u_i, t, l) \cdot x_{ijtl} \longrightarrow \max. \quad (14)$$

The conditions remain the same as in the previous paragraph, they are given by formulas (10), (11), (12).

Conclusions. This research has proposed a concept and models that can form the basis for a new tool aimed at expanding one's horizons, finding new professional and informal connections, and enhancing social skills. The application will offer an innovative approach to finding connections by randomly selecting a candidate for a brief meeting, such as for coffee. This solution will help avoid decision paralysis and significantly reduce the time spent on selection.

The research has achieved all of the set objectives:

- The system concept has been defined. It will be the CoffeeBreak mobile app-application, designed to easily provide users with a candidate for a short meeting.

- A conceptual model of the system has been created, reflecting all subsystems, the overarching system, and the application environment.

- Mathematical models for the meeting scheduling system have been developed. These models will form the basis of the algorithm for selecting candidates for meetings.

Further research in this area is advisable, as meeting scheduling algorithms may incorporate new criteria to improve how users are matched and distributed.

References

1. Primack B., Shensa A., Sidani J., Whaitte E., Lin L., Rosen D., Colditz J., Radovic A., Miller E. *Social Media Use and Perceived Social Isolation Among Young Adults in the U.S.* URL: <https://d-scholarship.pitt.edu/35420/> (access date: 28.10.2024).
2. Lanterman J., Bliethe S. *The Benefits, Challenges, and Disincentives of Interdisciplinary Collaboration.* URL: https://www.researchgate.net/publication/338048962_The_Benefits_Challenges_and_Disincentives_of_Interdisciplinary_Collaboration/fulltext/63923b4e484e65005bf49803/The-Benefits-Challenges-and-Disincentives-of-Interdisciplinary-Collaboration.pdf (access date: 28.10.2024).
3. *12 Apps That Will Actually Help You Make Friends.* URL: <https://www.cosmopolitan.com/sex-love/a24799641/best-friendship-apps/> (access date: 28.10.2024).
4. *15 Making Friends Apps to Meet New People.* URL: <https://agiletech.vn/top-making-friends-apps-to-meet-new-people/> (access date: 29.10.2024).
5. *If You Have No Idea How to Make Friends as an Adult, Download These Apps.* URL: <https://www.yahoo.com/lifestyle/11-apps-apos-11-friends-200300960.html> (access date: 29.10.2024).
6. Pettersen L., Karlsen F. *Strategic communication in digital ecosystems: A critical discourse analysis of dating applications. In Strategic communication – contemporary perspectives.* URL: <https://www.kristiania.no/en/research/research-projects/kuc/digilove> (access date: 15.10.2024).
7. Iyengar S. S., Lepper M. R. *When Choice is Demotivating: Can One Desire Too Much of a Good Thing?* URL: <https://business.columbia.edu/faculty/research/when-choice-demotivating-can-one-desire-too-much-good-thing> (access date: 15.10.2024).
8. Ziuziun V. *Analysis of the impact of information technologies for making management decisions, including project ones.* URL: https://www.researchgate.net/publication/371492759_Analysis_of_Aspects_of_Increasing_the_Efficiency_of_IT_Project_Management (accessed 30.10.2024).
9. Ziuziun V. *Analysis of Aspects of Increasing the Efficiency of IT Project Management.* URL: https://www.researchgate.net/publication/371492759_Analysis_of_

- Aspects_of_Increasing_the_Efficiency_of_IT_Project_Management (access date: 30.10.2024).
10. *Documentation Flutter*. URL: <https://flutter.dev/> (access date: 30.10.2024).
 11. *Documentation Flask*. URL: <https://flask.palletsprojects.com/en/3.0.x/> (access date: 30.10.2024).
 12. *Documentation MySQL*. URL: <https://www.mysql.com/> (access date: 30.10.2024).
- References (transliterated)**
1. Primack B., Shensa A., Sidani J., Whaithe E., Lin L., Rosen D., Colditz J., Radovic A., Miller E. *Social Media Use and Perceived Social Isolation Among Young Adults in the U.S.* Available at: <https://d-scholarship.pitt.edu/35420/> (accessed: 28.10.2024).
 2. Lanterman J., Blithe S. *The Benefits, Challenges, and Disincentives of Interdisciplinary Collaboration*. Available at: https://www.researchgate.net/publication/338048962_The_Benefits_Challenges_and_Disincentives_of_Interdisciplinary_Collaboration/fulltext/63923b4e484e65005bf49803/The-Benefits-Challenges-and-Disincentives-of-Interdisciplinary-Collaboration.pdf (accessed: 28.10.2024).
 3. *12 Apps That Will Actually Help You Make Friends*. Available at: <https://www.cosmopolitan.com/sex-love/a24799641/best-friendship-apps/> (accessed: 28.10.2024).
 4. *15 Making Friends Apps to Meet New People*. Available at: <https://agiletech.vn/top-making-friends-apps-to-meet-new-people/> (accessed: 29.10.2024).
 5. *If You Have No Idea How to Make Friends as an Adult, Download These Apps*. Available at: <https://www.yahoo.com/lifestyle/11-apps-apos-ll-friends-200300960.html> (accessed: 29.10.2024).
 6. Pettersen L., Karlsen F. *Strategic communication in digital ecosystems: A critical discourse analysis of dating applications. In Strategic communication – contemporary perspectives*. Available at: <https://www.kristiania.no/en/research/research-projects/kuc/digilove> (accessed: 15.10.2024).
 7. Iyengar S. S., Lepper M. R. *When Choice is Demotivating: Can One Desire Too Much of a Good Thing?* Available at: <https://business.columbia.edu/faculty/research/when-choice-demotivating-can-one-desire-too-much-good-thing> (accessed: 29.10.2024).
 8. Ziuziun V. *Analysis of the impact of information technologies for making management decisions, including project ones*. Available at: https://www.researchgate.net/publication/371492759_Analysis_of_Aspects_of_Increasing_the_Efficiency_of_IT_Project_Management (accessed 30.10.2024).
 9. Ziuziun V. *Analysis of Aspects of Increasing the Efficiency of IT Project Management*. Available at: https://www.researchgate.net/publication/371492759_Analysis_of_Aspects_of_Increasing_the_Efficiency_of_IT_Project_Management (accessed: 30.10.2024).
 10. *Documentation Flutter*. URL: <https://flutter.dev/> (accessed: 30.10.2024).
 11. *Documentation Flask*. Available at: <https://flask.palletsprojects.com/en/3.0.x/> (accessed: 30.10.2024).
 12. *Documentation MySQL*. URL: <https://www.mysql.com/> (accessed: 30.10.2024).

Received 04.11.2024

УДК 004.738.5

В. І. ЗЮЗИУН кандидат технічних наук (PhD), доцент, Київський національний університет імені Тараса Шевченка, доцент кафедри технологій управління, м. Київ, Україна; e-mail: vadyu.ziuziun@knu.ua; ORCID: <https://orcid.org/0000-0001-6566-8798>

Д. С. ОСОКА, Київський національний університет імені Тараса Шевченка, студент, м. Київ, Україна; e-mail: osokadania@knu.ua; ORCID: <https://orcid.org/0009-0004-8603-737X>

МАТЕМАТИЧНЕ ОБґРУНТУВАННЯ СТВОРЕННЯ ДОДАТКУ ДЛЯ ПРОВЕДЕННЯ ВИПАДКОВИХ ЗУСТРІЧЕЙ «COFFEE BREAK»

Сучасне суспільство зіштовхується зі зростаючою тенденцією соціальної ізоляції, оскільки люди все більше покладаються на соціальні мережі для взаємодії замість спілкування віч-на-віч. Відсутність особистого контакту часто призводить до почуття самотності та відірваності. У цьому дослідженні пропонується концепція мобільного додатку CoffeeBreak, розробленого для протидії цим тенденціям, пропонуючи користувачам платформу для організації коротких особистих зустрічей, наприклад, швидкої бесіди за кавою. Заохочуючи користувачів зустрічатися в реальному житті, програма спрямована на розвиток значущих соціальних зв'язків і боротьбу з почуттям ізоляції, поширеним у сучасному цифровому світі. Основна інновація додатку CoffeeBreak полягає в його унікальному підході до пошуку партнерів. Замість того, щоб надавати користувачам величезну кількість варіантів вибору, програма пропонує єдиний збіг протягом визначеного періоду часу, таким чином вирішуючи поширену проблему паралічу прийняття рішень, яка може виникнути, коли користувачам пропонується забагато варіантів. Завдяки спрощенню процесу CoffeeBreak дозволяє користувачам витратити менше часу на вибір і більше часу на спілкування з іншими. Цей підхід натхненний практиками, прийнятими у великих компаніях, де співробітники використовують ботів у робочих групах чату, щоб знайти партнера для короткої зустрічі. Ці взаємодії допомагають підвищити обізнаність про діяльність інших відділів і сприяють неформальним і професійним зв'язкам. Розширюючи цю практику на ширший суспільний рівень, додаток CoffeeBreak має на меті надати людям можливість спілкуватися за межами їхніх безпосередніх професійних кіл. В рамках дослідження було побудовано концептуальну модель системи та розроблено математичну структуру додатку. Зокрема, в дослідженні визначено концепцію мобільного додатку CoffeeBreak, окреслено модель системи з детальними підсистемами та взаємодією середовища, а також сформовано математичні моделі, які склали основу алгоритму відбору кандидатів. Модель забезпечує підбір користувачів таким чином, щоб сприяти взаємодії, оскільки кожен учасник може бути впевнений, що його підібраний партнер однаково мотивований для зустрічі. В подальшому додаток буде удосконалюватися. Він буде доповнений додатковими критеріями планування для підвищення якості збігів і охоплення все більш широкою аудиторією користувачів. Наприклад, якщо користувач відвідує зустріч протягом перших двох днів, він може розкрити потенціал для додаткових збігів до кінця тижня. Зрештою, CoffeeBreak має на меті розширити кругозір користувачів, допомогти їм створити нові професійні та неформальні зв'язки та покращити їхні соціальні навички. Результати цього дослідження закладають основу для нового інструменту, який заохочує особисту взаємодію, дозволяючи людям розширювати свої соціально-комунікативні навички.

Ключові слова: мобільний додаток для проведення випадкових зустрічей, соціалізація, нетворкінг, інформаційна система, концептуальна модель, математична модель.

Повні імена авторів / Author's full names

Автор 1 / Author 1: Зюзиун Вадим Ігорович / Ziuziun Vadym Ihorovych

Автор 2 / Author 2: Осока Даниїл Сергійович / Osoka Daniil Serhiiovych

Д. В. КАЛІНІН, аспірант кафедри системного аналізу та інформаційно-аналітичних технологій, Національний технічний університет «Харківський політехнічний інститут», м. Харків, Україна; e-mail: Denys.Kalinin@cs.khpi.edu.ua; ORCID: <https://orcid.org/0009-0004-4431-7728>

В. П. СЕВЕРИН, д-р техн. наук, професор, професор кафедри системного аналізу та інформаційно-аналітичних технологій, Національний технічний університет «Харківський політехнічний інститут», Харків, Україна; e-mail: Valerii.Severyn@khpi.edu.ua; ORCID: <https://orcid.org/0000-0002-2969-6780>

М. І. БЕЗМЕНОВ, кандидат технічних наук (PhD), доцент, професор кафедри системного аналізу та інформаційно-аналітичних технологій, Національний технічний університет «Харківський політехнічний інститут», м. Харків, Україна; e-mail: Mykola.Bezmenov@khpi.edu.ua; ORCID: <https://orcid.org/0000-0002-2995-2350>

МОДЕЛІ ПРИВАТНОСТІ ТА ТЕХНІКИ АНОНІМІЗАЦІЇ ТАБЛИЧНИХ МЕДИЧНИХ ДАНИХ

У сучасному світі питання приватності та захисту персональних даних набувають надзвичайної актуальності, особливо в медичній галузі, де використання великих обсягів даних для досліджень стає все більш поширеним. Використання персональних даних регулюється відповідними законами, які вимагають анонімізації даних для мінімізації ризиків ідентифікації осіб. Анонімізація є процесом, що дозволяє використовувати чутливі дані без ризику розкриття особистої інформації, зберігаючи при цьому їх корисність. У статті розглядаються основні моделі приватності та техніки анонімізації, що застосовуються для захисту табличних медичних даних. Моделі приватності включають k -анонімність (k -anonymity), l -диверсність (l -diversity) та t -близькість (t -closeness). Модель k -анонімності забезпечує щоб будь-яка комбінація квазіідентифікаторів була спільною для щонайменше k записів. Модель l -диверсності доповнює k -анонімність, вимагаючи наявності щонайменше l унікальних комбінацій значень чутливих атрибутів (SA) у кожному класі еквівалентності. Модель t -близькості враховує розподіл значень цих чутливих атрибутів, забезпечуючи, щоб відстань між розподілом SA у класі еквівалентності та загальним розподілом не перевищувала заданий поріг. Техніки анонімізації включають узагальнення (generalization), подавлення (suppression), перенесення (relocation), перестановку (permutation), пертурбацію (perturbation), розділення (slicing), диференційну приватність (differential privacy) та синтетичні дані (synthetic data). Узагальнення зменшує точність квазіідентифікаторів. Подавлення видаляє певні значення з набору даних для покращення його статистичних характеристик. Перенесення змінює обмежену кількість значень в даних з метою підвищення захисту. Перестановка зміщує значення квазіідентифікаторів між записами, зберігаючи при цьому загальні статистичні особливості набору даних. Пертурбація додає шум до даних, що підвищує приватність. Ідея диференційної приватності також полягає у додаванні шуму, але це виконується на етапі обробки запитів за даними. Генерація синтетичних даних дозволяє створювати нові набори даних, як схожі за характеристиками на оригінальні дані.

Ключові слова: обробка даних, моделі штучного інтелекту, класифікація, ідентифікація, моделі приватності, медичні дані, табличні дані, захист чутливих даних, анонімізація даних, техніки анонімізації даних, диференційна приватність.

Вступ. У сучасному світі зберігання та обробка персональних даних невід’ємно пов’язані з питаннями приватності та захисту таких даних. Для медичної галузі дані питання є особливо актуальними, оскільки все більш поширеною практикою стає використання великих об’ємів медичних даних для проведення різних досліджень та аналізу, у тому числі з використанням штучного інтелекту. Використання персональних даних для другорядних цілей регулюється конкретними законами GDPR [1] та HIPAA [2], порушення яких може призвести до серйозних наслідків, часом навіть незворотних. Відповідно до цих актів чутливі дані можуть бути використані тільки після спеціальної їх обробки, в результаті якої ідентифікація осіб з набору стає неможливою або ризик такої ідентифікації не перевищує допустимий рівень. Даний процес має назву *анонімізація* (anonymization), іноді – *деідентифікація*. *Основна задача анонімізації* – мінімізувати ризики розкриття особистої інформації, зберігаючи при цьому максимальну корисність вихідних даних. Для вирішення цієї задачі використовуються спеціальні техніки (методи), що забезпечують умови для певних моделей приватності (на практиці це часто комбінації технік).

Мета статті. Описати основні моделі приватності та техніки анонімізації і провести аналіз їх переваг та недоліків у контексті анонімізації табличних медичних даних.

Моделі приватності. У процесі анонімізації реляційних даних виділяють наступні категорії атрибутів: *прямі ідентифікатори* (direct identifiers, DI), *квазіідентифікатори* (quasi-identifiers, QI), *чутливі атрибути* (sensitive attributes, SA) та *нечутливі атрибути*. *Прямі ідентифікатори* – це атрибути, які дозволяють однозначно визначити конкретну особу. *Квазіідентифікатори* – атрибути, що лише у комбінації з іншими квазіідентифікаторами можуть бути використані для ідентифікації даних. Під *чутливими (захищеними) атрибутами* розуміють такі атрибути, які мають залишатися приватними та захищеними від потенційного зловмисника. *Нечутливі атрибути* у свою чергу можуть вважатися безпечними для публічного використання. Орієнтуючись на дану класифікацію, можна описати основні моделі приватності, які використовуються для анонімізації даних.

Модель приватності під назвою *k -анонімність* (k -anonymity) було запропоновано в роботі [3]. Дана модель вимагає, аби при будь-якій комбінації QIs щонайменше k записів мали однаковий набір значень квазіідентифікаторів, тим самим забезпечуючи анонімність у рамках відповідної групи (класу еквівалентності). Очевидно, що чим більше значення k , тим вищий рівень захисту ми отримуємо. Основними недоліками даної моделі є неефективність проти атак на основі однорідності (homogeneity attack) значень SAs в

© Калінін Д. В., Северин В. П., Безменов М. І., 2024



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.



межах класу еквівалентності, а також суттєве погіршення корисності анонімізованих даних при великих значеннях k .

Проблему беззахисності даних до атак на основі однорідності дозволяє вирішити модель *l*-диверсності (*l*-diversity), що була представлена як доповнення до моделі k -анонімності в [4]. Термін «*l*-диверсність» визначає, що в рамках кожного класу еквівалентності присутні хоча б l унікальних комбінацій значень для набору чутливих атрибутів. Маючи *l*-диверсність значень SAs, у загальному випадку можна стверджувати, що зловмисник буде мати не менше ніж l варіантів значень SAs при спробі отримати чутливу інформацію в рамках відповідного класу розміром k . Основною метою даної моделі приватності є захист чутливих атрибутів, але на практиці доволі часто виникають ситуації, коли *l*-диверсність не гарантує, що ці l варіантів є дійсно унікальними. Так, наприклад, деякі діагнози можуть мати декілька різних назв, але семантично бути еквівалентними, що призведе до розкриття чутливої інформації при формальному дотриманні вимог моделі. Також важливим недоліком даної моделі є її несприйнятливості до розподілу даних в SAs. Так, наприклад, якщо деякий набір значень SAs домінує в своєму класі еквівалентності, він тим самим збільшує шанси зловмисника отримати чутливу інформацію відповідно до наявного розподілу.

Модель *t*-близькості (*t*-closeness) була введена в рамках роботи [5], де було визначено, що клас еквівалентності називається таким, що має *t*-близькість, у випадку, коли відстань між розподілом чутливого атрибуту у своєму класі та його розподілом у повному наборі даних не перевищує заданий поріг t . Таким чином, можна казати про таку залежність, коли менше значення t відповідає більш високому рівню захисту чутливих даних. Для обчислення відстані між розподілами найчастіше використовують відстань Кульбака – Лейблера або відстань Васерштейна. Дана модель гармонійно доповнює моделі k -анонімності та *l*-диверсності, надаючи рішення проблеми викриття чутливих даних, що спричинена специфічним розподілом значень SAs у групі. Але звісно, що на практиці можуть виникати унікальні ситуації, коли та чи інша модель не зможе гарантувати задовільні результати.

Техніки анонімізації даних. Техніки, що розглядаються в даному розділі, базуються на вимогах вищеописаних моделей приватності: k -анонімності, *l*-диверсності та *t*-близькості, а також їх варіацій.

Техніка узагальнення (generalization) [3] використовується для зменшення точності певного QI. Так, наприклад, говорячи про конкретні дати (дата народження, дата відвідування лікаря тощо), дана техніка дозволяє змінити ступінь деталізації даних до необхідного рівня (місяця, року), тим самим забезпечуючи захист від ре-ідентифікації. Для атрибутів, значення яких беруться з деякої наперед відомої скінченної множини, зазвичай будується відповідна ієрархія, яка використовується для заміни даних менш конкретними узагальненнями. Важливою особливістю даної техніки є збереження правдивості даних. Серед недоліків можна виділити втрату корисності даних у випадках над-

мірного узагальнення. Для вирішення даної проблеми було запропоновано техніку адаптивного узагальнення (adaptive generalization) [6], що дозволяє мінімізувати негативний ефект на корисність даних.

Техніка подавлення даних (suppression) [7] часто застосовується у алгоритмах на базі k -анонімності. Дана техніка полягає у вилученні певних значень з набору даних з метою покращення статистичних характеристик цього набору. Так, наприклад, у випадку, коли один запис виділяється через аномальні значення певних атрибутів, для досягнення задовільного рівня захисту анонімізованих даних, виникає потреба у надмірному узагальненні таких атрибутів, що в результаті призводить до суттєвих втрат корисності всього набору даних. Вилучення необхідних значень чи відповідних записів з набору дозволяє вирішити дану проблему. З іншої сторони, надмірне застосування техніки вилучення також негативно впливає на корисність даних.

У роботі [8] було запропоновано **техніку перенесення (relocation)**, яка полягає у перенесенні або зміні обмеженої кількості значень набору даних для забезпечення більш високого рівня захисту від ре-ідентифікації. Очевидно, що недоліком даної техніки є надання анонімізованим даним певного рівню неправдивості, що може призвести до погіршення загальної їх якості.

Окремо можна виділити **техніку перестановки (permutation)**, яка представляє собою змішування значень певного квазіідентифікатора між різними записами, зберігаючи при цьому загальні статистичні особливості оригінальних даних та забезпечуючи високий рівень захисту. Треба зазначити, що зловживання даною технікою може призвести до порушень зв'язків між атрибутами, що у свою чергу може погіршити цінність таких анонімізованих даних для аналізу, чутливого до таких зв'язків.

Техніка пертурбації (perturbation) використовується для додавання адитивного або мультиплікативного шуму до оригінальних даних (частіше до SAs). У випадку адитивного шуму додається значення з фіксованого діапазону, в той час як для мультиплікативного випадково обирається коефіцієнт множення. Дана техніка може бути неефективною для анонімізації даних з аномаліями, а тому її часто використовують у комбінації з іншими техніками для забезпечення достатнього рівня приватності. Окрім цього, вибір невідлого діапазону чи коефіцієнту може призвести до суттєвого зниження корисності вихідних даних.

Техніка бакетизації (bucketization) [9] спрямована на забезпечення вимог моделі приватності *l*-диверсності. Спочатку дані розбиваються на групи (класи еквівалентності) за значеннями QIs, після чого до кожної групи застосовується вищезгадана техніка перестановки, яка перемішує значення чутливого атрибуту у межах своєї групи. Такий підхід дозволяє отримати вищий рівень корисності даних у порівнянні з технікою узагальнення, але при цьому він також має певні недоліки. Так, наприклад, бакетизація не захищає дані від атак на розкриття членства конкретної особи (membership disclosure) через те, що квазіідентифікатори залишаються незмінними.

На практиці забезпечити виконання умов моделей k -анонімності, l -диверсності та t -близькості разом буває дуже складно без суттєвих втрат корисності даних, а іноді й взагалі неможливо. Покращити процес анонізації в даному випадку допомагає використання *техніки мікроагрегації* (microaggregation) [10]. Дана техніка в загальному випадку застосовується до числових QIs, хоча також може бути використана для атрибутів на базі категорій при застосуванні конвертації в числовий формат [11]. Спочатку дані розбиваються на певні кластери таким чином, щоб дані записів в одному кластері були максимально між собою, а дані записів із різних кластерів мали максимальну відмінність. При цьому, кожен кластер буде мати щонайменше k записів. Далі, значення QIs в рамках конкретного кластера обробляються найбільш доцільним способом (наприклад, замінюються на середнє значення), в результаті чого забезпечується захист даних від ре-ідентифікації. Проблемою даного підходу є пошук оптимального рішення у випадку багатовимірних даних, оскільки такий алгоритм має клас складності NP та потребує значних ресурсів для обчислення.

Техніку розділення (slicing) було запропоновано в роботі [12], як альтернативу технікам узагальнення та бакетизації. Ідея техніки розділення представляє собою сегментацію набору даних одночасно по атрибутам (вертикальна) і по записам (горизонтальна). Вертикальна сегментація утворює нові колонки через групування сильно зв'язаних між собою атрибутів, в той час як горизонтальна – створює групи даних (buckets) на основі записів. Після цього застосовується перемішування або сортування даних у кожному стовпці для того, щоб приховати зв'язок з атрибутами з інших стовпців. У результаті анонізовані дані мають кращу корисність у порівнянні з застосуванням технік узагальнення та бакетизації саме за рахунок збереження зв'язків між зв'язаними атрибутами.

Техніка диференційної приватності (differential privacy, DP) [13] – це математична техніка, яка як і техніка пертурбації, базується на ідеї додавання шуму до даних. На відміну від техніки пертурбації, де шум інтегрується безпосередньо в набір даних, механізм диференційної приватності застосовується на рівні запитів за даними, надаючи певний рівень випадковості до кожного результату. Виділяють локальну та глобальну диференційні приватності, що відрізняються між собою архітектурними особливостями системи, яка обробляє запити. У роботі [14] також було запропоновано варіант реалізації DP із застосуванням глибоких нейронних мереж.

До технік анонізації даних також можна віднести генерацію синтетичних даних (synthetic data). На відміну від вищезазначених технік, генерація синтетичних даних надає абсолютно нові дані, які тільки схожі за структурою та розподілом на оригінальні дані, але при цьому не містять ніякої персональної інформації, оскільки вони не можуть бути асоційовані з конкретною реальною особою. І хоча проблема створення штучних даних не є новою, за останнє десятиріччя досягнення у сфері генеративного штучного

інтелекту сприяли поширенню використання моделей штучного інтелекту для вирішення даної задачі.

Коли мова йде про структуровані дані, для навчання та безпосередньо генерації нових даних використовуються такі групи моделей III, як варіаційні автокодувальники (Variational Autoencoders, VAEs) [15] та генеративні змагальні мережі (Generative Adversarial Networks, GANs) [16]. Дані моделі вивчають структуру та внутрішні залежності даних, на яких вони навчаються, а також, що важливо для задачі анонізації, аналізують правила розподілу цих даних. Варіаційні автокодувальники представляють собою групу генеративних моделей, які працюють на принципі стискання даних до формату представлення, що називається латентним простором. Даний простір надалі використовується для відтворення вхідних даних. У свою чергу модель генеративної змагальної мережі складається з двох нейронних мереж – мережі генератора (G) та мережі дискримінатора (D), між якими виникає антагоністична гра (гра з нульовою сумою). Мережа G створює зразки штучних даних, а мережа D намагається вгадати, чи є даний зразок штучним. При цьому, генератор постійно вдосконалює створювані штучні зразки таким чином, щоб дискримінатор не зміг відрізнити штучні дані від реальних, а дискримінатор вчиться точніше робити висновки.

Повний процес генерації синтетичних даних складається з таких етапів, як тренування моделі на реальних даних, безпосередньо генерація нових наборів даних, а також оцінка якості та корисності цих вихідних даних. Важливо зазначити, що ключовим фактором можливості використання синтетичних даних є ступінь коректності процесу генерації таких даних у контексті вирішення певної задачі. Так, наприклад, якщо деяка система потребує великої кількості даних певної структури з метою аналізу продуктивності цієї системи, то в такому випадку характеристики самих даних та їх розподіл не будуть мати великої різниці, а тому не має потреби вимагати від процесу генерації високого рівня правдоподібності вихідних наборів даних.

Якісні синтетичні дані дозволяють вирішити багато проблем, пов'язаних з обмеженнями використання та розповсюдження персональних даних, а в деяких випадках просто не мають альтернатив. На практиці синтетичні дані часто використовуються для тренування та перевірки моделей штучного інтелекту. Використання синтетичних даних також має ряд недоліків. Так, налаштування процесу генерації синтетичних даних є вкрай складною задачею, де особливу увагу треба приділяти оцінці згенерованих даних, їх якості та корисності відносно вирішення конкретної задачі.

Висновки. У роботі було розглянуто проблему зберігання та використання структурованих медичних даних, а також їх анонізацію, як механізм забезпечення приватності персональної інформації. У контексті анонізації було розглянуто три основні моделі приватності k -анонімність, l -диверсність та t -близькість. Були описані ролі даних моделей в забезпеченні анонімності, а також їх недоліки, що зумовили необхідність в створенні інших моделей приватності та

таких технік анонізації, як диференційна приватність та генерація синтетичних даних. Було також розглянуто техніки узагальнення, подавлення, перенесення, перестановки, пертурбації та розділення. Дані техніки використовуються для обробки даних таким чином, щоб виконати вимоги відповідних моделей приватності. Не дивлячись на те, що поєднання розглянутих моделей приватності та технік анонізації дозволяють гарантувати достатньо високий рівень захисту від реідентифікації, особливу увагу треба приділити оцінці корисності вихідних даних. Так, застосування неякісних анонізованих даних не тільки буде неефективним, а може навіть призвести до шкідливих наслідків. Такі техніки, як диференційна приватність та генерація синтетичних даних, відрізняються своїм підходом до забезпечення приватності даних та мають широке практичне застосування. Більш того, активний розвиток генеративних нейронних мереж відкрив нові можливості для генерації синтетичних даних, що робить даний напрям перспективним для подальшого дослідження.

Список використаної літератури

- Document 32016R0679. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance). *EUR-Lex. Access to European Union law*. URL: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32016R0679> (дата звернення: 30.10.2024).
- Summary of the HIPAA Privacy Rule*. USA: United States Department of Health and Human Services. URL: <https://www.hhs.gov/hipaa/for-professionals/privacy/laws-regulations/index.html> (дата звернення: 30.10.2024).
- Sweeney L. k-anonymity: a model for protecting privacy. *International Journal on Uncertainty, Fuzziness and Knowledge-based Systems*. 2002. Vol. 10, is. 5. P. 557–570. URL: https://epic.org/wp-content/uploads/privacy/reidentification/Sweeney_Article.pdf (дата звернення: 30.10.2024).
- Machanavajjhala A., Kifer D., Gehrke J., Venkatasubramanian M. L-diversity: Privacy beyond k-anonymity. *ACM Transactions on Knowledge Discovery from Data*. 2007. Vol. 1, is. 1. P. 3 – es.
- Li N., Li T., Venkatasubramanian S. t-Closeness: Privacy Beyond k-Anonymity and l-Diversity. *IEEE 23rd International Conference on Data Engineering*. 2007. P. 106–115. URL: <https://ieeexplore.ieee.org/document/4221659> (дата звернення: 01.11.2024).
- Majeed A., Ullah F., Lee S. Vulnerability- and Diversity-Aware Anonymization of Personally Identifiable Information for Improving User Privacy and Utility of Publishing Data. *Sensors*. 2017. Vol. 17, is. 5. Article 1059. URL: <https://www.mdpi.com/1424-8220/17/5/1059> (дата звернення: 01.11.2024).
- Samarati P., Sweeney L. Protecting Privacy when Disclosing Information: k-Anonymity and Its Enforcement through Generalization and Suppression. *Semantic Scholar*. 1998. URL: <https://www.semanticscholar.org/paper/Protecting-privacy-when-disclosing-information%3A-and-Samarati-Sweeney/7df12c498fecedac4ab6034d3a8032a6d1366ca6> (дата звернення: 30.10.2024).
- Nergiz M. E., Gök M. Z. Hybrid k-anonymity. *Computers & security*. 2014. Vol. 44. P. 51–63. URL: https://www.researchgate.net/publication/261139007_Hybrid_k-Anonymity (дата звернення: 30.10.2024).
- Martin D. J., Kifer D., Machanavajjhala A., Gehrke J., Halpern J. Y. Worst-Case Background Knowledge for Privacy-Preserving Data Publishing. *IEEE 23rd International Conference on Data Engineering*. 2007. P. 126–135. URL: https://www.academia.edu/1520959/Worst_Case_Background_Kno
- wledge_for_Privacy_Preserving_Data_Publishing (дата звернення: 30.10.2024).
- Domingo-Ferrer J., Mateo-Sanz J. M. Practical Data-Oriented Microaggregation for Statistical Disclosure Control. *IEEE Transactions on Knowledge and Data Engineering*. 2002. Vol. 14, no. 1. P. 189–201.
- Gal T. S., Tucker Th. C., Gangopadhyay A., Chen Z.. A data recipient centered de-identification method to retain statistical attributes. *Journal of biomedical informatics*. 2014. Vol. 50. P. 32–45.
- Li T., Li N., Zhang J., Molloy I. Slicing: A New Approach to Privacy Preserving Data Publishing. *IEEE Transactions on Knowledge and Data Engineering*. 2010. Vol. 24, is. 3. P. 561–574.
- Dwork C. Differential Privacy. *Automata, Languages and Programming*. 2006. P. 1–12.
- Abadi M., Chu A., Goodfellow I., McMahan H. B., Mironov I., Talwar K., Zhang L. Deep Learning with Differential Privacy. *CCS'16: Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*. 2016. P. 24–28.
- Kingma D. P., Welling M. Auto-Encoding Variational Bayes. *arXiv*. 2013. Article 1312.6114. URL: <https://www.semanticscholar.org/reader/5f5dc5b9a2ba710937e2c413b37b053cd673df02> (дата звернення: 01.11.2024).
- Goodfellow I. J., Pouget-Abadie J., Mirza M., Xu B., Warde-Farley D., Ozair Sh., Courville A., Bengio Y. Generative adversarial networks. *arXiv*. 2014. Article 1406.2661. URL: <https://dl.acm.org/doi/pdf/10.1145/3422622> (дата звернення: 01.11.2024).

References

- Document 32016R0679. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance). *EUR-Lex. Access to European Union law*. URL: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32016R0679> (accessed 30.10.2024).
- Summary of the HIPAA Privacy Rule*. USA: United States Department of Health and Human Services. URL: <https://www.hhs.gov/hipaa/for-professionals/privacy/laws-regulations/index.html> (accessed 30.10.2024).
- Sweeney L. k-anonymity: a model for protecting privacy. *International Journal on Uncertainty, Fuzziness and Knowledge-based Systems*. 20026 vol. 10, is. 5, pp. 557–570. URL: https://epic.org/wp-content/uploads/privacy/reidentification/Sweeney_Article.pdf (accessed 30.10.2024).
- Machanavajjhala A., Kifer D., Gehrke J., Venkatasubramanian M. L-diversity: Privacy beyond k-anonymity. *ACM Transactions on Knowledge Discovery from Data*. 2007, vol. 1, is. 1, pp. 3 – es.
- Li N., Li T., Venkatasubramanian S. t-Closeness: Privacy Beyond k-Anonymity and l-Diversity. *IEEE 23rd International Conference on Data Engineering*. 2007, pp. 106–115. URL: <https://ieeexplore.ieee.org/document/4221659> (accessed 01.11.2024).
- Majeed A., Ullah F., Lee S. Vulnerability- and Diversity-Aware Anonymization of Personally Identifiable Information for Improving User Privacy and Utility of Publishing Data. *Sensors*. 2017, vol. 17, is. 5, article 1059. URL: <https://www.mdpi.com/1424-8220/17/5/1059> (accessed 01.11.2024).
- Samarati P., Sweeney L. Protecting Privacy when Disclosing Information: k-Anonymity and Its Enforcement through Generalization and Suppression. *Semantic Scholar*. 1998. URL: <https://www.semanticscholar.org/paper/Protecting-privacy-when-disclosing-information%3A-and-Samarati-Sweeney/7df12c498fecedac4ab6034d3a8032a6d1366ca6> (accessed 30.10.2024).
- Nergiz M. E., Gök M. Z. Hybrid k-anonymity. *Computers & security*. 2014, vol. 44, pp. 51–63. URL: https://www.researchgate.net/publication/261139007_Hybrid_k-Anonymity (accessed 30.10.2024).
- Martin D. J., Kifer D., Machanavajjhala A., Gehrke J., Halpern J. Y. Worst-Case Background Knowledge for Privacy-Preserving Data Publishing. *IEEE 23rd International Conference on Data Engineering*. 2007, pp. 126–135. URL: https://www.academia.edu/1520959/Worst_Case_Background_Kno

- wledge_for_Privacy_Preserving_Data_Publishing (accessed 30.10.2024).
10. Domingo-Ferrer J., Mateo-Sanz J. M. Practical Data-Oriented Microaggregation for Statistical Disclosure Control. *IEEE Transactions on Knowledge and Data Engineering*. 2002, vol. 14, no. 1, pp. 189–201.
 11. Gal T. S., Tucker Th. C., Gangopadhyay A., Chen Z.. A data recipient centered de-identification method to retain statistical attributes. *Journal of biomedical informatics*. 2014, vol. 50, pp. 32–45.
 12. Li T., Li N., Zhang J., Molloy I. Slicing: A New Approach to Privacy Preserving Data Publishing. *IEEE Transactions on Knowledge and Data Engineering*. 2010, vol. 24, is. 3, pp. 561–574.
 13. Dwork C. Differential Privacy. *Automata, Languages and Programming*. 2006, pp. 1–12.
 14. Abadi M., Chu A., Goodfellow I., McMahan H. B., Mironov I., Talwar K., Zhang L. Deep Learning with Differential Privacy. *CCS'16: Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*. 2016, pp. 24–28.
 15. Kingma D. P., Welling M. Auto-Encoding Variational Bayes. *arXiv*. 2013, article 1312.6114. URL: <https://www.semanticscholar.org/reader/5f5dc5b9a2ba710937e2c413b37b053cd673df02> (accessed 01.11.2024).
 16. Goodfellow I. J., Pouget-Abadie J., Mirza M., Xu B., Warde-Farley D., Ozair Sh., Courville A., Bengio Y. Generative adversarial networks. *arXiv*. 2014, article 1406.2661. URL: <https://dl.acm.org/doi/pdf/10.1145/3422622> (accessed 01.11.2024).

Надійшло (received) 05.11.2024

UDC 004.67

D. V. KALININ, Postgraduate of Department of System Analysis and Information-Analytical Technologies, National Technical University "Kharkiv Polytechnic Institute", Kharkiv, Ukraine; e-mail: Denys.Kalinin@cs.khpi.edu.ua; ORCID: <https://orcid.org/0009-0004-4431-7728>

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; e-mail: Valerii.Severyn@khpi.edu.ua; ORCID: <https://orcid.org/0000-0002-2969-6780>

M. I. BEZMENOV, Candidate of Technical Sciences (PhD), Docent, Professor of the Department of System Analysis and Information-Analytical Technologies, National Technical University "Kharkiv Polytechnic Institute", Kharkiv, Ukraine; e-mail: Mykola.Bezmenov@khpi.edu.ua; ORCID: <https://orcid.org/0000-0002-2995-2350>

PRIVACY MODELS AND ANONYMIZATION TECHNIQUES FOR TABULAR HEALTHCARE DATA

In today's world, issues of privacy and personal data protection are becoming extremely relevant, especially in the healthcare field, where the use of large volumes of data for research is becoming increasingly common. The use of personal data is regulated by relevant laws that require data anonymization to minimize the risks of identifying individuals. Anonymization is a process that allows the use of sensitive data without the risk of disclosing personal information while maintaining its utility. This article discusses the main privacy models and anonymization techniques used to protect tabular healthcare data. Privacy models include k -anonymity, l -diversity, and t -closeness. The k -anonymity model ensures that any combination of quasi-identifiers is shared by at least k records. The l -diversity model complements k -anonymity by requiring at least l unique combinations of sensitive attribute (SA) values in each equivalence class. The t -closeness model considers the distribution of these sensitive attribute values, ensuring that the distance between the SA distribution in the equivalence class and the overall distribution does not exceed a specified threshold. Anonymization techniques include generalization, suppression, relocation, permutation, perturbation, slicing, differential privacy, and synthetic data. Generalization reduces the precision of quasi-identifiers. Suppression removes certain values from the dataset to improve its statistical characteristics. Relocation changes a limited number of values in the data to enhance protection. Permutation mixes the values of quasi-identifiers between records while preserving the overall statistical features of the dataset. Perturbation adds noise to the data, increasing privacy. The idea of differential privacy also involves adding noise, but this is done at the query processing stage. Generating synthetic data allows the creation of new datasets that are similar in characteristics to the original data.

Keywords: healthcare data, tabular data, data anonymization, privacy models, k -anonymity, l -diversity, t -closeness, data anonymization techniques, differential privacy.

Повні імена авторів / Author's full names

Автор 1 / Author 1: Калінін Денис Вікторович / Kalinin Denys Viktorovich

Автор 2 / Author 2: Северин Валерій Петрович / Severyn Valerii Petrovych

Автор 3 / Author 3: Безменов Микола Іванович / Bezmenov Mykola Ivanovych

О. М. КОНДРАТОВ, аспірант, старший викладач 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>

О. М. НИКУЛІНА, д-р техн. наук, професор, завідувачка кафедри інформаційних систем та технологій Національного технічного університету «Харківський політехнічний інститут», Харків, Україна; професор кафедри цифрових технологій та проєктно-аналітичних рішень ТОВ «ТУ «МЕТІНВЕСТ ПОЛІТЕХНІКА»», Запоріжжя, Україна; e-mail: elniknik02@gmail.com; ORCID: <https://orcid.org/0000-0003-2938-4215>

ПРОГРАМНА РЕАЛІЗАЦІЯ ІЗ ВИКОРИСТАННЯМ ТРАНСФОРМЕРА З ОПТИЧНИМ ПОТОКОМ ТА GEONET ДЛЯ ІДЕНТИФІКАЦІЇ ПАРАМЕТРІВ ДИНАМІЧНИХ ОБ'ЄКТІВ

Сьогодні міждисциплінарні дослідження в галузі комп'ютерних наук та інженерії стають все більш актуальними через зростаючу потребу обробляти дані в реальному часі для виявлення та відстеження об'єктів. Зокрема, ідентифікація параметрів динамічних об'єктів є критично важливою у таких сферах, як автономні транспортні системи, робототехніка, системи спостереження і моніторингу. Ефективне автоматичне отримання та обробка інформації з відеоджерел є перспективною галуззю для науковців і практиків, що працюють у суміжних областях. Метою цього дослідження є вдосконалення процесів виявлення та відстеження динамічних об'єктів шляхом розробки та впровадження інформаційної технології, заснованої на використанні сучасних методів машинного навчання, таких як DETR (Detection Transformer), оптичний потік та GeoNet. Методи дослідження включають розробку програмного забезпечення на основі мови програмування Python із застосуванням сучасних бібліотек та фреймворків для обробки зображень і відео. Для виявлення об'єктів було використано метод DETR, що дозволяє з високою точністю визначати положення об'єктів у кадрі. Оптичний потік застосовувався для визначення напрямку та швидкості їх переміщення, а GeoNet — для аналізу глибини сцени та геометричних параметрів. Запропонована технологія була протестована на різних відеозаписах, які відображають складні сценарії з динамічними умовами, включаючи зміну освітлення, перекриття об'єктів та швидкі зміни руху. Результати дослідження показали високу точність і надійність запропонованого підходу для ідентифікації параметрів динамічних об'єктів у різних умовах. Поєднання методів дозволило суттєво підвищити точність і стійкість системи виявлення та відстеження, особливо в умовах зміни середовища або низької якості відеозаписів. Висновки дослідження вказують на ефективність використання запропонованої інформаційної технології для практичного застосування у сфері автономних систем, робототехніки та відеоспостереження.

Ключові слова: Дистанційна ідентифікація динамічних об'єктів, виявлення об'єктів, оптичний потік, ідентифікація швидкості, глибоке навчання, згорткові нейронні мережі.

Вступ. Сьогодні швидка еволюція комп'ютерних технологій змінює наукові підходи до сучасних викликів і відкриває нові шляхи для створення нових методів або вдосконалення існуючих. Це особливо важливо для прикладних наук, де обчислювальні методи прискорюють розв'язання складних проблем, що вимагають обробки великих обсягів даних [1].

У галузі комп'ютерного зору досягнення в машинному навчанні та виявленні об'єктів революціонізували спосіб відстеження та аналізу динамічних об'єктів [2].

Одним з ключових досягнень є використання трансформера виявлення (DETR) для виявлення об'єктів та оптичного потоку (Optical Flow) для відстеження руху у відео послідовностях. Комбінація цих методів дозволяє точно ідентифікувати параметри динамічних об'єктів у різних середовищах. Спочатку моделі виявлення об'єктів мали обмежений спектр можливостей. Однак завдяки зростанню доступності відеоданих високої роздільної здатності та досягненням у глибокому навчанні моделі ідентифікації параметрів динамічних об'єктів значно еволюціонували, що дозволяє точно відстежувати і аналізувати об'єкти в реальному часі. Це розширило використання таких моделей у різних додатках, від автономних систем до спостереження, демонструючи критичну роль обчислювальних технологій у вирішенні сучасних проблем. Подальший розвиток комп'ютерних технологій, зокрема збільшення

обчислювальної потужності та зростання машинного навчання, відкрив нові напрями для науковців та інженерів у розширенні своїх методологій. Досягнення в області комп'ютерного зору призвели до створення моделей виявлення об'єктів, таких як DETR, яка дозволяє здійснювати виявлення об'єктів у реальному часі в відеопослідовностях, та Оптичний потік, який використовується для відстеження руху між кадрами. У динамічних середовищах точно ідентифікувати параметри об'єктів, такі як позиція, швидкість і траєкторія, є критично важливим для додатків, починаючи від автономних транспортних засобів і до спостереження. Характеристики реальних даних, такі як великий обсяг, швидкість і різноманіття, створюють труднощі для моделей ідентифікації параметрів динамічних об'єктів, оскільки їм необхідно швидко та точно обробляти великі обсяги відеоданих [3–4].

Сьогодні важливо використовувати моделі виявлення та відстеження об'єктів, які можуть адаптуватися до змінних умов середовища та працювати в реальному часі. Для ідентифікації параметрів динамічних об'єктів цей підхід є важливим для визначення таких характеристик, як розмір об'єкта, швидкість і напрямок. Дослідники в області комп'ютерного зору погоджуються, що інтеграція DETR та оптичного потоку відкриває нові можливості для міждисциплінарної співпраці між комп'ютерними науковцями та інженерами.

© Кондратов О. М., Нікуліна О. М., 2024



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.



Ключовим моментом для моделей виявлення об'єктів є їх здатність залишатися гнучкими та адаптованими до постійно змінюваних умов середовища. Багато галузей прагнуть впроваджувати такі технології для автоматизації та передової аналітики. Однак слід зазначити, що різні додатки можуть мати різні вимоги та цілі. В таких умовах цікаво дослідити, як різноманітні динамічні середовища впливають на ефективність моделей виявлення об'єктів.

Це дослідження має на меті відповісти на такі наукові питання.

1. Як можна покращити точність виявлення динамічних об'єктів у відео з використанням сучасних моделей машинного навчання? Це питання стосується ефективності різних моделей, таких як DETR для виявлення об'єктів, та аналізу їх застосування в реальних умовах, зокрема при обробці відео-потoku.

2. Яка роль оптичного потоку у відстеженні об'єктів і як він може допомогти в розпізнаванні руху в реальному часі? Вивчення того, як оптичний потік допомагає в аналізі руху, і як його можна використовувати для визначення напрямку та швидкості переміщення об'єктів в реальному часі, є важливим аспектом.

3. Як ефективно інтегрувати методи глибокого навчання, такі як DETR і GeoNet, для покращення точності та надійності систем виявлення і відстеження об'єктів у динамічних сценах?

4. Важливо розглянути, як поєднання DETR для виявлення об'єктів та GeoNet для глибокого навчання може підвищити точність виявлення та відстеження, зокрема в складних умовах, таких як зміни освітлення або перекриття об'єктів.

5. Які фактори впливають на ефективність моделей в реальних умовах, зокрема на точність виявлення та відстеження в умовах змінюваного оточення та швидкості? Вивчення того, як різні фактори, такі як зміна освітлення, туман, або швидкість руху об'єктів, можуть впливати на продуктивність моделей виявлення та відстеження, є важливою частиною цього дослідження.

Ця стаття представляє комплексний підхід до аналізу динамічних об'єктів у відеопотоках, поєднуючи DETR, оптичний потік та GeoNet для глибокої оцінки параметрів об'єктів, таких як положення, швидкість, напрямок руху та глибина. Розроблений конвеєр включає завантаження відео, обробку кадрів, виявлення об'єктів, аналіз руху та оцінку глибини. Система продемонструвала високу ефективність у різних сценаріях, включаючи складні фони та динамічні середовища [4–8].

Мета та задачі дослідження. Мета статті полягає в використанні методів ідентифікації параметрів динамічного об'єкту з використанням DETR та Optical Flow для побудови моделі в технології DETR.

Для досягнення мети поставлені задачі дослідження.

1. Описати аналітичну та реалізовану модель.
2. Описати реалізовану обробку відео-потoku.
3. Описати використання моделі DETR.
4. Описати використання оптичного потоку.

5. Описати розроблену модель з використанням GeoNet.

6. Аналіз результатів.

Аналітична модель. Схема ідентифікації параметрів динамічного об'єкту з використанням DETR та Optical Flow показана на рис. 1 і включає елементи. Обробка кадрів відеопотоку. Кожен кадр (1 кадр, 2 кадр і т. д.) аналізується з використанням Відео вхід. Детектор (DETR), що дозволяє визначити положення та межі об'єктів у кадрі. Застосування алгоритму оптичного потоку для визначення векторів руху об'єктів між послідовними кадрами. Аналіз параметрів об'єктів. На основі отриманих векторів руху проводиться аналіз параметрів об'єктів, включаючи швидкість, напрямок руху, траєкторію та інші характеристики [3, 6, 8, 9].

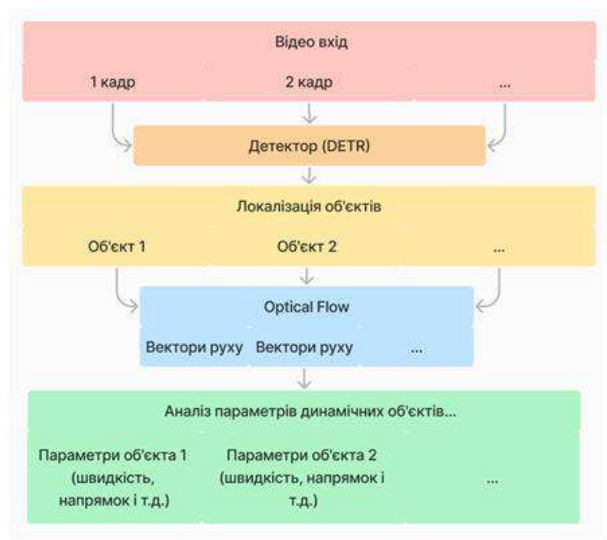


Рис. 1. Аналітична модель

Реалізована модель. На рис. 2, представлено реалізовану модель, яка включає обробку відео потоку, використання моделі DETR, використання оптичного потоку, формування і відображення результатів. Колорами відмічено різні елементи реалізованої моделі.



Рис. 2. Реалізована модель

Розширена реалізована модель. На рис. 3 представлено розширений опис реалізованої моделі. Яка включає наступні елементи у вигляді конвеєру:



Рис. 3. Розширений опис реалізованої моделі

1. Обробку відео потоку. Збереження відео файлу з відеопотоку. Екстракція кадрів з відеофайлу. Отримання кадрів з відео файлу для подальшої їх обробки.

2. Модель DETR. Кожен кадр обробляється моделлю для виявлення об'єктів, тобто виділення обмежувальних рамок об'єктів знайдених на кадрах. Отримання кадрів з рамками.

3. Обчислення оптичного потоку. Визначення векторів руху. Аналіз швидкості та напрямку. Отримання візуалізації руху.

Обробка відео-потоку. У розробленому застосунку здійснюється обробка відео-потоку з використанням моделі DETR для виявлення об'єктів та обчислення оптичного потоку між послідовними кадрами. Відео спочатку завантажується з YouTube (якщо файл не існує на диску), після чого аналізується кадр за кадром. Для кожного кадру проводиться виявлення об'єктів, відображення коробок навколо об'єктів та обчислення оптичного потоку для оцінки руху об'єктів. Застосунок використовує моделі на базі PyTorch і TensorFlow для максимізації продуктивності на GPU або TPU (Google Colab).

Опис використання моделі DETR. Модель DETR (Detection Transformer) використовується для виявлення об'єктів у кожному кадрі відео. Вона отримує зображення, попередньо оброблене через `DetrImageProcessor`, і проводить детекцію за допомогою трансформера. Виходи моделі включають логіти та координати предиктованих коробок для кожного виявленого об'єкта. Детекція здійснюється з використанням `softmax` для прогнозування класів об'єктів і їх точних координат. Виявлені об'єкти накладаються на оригінальне зображення через функцію `draw_boxes`, що дозволяє візуалізувати результати.

Використовується декодер, який має схожу структуру до кодера з деякими додатковими компонентами. Список шарів включає багатоголову самоувагу, суму та нормування, багатоголову увагу, суму та нормування, нейронну мережу прямого поширення, як і в кодері використовується багатощаровий перцептрон [3, 10].

Опис використання оптичного потоку. Оптичний потік обчислюється між двома зображеннями,

`prev_frame_gray` та `next_frame_gray`, за допомогою методу `Farneback`, який використовує параметри для покращення точності виявлення руху. Параметри включають використання пірамідного масштабу (0.5), трьох рівнів піраміди, вікно 15x15 пікселів для розрахунку потоку, три ітерації уточнення, поліноміальне розширення з розміром сусідства 5 пікселів, а також Гауссовий фільтр з відхиленням 1.2. Ці налаштування дозволяють отримати більш точні результати для аналізу руху об'єктів у відео, що важливо для відстеження динамічних об'єктів.

На рис. 4, представлено результат роботи використання оптичного потоку.

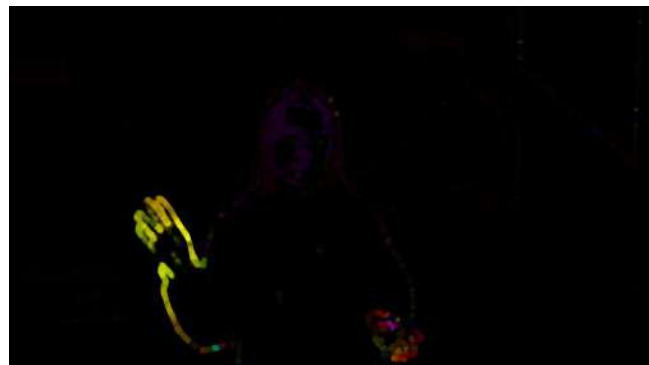


Рис. 4. Результат роботи використання оптичного потоку

Опис додаткової розробленої моделі з використанням GeoNet. Реалізована інтеграція GeoNet для обробки відео шляхом оцінки глибини зображень (`monocular depth estimation`). Що ідентифікує, ще один із найважливіших параметрів об'єкта, а саме глибину.

GeoNet працює із відеофраємами, які попередньо витягуються з відео, зберігаються у вигляді зображень та формуються для подальшої обробки. Входи обробляються через GeoNet, що передбачає отримання параметрів.

Етап підготовки даних. GeoNet здійснює попередню обробку кадрів відео з використанням скрипта. Цей етап включає зміну розмірів зображень, створення необхідних послідовностей кадрів (наприклад, для аналізу оптичного потоку або оцінки глибини) і

видалення статичних об'єктів для покращення результатів.

Інференс. Основний обчислювальний етап виконується за допомогою `geonet_main.py`, який аналізує дані та генерує результати, такі як глибинні карти, оптичний потік або оцінки позиції. Для цього використовуються попередньо натреновані моделі, збережені у вигляді контрольних точок (checkpoints).

Вивід результатів. Глибинні карти (depth maps) зберігаються у визначеному каталозі та візуалізуються для оцінки результатів роботи моделі.

На рис. 5, представлено реалізована модель з використанням GeoNet.

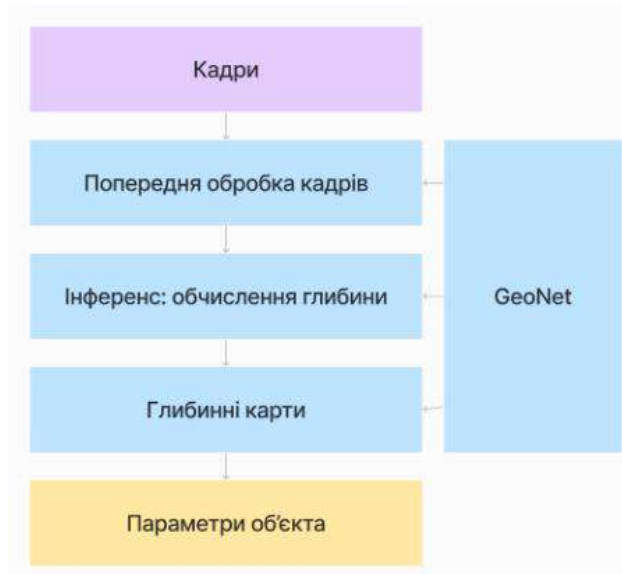


Рис. 5. Реалізована модель з використанням GeoNet

На рис. 6 представлено загальну реалізовану модель з використанням GeoNet. Яка реалізована для паралельного обчислення з DETR та оптичним потоком.



Рис. 6. Загальна реалізована модель з використанням GeoNet

Оцінка глибини (Depth Estimation). Результати роботи GeoNet продемонстрували можливість точного визначення глибини на основі кадрів відео. Згенеровані глибинні карти надають детальну інформацію про структуру сцени, що корисно для відстеження об'єктів у просторі та побудови 3D-моделей [8, 11–12].

Підготовка та обробка даних. Процес підготовки даних для GeoNet включає збереження кадрів із відео, їх форматування та видалення статичних об'єктів. Це дозволяє зменшити вплив шуму та покращити якість оцінки глибини.

Аналіз результатів. У результаті виконання обробки відео з використанням DETR, модель успішно виявляє об'єкти на кожному кадрі, і середній час обробки для одного кадру може бути вимірний, що дозволяє оцінити ефективність моделі.

В обчисленні оптичного потоку за допомогою методу Farneback на основі двох послідовних кадрів отримано точні дані про напрямок та швидкість руху об'єктів, що дозволяє покращити відстеження візуальних змін у кадрі.

Для кожного кадру відео вимірюється час обробки для детекції об'єктів та для обчислення оптичного потоку, що дозволяє оцінити ефективність кожного етапу обробки та оптимізувати продуктивність системи.

Інтеграція з іншими методами. У поєднанні з моделлю DETR результати глибинних карт використані для покращення розпізнавання об'єктів, а також для визначення їх положення та руху в просторі, що робить систему потужним інструментом для аналізу відео-потоків.

Для покращення реалізованої моделі та майбутнього представлення у вигляді цілісної системи доцільним є використання методів бустінгу та бегінгу. Бустінг дозволяє послідовно комбінувати кілька моделей, кожна з яких коригує помилки попередньої, що забезпечує високу точність виявлення та відстеження об'єктів. Застосування цього підходу може значно покращити продуктивність системи в умовах складних середовищ, таких як зміна освітлення або перекриття об'єктів. У поєднанні з DETR та оптичним потоком бустінг здатний підвищити ефективність аналізу динамічних об'єктів у відеопотоках.

Бегінг, у свою чергу, забезпечує зниження варіативності моделей шляхом паралельного навчання декількох базових моделей на різних підмножинах даних. Це сприяє підвищенню стійкості системи до шумів у вхідних даних та підвищує загальну надійність. Інтеграція цих підходів із застосуванням мета-алгоритму дозволяє створити ансамбль, який об'єднує результати всіх компонентів у єдину систему. Така архітектура забезпечить оптимальну взаємодію між DETR, GeoNet, оптичним потоком та іншими компонентами, що значно покращить ефективність і адаптивність системи в реальних сценаріях [13].

Висновки. Динамічні об'єкти є важливим елементом в багатьох сферах, таких як автономні автомобілі, робототехніка, моніторинг безпеки та медична діагностика. Ідентифікація їх параметрів, таких як швидкість, напрямок руху та інші відіграє ключову роль у розумінні та прогнозуванні їх поведінки.

Використання згорткових мереж для виявлення ознак з відеопослідовностей, інтеграція технік трансформерів для покращення точності виявлення та ідентифікації об'єктів, оптимізація методів навчання для обробки великих обсягів даних та динамічних змін у відеопослідовностях – і елементи допомагають у ство-

ренні ефективної системи для ідентифікації параметрів динамічних об'єктів з використанням переваг як згорткових нейронних мереж, так і трансформерів та оптичного потоку. Використання моделі, яка поєднує у собі DETR та метод Optical Flow для ефективної ідентифікації параметрів динамічного об'єкту. DETR є потужним інструментом для точного виявлення об'єктів на зображеннях.

Комп'ютерні технології суттєво змінюють підходи до вирішення сучасних проблем, зокрема в галузі комп'ютерного зору. Інтеграція методів, таких як DETR для виявлення об'єктів, оптичний потік для аналізу руху та GeoNet для оцінки глибини, відкриває нові можливості для точного відстеження динамічних об'єктів у реальному часі. Це дозволяє визначити такі параметри, як швидкість, траєкторія та глибина, що є критично важливим для застосувань у автономних транспортних системах, відеоспостереженні та інших галузях. Завдяки обчислювальним потужностям GPU і TPU реалізовані методи демонструють високу ефективність у складних середовищах.

Особливу увагу заслуговує аналітична та реалізована модель обробки відео, яка включає етапи екстракції кадрів, виявлення об'єктів за допомогою DETR і аналіз руху через оптичний потік. Використання попередньо натренованих моделей дозволяє отримати точні результати навіть у складних умовах, таких як зміни освітлення чи перекриття об'єктів.

Інтеграція глибинних карт GeoNet значно покращує здатність ідентифікації об'єктів у просторі, що є важливим для 3D-реконструкції сцен.

Усі ці методи були оптимізовані для роботи в режимі реального часу, що дозволяє використовувати їх на мобільних пристроях і в системах із обмеженими ресурсами.

Оцінка ефективності реалізованої системи базується на метриках, таких як час обробки кадрів, точність виявлення та стабільність результатів. Інтеграція DETR і GeoNet забезпечує високу адаптивність до змінних умов середовища, що робить ці методи універсальними для різних застосувань.

Подальший розвиток цього підходу сприятиме підвищенню точності та швидкості обробки, що відкриває перспективи для ширшого впровадження в міждисциплінарні проекти. Загалом, дослідження демонструє важливість інноваційних обчислювальних методів у вирішенні складних задач аналізу відеопотоків.

Реалізована модель, яка об'єднує сучасні алгоритми DETR, оптичний потік та GeoNet, продемонструвала високу ефективність для ідентифікації параметрів динамічних об'єктів у складних сценаріях. Інтеграція цих методів дозволила суттєво покращити точність, швидкість і стійкість обробки відеоданих, зокрема в умовах змінного освітлення, шуму та перекриття об'єктів. Геометрична інформація, отримана за допомогою GeoNet, забезпечила додатковий рівень деталізації, зокрема аналіз глибини сцени, що значно розширило можливості системи.

Результати дослідження підкреслюють перспективність міждисциплінарного підходу до розробки сис-

тем ідентифікації динамічних об'єктів. Застосування поєднаних методик має великий потенціал для практичного використання у таких сферах, як автономні транспортні системи, відеоспостереження та робототехніка. Подальший розвиток і оптимізація цих підходів сприятиме створенню більш досконалих рішень для аналізу відеоданих у реальному часі.

Список використаної літератури

1. Wang Z., Turko R., Shaikh O., Park H., Das N., Hohman F., Kahng M., Chau D. *CNN Explainer: Learning Convolutional Neural Networks with Interactive Visualization*. URL: <https://arxiv.org/abs/2004.15004> (дата звернення: 06.11.2024).
2. Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A., Kaiser L., Polosukhin I. *Attention Is All You Need*. URL: <https://arxiv.org/abs/1706.03762> (дата звернення: 06.11.2024).
3. Carion N., Massa F., Synnaeve G., Usunier N., Kirillov A., Zagoruyko S. *End-to-End Object Detection with Transformers*. URL: <https://arxiv.org/abs/2005.12872v3> (дата звернення: 06.11.2024).
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=IwAR0ILGAWTtWU-9-iH6lZyPFXyXA5JRWarM_XoSJ78QEhmm-txvr_iGEzCio (дата звернення: 06.11.2024).
5. 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: 06.11.2024).
6. Yin Z., Shi J. *GeoNet: Unsupervised Learning of Dense Depth, Optical Flow and Camera Pose*. URL: <https://arxiv.org/abs/1803.02276v2> (дата звернення: 06.11.2024).
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. Нікуліна О. М., Северин В. П., Кондратов О. М., Ольховий О. М. Моделі дистанційної ідентифікації параметрів динамічних об'єктів з використанням трансформерів виявлення та оптичного потоку. *Вісник НТУ «ХПІ»*. Серія: Системний аналіз, управління та інформаційні технології. Харків. НТУ «ХПІ», 2024. № 1 (11). С. 52–57.
9. Нікуліна О. М., Кондратов О. М. Модель ідентифікації параметрів динамічного об'єкту з використанням DETection TRansformer та Optical Flow. *Інформаційні технології: наука, техніка, технологія, освіта, здоров'я: Тези доповідей XXXII міжнародної науково-практичної конференції MicroCAD-2024, 22-24 травня 2024 р.* Харків, НТУ «ХПІ». 2024. С. 1047.
10. Нікуліна О. М., Кондратов О. М. Методи дистанційної ідентифікації динамічних параметрів об'єкта. *Інформаційні технології: наука, техніка, технологія, освіта, здоров'я: Тези доповідей XXXI міжнародної науково-практичної конференції MicroCAD-2023, 17-20 травня 2023 р.* Харків, НТУ «ХПІ». 2023. С. 1047.
11. Gracyk A., Chen X. *GeONet: a neural operator for learning the Wasserstein geodesic*. URL: <https://arxiv.org/abs/2209.14440> (дата звернення: 06.11.2024).
12. Inomata T., Kimura K., Hagiwara M. *Object Tracking and Classification System Using Agent Search*. URL: https://www.jstage.jst.go.jp/article/ieejagents/129/11/129_11_2065/_pdf/-char/ja (дата звернення: 06.11.2024).
13. Gavrylenko S., Chelak V., Hornostal O. Construction Method Of Fuzzy Decision Trees For Identification The Computer System State. *2022 XXXII International Scientific Symposium Metrology and Metrology Assurance (MMA)*. 2022. P. 1–5.

References (transliterated)

1. Wang Z., Turko R., Shaikh O., Park H., Das N., Hohman F., Kahng M., Chau D. *CNN Explainer: Learning Convolutional Neural Networks with Interactive Visualization*. URL: <https://arxiv.org/abs/2004.15004> (accessed: 06.11.2024).
2. Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A., Kaiser L., Polosukhin I. *Attention Is All You Need*. URL: <https://arxiv.org/abs/1706.03762> (accessed: 06.11.2024).

3. Carion N., Massa F., Synnaeve G., Usunier N., Kirillov A., Zagoruyko S. *End-to-End Object Detection with Transformers*. URL: <https://arxiv.org/abs/2005.12872v3> (accessed: 06.11.2024).
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=IwAR0ILGAWTtWU-9-iH6lZyPFXYXA5JRWarM_XoSJ78QEhmnn-txvr_iGEzCio (accessed: 06.11.2024).
5. 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: 06.11.2024).
6. Yin Z., Shi J. *GeoNet: Unsupervised Learning of Dense Depth, Optical Flow and Camera Pose*. URL: <https://arxiv.org/abs/1803.02276v2> (accessed: 06.11.2024).
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. Nikulina O. M., Severin V. P., Kondratov O. M., Olhovoy O. M. Models of remote identification of parameters of dynamic objects using detection transformers and optical flow. *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., no. 1 (11), pp. 52–57.
9. Nikulina O. M., Kondratov O. M. Dynamic object parameter identification model using DETection TRansformer and Optical Flow. *Information technologies: science, technology, technology, education, health: Abstracts of reports of XXII international scientific and practical conference MicroCAD-2024, May 22–24, 2024, Kharkiv, NTU "KhPI"*, 2024, p. 1047.
10. Nikulina O. M., Kondratov O. M. Methods of remote identification of dynamic object parameters. *Information technologies: science, technology, technology, education, health: Abstracts of reports of the XXI international scientific and practical conference MicroCAD-2023, May 17–20, 2023, Kharkiv, NTU "KhPI"*, 2023, p. 1047.
11. Gracyk A., Chen X. *GeONet: a neural operator for learning the Wasserstein geodesic*. URL: <https://arxiv.org/abs/2209.14440> (accessed: 06.11.2024).
12. 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: 06.11.2024).
13. Gavrylenko S., Chelak V., Hornostal O. Construction Method Of Fuzzy Decision Trees For Identification The Computer System State. *2022 XXXII International Scientific Symposium Metrology and Metrology Assurance (MMA)*. 2022, pp. 1–5.

Надійшла (received) 08.11.2024

UDC 004.8+004.9

O. M. KONDRATOV, Postgraduate, senior lecturer of Department Information Systems and Technologies National Technical University «Kharkiv Polytechnic Institute», Kharkiv, Ukraine; e-mail: kondratovolexiy@gmail.com; ORCID: <https://orcid.org/0000-0001-6367-9944>

O. M. NIKULINA, Doctor of Technical Sciences, Professor, Head of Department Information Systems and Technologies National Technical University «Kharkiv Polytechnic Institute», Kharkiv, Ukraine; Professor of Department IT, Analysis and Project Decisions, Technical University "METINVEST POLYTECHNICS", LLC, Zaporizhzhia, Ukraine; e-mail: elniknik02@gmail.com; ORCID: <https://orcid.org/0000-0003-2938-4215>

SOFTWARE IMPLEMENTATION USING TRANSFORMER WITH OPTICAL FLOW AND GEONET FOR IDENTIFYING PARAMETERS OF DYNAMIC OBJECTS

Today, interdisciplinary research in computer science and engineering has become increasingly relevant due to the growing demand for real-time data processing in object detection and tracking applications. The identification of dynamic object parameters plays a crucial role in various domains such as autonomous transportation systems, robotics, and surveillance. Effective automated acquisition and processing of video data represent a promising field for scientists and practitioners working in these interconnected disciplines. This research aims to enhance object detection and tracking processes by developing and implementing an information technology solution based on modern machine learning methods, including DETR (Detection Transformer), Optical Flow, and GeoNet. The research methodology involves designing software using Python programming language and modern libraries and frameworks for image and video processing. The DETR method was employed for precise object detection within video frames. Optical Flow was used to determine the direction and velocity of object movement, and GeoNet provided depth and geometric scene analysis. The proposed technology was tested on diverse video recordings depicting complex scenarios with dynamic conditions, such as varying lighting, object occlusions, and rapid motion changes. The results demonstrate the high accuracy and reliability of the proposed approach for identifying dynamic object parameters under various conditions. The integration of these methods significantly improved the precision and robustness of the detection and tracking system, particularly in challenging environments or low-quality video scenarios. The study concludes that the proposed information technology is effective and can be applied in practical fields such as autonomous systems, robotics, and video surveillance.

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: Кондратов Олексій Михайлович, Kondratov Oleksii Mikhailovich

Автор 2 / Author 2: Нікуліна Олена Миколаївна, Nikulina Olena Mykolaivna

V. O. SHAROV, Postgraduate of Department Information Systems and Technologies National Technical University "Kharkiv Polytechnic Institute", Kharkiv, Ukraine; e mail: wycypiy@gmail.com; ORCID: <https://orcid.org/0000-0003-3152-0650>
O. M. NIKULINA, Doctor of Technical Sciences, Professor, Head of Department Information Systems and Technologies National Technical University "Kharkiv Polytechnic Institute", Kharkiv, Ukraine; e mail: elniknik02@gmail.com; ORCID: <https://orcid.org/0000-0003-2938-4215>

STUDY OF COMPATIBILITY OF METHODS AND TECHNOLOGIES OF HIGH-LEVEL PROTOCOLS AND ERROR-CORRECTING CODES

Since the year 2000, the fields of error-correction codes and Virtual Private Networks (VPNs) have undergone significant advancements driven by technological demands for higher reliability and security in communication systems. In error-correction codes, the development of turbo codes and Low-Density Parity-Check (LDPC) codes reached new heights, with LDPC codes being adopted in standards like 5G and Wi-Fi 6 for their near-Shannon-limit performance. This period saw groundbreaking contributions from researchers like David MacKay and Radford Neal, who refined LDPC algorithms, and Erdal Arkan, who introduced polar codes in 2008. Polar codes have since been integrated into 5G systems due to their efficiency and low complexity, marking a milestone in modern coding theory. Advances in decoding methods, such as belief propagation and successive cancellation, further enhanced the utility of these codes in practical applications. Parallel to these developments, VPN technology evolved in response to the growing need for secure and private communication in an increasingly interconnected world. Enhanced encryption protocols such as IPsec and OpenVPN became widespread, supported by innovations in cryptography. Researchers like Hugo Krawczyk contributed to robust authentication mechanisms, such as the HMAC and IKEv2 protocols, ensuring the integrity and confidentiality of VPN tunnels. Meanwhile, the development of WireGuard in the mid-2010s, spearheaded by Jason A. Donenfeld, introduced a lightweight and highly secure VPN protocol, revolutionizing the way modern VPNs operate. These advancements addressed the escalating cyber threats and facilitated the secure exchange of data across global networks. The importance of studying error-correction codes and VPNs in the modern era cannot be overstated. Error-correction codes are integral to overcoming the challenges of high-noise environments, enabling reliable communication in technologies ranging from space exploration to massive IoT networks. Simultaneously, VPNs remain critical for preserving user privacy, securing corporate networks, and protecting sensitive data in the face of sophisticated cyberattacks. Emerging technologies like quantum computing and artificial intelligence introduce both opportunities and threats, necessitating continuous innovation in these fields. Exploring quantum error-correction codes and post-quantum cryptographic protocols represents a vital area for future research. By addressing these challenges, scientists and engineers can ensure the resilience and security of communication systems in an increasingly digital and interconnected world.

Keywords: VPN, FEC, ECC, CIA triad, common security model, cascade codes, data transmission channel

Introduction. In the modern era of digital communication, the demand for high-speed, reliable data transmission has never been greater. High-level protocols, error-correcting codes (ECC), also known as forward error correcting codes (FEC), are fundamental components of this ecosystem, enabling efficient and secure information exchange across diverse networks. However, the growing complexity of communication systems, driven by the proliferation of IoT devices, 5G networks, and emerging technologies like quantum computing, has highlighted the critical need for compatibility between these methods and technologies [1].

Studying the compatibility of methods and technologies of high-level protocols, among which VPN protocols are set, and error-correcting codes is highly valuable and relevant for several reasons:

- ensuring interoperability: modern communication systems are built on heterogeneous networks that integrate a wide range of devices and technologies. Ensuring compatibility between high-level protocols and ECCs is essential for seamless interoperability, reducing latency, and preventing data loss;
- optimizing network efficiency: compatibility directly impacts the efficiency of data transmission. By aligning protocols with the most suitable error-correcting codes, it is possible to reduce overhead, enhance throughput, and optimize bandwidth utilization, which is critical for applications requiring high data rates such as streaming services, online gaming, and real-time communications;

- enhancing security and reliability: error-correcting codes are crucial for mitigating data corruption, especially in noisy channels or environments with high interference. Compatibility with high-level protocols ensures that ECC mechanisms are effectively implemented, safeguarding data integrity and improving the reliability of communication systems;

- supporting emerging technologies: the evolution of new communication paradigms, such as 5G, 6G, and edge computing, introduces new challenges in protocol design and error correction. A thorough understanding of compatibility between these components allows for better adaptation to these innovations, supporting the development of robust and scalable network architectures;

- reducing implementation costs: mismatched or incompatible protocol and ECC implementations can lead to inefficiencies, increased error rates, and costly redesigns. By studying and ensuring compatibility early in the design process, organizations can reduce implementation [2].

The integration of VPN technologies with advanced error-correcting methods, such as FEC, is not just about improving individual components but about creating a holistic, resilient system capable of meeting the challenges of modern communication networks. By enhancing the compatibility and performance of these technologies, we can achieve more secure, efficient, and reliable data transmission across diverse and demanding environments. It becomes crucial when trying to deal with CIA triad and its prerequisites [3].

© Sharov V.O., Nikulina O.M., 2024



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 and the objectives of study. The purpose of study is in assessing the modern ways of connection between high-level OSI protocols and ECC, clarifying potential and options of unifying them in a concept of one technology.

Objectives can be defined as following:

- consider the possibilities of combining the use of error-correcting codes and VPN within a single outline of a general model for secure data transmission;
- study the possibility of combining error-correcting codes with VPN protocols;
- research the impact of combining error-correcting codes with VPN protocols.

Comprehensive Overview: combining VPN and Error-Correcting Codes (FEC). In modern data transmission, especially over VPN, issues such as unstable networks, high latency, and packet loss can reduce the quality and reliability of connections. By integrating Forward Error Correction (FEC) codes with VPN technologies, we can significantly enhance the stability and security of data transfer.

FEC codes add redundant data to the transmitted information, allowing the system to detect and correct errors without needing retransmissions. This is crucial for VPN scenarios where continuous, secure connections are vital. Below is a detailed analysis of different VPN types and their potential combinations with various FEC codes.

Tunneling in VPN: detailed description. Tunneling is the method of encapsulating original network traffic within another data transfer protocol, creating a secure communication channel between the client and the server.

Main Tunneling Types in VPN are presented below.

- PPTP (Point-to-Point Tunneling Protocol). Uses Layer 2 (Data Link) of the OSI model. Simple to set up but has outdated encryption methods, making it less secure.
- L2TP (Layer 2 Tunneling Protocol). Often combined with IPsec for encryption. It uses both Layer 2 and Layer 3 (Data Link and Network) and offers better security than PPTP due to IPsec protection.
- OpenVPN. Uses SSL/TLS for encryption, providing a high level of security. It works at the Transport layer (Layer 4) and can use various ports, making it flexible and harder to block.
- WireGuard. A modern protocol focused on performance and security, using advanced encryption. It operates at the Network layer (Layer 3) and has lower latency and higher speed than traditional VPNs.

Tunneling protocols secure network connections by encrypting traffic. PPTP is fast and simple but lacks modern security. L2TP with IPsec offers stronger encryption but can be slower due to double encapsulation. OpenVPN is highly secure, flexible, and excellent for bypassing restrictions but requires technical setup. WireGuard is fast, lightweight, and secure, using modern cryptography, though it lacks some advanced features.

Sources and Justifications. I rely on information that Paul Bischoff has written regarding different types of VPN protocols [4]. Also Aleksandar Kochovski's article, which assessed and approved by Aleksander Hougen and Simona Ivanovski was taken in sight [5, 6], results in table 1.

Table 1 – Comparison Table: Tunneling Protocols

Protocol	OSI Layer	Encryption	Security level	Performance
PPTP	Data Link (2)	MPPE (128-bit)	Low	High
L2TP/IPsec	Data Link (2) & Network (3)	IPsec (256-bit)	High	Moderate
OpenVPN	Transport (4)	SSL/TLS (AES-256)	Very High	Moderate to High
WireGuard	Network (3)	ChaCha20, Poly1305	Very High	Very High

1. PPTP (Point-to-Point Tunneling Protocol):

- security Level: PPTP utilizes the Microsoft Point-to-Point Encryption (MPPE) protocol with RC4 encryption, which is considered weak by modern standards. It has known vulnerabilities that can be exploited by attackers;
- performance: Due to its minimal encryption overhead, PPTP offers high performance and faster speeds.

2. L2TP/IPsec (Layer 2 Tunneling Protocol with IPsec):

- security Level: When combined with IPsec, L2TP provides robust security features, including strong encryption and authentication mechanisms;
- performance: The double encapsulation process can introduce additional overhead, leading to moderate performance.

3. OpenVPN:

- security Level: OpenVPN employs SSL/TLS protocols with support for various encryption standards, including AES-256, offering a high level of security;
- performance: Performance can vary based on configuration but generally provides a good balance between speed and security.

4. WireGuard:

- security Level: WireGuard uses modern cryptographic primitives like ChaCha20 for encryption and Poly1305 for data authentication, providing a high level of security;
- performance: Designed for efficiency, WireGuard offers high performance with low latency and high throughput.

Performance Evaluation Metrics:

Evaluating VPN performance involves measuring:

- latency (ms): the time taken for data to travel from source to destination;
- throughput (Mbps): the rate of successful data transfer over a network;
- CPU utilization (%): the amount of processing power required to handle VPN operations;
- packet loss (%): the percentage of data packets that are lost during transmission.

These metrics can be assessed using network performance tools and monitoring systems to determine the efficiency and impact of each VPN protocol on system resources.

Further it is very important to look through all benefits of combining FEC codes and VPN. It can be done in various combinations. For article purposes, it would be better to look through a few examples, bypassing some

connections, because of the amount of data. Benefits from combining FEC and VPN shall be balanced and reliable, in terms of practical usability.

Expected Results and Analysis both shown in table 2. The study established the following advantages provided by the use of FEC with various types of VPN protocols [7]. By combining strong sides from both FECs and VPNs with decreasing of influence of weaknesses it becomes possible to create stable and effective combinations.

Table 2 – Detailed Analysis of VPN and FEC Code Combinations

VPN Type	OSI Layer	Characteristics	Suggested FEC Code	Benefits of Combination
IPsec VPN	Network (3)	Encryption and packet integrity checks	LDPC, Reed-Solomon	Enhanced reliability and error correction for high-latency networks.
SSL VPN	Transport (4)	Protection via SSL/TLS, browser-based access	Convolutional, Reed-Solomon	Real-time error correction, improved stability in web applications.
L2TP/IPsec	Data Link (2) & Network (3)	Tunneling via L2TP with IPsec encryption	Turbo Codes, LDPC	Balanced low latency and high accuracy for remote access scenarios.
OpenVPN	Transport (4)	Open-source, supports multiple encryption protocols	Turbo Codes, Convolutional	Efficient handling of VoIP and streaming data with minimal retransmissions.

Due to table 3 low-density parity-check codes help recover lost data, reducing the need for retransmission and improving throughput, especially beneficial in high-latency environments like satellite networks. LDPC codes can be compared with other powerful coding schemes, e.g. turbo codes. From one side, bit error rate performance of turbo codes is influenced by low codes limitations. However, LDPC codes have any limitations of minimum distance that indirectly states that LDPC codes are more effective on large code rates. It is needed to highlight that LDPC codes, as well as turbo codes, are affected by error floor phenomenon and both have error floor region.

Table 3 – IPsec VPN and LDPC Codes

Metric	Without FEC	With LDPC Codes	Expected Improvement
Packet Loss (%)	20–30	10–15	Reduced by 40–50%
Average Latency (ms)	200	180	Reduced by 10%
Throughput (%)	80	90	Increased by 12%
Retransmissions	High	Low	Reduced retransmissions

Due to table 4 Reed-Solomon codes are highly effective in handling burst errors, making SSL VPN more reliable, especially for web access in mobile and satellite networks.

Table 4 – SSL VPN and Reed-Solomon Codes

Metric	Without FEC	With Reed-Solomon Codes	Expected Improvement
Packet Loss (%)	25–35	15–20	Reduced by 30–40%
Connection Establishment (s)	1.2	1	Reduced by 15%
Session Stability	Low	High	Increased stability
Load Errors (%)	High	Low	Reduced by 50%

Due to table 5 turbo codes use iterative decoding for high accuracy, effectively reducing packet loss and latency, improving the quality of video streams and remote access.

Table 5 – L2TP/IPsec VPN and Turbo Codes

Metric	Without FEC	With Turbo Codes	Expected Improvement
Packet Loss (%)	18–25	10–12	Reduced by 50%
Average Latency (ms)	180	160	Reduced by 11%
Video Quality	Medium	High	Improved by 20%
Throughput (%)	85	95	Increased by 12%

Due to table 6 convolutional codes correct real-time errors, ensuring stable and uninterrupted transmission for VoIP and streaming video, which is crucial for OpenVPN.

Table 6 – OpenVPN and Convolutional Codes

Metric	Without FEC	With Convolutional Codes	Expected Improvement
Packet Loss (%)	15–20	8–10	Reduced by 45%
VoIP Latency (ms)	150	130	Reduced by 13%
Audio Quality	Medium	High	Improved by 25%
Stream Stability	Low	High	Increased stability

Due to table 7 integrating FEC codes with different VPN types significantly improves the reliability and quality of data transmission, particularly in challenging network conditions such as mobile and satellite channels. The combined approach offers robust solutions for secure and stable communications, reducing packet loss, increasing throughput, and minimizing latency.

Concept. We have previously investigated the possibilities of using cascaded error-correcting codes and their impact on the triad CIA [8].

The use FEC has wide applicability, but they have long been a well-studied area of data transmission theory [9].

Table 7 – Summary Table: Advantages of Using FEC Codes with Various VPNs

VPN Type	FEC Code	Improved Metric	Expected Improvement
IPsec VPN	LDPC	Packet Loss, Throughput	-40% loss, +12% throughput
SSL VPN	Reed-Solomon	Session Stability, Latency	+30% stability, -15% latency
L2TP/IPsec	Turbo	Video Quality, Packet Loss	+20% quality, -50% loss
OpenVPN	Convolutional	VoIP Quality, Real-Time Errors	+25% VoIP quality, -45% errors

Using FEC and VPN in one algorithm aligns with the CIA triad principles (Confidentiality, Integrity, and Availability) and can be implemented effectively for secure data transmission, in more detail further:

1. Confidentiality:

- encryption at both the first and second stages ensures that data is unreadable to unauthorized entities. This aligns with the need to protect sensitive information during transmission;
- interference-resistant coding adds an extra layer of security by obfuscating the data structure, making it harder to breach.

2. Integrity:

- dual-stage coding ensures data is not tampered with during transmission. If any part of the data stream is altered, the decoding process would fail, signaling potential interference;
- cryptographic hash functions can be integrated into the algorithm to verify the integrity of transmitted data packets.

3. Availability:

- the robust structure of the VPN tunnel and error-resilient coding minimizes disruptions caused by noise or attacks, ensuring uninterrupted data access.
- implementing redundancy mechanisms within the coding process can further enhance data availability [10].

Implementation Steps:

1. Coding and Encryption:

- use state-of-the-art encryption algorithms such as AES-256 or ChaCha20, coupled with error-resilient codes like Reed-Solomon or LDPC (Low-Density Parity-Check);
- incorporate protocols like WireGuard or OpenVPN for modern, high-performance tunneling.

2. Interference Resistance:

- embed interference-resistant coding at both the first and second stages to mitigate the risk of errors caused by noisy channels or external interference.

3. Physical Transmission:

- utilize secure physical and virtual communication channels. Employ Secure Socket Layer (SSL)/Transport Layer Security (TLS) to safeguard data at the transport layer.

4. Error Handling and Decoding:

- implement mechanisms to identify and correct errors during the decoding process, ensuring data reliability.

5. Decryption:

- after verifying data integrity, decrypt the data at the recipient's end using secure cryptographic keys. Detailed algorithm is presented in fig.1 [11].

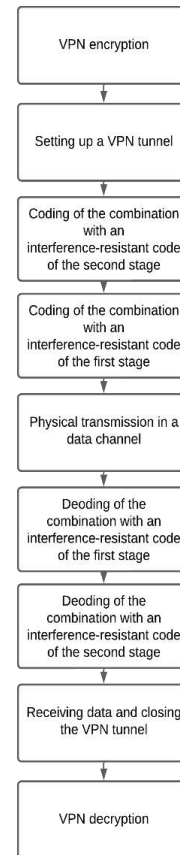


Fig. 1. Algorithm of using cascade codes with this model

As said, different types of ECC codes can be used, more on fig. 2 [12, 13].

Conclusions. We have already explored and analyzed several approaches to integrating Error Correction Codes (ECC) with VPN technology, uncovering their potential to enhance data transmission reliability and security. These studies have demonstrated how ECC, combined with Forward Error Correction (FEC) methods, can address critical challenges like packet loss, interference, and noise in modern communication systems. By embedding error-resilience directly into the transmission process, ECC and FEC reduce the dependency on retransmissions, ensuring smoother and faster communication over VPNs, even in adverse network conditions.

VPNs are essential for creating encrypted tunnels that protect data from eavesdropping and tampering. When combined with robust error-correction mechanisms, they provide an additional layer of reliability, ensuring that transmitted information remains accurate and undistorted. This is especially valuable in high-demand applications such as real-time video streaming, telemedicine, and remote work environments, where even minor transmission errors can disrupt critical processes.

Continued research in this domain is vital for the evolution of IT technologies. As networks become increasingly complex with the rise of 5G, IoT, and quantum

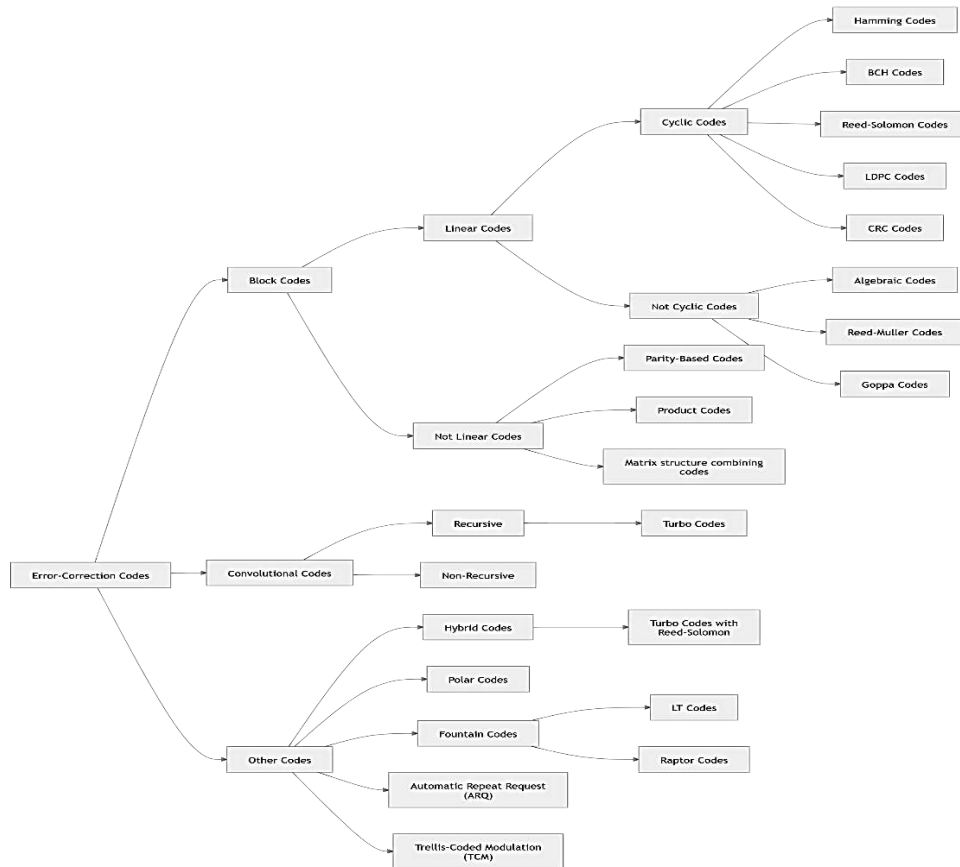


Fig. 2. ECC common types

computing, the demand for secure and efficient data transmission methods will grow exponentially. By studying ECC and its integration with VPNs, we can pave the way for innovative communication protocols that are not only faster and more reliable but also resilient against emerging cybersecurity threats. These advancements will lay the groundwork for the next generation of digital infrastructure, driving progress across industries and ensuring the seamless connectivity required in an interconnected world.

References

1. Stallings W. *Data and Computer Communications. 10th ed.* Pearson, 2013. 912 p.
2. Tanenbaum A. S., Wetherall D. J. *Computer Networks. 5th ed.* Pearson, 2010, 960 p.
3. Nieves M., Dempsey K., Pillitteri V. Y. *An Introduction to Information Security. National Institute of Standards and Technology, Special Publication 800 – 12 Revision 1*, June 2017. 101 p. URL: <https://nscarchive.gwu.edu/document/22632-document-07-michaelnieles-kelley-dempsey-and> (accessed 02.11.2024).
4. Bischoff P., *VPN protocols explained and compared.* Available at: <https://www.comparitech.com/vpn/protocols>. (accessed 02.11.2024).
5. Kochovski A., Hougen A., Ivanovski S. *A Full VPN Protocols List in 2024: Explained and Compared.* URL: <https://www.cloudwards.net/vpn-protocol-breakdown/> (accessed 02.11.2024).
6. Kochovski A., Hougen A., Ivanovski S. *PPTP vs OpenVPN: Differences, Advantages & Disadvantages in 2024.* URL: <https://www.cloudwards.net/pptp-vs-openvpn/> (accessed 02.11.2024).
7. Vincent R., Belkacem T., Christophe B., Tuan T. Cédric T. Less latency and better protection with AL-FEC sliding window codes: A robust multimedia CBR broadcast case study. *IEEE 13th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, October 2017, Italy. 9 p.
8. Шаров В. О., Нікуліна О. М. Дворівнева концепція для моделювання єдиної заводостійкої передачі цифрових даних. *Вісник Національного технічного університету «Харківський політехнічний інститут»*: зб. наук. пр. Темат. вип.: Системний аналіз, управління та інформаційні технології. Харків: НТУ «ХПІ», 2024. № 1 (11). С. 70–75.
9. Шаров В. О., Нікуліна О. М., Северин В. П. Моделювання та аналіз кодерів заводостійких каскадних кодів для динамічних систем. *Вісник Національного технічного університету «Харківський політехнічний інститут»*: зб. наук. пр. Темат. вип.: Системний аналіз, управління та інформаційні технології. Харків: НТУ «ХПІ», 2023. № 1 (9). С. 60–69.
10. Stallings W., Brown L. *Computer Security: Principles and Practice.* New York, Prentice Hall, 2008. 817 p.
11. Pfleeger C. P., Pfleeger S. L. *Security in Computing.* New Jersey, Prentice Hall, 2003. 746 p.
12. Tiller J. S. *The Ethical Hack: A Framework for Business Value Penetration Testing.* Auerbach Publications, 2005. 352 p.
13. Банкет В. Л., Іващенко В. Л., Іващенко М. О. *Заводостійке кодування в телекомунікаційних системах.* Одеса, ОНАЗ ім. О. С. Попова, 2011. 100 с.

References (transliterated)

1. Stallings W. *Data and Computer Communications. 10th ed.* Pearson, 2013. 912 p.
2. Tanenbaum A. S., Wetherall D. J. *Computer Networks. 5th ed.* Pearson, 2010. 960 p.
3. Nieves M., Dempsey K., Pillitteri V. Y. *An Introduction to Information Security. National Institute of Standards and Technology, Special Publication 800 – 12 Revision 1*, June 2017. 101 p. URL: <https://nscarchive.gwu.edu/document/22632-document-07-michaelnieles-kelley-dempsey-and> (accessed 02.11.2024).

4. Bischoff P., VPN protocols explained and compared. Available at: <https://www.comparitech.com/vpn/protocols>. (accessed 02.11.2024).
5. Kochovski A., Hougen A., Ivanovski S. *A Full VPN Protocols List in 2024: Explained and Compared*. Available at: <https://www.cloudwards.net/vpn-protocol-breakdown/>. (accessed 02.11.2024).
6. Kochovski A., Hougen A., Ivanovski S. *PPTP vs OpenVPN: Differences, Advantages & Disadvantages in 2024*. Available at: <https://www.cloudwards.net/pptp-vs-openvpn/> (accessed 02.11.2024).
7. Vincent R., Belkacem T., Christophe B., Tuan T. Cédric T. Less latency and better protection with AL-FEC sliding window codes: A robust multimedia CBR broadcast case study. *IEEE 13th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, October 2017, Italy. 9 p.
8. Sharov V. O., Nikulina O. M., Dvorivneva kontsepsiya dlya modelyuvannya yedynoyi zavadostiykoyi peredachi tsyfrovykh danykh [A two-level concept for simulating a single interference-resistant digital data transmission] *Bulletin of the National Technical University "KhPI". Series: System analysis, management and information technologies*. Kharkiv, NTU "KhPI" Publ., 2024, no. 1 (11), pp. 70–75. (In Ukr.).
9. Sharov V. O., Nikulina O. M., Severyn V. P. Modelyuvannya ta analiz koderiv zavadostiykykh kaskadnykh kodiv dlya dynamichnykh system [Modeling and analysis of encoders of interference-tolerant cascade codes for dynamic systems]. *Bulletin of the National Technical University "KhPI". Series: System analysis, management and information technologies*. Kharkiv, NTU "KhPI" Publ., 2023, no. 1 (9), pp. 60–69. (In Ukr.).
10. Stallings W., Brown L. *Computer Security: Principles and Practice*. New York, Prentice Hall, 2008. 817 p..
11. Pfleeger C. P., Pfleeger S. L. *Security in Computing*. New Jersey, Prentice Hall, 2003. 746 p.
12. Tiller J. S. *The Ethical Hack: A Framework for Business Value Penetration Testing*. Auerbach Publications, 2005. 352 p.
13. Banket V. L., Ivashchenko P. V., Ishchenko M. O. *Zavadostijke koduvannya v telekomunatsijnyh systemah* [Interference-resistant coding in telecommunication systems]. Odesa, ONAZ named O. S. Popova Publ., 2011. 100 p. (In Ukr.).

Received 05.11.2024

УДК 004.9

В. О. ШАРОВ, аспірант кафедри інформаційних систем та технологій Національного технічного університету «Харківський політехнічний інститут», Харків, Україна; e-mail: wycptiy@gmail.com; ORCID: <https://orcid.org/0000-0003-3152-0650>

О. М. НИКУЛИНА, 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>

ДОСЛІДЖЕННЯ СУМІСНОСТІ МЕТОДІВ І ТЕХНОЛОГІЙ ПРОТОКОЛІВ ВИСОКОГО РІВНЯ ТА КОДІВ ВИПРАВЛЕННЯ ПОМИЛОК

Починаючи з 2000 року, галузі кодів для виправлення помилок і віртуальних приватних мереж (VPN) зазнали значних успіхів, зумовлених технологічними вимогами до вищої надійності та безпеки систем зв'язку. У кодах для виправлення помилок розробка турбо-кодів і кодів перевірки парності з низькою щільністю (LDPC) досягла нових висот, коли коди LDPC були прийняті в таких стандартах, як 5G і Wi-Fi 6, через їхню продуктивність, близьку до меж Шеннона. У цей період були внесені новаторські внески таких дослідників, як Девід Маккей і Редфорд Ніл, які вдосконалили алгоритми LDPC, і Ердал Арикан, який представив полярні коди в 2008 році. Відтоді полярні коди були інтегровані в системи 5G завдяки їх ефективності та низькій складності, що стало важливою віхою в сучасній теорії кодування. Досягнення в методах декодування, таких як поширення переконань і послідовне скасування, ще більше підвищили корисність цих кодів у практичних застосуваннях. Паралельно з цими розробками технологія VPN розвивалася у відповідь на зростаючу потребу в безпечному та приватному спілкуванні у все більш взаємопов'язаному світі. Удосконалені протоколи шифрування, такі як IPsec і OpenVPN, отримали широке поширення, підтримані інноваціями в криптографії. Такі дослідники, як Хьюго Кравчик, зробили внесок у створення надійних механізмів автентифікації, таких як протоколи HMAS і IKEv2, які забезпечують цілісність і конфіденційність тунелів VPN. Тим часом розробка WireGuard у середині 2010-х років, яку очолив Джейсон А. Доненфельд, представила легкий і високозахисний протокол VPN, який революціонував роботу сучасних VPN. Ці досягнення спрямовані на вирішення ескалації кіберзагроз і сприяють безпечному обміну даними в глобальних мережах. Важливість вивчення кодів виправлення помилок і VPN у сучасну епоху неможливо переоцінити. Коди виправлення помилок є невід'ємною частиною подолання проблем середовищ із високим рівнем шуму, забезпечуючи надійний зв'язок у різних технологіях, від дослідження космосу до масивних мереж Інтернету речей. Водночас VPN залишаються критично важливими для збереження конфіденційності користувачів, безпеки корпоративних мереж і захисту конфіденційних даних перед обличчям складних кібератак. Нові технології, такі як квантові обчислення та штучний інтелект, створюють як можливості, так і загрози, що вимагає постійних інновацій у цих сферах. Вивчення квантових кодів корекції помилок і постквантових криптографічних протоколів є життєво важливою областю для майбутніх досліджень. Вирішуючи ці виклики, вчені та інженери можуть забезпечити стійкість і безпеку комунікаційних систем у все більш цифровому та взаємопов'язаному світі.

Ключові слова: VPN, FEC, ECC, триада CIA, єдина модель безпеки, каскадні коди, канал передачі даних.

Повні імена авторів / Author's full names

Автор 1 / Author 1: Шаров Владислав Олегович / Sharov Vladyslav Olegovich

Автор 2 / Author 2: Нікуліна Олена Миколаївна / Nikulina Olena Mykolaivna

DOI: 10.20998/2079-0023.2024.02.15
УДК 004.9

Н. Є. ХАЦЬКО, кандидат технічних, доцент, доцент кафедри інформаційних систем та технологій Національного технічного університету «Харківський політехнічний інститут», Харків, Україна; e-mail: natalia.khatsko@khpі.edu.ua; ORCID: <https://orcid.org/0000-0002-2543-0280>

М. В. СЛЕПУШКОВ, студент магістратури, Національний технічний університет «Харківський політехнічний інститут», м. Харків, Україна; e-mail: m.sliepushkov@gmail.com; ORCID: <https://orcid.org/0009-0001-0004-2820>

К. О. ХАЦЬКО, старший викладач кафедри інформаційних систем та технологій Національного технічного університету «Харківський політехнічний інститут», аспірант, Харків, Україна; e-mail: kurylo.khatsko@khpі.edu.ua; ORCID: <https://orcid.org/0000-0003-3315-1553>

Є. О. ШЕБАНОВ, аспірант, Національний технічний університет «Харківський політехнічний інститут», Харків, Україна; e-mail: yevhenii.shebanov@cs.khpі.edu.ua; ORCID: <https://orcid.org/0009-0006-9032-8764>

МОДИФІКОВАНИЙ АЛГОРИТМ РОЗГОРТАННЯ ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ З ВИКОРИСТАННЯМ БАГАТОПОТОЧНОСТІ

У статті представлено модифікований алгоритм розгортання програмного забезпечення з використанням багатопоточності в AWS CodeBuild, спрямований на оптимізацію часу збірки та зниження витрат на обчислювальні ресурси в хмарному середовищі. На основі скінченних автоматів, часових автоматів та мережі Петрі було змодельовано основні етапи процесу збірки, включаючи паралельне виконання тестів, аналіз розподілу завдань та управління обчислювальними ресурсами. Особлива увага приділялася ідентифікації та усуненню обмежень стандартних механізмів паралелізації AWS CodeBuild, які можуть призводити до неефективного використання ресурсів та збільшення тривалості збірки. Дослідження виявило, що стандартні механізми AWS CodeBuild не завжди здатні оптимально використовувати системні ресурси, особливо при обробці великих програмних проєктів із численними залежностями. Для подолання цих обмежень було запропоновано використання багатопоточності засобами Python, як зручного інструменту для розширення базового функціоналу. Запропонований підхід дозволив реалізувати гнучке керування потоками та розподіл завдань на рівні користувачьких сценаріїв, зменшивши загальний час збірки. Експериментальні результати показали значне скорочення часу виконання збірок у порівнянні зі стандартними налаштуваннями AWS CodeBuild. Це підтверджує ефективність використання запропонованого алгоритму для збільшення продуктивності та забезпечення високої масштабованості процесів збірки в хмарному середовищі. Розроблений алгоритм є особливо актуальним для великих програмних проєктів, які вимагають частих ітераційних збірок та тестування. Отримані результати можуть бути використані для вдосконалення автоматизованих процесів розгортання та управління обчислювальними ресурсами у хмарних середовищах.

Ключові слова: процес розгортання, інформаційна технологія, алгоритм, модель процесу розгортання програмного забезпечення, скінченний автомат, часовий автомат, мережі Петрі.

Вступ. У сучасному програмному середовищі, особливо у хмарних інфраструктурах, таких як AWS CodeBuild, ефективність збирання програмного забезпечення є важливим фактором, який впливає як на швидкість розгортання, так і на вартість використання ресурсів. Традиційний послідовний підхід до виконання збірки та тестування має обмежену продуктивність через необхідність послідовного виконання окремих етапів, що може спричинити тривалі затримки, особливо при роботі з великими проєктами. Оскільки CodeBuild оплачується відповідно до часу використання, оптимізація цього часу є першочерговою для зменшення витрат. Одним із способів вирішення цієї проблеми є впровадження багатопоточності, що дозволяє паралельно виконувати незалежні етапи. Водночас, недостатнє теоретичне обґрунтування механізмів паралельного виконання ускладнює оптимальне впровадження таких підходів. Необхідно розробити модель, яка враховує специфіку процесів збирання та дозволяє визначити оптимальні умови для використання паралельних потоків. Для постановки цієї проблеми доцільно застосувати формалізми такі як: скінченні автомати та мережі Петрі, що дозволяє описати процес збірки та змодельовати використання багатопоточності.

Мета дослідження полягає в розробці модифікованого алгоритму розгортання, що допоможе зменшити вартість використання хмарних ресурсів. План дослі-

дження складається з наступних етапів: аналіз традиційного алгоритму розгортання, модифікація алгоритму, приведення експерименту в якому вимірюються характеристики обох процесів розгортання, порівняння характеристик базового алгоритму розгортання та модифікованого.

У рамках дослідження було проаналізовано низку наукових робіт, що розкривають сучасні підходи до використання теорії автоматів у різних аспектах розробки програмних систем та їх оптимізації. У роботі [1] розглянуто застосування псевдопам'яті скінченних автоматів для створення криптографічних систем. Псевдопам'ять забезпечує швидкодію та гнучкість у шифруванні даних, що є перспективним для безпечного управління хмарними системами. Дослідження [2] пропонує ефективні методи індексації та компресії скінченних автоматів, які можуть бути використані для оптимізації зберігання та пошуку даних у великих хмарних середовищах, що підвищує ефективність обробки регулярних запитів. Інструмент ComVis, представлений у [3], є прикладом візуалізації автоматів, орієнтованої на освітні цілі, але підхід до інтерактивного моделювання може бути адаптований для управління складними процесами розгортання в хмарі [4]. У [5] запропоновано концепцію паралельних скінченних автоматів (parallel finite automata – PFAs), які підтримують моделювання багатопоточних систем із точним

© Хацько Н. Є., Слепушков М. В., Хацько К. О., Шебанов Є. О., 2024



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.



контролем станів. Це є актуальним для оптимізації паралельних потоків у процесі зборки програмного забезпечення. Автори демонструють, як PFAs допомагають уникати нескінченних просторів станів, притаманних деяким мережам Петрі, зберігаючи при цьому можливість моделювати паралельні процеси. У роботі показано, що PFAs, хоч і еквівалентні детермінованим скінченним автоматам за мовою, але структурно простіші. Таким чином PFAs кращі для моделювання багатопоточності.

Дослідження [6] вводить новий клас мереж Петрі – Gadaga, які використовуються для уникнення взаємоблокувань у багатопотокових середовищах, що є особливо важливим для забезпечення стабільності хмарних систем. Робота [7] демонструє підходи до перевірки моделей багатопотокових програм за допомогою асинхронних методів, які забезпечують стабільну роботу паралельних процесів і знижують ризик збоїв у системах. Інструмент Ballet, описаний у [8], автоматизує координацію децентралізованої переконфігурації в хмарних середовищах, дозволяючи значно знизити залежність від централізованого управління та збільшити швидкість і надійність процесів. У [9] описано ефективність використання фреймворка для обробки автоматів у хмарних середовищах, що забезпечує високу пропускну здатність і гнучкість конфігурації для складних систем. Особливо актуально це для розгортання додатків, які мають мікросервісну архітектуру [10]. Дослідження [11] підкреслює можливість застосування кінцевих автоматів для аналізу питань довіри у хмарних сервісах, що дозволяє структурувати управління станами та забезпечувати безпеку даних. У роботі [12] представлено інструмент для симуляції та тестування детермінованих скінченних автоматів, що дозволяє спростити відлагодження і перевірку програм, побудованих на основі автоматів. Такий інструмент міг бути корисним для дослідження, що проводиться, але не застосований через відсутність підтримки симуляції багатопоточності.

В статті [13] вперше була описана концепція часових автоматів – це різновид скінченних автоматів, який включає додаткові змінні, відомі як годинники, що дозволяють враховувати час між переходами між станами. На відміну від традиційних скінченних автоматів, де перехід відбувається миттєво при спостереженні певних подій або умов, часові автомати дозволяють моделювати системи, в яких переходи залежать від проміжку часу, що минув з моменту останнього переходу або події. Це робить їх особливо корисними для опису і верифікації систем з реальним часом, таких як вбудовані системи, програмне забезпечення з часовими обмеженнями тощо. Система розгортання, що досліджується, має часові обмеження, тому часові автомати доцільно використати.

Таким чином, в дослідженні скористаємось концепцією часових скінченних автоматів та мережами Петрі для уникнення взаємоблокувань у багатопотокових середовищах, що відповідає нашій меті теоретичного моделювання паралельних потоків на одному з етапів збірки.

Огляд наукових статей та результати попередніх досліджень різних авторів підтверджує актуальність використання теорії автоматів для розв'язання задач моніторингу та оптимізації процесів розгортання у хмарному середовищі є актуальним.

Методи дослідження. Теорія автоматів представляє собою математичний апарат, який дозволяє описувати та аналізувати поведінку дискретних систем. У контексті розгортання програмних систем прийнятні скінченні автомати та їх розширення. Часовий автомат – це розширення класичного скінченного автомата, яке дозволяє враховувати часові обмеження для переходів між станами або перебування в певних станах. Він широко використовується для моделювання систем реального часу, де час є важливим фактором (наприклад, керування транспортними системами, аналіз протоколів зв'язку, хмарні обчислення). Часовий автомат розширює класичний автомат, додаючи додаткові поняття. Деякі з них:

- годинники – набір змінних, які завжди збільшуються зі швидкістю 1 (реальне значення часу). Їх можна обнуляти при виконанні певних дій;
- стан – описує поточний стан системи з додаванням часових умов;
- переходи – вказують реакцію автомату на впливи, мають часові умови (гвардії), які визначають, коли перехід дозволено;
- часові обмеження – умови, накладені на годинники. Окрім гвардій, які повинні виконуватися для здійснення переходу, існують інваріанти, що визначають час знаходження системи в певному стані.

Використання часових автоматів при формалізації процесу збирання програмного забезпечення за допомогою AWS CodeBuild є абсолютно доречним та корисним, оскільки вони дозволяють точно моделювати часові обмеження на кожному етапі збірки, враховуючи послідовність виконання завдань і можливість паралельного виконання. Завдяки цим моделям можна формалізувати часові характеристики етапів, таких як завантаження вихідного коду, встановлення залежностей, виконання збірки та тестів.

AWS CodeBuild має чітко визначені етапи збірки, які в автоматі представлені як стани. Час кожного етапу вимірюється, крім того ми маємо обмеження на загальний час збірки. В разі перевищення цього обмеження збірка буде завершена.

Опишемо етапи збірки (стани автомату).

1. SUBMITTED – задачу збірки створено, і вона очікує на запуск у черзі.
2. QUEUED – задача перебуває в черзі на виконання.
3. PROVISIONING – CodeBuild налаштовує середовище збірки, включаючи створення обчислювальних ресурсів.
4. DOWNLOAD_SOURCE – завантаження вихідного коду з визначеного джерела.
5. INSTALL – встановлення необхідних залежностей та інструментів, визначених у файлі конфігурації збірки (buildspec.yml).

6. PRE_BUILD – підготовка до основного процесу збірки, наприклад, виконання скриптів ініціалізації чи підготовчих дій.

7. BUILD – основний етап збірки, на якому виконуються всі завдання, необхідні для створення продукту (компіляція, збирання, тестування, генерація результатів).

8. POST_BUILD – завершальні дії, упаковка артефактів, створення Docker-образів тощо.

9. UPLOAD_ARTIFACTS – завантаження результатів збірки у місце зберігання, наприклад, Amazon S3, Docker репозиторій.

10. ERROR – процес завершився з помилкою; збірку не вдалося виконати.

11. FINALIZING – завершення процесу, очищення середовища збірки, підготовка звітів.

12. COMPLETED – завдання завершено.

Функція переходів у теорії часових автоматів визначає, як і за яких умов відбувається перехід між станами. В контексті нашого завдання вона може бути записана як:

$$\delta : L \times B(C) \rightarrow 2^L,$$

де L – множина станів, що описані вище; $B(C)$ – булеві вирази (гвардії), що задають умови переходу на основі відліку часу; 2^L – множина підмножин станів множини L , куди можливий перехід.

Визначимо функцію переходів δ .

1. Початковий стан:

$$\delta(\text{SUBMITTED}, x \leq t_1) = \{\text{QUEUED}\}.$$

2. Перехід між етапами: Для кожного стану визначається наступний стан та часова гвардія $x \leq t_i$, яка гарантує, що перехід відбувається лише в допустимий проміжок часу:

$$\delta(\text{SUBMITTED}, x \leq t_2) = \{\text{PROVISIONING}\},$$

$$\delta(\text{PROVISIONING}, x \leq t_3) = \{\text{DOWNLOAD_SOURCE}\},$$

$$\delta(\text{DOWNLOAD_SOURCE}, x \leq t_4) = \{\text{INSTALL}\}.$$

І так далі, до завершального етапу.

3. Перехід у стан ERROR – якщо годинник перевищує допустимий час t_i , автомат переходить до стану ERROR:

$$\delta(l, x > t_i) = \{\text{ERROR}\},$$

де l – один з перших дев'яти станів, що описані вище та на рис. 1.

4. Завершення роботи – зі стану FINALIZING автомат завжди переходить до COMPLETED:

$$\delta(\text{FINALIZING}, \text{true}) = \{\text{COMPLETED}\}.$$

Таким чином, функцію переходів визначаємо як:

$$\delta(l, g) = \begin{cases} \{l'\}, & \text{якщо } g = (x \leq t_i), \\ \{\text{ERROR}\}, & \text{якщо } g = (x > t_i), \\ \{\text{COMPLETED}\}, & \text{якщо } g = \text{true}, \end{cases} \quad (1)$$

де l' – наступний стан; $g = (x \leq t_i)$ – гвардія для кожного переходу, визначає максимальний час t_i , допустимий для поточного стану; ERROR – стан, що позначає виникнення помилки, якщо час у стані перевищує допустимий t_i .

Функція переходів δ дозволяє чітко описати логіку переходів у часовому автоматі збірки програмного забезпечення (табл. 1).

Графічно автомат виглядає як орієнтований граф, де кожен стан представлений вузлом, а переходи між станами – стрілками (див. рис. 1).

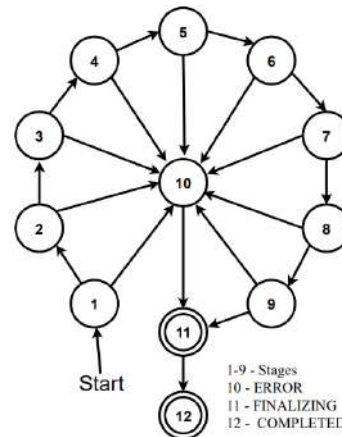


Рис. 1. Скінченний автомат збірки програмного забезпечення в AWS CodeBuild

Таблиця 1 – Таблиця визначення функції переходів

Номер	Поточний стан (l)	Наступний стан (l')	Гвардія (g)	Дія (Σ)
1	SUBMITTED	QUEUED	$x \leq t_1$	Завдання додано до черги.
2	QUEUED	PROVISIONING	$x \leq t_2$	Початок налаштування середовища.
3	PROVISIONING	DOWNLOAD_SOURCE	$x \leq t_3$	Завантаження джерел коду.
4	DOWNLOAD_SOURCE	INSTALL	$x \leq t_4$	Інсталяція залежностей.
5	INSTALL	PRE_BUILD	$x \leq t_5$	Підготовчі дії перед збіркою.
6	PRE_BUILD	BUILD	$x \leq t_6$	Початок основного процесу збірки.
7	BUILD	POST_BUILD	$x \leq t_7$	Завершення збірки.
8	POST_BUILD	UPLOAD_ARTIFACTS	$x \leq t_8$	Завантаження артефактів у сховище.
9	UPLOAD_ARTIFACTS	FINALIZING	$x \leq t_9$	Завершення роботи.
11	FINALIZING	COMPLETED	$x \leq t_{10}$	Завдання завершено успішно.
	ANY_STATE	ERROR	$x > t_{\max}$	Перехід до обробки помилки.
10	ERROR	FINALIZING	Завжди	Очищення та завершення роботи.

У процесі збирання програмного забезпечення кожен етап зазвичай слідує за попереднім, але існують завдання, які можуть виконуватися незалежно та паралельно. Це дозволяє ефективніше використовувати доступні ресурси і скорочувати загальний час збірки. Наприклад, завантаження залежностей і виконання юніт-тестів є завданнями, які часто не залежать один від одного і можуть виконуватися одночасно. Так, всередині 7-го стану (BUILD), як правило, крім збірки виконуються ще й юніт-тести. Для моделювання їх паралельного виконання ми застосуємо мережі Петрі, які добре підходять для формалізації систем, що включають одночасні, залежні та незалежні процеси.

При побудові моделі тестування за допомогою мереж Петрі кожен тест можна уявити як окремих процес, представлений місцем у мережі. Переходи між цими місцями позначають виконання тестів, а загальний результат тестування стає доступним лише після завершення всіх окремих процесів. Це дозволяє побудувати схему, яка відображає як паралельне виконання тестів, так і їх об'єднання в єдиний завершений результат.

Мережу Петрі описують як кортеж:

$$PN = (P, T, f, M),$$

де $P = \{p_0, p_1, p_2, \dots, p_n, p_{done}\}$ – множина місць. p_0 відповідає початку процесу, p_1, p_2, \dots, p_n – місця для виконання окремих тестів, p_{done} – місце завершення тестування;

$T = \{t_1, t_2, \dots, t_n, t_{finish}\}$ – множина переходів, t_1, t_2, \dots, t_n відповідають запуску і завершенню кожного тесту, t_{finish} відповідає завершенню тестування після виконання всіх тестів;

F – множина дуг, яка задає відношення між місцями і переходами; з рис. 2 виходить, що

$$F = \{(p_0, t_1), (t_1, p_1), \dots, (p_0, t_n), (t_n, p_n), (p_1, t_{finish}), \dots, (p_n, t_{finish}), (t_{finish}, p_{done})\}.$$

M – множина важелів дуг. В контексті цієї моделі це стало значення, що не має впливу.

Паралельне виконання тестів змодельовано та представлено на рис. 2. Мережа Петрі моделює процес паралельного виконання тестування в системі збирання програмного забезпечення.

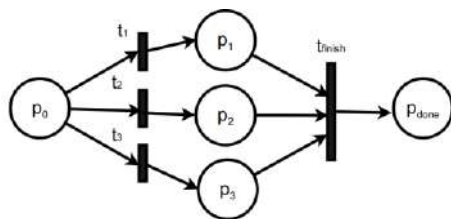


Рис. 2. Схема паралельних процесів з використанням Мережі Петрі

На рис. 2 модель містить описані нижче елементи.

1. Місця:

- p_0 – початковий стан, тести ще не запуснені;
- p_1, p_2, p_3 – місця (стани), які відповідають активним станам виконання окремих тестів;
- p_{done} – кінцевий стан, всі тести завершені.

2. Переходи:

- t_1, t_2, t_3 – переходи, які запускають виконання кожного з тестів;
- t_{finish} – перехід після завершення всіх тестів.

Логіка роботи мережі Петрі в контексті паралельного тестування наступна.

1. В стані системи p_0 всі тести завантажені.

2. Переходи t_1, t_2, t_3 активуються одночасно, переміщуючи виконання тестів до станів p_1, p_2, p_3 .

3. Після завершення тестів активується перехід t_{finish} , який переміщує тести до стану p_{done} , що сигналізує завершення тестування.

Формула часу для послідовного виконання тестів:

$$T_{посл} = T_{синх} + \sum_{i=1}^n T_i + T_{заверш},$$

де $T_{синх}$ – час синхронізації перед запуском тестування;

T – час виконання i -го тесту; $T_{заверш}$ – час завершального етапу тестування.

Формула для часу для паралельної моделі:

$$T_{пар} = t_{синх} + \max(t_1, t_2, \dots, t_n) + t_{заверш}, \quad (2)$$

де $t_{синх}$ – час синхронізації на початку тестування, що необхідний для розподілу задач між потоками; t_1, t_2, \dots, t_n – час виконання кожного з паралельних тестів; $t_{заверш}$ – час синхронізації системи після завершення тестів.

Паралельна природа процесу означає, що загальний час визначається найдовшим з усіх паралельних тестів.

Експеримент. Для перевірки теоретичних припущень було створено проєкт на AWS CodeBuild та написано програму, до якої застосовувалися тести.

Коректність моделі роботи системи (рис. 1, 2) була перевірена шляхом створення помилок на різних етапах збірки продукту. А саме на рис. 3 та рис. 4 зображені стани, що завершилися помилками, після яких збірка не продовжувалась, а процес переходив до стану FINALIZING та COMPLETED. Таким чином, зв'язки попередніх станів зі станом ERROR були підтвержені. На рис. 5–7 продемонстровано успішне виконання, коли пройдено всі етапи збірки без помилок. Кожен етап має свій час виконання, а також обмеження цього часу, в разі перевищення ліміту часу збірка перепивається. Що і враховано в (1), (2).

Окремого етапу для тестування в CodeBuild не передбачено тому тести були виконані на етапі BUILD.

Спочатку було виконано тестування в однопоточному режимі. Повторивши тест в циклі 100 разів, час етапу BUILD склав 211 секунд (рис. 5).

На наступному етапі було запуснено виконання ста тестів в паралельному режимі засобами AWS CodeBuild, і результати виявились неочікуваними. Час виконання в стані BUILD збільшився до 765 секунд. Для вирішення ситуації з CodeBuild одним з варіантів розв'язку проблеми є застосування алгоритмів штучного інтелекту [14], але швидший спосіб використати оптимізацію на рівні операційної системи засобами Python.

Name	Status	Context	Duration
SUBMITTED	✔ Succeeded	-	<1 sec
QUEUED	✔ Succeeded	-	<1 sec
PROVISIONING	✔ Succeeded	-	4 secs
DOWNLOAD_SOURCE	✘ Failed	CLIENT_ERROR: authorization failed for primary source	95 secs
FINALIZING	✔ Succeeded	-	<1 sec
COMPLETED	✔ Succeeded	-	-

Рис. 3. Помилка в стані DOWNLOAD_SOURCE

Name	Status	Context	Duration
SUBMITTED	✔ Succeeded	-	<1 sec
QUEUED	✔ Succeeded	-	<1 sec
PROVISIONING	✔ Succeeded	-	5 secs
DOWNLOAD_SOURCE	✔ Succeeded	-	4 secs
INSTALL	✔ Succeeded	-	<1 sec
PRE_BUILD	✘ Failed	COMMAND_EXECUTION_ERROR: Error while executing command: pip install -r requirements.txt. Reason: exit status 1	23 secs
FINALIZING	✔ Succeeded	-	<1 sec
COMPLETED	✔ Succeeded	-	-

Рис. 4. Помилка в стані PRE_BUILD

Name	Status	Context	Duration
SUBMITTED	✔ Succeeded	-	<1 sec
QUEUED	✔ Succeeded	-	<1 sec
PROVISIONING	✔ Succeeded	-	5 secs
DOWNLOAD_SOURCE	✔ Succeeded	-	139 secs
INSTALL	✔ Succeeded	-	1 sec
PRE_BUILD	✔ Succeeded	-	<1 sec
BUILD	✔ Succeeded	-	211 secs
POST_BUILD	✔ Succeeded	-	<1 sec
UPLOAD_ARTIFACTS	✔ Succeeded	-	<1 sec
FINALIZING	✔ Succeeded	-	<1 sec
COMPLETED	✔ Succeeded	-	-

Рис. 5. Послідовне виконання тестів

Name	Status	Context	Duration
SUBMITTED	✔ Succeeded	-	<1 sec
QUEUED	✔ Succeeded	-	<1 sec
PROVISIONING	✔ Succeeded	-	5 secs
DOWNLOAD_SOURCE	✔ Succeeded	-	163 secs
INSTALL	✔ Succeeded	-	3 secs
PRE_BUILD	✔ Succeeded	-	<1 sec
BUILD	✔ Succeeded	-	75 secs
POST_BUILD	✔ Succeeded	-	<1 sec
UPLOAD_ARTIFACTS	✔ Succeeded	-	<1 sec
FINALIZING	✔ Succeeded	-	<1 sec
COMPLETED	✔ Succeeded	-	-

Рис. 6. Паралельне виконання тестів засобами Python

Для реалізації було використано альтернативний спосіб запуску тестів. До процесу тестування додана

можливість запуску в багатопроесорному режимі і запущено той же тест 100 разів. Результат тестування виправдав очікування – час етапу Build зменшився до 73 секунд (рис. 6).

Висновки. У роботі змодельовано основні етапи процесу збірки, включаючи паралельне виконання тестів. Модель побудовано засобами скінченного часового автомату та мережі Петрі. Представлено модифікований алгоритм розгортання програмного забезпечення з використанням багатопоточності, що дозволяє оптимізувати процес збірки та тестування в середовищі AWS CodeBuild, та усунути обмеження стандартних механізмів паралелізації AWS CodeBuild.

Результати дослідження показали, що використання стандартних засобів паралельного тестування в AWS CodeBuild може призводити до збільшення часу виконання. Натомість впровадження модифікованого алгоритму з багатопоточністю дозволяє значно скоротити загальний час розгортання.

Розроблений алгоритм є особливо актуальним для великих програмних проектів, які вимагають частих ітераційних збірок та тестування. Отримані результати можуть бути використані для вдосконалення автоматизованих процесів розгортання та управління обчислювальними ресурсами у хмарних середовищах.

Список використаної літератури

- Shakhmetova G., Saukhanova Z., Udzir N. I., Sharipbay A., Saukhanov N. Application of Pseudo-Memory Finite Automata for Information Encryption. *IntelliSIS*. 2021. P. 330–339.
- Cotumaccio N., Prezza N. On indexing and compressing finite automata. *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms (SODA)*. Philadelphia: Society for Industrial and Applied Mathematics, 2021. P. 2585–2599.
- Jovanović N., Miljković D., Stamenković S., Jovanović Z., Chakraborty P. Teaching concepts related to finite automata using ComVis. *Computer Applications in Engineering Education*. 2021. Vol. 29, no. 5. P. 994–1006.
- Nikulina O. M., Khatsko K. O. Method of converting the monolithic architecture of a front-end application to microfrontends. *Вісник Національного технічного університету «ХПІ»*. Серія: Системний аналіз, управління та інформаційні технології. Харків: НТУ «ХПІ», 2023. № 2 (10). С. 79–84. DOI: doi.org/10.20998/2079-0023.2023.02.12.
- Stotts P. D., Pugh W. Parallel finite automata for modeling concurrent software systems. *Journal of Systems and Software*. 1994. Vol. 27, no. 1. P. 27–43.
- Liao H., Wang Y., Cho H. K., Stanley J., Kelly T., Lafortune S., et al. Concurrency bugs in multithreaded software: Modeling and analysis using Petri nets. *Discrete Event Dynamic Systems*. 2013. Vol. 23. P. 157–195.
- Sen K., Viswanathan M. Model checking multithreaded programs with asynchronous atomic methods. *Computer Aided Verification: 18th International Conference, Seattle, WA, USA, August 17–20, 2006*. Springer, 2006. P. 300–314.
- Philippe J., Omond A., Coullon H., Prud'Homme C., Raïs I. Fast Choreography of Cross-DevOps Reconfiguration with Ballet: A Multi-Site OpenStack Case Study. *International Conference on Software Analysis, Evolution and Reengineering*. 2024. P. 1–11. DOI: doi.org/10.1109/SANER60148.2024.00007.
- Bo C., Dang V., Xie T., Wadden J., Stan M., Skadron K. Automata processing in reconfigurable architectures: In-the-cloud deployment, cross-platform evaluation, and fast symbol-only reconfiguration. *ACM Transactions on Reconfigurable Technology and Systems*. 2019. Vol. 12, no. 2. P. 1–25.
- Zamkovyi M., Gavrylenko S., Khatsko K., Khatsko N. Algorithmic support for building a distributed iot system in a cloud service. *4th KhPI Week on Advanced Technology*, Kharkiv, Ukraine, 2023. P. 1–6. DOI: doi.org/10.1109/KhPIWeek61412.2023.1031299.

11. Zimba A., Hongsong C. Analyzing trust concerns in public clouds using finite state automata. *2nd International Conference on Cloud Computing and Internet of Things*. 2016. P. 25–29.
12. Vayadande K. B., Sheth P., Shelke A., Patil V., Shevate S., Sawakare C. Simulation and testing of deterministic finite automata machine. *International Journal of Computer Sciences and Engineering*. 2022. Vol. 10, no. 1. P. 13–17.
13. Alur R., Dill D. L. A theory of timed automata. *Theoretical Computer Science*. 1994. Vol. 126, no. 2. P. 183–235.
14. Gavrylenko S., Zozulia V., Khatsko N., Methods for improving the quality of classification on imbalanced data. *2023 IEEE 4th KhPI Week on Advanced Technology*, Kharkiv, Ukraine, 2023. P. 1–5. DOI: doi.org/10.1109/KhPIWeek61412.2023.10312879.
6. Liao H., Wang Y., Cho H. K., Stanley J., Kelly T., Lafortune S., et al. Concurrency bugs in multithreaded software: Modeling and analysis using Petri nets. *Discrete Event Dynamic Systems*. 2013, vol. 23, pp. 157–195.
7. Sen K., Viswanathan M. Model checking multithreaded programs with asynchronous atomic methods. *Computer Aided Verification: 18th International Conference, Seattle, WA, USA, August 17–20, 2006*. Springer, 2006, pp. 300–314.
8. Philippe J., Omond A., Coullon H., Prud'Homme C., Raïs I. Fast Choreography of Cross-DevOps Reconfiguration with Ballet: A Multi-Site OpenStack Case Study. *International Conference on Software Analysis, Evolution and Reengineering*. 2024, pp. 1–11. DOI: doi.org/10.1109/SANER60148.2024.00007.
9. Bo C., Dang V., Xie T., Wadden J., Stan M., Skadron K. Automata processing in reconfigurable architectures: In-the-cloud deployment, cross-platform evaluation, and fast symbol-only reconfiguration. *ACM Transactions on Reconfigurable Technology and Systems*. 2019, vol. 12, no. 2, pp. 1–25.
10. Zamkovyi M., Gavrylenko S., Khatsko K., Khatsko N., Algorithmic support for building a distributed iot system in a cloud service. *4th KhPI Week on Advanced Technology*, Kharkiv, Ukraine, 2023, pp. 1–6. DOI: doi.org/10.1109/KhPIWeek61412.2023.1031299

References (transliterated)

1. Shakhmetova G., Saukhanova Z., Udzir N. I., Sharipbay A., Saukhanov N. Application of Pseudo-Memory Finite Automata for Information Encryption. *IntelTISIS*. 2021, pp. 330–339.
2. Cotumaccio N., Prezza N. On indexing and compressing finite automata. *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms (SODA)*. Philadelphia: Society for Industrial and Applied Mathematics, 2021, pp. 2585–2599.
3. Jovanović N., Miljković D., Stamenković S., Jovanović Z., Chakraborty P. Teaching concepts related to finite automata using ComVis. *Computer Applications in Engineering Education*. 2021, vol. 29, no. 5, pp. 994–1006.
4. Nikulina O. M., Khatsko K.O. Method of converting the monolithic architecture of a front-end application to microfrontends. *Visnyk Natsional'noho tekhnichnoho universytetu "KhPI". Seriya: Systemnyy analiz, upravlinnya ta informatsiyni tekhnolohiyi*. [Bulletin of the National Technical University "KhPI". Series: System analysis, control and information technology]. Kharkov, NTU "KhPI" Publ., 2023, no. 2 (10), pp. 79–84. DOI: doi.org/10.20998/2079-0023.2023.02.12.
5. Stotts P. D., Pugh W. Parallel finite automata for modeling concurrent software systems. *Journal of Systems and Software*. 1994, vol. 27, no. 1, pp. 27–43.
11. Zimba A., Hongsong C. Analyzing trust concerns in public clouds using finite state automata. *2nd International Conference on Cloud Computing and Internet of Things*. 2016, October, pp. 25–29.
12. Vayadande K. B., Sheth P., Shelke A., Patil V., Shevate S., Sawakare C. Simulation and testing of deterministic finite automata machine. *International Journal of Computer Sciences and Engineering*. 2022, vol. 10, no. 1, pp. 13–17.
13. Alur R., Dill D. L. A theory of timed automata. *Theoretical Computer Science*. 1994, vol. 126, no. 2, pp. 183–235.
14. Gavrylenko S., Zozulia V., Khatsko N., Methods for improving the quality of classification on imbalanced data. *2023 IEEE 4th KhPI Week on Advanced Technology*, Kharkiv, Ukraine, 2023, pp. 1–5. DOI: doi.org/10.1109/KhPIWeek61412.2023.10312879.

Hadziuzna (received) 05.12.2024

UDC 004.9

N. Y. KHATSKO, Candidate of Technical Sciences, Associate Professor, Associate Professor of the Department of Information Systems and Technologies of the National Technical University "Kharkiv Polytechnic Institute", Kyiv, Ukraine; e-mail: nataliia.khatsko@kphi.edu.ua; ORCID: <https://orcid.org/0000-0002-2543-0280>

M. V. SLIEPUSHKOV, Master's Student, National Technical University "Kharkiv Polytechnic Institute", Kharkiv, Ukraine; e-mail: m.sliepushkov@gmail.com; ORCID: <https://orcid.org/0009-0001-0004-2820>

K. O. KHATSKO, Senior Lecturer of the Department of the Department of Information Systems and Technologies of the National Technical University "Kharkiv Polytechnic Institute", Graduate Student, Kharkiv, Ukraine; e-mail: kyrylo.khatsko@kphi.edu.ua; ORCID: <https://orcid.org/0000-0003-3315-1553>

Y. O. SHEBANOV, Graduate Student, National Technical University "Kharkiv Polytechnic Institute", Kharkiv, Ukraine; e-mail: yevhenii.shebanov@cs.kphi.edu.ua; ORCID: <https://orcid.org/0009-0006-9032-8764>

MODIFIED SOFTWARE DEPLOYMENT ALGORITHM USING MULTI-THREADING

The article presents a modified deployment algorithm for software systems using multithreading in AWS CodeBuild, aimed at optimizing build time and reducing computational resource costs in cloud environments. The key stages of the build process, including parallel test execution, task allocation analysis, and resource management, were modeled using finite automata, timed automata, and Petri nets. Particular attention was given to identifying and addressing the limitations of AWS CodeBuild's standard parallelization mechanisms, which can lead to inefficient resource utilization and extended build durations. The study revealed that AWS CodeBuild's default mechanisms are not always capable of optimally leveraging system resources, especially when handling large software projects with numerous dependencies. To overcome these limitations, the use of Python's multithreading capabilities was proposed as a convenient tool for extending the platform's base functionality. The proposed approach enabled flexible thread management and task distribution at the user scenario level, significantly reducing overall build time. Experimental results demonstrated substantial reductions in build execution time compared to the default AWS CodeBuild settings, confirming the effectiveness of the proposed algorithm in enhancing performance and ensuring high scalability for build processes in cloud environments. The developed algorithm is particularly relevant for large software projects requiring frequent iterative builds and testing. The findings can be utilized to improve automated deployment processes and computational resource management in cloud ecosystems.

Keywords: deployment process, information technology, algorithm, software deployment process model, timed finite automaton, Petri nets.

Повні імена авторів / Author's full names

Автор 1 / Author 1: Хацько Наталія Євгенівна / Khatsko Nataliia Yevgenivna

Автор 2 / Author 2: Слєпушков Микола Васильович / Sliepushkov Mykola Vasylovych

Автор 3 / Author 3: Хацько Кирило Олександрович, Khatsko Kyrylo Olexandrovych

Автор 4 / Author 4: Шебанов Євгеній Олександрович / Shebanov Yevhenii Olexandrovych

В. О. КОЛБАСІН, кандидат технічних наук (PhD), доцент, доцент кафедри системного аналізу інформаційно-аналітичних технологій Національного технічного університету «Харківський політехнічний інститут», м. Харків, Україна; e-mail: viacheslav.kolbasin@khp.edu.ua; ORCID: <https://orcid.org/0009-0005-6367-7574>

ОРКЕСТРУВАННЯ РОБОЧИХ ПРОЦЕСІВ ВЗАЄМОДІЇ З КЛІЄНТАМИ В КОРПОРАТИВНИХ ЗАСТОСУНКАХ

Автоматизація процесів є важливим чинником розвитку сучасних корпоративних систем, зокрема систем електронної комерції. Однією з задач, що треба вирішувати при побудові подібних систем є оркестрування процесів взаємодії з клієнтом, які за своєю природою є асинхронними. Сучасні популярні фреймворки для оркестрування, зокрема Airflow, не дуже пристосовані для одночасного очікування відповіді від клієнта великою кількістю робочих процесів. Це призводить до зайвих витрат обчислювальних ресурсів та зниження економічної ефективності системи. В даній роботі розглядається задача побудови системи оркестрування робочих процесів взаємодії з клієнтом на основі подійно-орієнтованого підходу. Замість умовно послідовного централізованого виконання графової моделі робочого процесу пропонується представити робочий процес у вигляді послідовності подій та реакції системи на них, що дозволяє позбутися потреби в явному очікуванні відповіді від клієнта. В такій моделі операції робочого процесу будуть виконуватися під час обробки подій, а перехід до наступної операції відбуватиметься за рахунок відправлення нового повідомлення з описом цієї операції. Розглянуто централізований та розподілений підходи для виконання переходів між операціями робочого процесу, показані переваги та недоліки кожного з них. Також розглянуто реалізацію механізму очікування відповіді від клієнта, для чого пропонується перед початком взаємодії з останнім зберігати повідомлення у спеціалізованому сховищі, прив'язане до ідентифікатору сеансу взаємодії з клієнтом. А після надходження даних від клієнта – додавати їх до повідомлення і відправляти його назад в чергу повідомлень для продовження виконання робочого процесу. Для опису моделі робочого процесу пропонується використовувати формат JSON, зважаючи на можливість використання обробників, написаних на різних мовах програмування. Описано один з підходів до побудови такого формату опису та продемонстровано його використання для реалізації прикладу робочого процесу. Результати дослідження можуть бути корисними при створенні систем оркестрування робочих процесів взаємодії з клієнтом.

Ключові слова: оркестрування робочого процесу, подійно-орієнтований підхід, взаємодія з клієнтом, асинхронні операції.

Вступ. Здатність зручно та економічно організувати взаємодію з клієнтами є однією з конкурентних переваг сучасної компанії, орієнтованої на взаємодію з кінцевим споживачем. Для зниження витрат компанії треба збільшувати рівень автоматизації та переводити все більше і більше процесів в режим з найменшим залученням людини. З іншого боку клієнти не завжди радіють такій автоматизації, бо відчутна частина їх запитів не підтримується програмним забезпеченням системи і вони не можуть бути опрацьовані.

Використання чат-ботів з підтримкою штучного інтелекту покращує задоволеність клієнтів процесом спілкування, але слабка здатність системи вирішувати нетипові запити все ще є проблемою [1, 2]. Кожен нетиповий запит потребує певного аналізу людиною та можливо отримання дозволів на його виконання. А після того, як він стає одним з типових сценаріїв – він має бути доданий до системи.

Обробка запиту та інші бізнес-процеси компанії зазвичай описуються робочим процесом (workflow). Робочий процес зазвичай складається з певної послідовності типових операцій. Це можуть бути операції виконання певних запитів до інших систем, очікування додаткової інформації від клієнта, операції вибору подальшого шляху обробки запиту та інші подібні операції.

Завдяки тому, що базові операції повторюються в різних робочих процесах, їх можна реалізувати окремо та викликати в ході виконання робочого процесу. Таким чином, реалізація бізнес-процесів буде виглядати як послідовний виклик стандартних операцій під управлінням певного управляючого процесу – оркест-

ратора. Його задачею є управління послідовністю виклику операцій робочого процесу та обробка особливих ситуацій під час його виконання.

Найпоширенішою моделлю робочого процесу є направлений ациклічний граф (directed acyclic graph, DAG), в якому вузлами графу є операції, а гілками графу – переходи між ними. Вузли можуть виконувати як бізнес-операції (виконати запит до платіжної системи або надіслати інформацію користувачеві), так і службові операції або перевірки (чи відповідає користувач, або його замовлення певним умовам – наприклад чи перший раз він використовує промокод). Гілки ж задають допустимі послідовності операцій.

Типовими реалізаціями оркестраторів графових моделей робочого процесу є платформи Apache Airflow [3] та Step Functions в хмарній платформі Amazon Web Services [4]. Вони дозволяють реалізувати довільний ациклічний граф робочого процесу з базових та допоміжних операцій. Кількість моделей робочих процесів та процесів, що одночасно виконуються є доволі великою, але вона обмежена ресурсами інфраструктури виконання.

З ростом кількості моделей робочих процесів та збільшенням відсотка операцій очікування, зручність використання наведених вище засобів оркестрування знижується. Особливо це відчутно при їх використанні для автоматизації процесів взаємодії з клієнтом, де система може доволі довго чекати на відповідь. Причиною таких обмежень є те, що оркестратор має відслідковувати стан виконання процесу та зберігати його контекст, тобто дані, що мають передаватися між

© Колбасін В.О., 2024



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.



операціями. А для цього потрібні певні ресурси процесора та оперативної пам'яті.

Таким чином використання найбільш поширених платформ для оркестрування виконання робочих процесів має обмеження за рахунок наявності в них централізованого відстеження ходу виконання процесу та зберігання його контексту. Зменшення потреби в обчислювальних ресурсах та розробка більш гнучкого механізму оркестрування робочих процесів взаємодії з клієнтом в корпоративних застосунках є актуальною задачею, яка розглядається в даному дослідженні.

Виконання робочого процесу. Графова модель робочого процесу задається спеціальним описом графа у форматі JSON для платформи AWS Step Functions та програмою спеціального виду на мові програмування Python для платформи Airflow [3, 4]. Дана модель зберігається у централізованому сховищі платформи оркестрування та зчитується з нього при запуску процесу. Платформа відстежує результат виконання поточної операції процесу та залежно від нього виконує перехід на виконання наступної операції.

Процес взаємодії з клієнтом є за своєю природою асинхронним і очікує різної поведінки сторін. Клієнт може відповісти на питання системи або надати додаткову інформацію в будь-який момент часу протягом певного періоду очікування. Якщо він не відповідає – система може йому нагадати про потребу надати дані. Якщо ж клієнт на взаємодії з системою протягом певного періоду часу – сеанс взаємодії буде завершений та виконання бізнес-процесу перервано. Від системи навпаки очікується швидка відповідь та здатність спілкуватися з клієнтом в будь-який час доби.

Приклад робочого процесу взаємодії з клієнтом для надання йому інформації про стан рахунку наведено на рис. 1, де жовтим позначені звичайні операції робочого процесу, синім – операції очікування та червоним – кінцеві операції процесу.



Рис. 1. Робочий процес отримання стану рахунку.

При такому сценарії взаємодії більшість часу виконання робочого процесу буде витратитися на очікування відповіді від клієнта. Але пряма його реалізація базовими засобами того ж Airflow призведе до блокування одного зі слотів виконання на весь час очікування. Реалізація через відкладені оператори та тригери [5] є економічнішою, але вона є схильною до помилок та також використовує ресурси відповідного обмеженого компоненту платформи.

Відмовитися від явної операції очікування можна зміною моделі виконання робочого процесу з умовно послідовного виконання операцій на графі на модель, засновану на обробці подій [6, 7]. Подія представляється повідомленням, що несе інформацію про подію – в даному випадку це може бути відповідь від клієнта та запитані дані, або результат обробки

даних клієнта системою. Подія передається споживачу за допомогою брокера повідомлень, такого як RabbitMQ [8] або Apache Kafka [9]. Споживач подій виконує обробку повідомлень та передає результат у вигляді іншого повідомлення в певну чергу.

В цій моделі операції робочого процесу будуть виконуватися під час обробки подій, а перехід до виконання наступної операції буде відображатися надсиланням повідомлення з описом цієї операції.

Модель робочого процесу може виконуватися централізовано, і тоді після закінчення виконання кожної операції робочого процесу повідомлення з його результатами має надсилатися диспетчеру (споживач спеціального типу), який буде відповідати за аналіз результатів та вибір наступної операції. Іншим варіантом є покладення обов'язків робити вибір наступної операції робочого процесу на споживачів, що виконують операції бізнес-процесу.

Використання другого підходу є компромісним і разом з очевидними перевагами у вигляді зменшення числа елементів системи містить і певні ризики, бо вимагає строго однакової обробки моделі процесу всіма обробниками.

Приймаючи до уваги, що процес взаємодії з користувачем може бути доволі тривалим і має потребу в постійному покращенні та адаптації під потреби підприємства та клієнтів, ймовірно, що одномоментно будуть існувати різні версії одного й того ж самого робочого процесу. Для забезпечення послідовності виконання існуючих процесів різних версій при централізованому підході до зберігання моделі треба буде зберігати різні версії моделі робочого процесу і у повідомленнях передавати ідентифікатор версії. З іншого боку, враховуючи не дуже великий розмір типових робочих процесів взаємодії з клієнтом, в даній роботі пропонується зберігати всю модель робочого процесу як частину повідомлення.

Окрім операції можуть виконуватися як за допомогою загальних обробників повідомлень, так і з використанням спеціалізованих обробників, що можуть відповідати за розпізнавання мови та зображень (вони потребують специфічного апаратного забезпечення) або знаходитися всередині внутрішньої захищеної мережі компанії. У випадку використання спеціалізованих обробників треба забезпечити можливість диспетчеризації повідомлень залежно від типу операції.

Очікування відповіді від клієнта є частиною робочого процесу та має описуватися окремим повідомленням. На час очікування це повідомлення має бути збережено в певному сховищі, а потім при появі відповіді клієнта, має бути надіслано обробнику.

Зазвичай інформацію від клієнта система отримує через механізм зворотного виклику (callback), який не є специфічним для клієнта – тобто повідомлення від всіх клієнтів приходять на одну точку входу. Більш того, повідомлення від клієнта може бути початковим для старту взаємодії з системою. Такі повідомлення треба розрізняти від тих, що є відповіддю клієнта. А для останніх треба отримати робочий процес, що чекає відповіді, та збережене повідомлення. Це можна

зробити за рахунок використання ідентифікатора взаємодії з клієнтом, який може бути як випадковим ідентифікатором сесії у випадках веб-чату, так і постійним ідентифікатором користувача у месенджері.

З урахуванням всього наведеного вище структурна схема виконання робочого процесу буде виглядати як наведено на рис. 2.

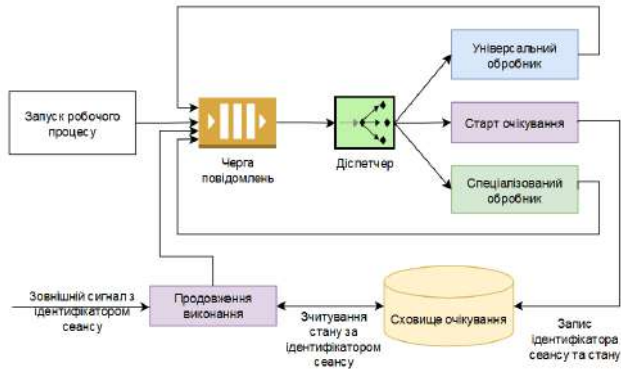


Рис. 2. Структурна схема виконання робочого процесу.

Виконання робочого процесу запускається зовнішнім надсиланням відповідного повідомлення з даними та моделлю процесу у чергу фреймворку. Далі відбувається виконання операцій процесу загальними та спеціалізованими обробниками, кожен з яких після закінчення виконання операції обчислює наступний стан та надсилає повідомлення зі змінним станом процесу назад до черги. Так повторюється до тих пір, поки виконання процесу не закінчиться.

Для програмної реалізації даного підходу треба створити відповідні обробники та модулі взаємодії з зовнішніми системами взаємодії з клієнтом. Механізм диспетчеризації повідомлень також має бути реалізований програмно.

При використанні мови програмування Python для створення системи оркестрування є сенс розглянути фреймворк Celery [10], який реалізує розподілену чергу задач, яка управляється повідомленнями. Він дозволяє робити ациклічні моделі робочих процесів на рівні Python коду, але не підтримує напряму зупинки обробки задачі та її подовження. Тобто його можна використати для виконання лінійних частин графу без очікування. Також одним з напрямків подальших досліджень є вивчення можливостей адаптації даного фреймворку для реалізації очікування зовнішнього сигналу, що управляється повідомленнями [11].

Структура повідомлень. Повідомлення має містити всю необхідну інформацію для виконання поточної та всіх наступних операцій робочого процесу. Це означає, що воно має зберігати модель робочого процесу, ідентифікатор поточної операції та контекст виконання. Всі ці дані мають бути розташовані у повідомленні, для серіалізації якого в даній роботі пропонується використати формат JSON [12].

Контекст виконання робочого процесу потрібен для забезпечення зв'язку по даним між операціями. Оскільки оркестратор не може знати всі можливі поля даних та їх комбінації – контекст має підтримувати довільний формат даних, що зручно JSON об'єктом.

Кожна операція повинна мати свій власний унікальний ідентифікатор в моделі, також вона має містити всю необхідну інформацію для виклику операції обробником. Пропонується використати наступний опис операції:

- **handler** – опис обробника, визначає який обробник має виконати операцію;
- **function** – функція, яка має бути виконана обробником, містить посилання на код функції;
- **parameters** – параметри виконання функції;
- **next** – операція, що має бути виконана після виконання даної.

Будь-яка функція отримує два параметри – контекст виконання та параметри самої функції. Обидва параметри є словниками, отриманими через десеріалізацію відповідних JSON об'єктів.

Службові операції на кшталт вибору гілки графу або запуску паралельних операцій помічаються спеціальним значенням типу обробника **system** і вони мають бути опрацьовані тим самим обробником, що і операція перед ними.

З використанням даного підходу до опису робочого процесу, приклад процесу отримання стану рахунку, наведений на рис. 1, буде виглядати так:

```
{
  "checkAuth": {
    "function": "com.company.CheckAuth",
    "parameters": {
      "result": "authStatus"
    },
    "next": "authChoice",
  },
  "authChoice": {
    "handler": "system",
    "function": "choice",
    "parameters": {
      "var": "context.authStatus",
      "options": {
        "1": "askAccount",
        "0": "startAuth"
      }
    }
  },
  "startAuth": {
    "function": "com.company.Request2FA",
    "next": "check2FA"
  },
  "check2FA": {
    "function": "com.company.Check2FA",
    "parameters": {
      "result": "authResult"
    },
    "next": "2FAChoice"
  },
}
```

Рис. 3. Фрагмент опису процесу отримання стану рахунку.

Як можна побачити на цьому описі, інформація про те, чи потрібна аутентифікація, отримується окремою функцією та її результат записується в змінну **authStatus** контексту. Далі цей самий обробник перевіряє значення змінної контексту та залежно від нього надсилає у чергу повідомлення з активним станом **askAccount** або **startAuth**.

Операція **startAuth** надсилає СМС клієнту та очікує його відповіді. Вона має спочатку зберегти стан моделі в сховищі і тільки після цього надіслати повідомлення клієнту для запобігання виникнення стану гонки (race condition).

Після отримання повідомлення від клієнта, обробник зворотного виклику зчитує повідомлення за ідентифікатором сеансу, додає до контексту отримане від клієнта значення та надсилає повідомлення в чергу. Обробник перевіряє наявність отриманого від клієнту значення і надсилає в чергу повідомлення з активним станом **check2FA**.

Висновки. У ході дослідження було розглянуто оркестрування робочих процесів взаємодії з клієнтом та було виявлено, що популярні платформи не дуже економічно виконують очікування відповіді від клієнта. Було запропоновано підхід до оркестрування таких процесів на основі архітектури обробки подій та визначені головні структурні елементи оркестратора та сформульовано базовий формат опису моделі процесу. Результати дослідження можуть знайти використання при розробці корпоративних інформаційних систем, зокрема систем електронної комерції.

Список використаної літератури

- Uzoka A., Cadet E. Ojukwu P.U. Leveraging AI-Powered chatbots to enhance customer service efficiency and future opportunities in automated support. *Computer Science & IT Research Journal*. 2024. Vol. 5(10). P. 2485–2510. DOI: 10.51594/csitrj.v5i10.1676.
- Hsu C.-L., Lin J.C.-C. Understanding the user satisfaction and loyalty of customer service chatbots. *Journal of Retailing and Consumer Services*, Vol. 71, 103211. DOI: 10.1016/j.jretconser.2022.103211.
- Ruiter J., Harenslak B. *Data Pipelines with Apache Airflow*. Manning, 2021. 480 p.
- Sbarski P., Cui Y., Nair A. *Serverless Architectures on AWS, Second Edition*. Manning, 2022. 256 p.
- Deferrable Operators & Triggers*. URL: <https://airflow.apache.org/docs/apache-airflow/stable/authoring-and-scheduling/deferring.html> (дата звернення: 10.12.2024).
- Hohpe G., Woolf B. *Enterprise Integration Patterns: Designing, Building, and Deploying Messaging Solutions*. Addison-Wesley, 2003. 736 p.
- Stopford B. *Designing Event-Driven Systems*. O'Reilly Media Inc., 2018. 171 p.
- Roy G. *RabbitMQ in Depth*. Manning, 2017. 264 p.
- Shapira G, Palino T., Sivaram R., Petty K. *Kafka: The Definitive Guide, 2nd Edition*. O'Reilly Media Inc., 2021. 485 p.
- Zhang C. *Celery. A detailed guide on how Celery asynchronous task queue works*. URL: <https://python.plainenglish.io/asynchronous-task-queueing-with-celery-d9709364e686>. (дата звернення: 10.12.2024)
- Custom Message Consumers. *Celery documentation*. URL: <https://docs.celeryq.dev/en/latest/userguide/extending.html#custom-message-consumers>. (дата звернення: 10.12.2024)
- Bray T. *The JavaScript Object Notation (JSON) Data Interchange Format*. IETF, 2017. DOI: 10.17487/RFC8259.

References (transliterated)

- Uzoka A., Cadet E. Ojukwu P.U. Leveraging AI-Powered chatbots to enhance customer service efficiency and future opportunities in automated support. *Computer Science & IT Research Journal*. 2024. Vol. 5(10). P. 2485–2510. DOI: 10.51594/csitrj.v5i10.1676.
- Hsu C.-L., Lin J.C.-C. Understanding the user satisfaction and loyalty of customer service chatbots. *Journal of Retailing and Consumer Services*, Vol. 71, 103211. DOI: 10.1016/j.jretconser.2022.103211.
- Ruiter J., Harenslak B. *Data Pipelines with Apache Airflow*. Manning, 2021. 480 p.
- Sbarski P., Cui Y., Nair A. *Serverless Architectures on AWS, Second Edition*. Manning, 2022. 256 p.
- Deferrable Operators & Triggers*. URL: <https://airflow.apache.org/docs/apache-airflow/stable/authoring-and-scheduling/deferring.html> (access date: 10.12.2024).
- Hohpe G., Woolf B. *Enterprise Integration Patterns: Designing, Building, and Deploying Messaging Solutions*. Addison-Wesley, 2003. 736 p.
- Stopford B. *Designing Event-Driven Systems*. O'Reilly Media Inc., 2018. 171 p.
- Roy G. *RabbitMQ in Depth*. Manning, 2017. 264 p.
- Shapira G, Palino T., Sivaram R., Petty K. *Kafka: The Definitive Guide, 2nd Edition*. O'Reilly Media Inc., 2021. 485 p.
- Zhang C. *Celery. A detailed guide on how Celery asynchronous task queue works*. URL: <https://python.plainenglish.io/asynchronous-task-queueing-with-celery-d9709364e686>. (access date: 10.12.2024).
- Custom Message Consumers. *Celery documentation*. URL: <https://docs.celeryq.dev/en/latest/userguide/extending.html#custom-message-consumers>. (access date: 10.12.2024)
- Bray T. *The JavaScript Object Notation (JSON) Data Interchange Format*. IETF, 2017. DOI: doi:10.17487/RFC8259.

Надійшла (received) 06.12.2024

UDC 004.9

V. O. KOLBASIN, Candidate of Technical Sciences (PhD), Docent, Associate Professor at the Department of System Analysis and Information-Analytical Technologies, National Technical University «Kharkiv Polytechnic Institute», Kharkiv, Ukraine; e-mail: viacheslav.kolbasin@khp.edu.ua; ORCID: <https://orcid.org/0009-0005-6367-7574>

ORCHESTRATION OF CUSTOMER INTERACTION WORKFLOWS IN ENTERPRISE APPLICATIONS

Process automation is an important factor in the development of modern corporate systems, in particular e-commerce systems. One of the tasks that must be solved during development of such systems is the orchestration of customer interaction processes, which are asynchronous in nature. Modern popular orchestration frameworks (Airflow for example), are not very well adapted to simultaneously waiting for a response from the client for a large number of workflows. This leads to unnecessary costs of computing resources and a decrease in the economic efficiency of the system. This paper considers the task of building a system for orchestrating customer interaction workflows based on an event-driven approach. Instead of a sequential-like centralized execution of the graph model of the workflow, it is proposed to present the workflow as a sequence of events and the system's reactions to them, which eliminates the need to explicitly wait for a response from the client. In such model workflow operations will be performed during event processing, and the transition to the next operation will occur by sending a new message with a description of the operation. Centralized and distributed approaches for performing transitions between workflow operations are considered, the advantages and disadvantages of each of them are shown. The implementation of waiting for client response is also considered, for which it is proposed to store the message tied to the client session identifier in a specialized storage before starting interaction with client. And after receiving data from the client, add them to the message and send it back to the message queue to continue the workflow. Taking into account the possibility of using handlers written in different programming languages, JSON format based description of workflow model is proposed to be used. One of the approaches to building model description format is described and its use for demonstrating example of a workflow. The results of the study can be useful when creating systems for orchestrating workflows of interaction with the client.

Keywords: workflow orchestration, event-driven approach, client interaction, asynchronous operations.

Повне ім'я автора / Author's full name

Колбасін Вячеслав Олександрович / Kolbasin Viacheslav Oleksandrovych

Є. В. БОДЯНСЬКИЙ, доктор технічних наук, професор, Харківський національний університет радіоелектроніки, професор кафедри штучного інтелекту, м. Харків, Україна; e mail: yevgeniy.bodyanskiy@nure.ua, ORCID: <https://orcid.org/0000-0001-5418-2143>

О. С. ЧАЛА, Харківський національний університет радіоелектроніки, аспірант кафедри штучного інтелекту; м. Харків, Україна, e mail: olha.chala@nure.ua, ORCID: <https://orcid.org/0000-0002-7603-1247>

ІНТЕЛЕКТУАЛЬНА ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ ШВИДКОЇ КЛАСИФІКАЦІЇ ЗА УМОВ ПЕРЕТИННИХ КЛАСІВ

Предметом дослідження є процес швидкої класифікації даних за умов перетинних класів. Швидка класифікація виконується в режимі реального часу або близькому до нього. Мета роботи полягає у розробці інтелектуальної інформаційної технології швидкої класифікації в online та nearline режимах за умов перетинних класів. Досягнення мети дає можливість врахувати нестационарність вхідних даних та дисбаланс класів за умов потокового надходження даних. Задачі компенсації шумів у вхідних даних та зміну розподілу вхідних даних внаслідок нестационарності таких даних, а також задача компенсації дисбалансу класів пов'язані між собою при класифікації в умовах перетинних класів та потребують розробки комплексного рішення. Для досягнення мети вирішуються такі задачі: структуризація підходів до класифікації перетинних класів з урахуванням нестационарності вхідних даних та дисбалансу класів; розробка інтелектуальної технології класифікації в online та nearline режимах. Запропоновано інтелектуальну інформаційну технологію швидкої класифікації за умов перетинних класів. Технологія містить етапи попередньої класифікації з урахуванням шуму у вхідних даних, класифікації з урахуванням дисбалансу класів та класифікації з урахуванням зміни закономірностей у вхідних даних. Технологія передбачає послідовне використання нео-фаззі системи, адаптивної нейро-фаззі системи та багатопшарової нейронної мережі з ядерними дзвонуватими функціями активації. Нео-фаззі система використовує нео-нечіткі нейрони, що забезпечує стійкість до шуму. Адаптивна нейро-фаззі система враховує відстані між вхідними даними та центром класу у просторі ознак, що забезпечує класифікацію в умовах дисбалансу класів. Багатопшарова нейронна мережа з ядерними дзвонуватими функціями активації використовує рекурентний алгоритм навчання, що забезпечує адаптацію до нових даних з новим розподілом. Технологія дає можливість забезпечити швидке ітеративне уточнення рішень з класифікації згідно змін характеристик вхідних даних.

Ключові слова: швидка класифікація, перетинні класи, online режим, nearline режим, нейронна мережа, нейро-фаззі система, нео-фаззі система, адаптивне навчання, нечітка логіка.

Вступ. Класифікація даних, тобто віднесення даних до одного із відомих класів, знаходить широке застосування в різних галузях, включаючи розпізнавання образів, аналіз відеоданих та автоматизацію процесів підтримки прийняття рішень [1]. Задача швидкої класифікації, передбачає обробку даних у режимі online або близькому до нього (nearline). Швидка класифікація є актуальною у системах, що мають швидко реагувати на зміну навколишнього середовища, наприклад, у транспортних системах, системах моніторингу безпеки тощо [2].

Швидка класифікація потребує досягнення заданого рівня точності при обмеженнях на обчислювальні ресурси, пов'язаних із обмеженнями на час прийняття рішень в системі штучного інтелекту.

Швидка класифікація, як правило, реалізується в умовах перетинних класів, тобто коли дані з різних класів перетинаються у просторі ознак, оскільки вони мають близькі характеристики [3, 4].

Реалізація швидкої класифікації в умовах перетинних класів потребує додаткового вирішення задач усунення шуму та адаптації до змін у розподілі даних (concept drift) внаслідок нестационарності вхідних даних, усунення дисбалансу класів а також уточнення границь між класами. Так, дисбаланс класів може впливати на точність класифікації внаслідок суттєво різної кількості елементів даних у класах. Шум представляється шумовими точками у просторі ознак, що не відповідають загальним закономірностям даних. Відповідно, шум може знижувати точність класифікації.

Адаптація до concept drift забезпечує актуальність моделі в умовах динамічних змін у середовищі [5]. Уточнення границь перетину класів дозволяє підвищити точність класифікації за умов складного розподілу даних.

Таким чином, швидка класифікація потребує комбінування підходів до послідовного вирішення набору задач, пов'язаних із особливостями реальних даних та особливостями функціонування інформаційних систем, що генерують такі дані.

Аналіз останніх досліджень і публікацій.

Методи швидкої класифікації, зокрема в умовах потокової обробки даних, виконують адаптацію до нових даних у режимі реального, або близького до реального часу. Така адаптація виконується з використанням інкрементного навчання [6].

Еволюційні нейро-фаззі системи орієнтовані на роботу з перетинними класами з використанням нечіткої логіки. Такий підхід дає можливість оцінювати рівень належності даних до можливих класів, коли класи частково перекриваються у просторі ознак [7, 8].

Для вирішення проблеми дисбалансу класів у online-режимі використовуються базові стратегії створення додаткових проміжних даних для менш представленого класу та вилучення частини даних для домінуючого класу [9]. Вказані стратегії дають можливість підвищити точність класифікації без значних втрат вхідної інформації, проте вони не можуть бути використані при потоковій обробці даних.

© Є.В. Бодяньський, О.С. Чала, 2024



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.



В умовах нестаціонарності, коли вхідні дані містять шум, використовуються методи видалення шумових точок [10], що дає можливість зменшити вплив аномальних даних на результати класифікації, що підвищує загальну точність системи.

Класифікація в умовах перетинних класів з можливостями адаптації в умовах нестаціонарності, в тому числі при зміні розподілу вхідних даних, виконується з використанням комбінованого навчання в нейро-фаззі [7, 11, 12] та нео-нечітких системах [13].

Таким чином, сучасні підходи до класифікації в online та nearline режимах зазвичай орієнтовані на вирішення окремо задач компенсації нестаціонарності вхідних даних, дисбалансу класів, короткої вибірки. Проте ці задачі пов'язані між собою при класифікації в умовах перетинних класів та потребують єдиного інтегрованого рішення, що свідчить про актуальність розробки інтелектуальної інформаційної технології швидкої класифікації.

Мета та задачі дослідження. Мета роботи полягає у розробці інтелектуальної інформаційної технології швидкої класифікації в online та nearline режимах за умов перетинних класів. Вирішення цієї задачі дає можливість врахувати нестаціонарність вхідних даних та дисбаланс класів при потоковому надходженні вхідних даних.

Для досягнення поставленої мети вирішуються наступні задачі:

- структуризація нейро-фаззі, нео-фаззі підходів до класифікації перетинних класів з урахуванням нестаціонарності вхідних даних та дисбалансу класів;
- розробка інтелектуальної технології класифікації в online та nearline-режимах.

Структуризація підходів до класифікації з урахуванням нестаціонарності вхідних даних та дисбалансу класів.

Запропонована структуризація базується на порівнянні можливостей нео-фаззі, нейро-фаззі систем та багатошарової нейронної мережі з ядерними дзвонуватими функціями активації [12, 13].

Нео-фаззі система поєднує радіально-базисну нейронну мережу з нео-нечітким нейроном. Система забезпечує високу швидкість навчання завдяки використанню нелінійних синапсів. Вихідні сигнали залишаються лінійно залежними від налаштованих параметрів, що забезпечує можливість роботи в режимі онлайн. Система використовує комбіноване навчання, що передбачає: навчання з учителем, самонавчання та «лінійне навчання». Такий підхід забезпечує можливість адаптації до змін у вхідних даних у реальному часі з використанням обмежених вибірок. Навчання з учителем відповідає за налаштування синаптичних ваг системи.

Нео-нечіткий нейрон використовує нечіткі функції належності, що дозволяє обробляти невизначену інформацію та забезпечує гнучкість у моделюванні складних залежностей у даних.

Система має стійкість до шуму завдяки використанню функцій належності: у процесі класифікації активуються дві сусідні функції належності одночасно, що зменшує вплив шумових даних. Можливість адап-

тивного налаштування центрів та кількості функцій належності зменшує вплив змін у вхідних даних.

Адаптивна нейро-фаззі система дає можливість використати дані не лише у векторній, а й у матричній формі. Система складається з чотирьох шарів. Перший шар містить нейрони, що враховують відстані між вхідними даними та центром класу у просторі ознак. Другий шар складається з суматорів для кожного класу, які обчислюють оцінки Парзена щільності розподілу ймовірностей. Третій шар коригує ці оцінки вартості помилок класифікації. Вихідний шар визначає рівні належності вхідних об'єктів до класів.

Система використовує комбінацію методів навчання: «лінійне навчання»; самоорганізацію згідно принципу «переможець забирає все»; нечітка кластеризація. Лінійне навчання базується на концепції «нейрони в точках даних». Така концепція дає можливість адаптувати центри функцій відповідно до вхідних даних.

Задачу усунення дисбалансу класів вирішується за рахунок налаштування центрів активіаційних функцій при навчанні нейро-фаззі системи.

Багатошарова нейронна мережа з ядерними дзвонуватими функціями активації складається з трьох шарів. Вхідний шар може отримувати дані не лише у вигляді векторів, а й у вигляді матриць, що дає можливість зберегти просторову структуру даних. Прихований шар містить нейрони з ядерними дзвонуватими функціями активації. Ці функції моделюють складні нелінійні границі між класами. Рівень активації нейрона визначається відстанню між вхідними даними та центром нейрона, що дозволяє адаптивно налаштувати рецептивні поля нейронів і, відповідно, уточнити границі між класами. Вихідний шар формує результат класифікації на основі лінійної комбінації виходів прихованого шару. Мережа використовує рекурентний онлайн-алгоритм навчання, що забезпечує оновлення параметрів у реальному часі або близькому до реального часу. Такий алгоритм забезпечує адаптацію до нових даних, з новим розподілом, без необхідності повторного навчання на всьому наборі даних. Адаптивне налаштування параметрів функцій активації виконується з використанням градієнтного методу. При відмінності поточного виходу мережі від бажаного вагові коефіцієнти коригуються з метою мінімізувати різницю. На даному етапі виконується уточнення границь між класами за рахунок використання ядерних функцій активації, що дає можливість ефективно розрізняти класи навіть при наявності складних перетинних границь. Швидке уточнення параметрів забезпечує можливість адаптації до змін у розподілі вхідних даних.

Інтелектуальна інформаційна технологія швидкої класифікації в умовах перетинних класів

Розроблена інтелектуальна інформаційна технологія містить такі етапи (рис. 1).

Етап 1. Швидка класифікація в online режимі з урахуванням шуму у вхідних даних з використанням нео-фаззі системи.

Етап 2. Швидка класифікація в online режимі або nearline режимі з урахуванням дисбалансу класів з використанням адаптивної нейро-фаззі система.

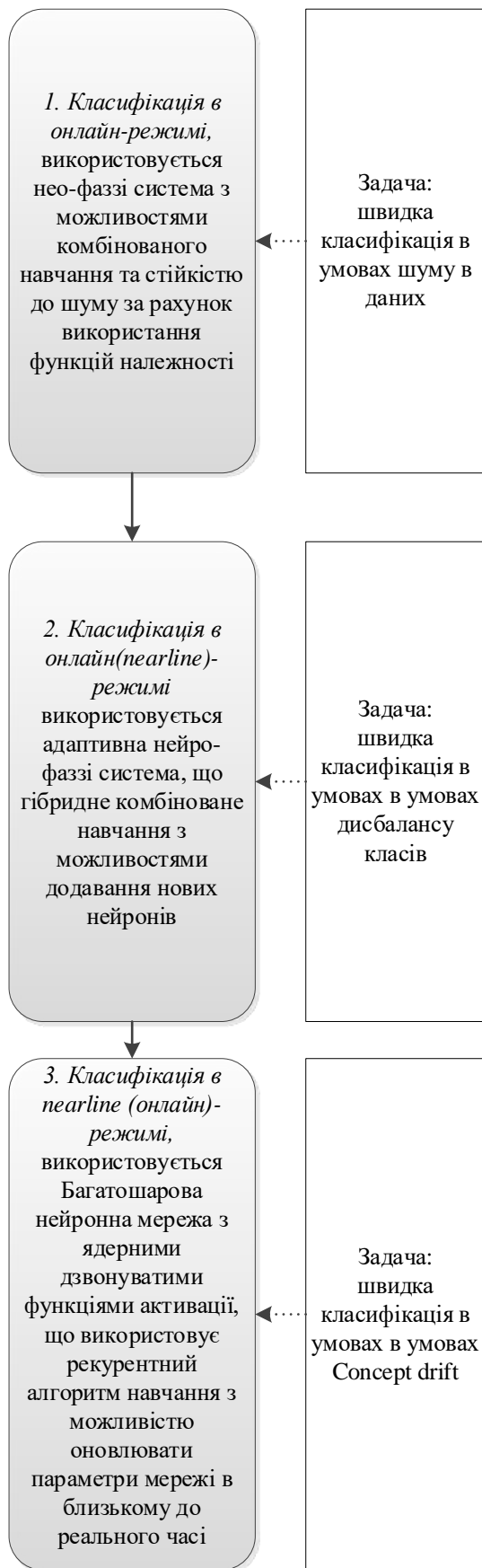


Рис. 1. Послідовність етапів інтелектуальної інформаційної технології швидкої класифікації за умов перетинних класів

Етап 3. Швидка класифікація в online режимі в online режимі або nearline режимі урахуванням зміни розподілу вхідних даних (concept drift) з використанням багатошарової нейронної мережі з ядерними дзвонуватими функціями активації.

Розглянемо послідовність використання розробленої технології на прикладі системи відеоспостереження з використанням двох потоків даних: основного та додаткового.

Основний потік використовується для детального аналізу потенційно небезпечних ситуацій. У даному потоці використовуються зображення з високою роздільною здатністю, зазвичай максимальною роздільною здатністю, доступною з відеокамер. Додатковий потік, на відміну від основного, має суттєво нижчу якість зображення з тим, щоб зменшити навантаження на мережеве обладнання.

Цей потік зазвичай використовує стандартні розширення з низькою роздільною здатністю, такі як QCIF, CIF, 2CIF, DCIF, 4CIF (D1).

Основне призначення додаткового потоку полягає в підтримці постійного віддаленого перегляду та моніторингу, в тому числі на мобільних пристроях та в умовах обмеженої пропускної здатності мережі.

Класифікація за наявності шуму у вхідних даних здійснюється в онлайн-режимі для додаткового потоку відеоданих. Цей потік має менший обсяг даних, що дає можливість здійснювати швидку класифікацію з низькою затримкою.

Уточнення границь перетинних класів та вирішення проблеми concept drift здійснюється в nearline-режимі для основного потоку (або в online-режимі при низькій частоті кадрів в системі відеоспостереження). Основний потік містить більше даних, необхідних для точного визначення границь класів та виявлення змін у розподілі даних. Також дана обробка потребує більш складних обчислень та, відповідно, є більш ресурсоємною.

Поєднання online та nearline-режимів дає можливість раціонально розподілити обчислювальні ресурси у процесі розпізнавання зображень у системі відеоспостереження.

На першому етапі виконується обробка зображень низької роздільної здатності у форматі QCIF. При бінарній класифікації виділяються класи зображень з пустою платформою та платформою з пасажиром. На даному етапі виконується попередня класифікація потоку зображень з урахуванням шумів у даних. На рис. 2 представлено приклад зображення низької роздільної здатності із шумом у даних, що класифікується на даному етапі.



Рис. 2. Приклад зображення низької роздільної здатності, що класифікується на етапі 1 інтелектуальної інформаційної технології

Шум, зокрема, пов'язаний із відблисками сонця на зображенні платформи. У прикладі використані зображення із датасету [14].

На другому етапі виконується класифікація зображень більшої роздільної здатності (рис. 3) із урахуванням дисбалансу класів.

Дисбаланс пов'язаний з тим, що кількість зображень пустої платформи суттєво перевищує кількість зображень платформи з пасажиром.



Рис. 3. Приклад зображення високої роздільної здатності, що класифікується на етапі 2 інтелектуальної інформаційної технології

На третьому етапі виконується класифікація зображень у випадку concept drift. Приклад періодичних змін закономірностей вхідних даних представлено на рис. 4.

В даному випадку на зображенні з'являється новий клас «забуті речі».

Результатом використання технології є класифікація зображень з урахуванням нестационарності вхідних даних та дисбалансу класів.



Рис. 4. Приклад зображення зі зміною структури вхідних, що класифікується на етапі 3 інтелектуальної інформаційної технології

Висновки. Запропоновано інтелектуальну інформаційну технологію швидкої класифікації за умов перетинних класів. Технологія містить етапи попередньої класифікації з урахуванням шуму у вхідних даних, класифікації з урахуванням дисбалансу класів та класифікації з урахуванням зміни закономірностей у вхідних даних.

У практичному аспекті запропонована технологія дає можливість забезпечити швидке ітеративне уточнення рішень з класифікації згідно характеристик вхідних даних та можливих непередбачуваних впливів на ці дані. При використанні даної технології базовим є перший етап, який забезпечує швидку online-класифікацію у типових умовах.

При наявності дисбалансу класів має бути реалізований етап 2, а при появі нових класів – етап 3 інтелектуальної інформаційної технології.

Список використаної літератури

- Engelbrecht Andries P. *Computational Intelligence: An Introduction*. NJ: John Wiley & Sons, 2007. 632 p.
- Howard A.G., Zhu M., Chen B., Kalenichenko D., Wang W., Weyand T., Adam H. *Mobilenets: Efficient convolutional neural networks for mobile vision applications*. URL: <https://arxiv.org/abs/1704.04861> (дата звернення 14.10.2024).
- Xiong H., Wu J., Liu L. Classification with class overlapping: A systematic study and a new measure of class separability. *IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans*, 2010. Vol. 40(4). P. 973–985.
- Vuttipittayamongkol P., Zhang X., Yang H. Overlap-based undersampling for improving classification performance on imbalanced data with overlapping classes. *Knowledge-Based Systems*, 2020.
- Krawczyk B., Minku L.L., Gama J., Stefanowski J., Woźniak M. Ensemble learning for data stream analysis: A survey. *Information Fusion*, 2017. Vol. 37. P. 132–156.
- Bifet A., Holmes G., Kirkby R., Pfahringer B. MOA: Massive Online Analysis. *Journal of Machine Learning Research*, 2010. Vol. 11. P. 1601–1604.
- Bodyanskiy Y.V., Tyshchenko O.K., Deineko A.O. An Evolving Neuro-Fuzzy System with Online Learning/Self-learning. *International Journal of Modern Education and Computer Science*, 2015. Vol. 7. no. 2. P. 1-7.
- Bodyanskiy Y., Pliss I., Chala O., Deineko A. Evolving fuzzy-probabilistic neural network and its online learning. *Proceedings of the 10th International Conference on Advanced Computer Information Technologies*. Deggendorf, Germany, 2016. URL: <https://doi.org/10.1109/ACIT49673.2020.9208904> (дата звернення 14.10.2024).
- Wang S., Minku L.L., Yao X. Dealing with Multiple Classes in Online Class Imbalance Learning. *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, 2016. P.1–7.
- Batista G.E.A.P.A., Prati R.C., Monard M.C. A Study of the Behavior of Several Methods for Balancing Machine Learning Training Data. *SIGKDD Explorations Newsletter*, 2004. Vol. 6(1). P. 20–29.
- Bodyanskiy Y., Chala O.V. et al. Evolving Stacking Neuro-Fuzzy Probabilistic Networks and Their Combined Learning in Online Pattern Recognition Tasks. *Artificial Intelligence in Control and Decision-making Systems*, 2023. P. 95–123.
- Bodyanskiy Y., Deineko A., Pliss I., Chala O. Fast probabilistic neuro-fuzzy system for pattern classification task. *Information Technology and Management Science*, 2020. Vol. 23. P. 12–16.
- Bodyanskiy Y., Chala O. Enhanced multidimensional neo-fuzzy classification system and its learning for the video classification task. *Management Information System and Devices*, 2024. Vol. 181. P. 42–50. URL: https://www.ewdtest.com/asu/wp-content/uploads/2024/09/ASUtaPA_181_42_50.pdf (дата звернення 14.10.2024).
- Wang Y., Jodoin P.-M., Porikli F., Konrad J., Benzeneth Y., Ishwar P. CDnet 2014: An expanded change detection benchmark dataset. *Proceedings of the IEEE Workshop on Change Detection (CDW-2014)* at CVPR-2014, 2014. P. 387–394. URL: <http://jacarini.dinf.usherbrooke.ca/dataset2014> (дата звернення 14.10.2024).

References (transliterated)

- Engelbrecht Andries P. *Computational Intelligence: An Introduction*. NJ: John Wiley & Sons, 2007. 632 p.
- Howard A.G., Zhu M., Chen B., Kalenichenko D., Wang W., Weyand T., Adam H. *Mobilenets: Efficient convolutional neural networks for mobile vision applications*. Available at: <https://arxiv.org/abs/1704.04861> (accessed 14.10.2024).
- Xiong H., Wu J., Liu L. Classification with class-overlapping: A systematic study and a new measure of class separability. *IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans*. 2010, vol. 40(4), pp. 973–985.
- Vuttipittayamongkol P., Zhang X., Yang H. Overlap-based undersampling for improving classification performance on imbalanced data with overlapping classes. *Knowledge-Based Systems*. 2020.
- Krawczyk B., Minku L.L., Gama J., Stefanowski J., Woźniak M. Ensemble learning for data stream analysis: A survey. *Information Fusion*. 2017, vol. 37, pp. 132–156.
- Bifet A., Holmes G., Kirkby R., Pfahringer B. MOA: Massive Online Analysis. *Journal of Machine Learning Research*. 2010, vol. 11, pp. 1601–1604.
- Bodyanskiy Y.V., Tyshchenko O.K., Deineko A.O. An Evolving Neuro-Fuzzy System with Online Learning/Self-learning. *International Journal of Modern Education and Computer Science*. 2015, vol. 7, no. 2, pp. 1–7.
- Bodyanskiy Y., Pliss I., Chala O., Deineko A. Evolving fuzzy-probabilistic neural network and its online learning. In *10th International Conference on Advanced Computer Information Technologies*. Deggendorf, Germany. 2016. Available at: <https://doi.org/10.1109/ACIT49673.2020.9208904> (accessed 14.10.2024).
- Wang S., Minku L.L., Yao X. Dealing with Multiple Classes in Online Class Imbalance Learning. *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*. 2016, pp. 1–7.
- Batista G.E.A.P.A., Prati R.C., Monard M.C. A Study of the Behavior of Several Methods for Balancing Machine Learning Training Data. *SIGKDD Explorations Newsletter*. 2004, vol. 6(1), pp. 20–29.
- Bodyanskiy Y., Chala O.V. et al. Evolving Stacking Neuro-Fuzzy Probabilistic Networks and Their Combined Learning in Online Pattern Recognition Tasks. *Artificial Intelligence in Control and Decision-making Systems*. 2023, pp. 95–123.
- Bodyanskiy Y., Deineko A., Pliss I., Chala O. Fast probabilistic neuro-fuzzy system for pattern classification task. *Information Technology and Management Science*. 2020, vol. 23, pp. 12–16.
- Bodyanskiy Y., Chala O. Enhanced multidimensional neo-fuzzy classification system and its learning for the video classification task. *Management Information System and Devices*. 2024, vol. 181, pp. 42–50. Available at: https://www.ewdtest.com/asu/wp-content/uploads/2024/09/ASUtaPA_181_42_50.pdf (accessed 14.10.2024).
- Wang Y., Jodoin P.-M., Porikli F., Konrad J., Benzeeth Y., Ishwar P. CDnet 2014: An expanded change detection benchmark dataset. In *Proceedings of the IEEE Workshop on Change Detection (CDW-2014) at CVPR-2014*. 2014, pp. 387–394. Available at: <http://jacarini.dinf.usherbrooke.ca/dataset2014> (accessed 14.10.2024).

Надійшла (received) 30.10.2024

UDC 004.8:004.9

E. V. BODIANSKY, Doctor of Technical Sciences, Professor, Kharkiv National University of Radio Electronics, Professor of the Department of Artificial Intelligence, Kharkiv, Ukraine; e-mail: yevgeniy.bodyanskiy@nure.ua; ORCID: <https://orcid.org/0000-0001-5418-2143>

O. S. CHALA, Kharkiv National University of Radio Electronics, graduate student of the Department of Artificial Intelligence; Kharkiv, Ukraine; e mail: olha.chala@nure.ua; ORCID: <https://orcid.org/0000-0002-7603-1247>

INTELLIGENT INFORMATION TECHNOLOGY FOR RAPID CLASSIFICATION UNDER CONDITIONS OF OVERLAPPING CLASSES

The subject of this research is the process of rapid data classification under conditions of overlapping classes. Rapid classification is performed in real-time or near-real-time mode. The aim of the work is to develop an intelligent information technology for rapid classification in online and nearline modes under conditions of overlapping classes. Achieving this goal allows for the consideration of non-stationarity in input data and class imbalance under conditions of streaming data. The tasks of compensating for noise in input data and changes in input data distribution due to non-stationarity, as well as the task of compensating for class imbalance, are interconnected when classifying under conditions of overlapping classes and require the development of a comprehensive solution. To achieve the goal, the following tasks are addressed: structuring approaches to classification of overlapping classes considering non-stationarity in input data and class imbalance; developing an intelligent technology for classification in online and nearline modes. An intelligent information technology for rapid classification under conditions of overlapping classes is proposed. The technology includes stages of preliminary classification considering noise in input data, classification considering class imbalance, and classification considering changes in input data patterns. The technology involves sequential use of a neo-fuzzy system, an adaptive neuro-fuzzy system, and a multilayer neural network with kernel bell-shaped activation functions. The neo-fuzzy system uses neo-fuzzy neurons, ensuring resistance to noise. The adaptive neuro-fuzzy system considers distances between input data and class centers in feature space, ensuring classification under class imbalance conditions. The multilayer neural network with kernel bell-shaped activation functions uses a recurrent learning algorithm, ensuring adaptation to new data with a new distribution. The technology enables rapid iterative refinement of classification decisions according to changes in input data characteristics.

Keywords: rapid classification, overlapping classes, online mode, nearline mode, neural network, neuro-fuzzy system, neo-fuzzy system, adaptive learning, fuzzy logic.

Повні імена авторів / Author's full names

Автор 1 / Author 1: Бодянський Євгеній Володимирович / Bodiansky Yevhenii Volodymyrovych

Автор 2 / Author 2: Чала Ольга Сергіївна / Chala Olga Serhiivna

ЗМІСТ

СИСТЕМНИЙ АНАЛІЗ І ТЕОРІЯ ПРИЙНЯТТЯ РІШЕНЬ.....	3
<i>Pavlov A. A., Holovchenko M. N., Drozd V. V.</i> Modification of the decomposition method of constructing multivariate polynomial regression which is linear with respect to unknown coefficients	3
<i>Lavshchenko R. R., Lvov G. I.</i> Analysis of the applications of the data-driven approach in evaluating the thermal-physical properties of composites	11

УПРАВЛІННЯ В ТЕХНІЧНИХ СИСТЕМАХ	18
<i>Полярус О. В., Хоменко Ю. С.</i> Метод виявлення наземних орієнтирів для навігації автономних мобільних роботів з використанням особливостей розподілу середньої інтенсивності кольорів	18
<i>Kutsenko O. S., Alforov M. E., Alforov A. E.</i> Synthesis of design parameters of multi-purpose dynamic systems	25
УПРАВЛІННЯ В ОРГАНІЗАЦІЙНИХ СИСТЕМАХ	30
<i>Kopp A. M., Liutenko I. V., Yamburenko V. V., Pashniev A. A.</i> A software solution for real-time collection and processing of medical data for epilepsy patients	30
<i>Козуля М. М., Солдатко О. В.</i> Проектування інформаційної CRM-системи у сфері стоматологічного адміністрування	38
МАТЕМАТИЧНЕ І КОМП'ЮТЕРНЕ МОДЕЛЮВАННЯ	42
<i>Петров К. Е., Чалий Т. В.</i> Ситуаційна модель медичного бізнес-процесу	42
ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ	46
<i>Kopp A. M., Orlovskiy D. L., Gamayun I. P., Sapozhnykov I. V.</i> Software development and research for machine learning-based structural errors detection in BPMN models	46
<i>Rudskiy O. V., Kopp A. M., Goncharenko T. Ye., Gamayun I. P.</i> Intelligent technology for semantic completeness assessment of business process models	56
<i>Ziuziun V. I., Petrenko N. A.</i> Application of optical character recognition and machine learning technologies to create an information system for automatic verification of offline testing	66
<i>Ziuziun V. I., Osoka D. S.</i> Mathematical rationale for creating an application for conducting random meetings «Coffee Break»	76
<i>Калінін Д.В., Северин В.П., Безменов М.І.</i> Моделі приватності та техніки анонімізації табличних медичних даних	81
<i>Кондратов О. М., Нікуліна О. М.</i> Програмна реалізація із використанням трансформера з оптичним потоком та GEONET для ідентифікації параметрів динамічних об'єктів	86
<i>Sharov V. O., Nikulina O. M.</i> Study of compatibility of methods and technologies of high-level protocols and error-correcting codes	92
<i>Хацько Н. Є., Слепушков М. В., Хацько К. О., Шебанов Є. О.</i> Модифікований алгоритм розгортання програмного забезпечення з використанням багатопоточності	98
<i>Колбасін В. О.</i> Оркестрування робочих процесів взаємодії з клієнтами в корпоративних застосунках	104
<i>Бодяньський Є. В., Чала О. С.</i> Інтелектуальна інформаційна технологія швидкої класифікації за умов перетинних класів	108

CONTENT

SYSTEM ANALYSIS AND DECISION-MAKING THEORY	3
<i>Pavlov A. A., Holovchenko M. N., Drozd V. V.</i> Modification of the decomposition method of constructing multivariate polynomial regression which is linear with respect to unknown coefficients	3
<i>Lavshchenko R. R., Lvov G. I.</i> Analysis of the applications of the data-driven approach in evaluating the thermal-physical properties of composites	11
CONTROL IN TECHNICAL SYSTEMS	18
<i>Poliarus O. V., Khomenko Y. S.</i> Method of detecting landmarks for navigation of autonomous mobile robots using features of average color intensity distribution	18
<i>Kutsenko O. S., Alforov M. E., Alforov A. E.</i> Synthesis of design parameters of multi-purpose dynamic systems	25
MANAGEMENT IN ORGANIZATIONAL SYSTEMS	30
<i>Kopp A. M., Liutenko I. V., Yamburenko V. V., Pashniev A. A.</i> Synthesis of design parameters of multi-purpose dynamic systems	30
<i>Kozulia M. M., Soldatko O. V.</i> Design of CRM information system in the field of dental administration	38
МАТЕМАТИЧНЕ І КОМП'ЮТЕРНЕ МОДЕЛЮВАННЯ	42
<i>Petrov K. E., Chalyi T. V.</i> Situational model of a medical business process	42
ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ	46
<i>Kopp A. M., Orlovskiy D. L., Gamayun I. P., Sapozhnykov I. V.</i> Software development and research for machine learning-based structural errors detection in BPMN models	46
<i>Rudskiy O. V., Kopp A. M., Goncharenko T. Ye., Gamayun I. P.</i> Intelligent technology for semantic completeness assessment of business process models	56
<i>Ziuziun V. I., Petrenko N. A.</i> Application of optical character recognition and machine learning technologies to create an information system for automatic verification of offline testing	66
<i>Ziuziun V. I., Osoka D. S.</i> Mathematical rationale for creating an application for conducting random meetings «Coffee Break»	76
<i>Kalinin D. V., Severyn V. P., Bezmenov M. I.</i> Privacy models and anonymization techniques for tabular healthcare data	81
<i>Kondratov O. M., Nikulina O. M.</i> Software implementation using transformer with optical flow and geonet for identifying parameters of dynamic objects	86
<i>Sharov V. O., Nikulina O. M.</i> Study of compatibility of methods and technologies of high-level protocols and error-correcting codes	92
<i>Khatsko N. Y., Slipeushkov M. V., Khatsko K. O., Shebanov Y. O.</i> Modified software deployment algorithm using multi-threading	98
<i>Kolbasin V. O.</i> Orchestration of customer interaction workflows in enterprise applications	104
<i>Bodiansky E. V., Chala O. S.</i> Intelligent information technology for rapid classification under conditions of overlapping classes	108

НАУКОВЕ ВИДАННЯ

**ВІСНИК НАЦІОНАЛЬНОГО ТЕХНІЧНОГО УНІВЕРСИТЕТУ «ХПІ».
СЕРІЯ: СИСТЕМНИЙ АНАЛІЗ, УПРАВЛІННЯ ТА ІНФОРМАЦІЙНІ
ТЕХНОЛОГІЇ**

Збірник наукових праць

№ 2 (12) 2024

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@khi.edu.ua

Підп. до друку 27.12.2024 р. Формат 60×84 1/8. Папір офсетний.
Друк офсетний. Гарнітура Times New Roman. Умов. друк. арк. 9,5. Облік.-вид. арк. 10.
Тираж 100 пр. Зам. № 6/12/24. Ціна договірна.

Виготовлювач: ФОП Панов А. М. Свідоцтво серії ДК № 4847 від 06.02.2015 р.
м. Харків, вул. Жон Мироносиць, 10, оф. 6, тел. +38(057)714-06-74, +38(050)976-32-87
copy@vlavke.com